

Subnational Leaders and Economic Growth

Evidence from Chinese Cities*

Yang Yao[†] and Muyang Zhang[‡]

First Version: January 14, 2011

This Version: August 18, 2014

Abstract

This paper studies the role of subnational leaders in economic growth using a unique city-leader matched dataset of Chinese cities for the period 1994-2010. A unique feature of China's institutional setups is that local leaders are often moved from one city to another. This allows us to compare leaders across cities. Adopting the decomposition method applied to employer-employee matched data we find that leaders matter for local economic growth. We also study how leaders' contributions to local economic growth impact on their chances of promotion in the Chinese hierarchy. We find that personal abilities become more important as a leader gets older and this effect is the most pronounced around the median age in the sample. Those results remain robust when we conduct full-information ML estimation of a system of equations consisting of both the growth and promotion equations.

Keywords: local leaders, economic performance, promotion tournament

JEL Classification: H11, M51, O53, P26

1 Introduction

Empirical studies in the literature have found that national leaders play an important role in economic growth. For example, Glaeser et al (2004) find that poor countries in the 1960s, mostly governed by dictators, got out of poverty through good policies before they improved their political institutions. Using the sudden death of a leader as an exogenous shock, Jones and Olken (2005) find that the change of leaders has a significant impact on a country's economic growth. These findings can be contrasted with the thesis that institutions are the more

*We thank Yuen-Yuen Ang, Michael Carter, Steven Durlauf, Ethan Kaplan, Karl Scholz, Li-An Zhou, three anonymous referees and seminar participants in Peking University, Shangdong University, Stanford University, the University of California-Davis, the University of Chicago, the University of Washington, the University of Wisconsin-Madison, the Third Doctoral Student Forum (Xiamen, December 2011), the Eleventh China Economic Annual Meeting (Shanghai, December 2011), the Second Annual CUHK-Fudan-Tsinghua Conference on the Chinese Economy (Beijing, December 2011), and the First Conference on Organizational Economics (Beijing, June 2012) for their helpful comments. The study is supported by the Major Project Program of the National Social Science Foundation of China (No. 09&ZD020).

[†]China Center for Economic Research and National School of Development, Peking University.

[‡]Corresponding author. China Public Finance Institute and School of Public Economics and Administration, Shanghai University of Finance and Economics.

fundamental cause for economic growth (e.g., North and Thomas, 1973; Acemoglu, Johnson, and Robinson, 2006). This paper extends the above literature to studying subnational leaders using data collected on 312 Chinese cities for the period 1994-2010. The advantage of studying subnational leaders is that these leaders face the same national institutional setup so their role can be isolated from the role of national institutions. As a result, such a study can provide more decisive evidence for the question whether leaders matter for economic growth.

In most countries, though, subnational leaders do not serve in more than one locality, making comparisons among them impossible because their observed performances are results of the combination of their own abilities and local conditions. In this regard, China provides a unique opportunity for the study of subnational leaders. In the country, local leaders are engaged in a promotion tournament in which they compete with each other for promotion (Li and Zhou, 2005; Xu, 2011). In the process, a large number of them are periodically shuffled between localities; those movers "connect" leaders serving in different localities in a way that researchers are made able to compare all the leaders in the connected cities, regardless when they served there. Tracing the leaders who served in more than one city, we construct "connected" subsamples of cities that had leaders moving between them. With the largest of those subsamples, we are able to adopt the decomposition method developed by labor economists for employer-employee matched data (Abowd, Kramarz, and Margolis, 1999; Bertrand and Schoar, 2003) to estimate leaders' relative contributions (leader effects) to local economic growth and compare them across cities and over time. This is an improvement to the Jones and Olken test, which only accounts for within-locality variations.

One potential problem of our decomposition exercise is that the leader effects thus measured may only pick up the heteroskedastic shocks that the cities received during their tenures. To rule out this possibility, we conduct several robustness checks including a study of the residuals first applied by Bertrand and Schoar (2003), a placebo test that perturbs leaders' tenures and introduction of city-specific AR(1) processes.

In addition to studying local leaders' contribution to economic growth, we explore the linkage between the estimated leader effects and leaders' chances of getting promoted. We keep in mind two purposes of doing this. One is to check the consistency of our first-step results. To the extent that the Chinese hierarchy — one that is modeled on the Soviet *nomenklatura* — promotes local leaders based on their personal abilities (Xu, 2011), we should find a positive correlation between leader effects and leaders' chances of promotion. We can then be more confident that our estimates of the leader effects reflect leaders' true abilities if these effects have good predicting power for their chances of getting promoted.

However, studying leaders' promotion introduces a complication to our econometric strategy. To the extent that leaders' abilities affect both local economic growth and their chances of promotion, estimating their leader effects and promotion separately would suffer from the

problem of simultaneity biases. Inspired by Abowd et al (2006) and Buchinsky et al (2010), we then take a system of equations approach to estimate the leader effects and their impacts on promotion together.

Our other aim to study promotion is to improve on the existing literature of the Chinese promotion tournament. We make two methodological contributions in this regard. First, the existing studies use the average growth rate of a leader's tenure to predict his promotion. However, the growth rate may not entirely reflect a leader's own capabilities which presumably are the factor the government's organizational department looks at. We are able to infer leader effects from local economic growth rates and use them to predict their promotion; therefore, we provide a more informative test for the relationship between leader performance and promotion. Second, the existing studies all take a single equation approach and thus can suffer from the simultaneity biases. Our system of equations approach can correct it.

In addition to the two methodological contributions, our data enable us to provide more reliable estimates than those found in the current literature. The existing studies (e.g., Li and Zhou, 2005; Xu, Wang, and Shu, 2007; Wang, Xu and Li, 2009) have all studied provincial leaders. One problem of this approach is that it has a limited number of observations. In addition, the promotion of provincial leaders to the central government is often influenced by political factors (Oppen and Brehm, 2007; Shih, Adolph, and Liu, 2012) whereas the promotion of city leaders to the provincial offices is less likely so.

The rest of the paper is organized as follows. Section 2 gives a description of the sources and structure of the data. Section 3 studies leaders' contribution to local economic growth. We first lay out the baseline econometric specification linking local economic growth with the leader effects and describe the problem of indeterminacy implied by the specification. Then we show how a connected sample allows us to overcome the problem. Lastly, we present the empirical results and several robustness tests dealing with the potential heteroskedasticity in the data. Section 4 studies how leader effects affect leaders' promotion. We first adopt a single-equation approach, and then estimate leader effects and leaders' promotion in a system of equations. Section 5 concludes the paper.

2 The Data

China has a highly decentralized fiscal system despite its one-party political system (Che, Qian, and Weingast, 2005; Xu, 2011).¹ There are five levels of government in the country: central, provincial, municipal, county/district, and township. Each level of government has its own independent budgets and independent or shared revenue sources. At each level, two officials

¹Xu (2011) calls it a regionally decentralized authoritarian (RDA) regime. He provides a comprehensive survey of this regime, explaining its historical roots, the way it functions, and its implications for economic reform and growth in China.

assume the highest offices. One is the secretary of the local communist party committee, and the other is the head of the executive branch (e.g., governor at the provincial level and mayor at the city level). In theory, the party secretary is elected by the local party congress and the executive officer is elected by the local People's Congress, the legislative body; in reality, most of them are appointed by the organizational department of the Communist Party in the government one level higher. As described in Li and Zhou (2005) and Xu (2011), the process of moving up in the government/party hierarchy can be best characterized by a tournament. That is, lower-level (e.g., county) government officials compete with each other, and those picked by the organizational department enter another round of elimination game at a higher-level (e.g., city), and so on, until one reaches the highest level, the standing committee (in the latest round, consisted of seven people) of the political bureau in the Party's central committee. As such, the competition becomes more and more intense when one moves up the hierarchy.

While political connections are an important determining factor in the promotion tournament, personal abilities are equally important, especially at the lower levels. Although the central government has been trying to include more goals in the criteria of evaluation, economic growth is the dominant goal for most officials. From the perspective of the multitasking theory, this is hardly surprising because economic growth is much easier to measure than other goals. The literature has documented both systematic and anecdotal evidence for the importance of economic growth in the promotion tournament; some authors (e.g., Xu, 2011) attribute China's high economic growth rates to competition among local leaders.² On the other hand, higher growth rates of the local economy serve as a good predictor for a leader's chances of promotion (Li and Zhou, 2005). Presumably, the organizational department is using the record of economic growth in a leader's tenure to gauge his personal abilities.

In this paper, we study the party secretaries and mayors in the cities. There are four kinds of cities: the four provincial cities (Beijing, Shanghai, Tianjin and Chongqing), sub-provincial cities (provincial capitals and one or two other big cities in the province), prefectural cities, and county-level cities. The four provincial cities are clearly outliers as they are actually a special form of provincial unit, and the numerous county cities are of lower level than the prefectural cities in the hierarchy. Hence, we study sub-provincial and prefectural cities in this paper, as they are more homogeneous in terms of size and the rank of government officials. There were 333 such cities in the country as of 2012.

By law, the mayor is the executive officer of the municipal government; at the same time, the law also says that the mayor is under the guidance of the city communist party committee

²The authors' own interviews in the summer of 2011 found that city and county governments used score boards to evaluate subordinate government officials, every month in some cases. The goals on top of the list were ubiquitously related to economic growth, such as the growth of tax revenues, the progress of road building, conversion of village land to industrial parks, etc., as well as GDP figures themselves. The officials who were ranked at the bottom faced the risk of being sacked (in one county, it was made clear that a leader who ranked as the last one twice in a row would be removed from his/her position).

for which the party secretary is the head. In most cities, the party secretary clearly is the number-one person because important decisions are made in the party committee. However, his power is checked by the mayor because in theory executive orders should be delivered through the mayor. In the end, the party secretary and the mayor share power in a city. A division of labor which emerges from the reality is that the party secretary is in charge of the personnel and other political duties such as maintaining social stability while the mayor is in charge of the daily operation of the government for which economic growth is a top priority. To the extent that the mayor has to rely on the bureaucracy to manage the economy, his contribution to local economic growth is tied to the party secretary's efforts to select more capable subordinates. In reality, the interaction between the party secretary and the mayor takes many forms and the pattern of their contributions to local economic growth cannot be readily parameterized. In our empirical study, we will take a simple approach by treating them as making separate contributions to local economic growth.³

The period covered by our study is 1994-2010. This period was chosen primarily because of the availability of data. It is difficult to get data on city leaders before 1994, and prefecture-level macroeconomic and demographic data beyond 2010 were not made public when this paper was written. The year 1994 was chosen as the starting year also because China started a new revenue-sharing system in that year. Before that year, the central government shared revenue with provincial governments based on negotiation; since that year, revenue has been shared under preset rules, similar to the federal system adopted in the United States.

Information on the party secretaries and mayors was collected from *The China Yearbook of Municipalities*, provincial yearbooks and reports from the media, especially the Internet. We then match the leaders to annual macroeconomic data collected from provincial yearbooks by the following rules⁴:

1. Each city-year observation is matched with one secretary and one mayor.
2. If there was a turnover within a year, we take the leader who stayed for over 6 months in that year.
3. If there were multiple turnovers in a year and no leader stayed for over 6 months, we take the leader with the longest stay in that year.

Due to the limitation of data sources, we were able to collect an unbalanced panel of 2,138 leaders in 312 cities with the starting year varying between 1994 and 1998.⁵ For information

³Including both the secretaries and mayors in our analysis can substantially increase the sample size of our largest connected sample. On the other hand, though, this causes an identification problem. We will deal with it in the next section.

⁴These rules are similar to those applied in Li and Zhou (2005)

⁵The sample covers all prefectural cities except those in Xinjiang and Tibet, two most remote ethnic autonomous regions.

on promotion, however, the starting year varies from 1998 to 2001, depending on data availability. Subsequently, we will call the whole sample as "the long sample" and the sample with information of promotion as "the short sample". Table A1 in the appendix lists the names of cities in our dataset as well as the starting years of detailed information; Figure 1 then maps them to a Chinese map. The shaded cities are our sample cities, and the heavily shaded cities are those in our largest connected sample.⁶

[Figures 1 inserted here.]

Among all the 2,138 leaders in the long sample, 1,817 served in only one city, 282 had one switch, 34 had two switches, and the remaining five had three switches (Table A2). We call those who served in more than one city in our sample period "the movers". The total number of movers was 321, or about 15% of the total number of leaders. Figure 2 shows the distribution of all leaders' tenures in one city. The average was 3.74 years, lower than the designated tenure of five years, also lower than the average tenure of the provincial leaders during the period 1978-2005 which was almost four years (Wang and Xu, 2008). The median tenure was even shorter, only three years.⁷

[Figure 2 inserted here.]

In a dataset like ours, attrition is unavoidable. Table A3 presents the distribution of the number of years a leader appears in our sample. Half of the leaders appear for less than three years while only one quarter appear for more than five years. There are generally three ways for a leader to leave our sample: promotion to the provincial or central government, being moved to a city not covered by our sample, and retirement. We follow Li and Zhou (2005) to define promotion in the following way:

- From a mayor to a party secretary in any city.
- From an ordinary city to a mayor/party secretary in a sub-provincial city.
- From any city to a post in the central government or to the provincial government as party secretary, governor, vice secretary or vice governor.
- From an ordinary city to the head of a department in the provincial government.

Same as in Li and Zhou (2005), we treat being moved to the city or provincial legislative bodies (People's Congress and People's Political Consultation Conference) as retirement in

⁶The composition of connected samples will be described in detail in the next section.

⁷There are several reasons why the tenure is so short. One is to limit the chances of corruption; another is to prevent local leaders from building their own power bases; the most important, though, is the promotion tournament itself — to move from the bottom to the very top, it would take a person several decades, so his tenure at one post has to be short enough to allow that happen before he gets to the retirement age.

addition to regular retirement because this kind of moves tends to signal a loss of power and influence within the Chinese system.

Using the sample with non-missing promotion information in the short sample, we can get a sense of the distribution of attrition. Among the 2,378 leader-term pairs, 1,326 of them (55.8%) ended with promotion while 369 (15.5%) ended with retirement. Among the 2,138 leaders in the short sample, 1,562 (73.1%) left our sample before 2010. While the determination of each of the three ways of attrition is not likely to be random, what is pertinent to our study is whether the group of leaders who leave the sample as a whole is systematically different from the group staying in our sample. If it is not, then attrition can be treated as random draws from the existing sample.

3 Contribution of Leaders to Economic Growth

3.1 The Growth Equation

Our main purpose is to compare leaders by their personal abilities. Obviously there is no direct way for us to do that, especially when we do not have information about their education levels and career backgrounds for most of the time. The method we adopt is the fixed-effect method Bertrand and Schoar (2003) use to measure the contributions of CEOs in different companies. To be sure, what we will get may not reflect leaders' true abilities, and when they are correctly measured, they only reflect leaders' abilities to promote economic growth. So instead of "personal abilities" we use "leader effects" to describe what we aim at obtaining.⁸ To begin with, we note that the economic growth rate of a city in a particular year is related to four unobserved factors in addition to observed covariates, namely, the year fixed effect, the city fixed effect, and the party secretary and mayor's leader effects. The year fixed effect is orthogonal to the other three effects and can be identified by including the year dummies in a panel regression with the growth rate as the dependent variable. On the other hand, the city fixed effect and the two leader effects share the same dimension of data and cannot be readily identified. Our aim is to find a strategy to disentangle these three effects. For that, we start with a discussion of the relationship between the party secretary and the mayor's leader effects.

As we pointed out in the last section, the party secretary and the mayor share power in a city. With no *prior* knowledge on how they interact with each other — there would be tremendous variations across cities even if we did have that knowledge — any specification that requires them to complement or substitute with each other would yield biased estimates.

⁸To be precise, here leader effects include both leaders' personal abilities and their efforts to promote economic growth. For that matter, our study does not provide direct evidence to distinguish between the selection effect and the incentive effect, which are often studied by political scientists for government officials.

⁹ In practice, we assume that they contribute to local growth separately. Technically, we treat them as if they worked in two different but identical cities, so they enter our analysis as two different observations. With this in mind, our econometric specification is the following three-way fixed-effect model: ¹⁰

$$y_{i(jt)} = X_{i(jt)}\beta + \theta_i + \psi_j + \gamma_t + \epsilon_{i(jt)} \quad (1)$$

where $y_{i(jt)}$ is the real GDP growth rate (in decimal form) of city j in year t under leader i 's tenure, $X_{i(jt)}$ is a set of time-varying controls, θ_i is the personal (fixed) effect of leader i (either a party secretary or a mayor) of city j in year t , ψ_j is city j 's fixed effect, γ_t is the fixed effect for year t , and $\epsilon_{i(jt)}$ is the random disturbance for city j 's growth in year t . In $X_{i(jt)}$ we include per capita GDP j in the starting year of leader i 's tenure and city population of year t , both in logarithm forms, and the inflation rate (in decimal form) of year t .¹¹ Under the maintained assumption of exogeneity

$$E(\epsilon_{i(jt)} | X_{i(jt)}, \theta_i, \psi_j, \gamma_t) = 0$$

Equation (1) is a revised version of the regular growth equation.

Note that by using equation (1), the GDP growth rate of a city in a particular year, as well as the corresponding right-hand-side variables other than the leader effect, appears twice in the dataset: one for the party secretary and the other for the mayor. In effect, we are stacking together the data of two separate regressions for party secretaries and mayors. The main gain of stacking the data is that it substantially increases the size of the largest connected sample. The size of a connected sample is a convex function of the number of leaders moving between cities. If we estimate mayors and party secretaries separately, the number of movers in each sample is about half of the number of movers in the combined sample, but the size of each

⁹For example, a sensible way to handle the two effects seems to assume that they are additive. However, this will lead to two problems. One is that it implies that the mayor and the party secretary substitute for each other, which is a strong assumption and hard to substantiate. The other is that it will make our identification more difficult because we need to disentangle three effects, the city fixed effect, the party secretary effect and the mayor effect, that share the same dimension of data. While we can still construct connected samples including both mayors and party secretaries, the sizes of those samples are drastically reduced. This is because to build a connected sample of mayors (party secretaries), we need first divide all the mayors (party secretaries) into small groups such that in each group, the mayors (party secretaries) have worked with the same party secretary (mayor). Then, to connect any pair of these groups, we need at least one mayor (party secretary) who has worked with both party secretaries (mayors) who respectively define the two groups. Obviously, this is a very demanding condition in reality.

¹⁰Recently, random effect models are becoming popular in the analysis of linked-employer-employee data (Card et al, 2013). Here we stick to the tradition fixed effect model for two reasons. First, compared to the firm-employee data, which have as many as millions of observations, the size of our data is much smaller. We only have 1,196 leaders and 5,403 observations in the connected sample. With this sample size, it is unrealistic to assume that the error components and the observables are uncorrelated, so the fixed effect model is more reliable than the random effect model. Second, we not only examine whether leaders matter, but also use the estimated personal fixed effect as a predictor for promotion. This also requires us to apply the fixed effect model, so that leader effects can be recovered.

¹¹It is defined as the growth rate of the provincial GDP deflator. We could find the city GDP deflators only after 2000; we use the provincial deflator for all years for the sake of consistency. We have tried replacing the provincial deflator with the city deflator for years since 2001 onward, and the result is not changed much.

sample is reduced to less than half of the size of the combined sample. In fact, the size of the separate samples can be very small depending on the actual circumstances, such as in our case. Later when we conduct robustness checks, we will re-estimate Equation (1) by giving the mayor and party secretary a set of *ad hoc* weights for their contributions.

Note also that equation (1) is built on the idea of decomposition of variance. We rely on the different tenures of leaders having served in the same city to attribute that city’s economic growth in the relevant periods to these leaders’ personal contributions. In particular, the paired party secretary and mayor are assumed to make separate contributions. Therefore, if a party secretary and a mayor had their tenures perfectly coincided (i.e., they either worked in exactly the same years in the same city or moved together to another city), their estimated leader effects would be the same. Fortunately, we do not have any pair of such party secretaries and mayors.

In Equation (1) the growth rates of the two identical cities that the paired mayor and party secretary work for are assumed to be independently drawn from the same distribution. As a robustness check, we can also assume that they are drawn from a bivariate distribution. The mayor and party secretary are still assumed to make separate contributions, but their contributions are correlated. In effect, we estimate the following system of equations:

$$\begin{aligned} y_{i(jt)} &= X_{i(jt)}\beta + \theta_i + \psi_j + \gamma_t + \epsilon_{i(jt)} \\ y_{i'(jt)} &= X_{i'(jt)}\beta + \theta_{i'} + \psi'_j + \gamma_t + \epsilon'_{i'(jt)} \end{aligned} \quad (2)$$

Where i is the index for mayors and i' is the index for party secretaries. We assume that $\epsilon_{i(jt)}$ and $\epsilon'_{i'(jt)}$ follow a bivariate normal distribution with zero means. Here we allow the city fixed effects in the two equations to be different. Note that by this specification, we need to impose the assumption that the leader effects have an equal mean across mayors and party secretaries to maintain the connected sample. This assumption may be even stronger than the assumptions maintained when we apply the benchmark estimation based on Equation (1).

3.2 Identification and Test Strategies

In most panel-data analysis, researchers focus on the coefficients of covariates and add fixed effects only as controls to eliminate unobservable within-group-invariant factors. In this paper, we care about the fixed effects themselves. However, we have three sets of fixed effects to estimate while the data on economic performance only have two dimensions, i.e., city-leader pair and the calendar year, so there is indeterminacy between the city and leader fixed effects. This is evidently reflected by Equation (1). The dependent variable y_{jt} is indexed by j that identifies cities and t that identifies the years. In the meantime, the leader effect θ_i enters the equation to coincide with the city fixed effect ψ_j for a number of years. That is, the data points corresponding to θ_i are a subset of the data points corresponding to ψ_j if leader i only

worked in one city. As a result, the best one can do for this kind of leaders is to estimate the sum of θ_i and ψ_j . However, the literature of employer-employee matched data in labor economics (Abowd et al, 1999; Abowd et al, 2002; Abowd and Kramarz, 2006; Cornelissen, 2008) provides guidance for us to address this problem. Following that literature, we can build "connected samples" created by leaders who moved between cities and estimate a relative order for θ_i 's and ψ_j 's separately.

Figure 3 provides a simple illustration with only one leader at a time in each city. In the figure there are three cities and six leaders. Leaders 1 and 2 only served in city A, leader 3 served in both city A and city B, leader 4 only served in city B, and leaders 5 and 6 only served in city C. City 1 and city 2 are connected by leader 3 who served in both cities. Because of that, all the four leaders having served in the two cities are also connected. We then call cities A and B and leaders 1 to 4 a connected group. In contrast, city C does not have a leader switching to the other two cities, nor does it have a leader coming from the other two cities. So city C and leaders 5 and 6 form a separate group.

[Figure 3 inserted here.]

As we pointed out previously, normally we can only identify the sum of the leader fixed effect and the city fixed effect $\omega_{ij} = \theta_i + \psi_j$. As a result, the fixed effect of city C ψ_C cannot be separated from the fixed effects of leaders 5 and 6, θ_5 and θ_6 . So ψ_C cannot be identified. However, the difference of θ_5 and θ_6 can be identified because it is equal to $\omega_{6C} - \omega_{5C}$.

In the connected group, we can do more using the connection created by the "mover", leader 3. Firstly, subtracting ω_{3A} from ω_{3B} we get the difference between city A and city B's fixed effects $\psi_B - \psi_A$. Then subtracting ω_{3A} from ω_{1A} and ω_{2A} we get the difference between leaders 1 and 3 and the difference between leaders 2 and 3, respectively. Finally, subtracting ω_{3B} from ω_{4B} we get the difference between leader 4 and leader 3. With that, we can finally compare all the four connected leaders. However, the values of θ_i and ψ_j are not unique. For example, we can add 1 to ψ_j for all j and subtract 1 from θ_i for all i and leave $\omega_{ij} = \theta_i + \psi_j$ unchanged.

The study of both the mayor and the party secretary in a city complicates the identification problem because it adds one more dimension. In this regard, our specification (1) helps us out. By this specification, we can treat a pair of party secretary and mayor working in the same city in the same year as if they were working in two separate albeit identical cities. Because they share the same city fixed effect, they can be treated as if working in the same city, so the size of the connected group is greatly increased. In the meantime, because their tenures did not perfectly overlap with each other in our data and they are treated and estimated separately in our econometric setup, their ω_{ij} have different values despite sharing the same city fixed effect as a component, and we can separate their leader effects from each other.

In conclusion, we can identify the differences of fixed effects between leaders as well as between cities within a connected group. Moving local officials from city to city increases the size of the connected group, which makes it feasible to compare personal qualities within a larger amount of cities. This may be one of the key reasons why the central government in China keeps moving officials across cities and provinces.

In our long sample of 2,138 leaders, the 321 movers allow us to identify 20 connected groups plus 38 isolated cities and 253 isolated leaders. Table A4 provides detailed information on these groups. As one can see, among the connected groups, many groups are small. But Group 1 is sufficiently large for our analytical purposes. This group consists of 175 cities, 1,196 leaders (among which 218 are movers) and 5,403 leader-year pairs (observations); that is, it accounts for 58% of the whole sample. Thereafter we will mainly analyze this group and simply call it "the connected sample". In Figure 1 the connected sample of cities is shaded in a darker color contrasting to other sample cities in a lighter color.

Based on the connected sample, we can estimate equation (1) to test whether leaders matter. As noted above, we can only identify the differences between cities and between leaders. But to save notation, we set the mean of θ_i 's to zero and still use ψ_j and θ_i to denote those differences. Then an F test on the joint significance of θ_i suffices for our purpose. This test shares the same idea as Jones and Olken's, i.e., relying on the variation among leaders to answer the questions whether leaders matter. However, the Jones and Olken test is a stronger test than ours. This is primarily because the Jones and Olken test, if it were to be applied to our case, would only consider within-city variations, but our test considers both within-city and between-city variations, so a rejection of the null by the Jones and Olken test definitely implies a rejection of the null by our test. Conversely, if the Jones and Olken test cannot reject the null, it does not mean that leaders do not matter, as Jones and Olken themselves have noticed, because leaders in different cities may perform differently. Failure in rejecting the null by our test does not mean that leaders do not matter either. This is because we can only estimate the differences among leaders and cannot estimate the absolute values of their leader effects. That is, failure in rejecting the null can imply that leaders are equally capable. Because of the dimensional limitation involving leaders and cities, this is by far the best result one can achieve. Our improvements on Jones and Olken's test thus are two folds. One is that our test has a smaller probability of making the Type II error than Jones and Olken's, and the other is that we use more information and failure in rejecting the null is a more decisive indicator that leaders are all the same and do not matter for local economic growth in the sense that shuffling them around does not have any effect.

To see how our tests would differ from that in Jones and Olken (2005), we also compose a χ^2 -test similar to theirs using only within-city variations. Our data have multiple turnovers of leaders in a single city so the *PRE* and *POST* dummies in their paper are not clearly

defined in our case. But consider a city A with 3 consecutive leaders with fixed effects $\theta_{1,A}$, $\theta_{2,A}$ and $\theta_{3,A}$. In the first turnover, $POST - PRE = \theta_{2,A} - \theta_{1,A}$; while in the second turnover, $POST - PRE = \theta_{3,A} - \theta_{2,A}$. So the parallel test is $\theta_{2,A} = \theta_{1,A}$ and $\theta_{3,A} = \theta_{2,A}$.

In practice, we can estimate equation (1) using the whole long sample and construct a J statistic similar to that in Jones and Olken (2005).

$$J = \frac{1}{Z} \sum_{i=1}^Z \frac{(\theta_i - \theta_{i'})^2}{\frac{2\sigma_{\epsilon_i}^2}{(T_i + T_{i'})/2}}$$

where Z is the total number of leaders, $\theta_i - \theta_{i'}$ is the difference in the leader effects of two consecutive leaders for the same city-position, $\sigma_{\epsilon_i}^2$ is the variance of the error term for city i , and $(T_i + T_{i'})/2$ denotes the average tenure length of the consecutive terms. $\theta_i - \theta_{i'}$ and $\sigma_{\epsilon_i}^2$ are replaced with $\hat{\theta}_i - \hat{\theta}_{i'}$ and $\hat{\sigma}_{\epsilon_i}^2$ from Equation (1), respectively.

We then execute a χ^2 -test with Z degrees of freedom. If we reject the null, then we can conclude that leaders differ even within a city, as done by Jones and Olken; if we cannot reject the null, then we conclude that leaders do not outperform each other within a city.

3.3 Baseline empirical results

We first mimic the Jones and Olken's original J test by using retirement as an exogenous shock. ¹² The premise is that if retirement is sufficiently random, we would observe significantly different performance records of the new leaders if leaders did matter for growth. The two groups of comparison are the terms of retired leaders and the terms of their immediate successors who themselves did not retire in those terms. We first run a simple regression and find that after a city leader retires, his successor can boost local economic growth by 1.0 percentage point. When we repeat Jones and Olken's J test, the change is also significant. However, we need to be cautious about the validity of this result because retirement may not be exogenous. As will be uncovered later in the paper, city leaders who retire are those who fail to get promotion, and their personal abilities are systematically lower than leaders who get promoted.

We then conduct the modified Jones and Olken test on all the leaders in a city. We do this twice, one using the whole long sample and the other using the connected long sample. For the whole sample, the J statistic is 2.975 and the p-value is 1.000 under 2,363 degrees of freedom. As for the connected sample, the statistic is 3.333, and the p-value is 0.999 under 1,404 degrees of freedom. Thus we cannot reject the null hypothesis with a large margin and conclude that within-city variations are not large enough to justify the contribution of leaders. That is, the Jones and Olken test fails in our dataset. One of the reasons is probably that changing leaders at the national level is a more dramatic event than changing leaders in a city. In addition, government operation at the city level may be more routine and relying more on bureaucracy

¹²We thank a referee for suggesting this to us.

than government operation at the national level. This is consistent with Jones and Olken’s finding that a change of leaders only matters in non-democracies, not in democracies. One of the explanations there is that procedures are more robust in democracies so their performance is more stable than in non-democracies.

As we pointed out before, we may make the Type II error if we conclude from the Jones and Olken test that leaders do not matter for local economic growth because leaders’ leader effects may vary between cities. Next we conduct our F-test by estimating equation (1) using the connected long sample of 1,196 leaders and a total of 5,403 observations of leader-year pairs. Following the solution in Cornelissen (2008), we impose the following zero-mean constraint in our estimation

$$\sum \theta_i = 0.$$

By imposing this constraint, it is straightforward to apply the standard F-test that θ_i ’s are all zero. The regression results are presented in Table 1, with ordinary, White heteroskedasticity-robust and province- and city-year spell clustered standard errors. The resulting F-static is $F(1,195, 4,014)=1.93$, and the p-value is less than 0.001.¹³ That is, the null hypothesis that leaders do not matter is rejected with a large margin. leader effects

[Table 1 here]

The estimates for population and GDP per capita both return negative coefficients although the coefficient for population is not significant when the standard errors are clustered for the provinces. The result of the initial GDP per capita fits into the story of convergence, and the other result implies a certain burden of a larger population. The estimated coefficient for the Inflation rate is also negative and significant, showing that inflation is bad for economic growth.

The role of leaders can also be shown by an analysis of variance, as is done in Graham et al. (2012). Table 2 shows the shares of variance of real GDP growth that the city, year, and personal dummies respectively explain in the connected sample. The city dummies alone explain only 8% of the total variation, the year dummies explain almost 12%, and the personal dummies explain 26%. If we have correctly measured the leader effects, this shows that leaders have played a significant role in explaining local economic growth.

[Table 2 inserted here.]

Figure 4 presents the kernel density of the estimated individual leader effects. Table 3 provides summary statistics of the distribution of all leaders. The standard deviation is relatively small and the kurtosis is large, but the distribution is skewed left, indicating that

¹³Note that using different kinds of standard errors does not affect the F test.

there is a group of leaders with relatively low personal abilities. Figure 5 then separates the party secretaries and the mayors. The distribution of the mayors slightly dominates the party secretaries' although the gap is not statistically significant. We also compare the distribution of leaders who are observed to have left our sample and the distribution of the whole sample. We do not find any significant difference between them. Thinking back, this should be an expected result because almost all leaders (except those who were still in office by 2010) eventually left our sample. Therefore, biased attrition should not be a problem for our test.

[Figures 4 and 5 and Table 3 inserted here.]

Finally, we estimate Equation (2) as a SUR model. Then we conduct two F-tests for mayors' leader effects to be jointly zero and for party secretaries' leader effects to be jointly zero, respectively. The resulting F-statistics are 2.19 and 1.84, respectively, whose p-values are both less than 0.001. This confirms our results based on Equation (1). In subsequent analysis, we will thus focus on Equation (1).

3.4 Robustness tests

By our identification method, the leader effects are essentially estimated by cities' average economic growth rates during the respective leaders' tenures. The null hypothesis is that the residuals of these average growth rates after controlling the right-hand-side variables except the leader fixed effects should be random draws from the same distribution. Our baseline model in Equation (1) parameterizes these residuals by the leader fixed effects and our F-test finds that they are not random draws from the same distribution. However, this positive result may arise only because cities received heteroskedastic or spurious time-persistent shocks during leaders' specific terms (Easterly and Pennings, 2014). If that were the case, our estimates of the leader effects would be incidental than reflecting leaders' true abilities. To rule out this possibility, we conduct several robustness checks in this subsection.

The first is to follow Bertrand and Schoar (2003) to study the residuals attributable for leaders. Specifically, we first estimate Equation (1) without the leader fixed effects, and then collapse the residuals at the leader-city level (i.e., by leaders' tenures in different cities). If leaders' fixed effects only picked up city-specific heteroskedastic shocks, there should not be any correlation between these collapsed residuals — not so even between different tenures of the same leader who moved between cities. Consequently, we pick up the movers and regress the collapsed residuals of their terms before and after the move plus a constant.¹⁴ This regression returns a coefficient of 0.356, which is statistically significant at the 5% significance level. This result means that leaders did lay consistent marks on the cities they have served; our estimates of the leader effects are not incidental.

¹⁴For leaders with multiple switches, we treat each switch separately. In addition, leaders whose tenures appeared in our dataset for only one year are deleted.

To further validate this result, we regress the collapsed residuals of the movers' last tenure on the corresponding cities' average residuals of the three years before the leader turnover. The coefficient thus obtained is 0.151 and highly insignificant. That is, a city's economic performance does not have a medium-term persistent component that is passed from a leader's tenure to his successor's. This is contrasted with our first result that the error terms of a leader's different tenures are highly correlated and thus reinforces our confidence that our estimates of the leader effects are not incidental.

Our third robustness check is a placebo test that permutes leaders' tenures. If our estimates of leader effects only picked up heteroskedastic shocks, we would have no reason to believe that these shocks would form a consistent pattern that follows the cycle of leaders' tenures; instead, they may still exhibit a pattern, i.e., significantly different across different leaders, even when we fit them into a cycle that is sufficiently different from the cycle defined by leaders' real tenures. For that, we randomly permute each leader's tenures across years within the same city and re-estimate Equation (1).¹⁵ The number of possible permutations is extremely large, so a full permutation is not computationally feasible; nor is it necessary. In practice, we conduct several rounds of permutation with each round consisting of 1,000 permutations. We find that the F-statistic from the true data is either No. 1 or No. 2 among the F-statistics from any round of permutation. This gives us more confidence in our baseline results.

Our fourth robustness test is to re-estimate Equation (1) by introducing two city-specific AR(1) processes, one for mayors and the other for party secretaries, to its error term. This exercise allows us to explicitly address heteroskedasticity while estimating the leader fixed effects. Neither the regression result, nor the test result, though, changes in a meaningful way.

As a last robustness test, we re-estimate Equation (1) by assigning weights to the co-working mayor and the party secretary's contributions. Specifically, let the share of growth contributed by the mayor be m , so the share of the party secretary is $1 - m$. Then, we replace $y_{i(jt)}$ by $my_{i(jt)}$ when the observation is for a mayor or $(1 - m)y_{i(jt)}$ when the observation is for a party secretary, and re-estimate Equation (1) four times with m set to be 40%, 45%, 55% and 60%, respectively. Compared with the benchmark, which is equivalent to $m = 50%$, the F statistics are larger and p-values are smaller. That is, our baseline result is robust to different weights assigned to the contributions of mayors and party secretaries.

¹⁵Since leaders' terms are no longer continuous, we drop "initial per capita GDP" in the set of explanatory variables.

4 Leader effects and Promotion

4.1 The Promotion Equation

As we pointed out in the introduction, we have two purposes in mind to study how leader effects affect promotion. One is to test the sensitivity of our estimates of the leader effects and the other is to improve on the literature of promotion tournaments. Li and Zhou (2005) find that the average GDP growth rate of a provincial leader's tenure is a good predictor for his probability of promotion and retirement. However, Opper and Brehm (2007), Wang and Xu (2008), and Shih, Aldolph and Liu (2012) provide different results. Opper and Brehm (2007) define an index of political connections and find that it is a strong predictor for the promotion of provincial leaders while the local growth rate has no predictor power. Shih, Aldolph and Liu (2012) use a more comprehensive set of indexes for political connections and find that political connections are important for a person to enter the central committee of the CCP. Yet Wang and Xu (2008)'s results seem to reject the political connection story. While they find that provincial party secretaries and governors who are later promoted to the central government do not significantly outperform others, they also find that the provincial leaders who come from and then go back to the central government underperform the average leader. The controversy may have a lot to do with many studies' directly using the GDP growth rate as the predictor. The GDP growth rate may not be a good indicator for a leader's personal ability because it is highly correlated with local conditions, some of which change over time and cannot be accounted for by the provincial fixed effect. One of the regularities concerning the promotion is that almost all the members of the standing committee of the politburo, the center of the CCP leadership, have been either directly promoted from or have worked in the few most advanced provinces as well as the three big cities, Beijing, Shanghai and Tianjin. Because those localities enjoy preferred economic policies during various periods, Li and Zhou (2005)'s significant results may well reflect some particular features of the data rather than general links between promotion and performance. One component of Opper and Brehm (2007)'s index of political connection is whether a leader has worked in provinces that a member of the standing committee has worked for. So their results may suffer the same problem of incidental correlation.¹⁶ Lastly, ministry-level officials in the central government may be sent to provinces only for them to gain local experience, which could lead to Wang and Xu (2008)'s findings.

Our data and identification strategy allow us to improve on the existing literature. On the one hand, the leader effects estimated from equation (1) reflect leaders' own contributions to cities' economic performance; on the other hand, the promotion of city leaders is less subjected

¹⁶Another component of their index is whether a leader went to the same university that a member of the standing committee went to. Because most of the top leaders graduated from the few top universities, this can also cause an incidental correlation.

to political considerations and can be more linked with leaders' personal abilities. Therefore, we can be more confident that our estimates of the leader effects reflect leaders' real abilities if we find that they are better predictors for leaders' chances of promotion than GDP growth rates.

Note that we only have data of promotion with the starting year varying from 1998 to 2001 and can only estimate the leader effects of leaders in the connected sample. So our estimation of the relationship between leader effects and promotion is performed for the period 1998-2010 using the connected sample. This allows us to examine 995 leaders.

As a first start, we realize that at any point of time there are three possible states for a leader: promotion, retirement, and staying as a city leader (including moving to another city). We thus adopt two econometric methods to conduct our study. One is to use the linear probability model (LPM) to compare promotion and the other two options for each calendar year.¹⁷ The LPM has the advantage of being able to provide straightforward and unconditional predictions based on individual estimates alone. The other estimation method we adopt is the ordered probit model (OPM), also for each calendar year. The OPM has the advantage of treating all the three states simultaneously and thus avoiding potential biases in the estimates of the LPM due to the failure of accounting for the correlation between states. However, it does not provide direct unconditional predictions.¹⁸ When the OPM is applied, the observed outcome variable is valued as 1 for promotion, 0 for staying, and -1 for retirement.

When analyzing the probability of promotion, we need also to be concerned about the spatial scope of comparison. While there have been leaders moving to cities out of their own provinces, it has been more common that city leaders are promoted or shuffled within the same province. Among the 321 movers in the connected long sample, only 34 served in two different provinces; among the 689 leaders who enjoyed promotion, only 29 were promoted outside the province. That is, competition among leaders is mostly restricted within the same province. On the other hand, comparing leaders only within the same city would apparently be too restrictive. Therefore, we add the provincial dummies in our regression analyses to account for the fact that city leaders compete within the same province.

In summary, let $p_{i(jt)}$ be a notional variable for leader i 's latent chances to get promoted in year t when he served in city j , we then specify the promotion equation as

¹⁷We have also conducted our study based on leaders' tenures and their whole careers recorded by our sample. The results are broadly similar to those obtained based on calendar years.

¹⁸It is noteworthy that neither LPM nor OPM accounts for the time dependency of promotion. In this regard, the competing risks model is a more appropriate method. However, the difficulty is to decide the time path of the competing risks. A person's tenure as a city leader seems to be a natural choice because being a city leader longer increases both the risks of promotion and retirement, but age can also be an equally qualified candidate because the risks of promotion and retirement can follow significant patterns as one's age increases. As a result, we do not adopt the competing risks model. We add age, city tenure, and year dummies in both the LPM and OPM to address time dependency.

$$p_{i(jt)} = Z_{i(jt)}\delta + \alpha\theta_i + v_k + \eta_t + u_{i(jt)} \quad (3)$$

where θ_i is the leader effect as in Equation (1), $Z_{i(jt)}$ is a set of controls, v_k is the provincial fixed effect for province k , η_t is the year fixed effect, and $u_{i(jt)}$ is a random disturbance. In both the LPM and the OPM, $Z_{i(jt)}$ includes the following variables: leader age, number of years since a person became a city leader (city tenure), and a dummy variable indicating whether a leader worked in the provincial government or not (provincial experience).

Age could be the most important factor in addition to ability determining a leader's chances of promotion. The promotion tournament constantly eliminates people who reach the age limit for each level in the hierarchy,¹⁹ so being young can be a big advantage if one wants to move upward in the bureaucratic hierarchy.

Besides age's direct effect, the effect of leader effects on promotion can also vary by age. All leaders must stay somewhere in the government at different stage of their political careers. Although we are unable to observe their performance before they serve as city leaders, the department of organization can observe and evaluate their performances all the way along their career tracks. Intuitively, personal ability is revealed more clearly the longer a leader serves in the government, and the estimated leader effect reflects more of the true ability. In this regard, we add an interaction of the leader effect and age in the promotion equation.

Lastly, provincial experience is meant to capture the influence of political affiliation. Having worked in the provincial government should increase a person's chances of getting promoted if political connection is important.²⁰ However, it could also be the case that coming down from the provincial government to work in a city represents a step downward because it means that he cannot get promoted in the provincial government.

4.2 Empirical Results for Single-equation Estimation

Table 4 shows the results of the LPM estimated on Equation (2) when we compare the promoted leaders with those staying and retired. In column 1, we find that the leader effect has no significant effect on promotion, and age is negatively correlated with promotion although the effect is not economically significant. Instead, provincial experience is found to be helpful for promotion. Holding other things equal, a leader coming from a provincial position enjoys a higher probability of promotion by 5.1 percentage points on average. This seems to imply that political connections in the government system matter for promotion. Longer tenure as a city leader also helps. One more year of tenure increases a person's chances of promotion by 2.5

¹⁹For city leaders, 60 years old is the age limit. With few exceptions, city leaders who reach 60 should retire unless they are promoted to the provincial (or ministry) level.

²⁰By saying that a leader has worked in the provincial government, we do not mean that they work as provincial leaders. Most probably, he/she works as the head or deputy head of certain department, which is of equal or lower level of a city leader in the hierarchy.

percentage points.

As we pointed out in the last section, leader effects may be heterogeneous across age. We then interact the leader effect with age and rerun the regression. Column 2 of Table 4 presents the results. Now, the impact of the leader effect grows with age. That is, the older the leader, the more significant the role the leader effect plays in predicting his chances of promotion. By the point estimate for the interaction term, if the most capable person in the sample gets one year older, his chances of promotion increase by 2.6 percentage points over those of the least capable person in the sample. One of the causes behind this result is perhaps related to the way we estimate the leader effect. It is estimated for a leader's whole career appearing in our sample. That is, in a sense, it is the life-time ability of a leader. Therefore, as a leader becomes older and, for that matter, has worked in the government longer, more information is revealed to the organizational department who can then make more credible inference about the leader's ability. Another possible reason for the result is that competition becomes more intense when leaders become older, which could increase the value of performance in their promotion. As a matter of fact, younger leaders, from the day they assumed the municipal positions, are usually designated as political hopefuls who would one day rise up in the party hierarchy. Consequently, subsequent performance may not be a decisively factor determining their promotion.

To capture possible nonlinearity of age, we replace age by a threshold. We choose 49 years old to 52 years old (around the median age of 50) as the age threshold and present the results in columns 3 through 6 in Table 4. While the impact of the leader effect on leaders younger than the threshold age is unstable, leaders older than the threshold age do get larger chances of promotion when their leader effects become larger. When the threshold is 49 years old, among the older group of leaders, the most capable person has a chance of promotion 22.4 percentage points higher than the least capable person. This gap becomes 18.4 percentage points and 29.1 percentage points when the threshold is raised to 50 and 51 years old, respectively. However, it drops to 15.1 percentage points and becomes statistically insignificant when the threshold is raised to 52 years old. We have also tried thresholds less than 49 years old and larger than 52 years old and found that the significance of personal abilities declines. Therefore, we conclude that personal abilities have the largest and most significant impacts around the median age of the leaders in our sample.

[Table 4 inserted here.]

Table 5 presents the results of the OPM. The results do not change qualitatively compared with those of the LPM in all coefficients. However, the impact of the leader effect has become more significant in both statistical and economic terms. As pointed out by Ai and Norton (2003), the magnitude of the interaction effect in nonlinear models does not equal the marginal

effect of the interaction term. An algorithm of marginal effects for interaction terms in nonlinear models with either continuous or dummy variables is developed in their subsequent paper (Norton et al, 2004). Using that method, we derive the marginal effects for the interaction term between the leader effect and the age threshold dummy with other covariates measured at their means. These marginal effects are 36.8, 31.8, 47.5 and 27.3 percentage points when the age threshold is 49 to 52 years old, respectively, with all but the last being statistically significant. To the extent that the OPM provides a more reliable set of estimates, more weights should be given to these results as compared with those generated by the LPM.

[Table 5 inserted here.]

4.3 Joint Estimation of Growth and Promotion

Although the single-equation estimation in the last subsection is straightforward, it may encounter a problem if the error terms in Equation (1) (the growth equation) and Equation (2) (the promotion equation) are correlated. In particular, the promotion equation uses the leader effects estimated from the growth equation to predict promotion and thus will return a biased estimate if the two error terms are correlated. On the other hand, the estimates of the leader effects can be equally biased because they affect attrition for which promotion is one of its forms. In this subsection, we would like to form an econometric strategy to integrate the estimation of the two equations in a simultaneous-equation system. Unlike in the baseline estimation where leaders' leader effects are estimated from the growth equation and used in the promotion equation, now we estimate the growth and promotion equations simultaneously to take into consideration the correlation between the two error terms of the two equations. To do so, we start with the following unrestricted model that combines Equations (1) and (3) in a system of equations:

$$\begin{cases} y_{i(jt)} = X_{i(jt)}\beta + \theta_i + \psi_j + \gamma_t + \epsilon_{i(jt)} \\ p_{i(jt)} = Z_{i(jt)}\delta + \alpha\theta_i + v_k + \eta_t + u_{i(jt)} \end{cases} \quad (4)$$

Because both equations are linear, we will call the model in (4) the linear-linear model, or simply the L-L model. We assume that $\epsilon_{i(jt)}$ and $u_{i(jt)}$ follow a joint Normal distribution:

$$\begin{pmatrix} \epsilon_{i(jt)} \\ u_{i(jt)} \end{pmatrix} = \mathcal{N}(0, \Sigma), \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{21} \\ \sigma_{12} & 1 \end{pmatrix}$$

Note that the variables in $X_{i(jt)}$ appear only in the first equation of the equation system (4) so they serve as the "instruments" for identification. Our identification assumption is that these variables (i.e., initial GDP per capita, city population and inflation) only affect a leader's promotion by affecting his personal contribution to local economic growth. To the extent that a leader's ability to promote economic growth is the most significant factor determining

his chances of promotion and those variables are closely related to economic growth, this assumption is reasonable.

As we noted before, we only have data for promotion starting around 1998-2001. In our baseline regressions, we can first estimate the leader effects using the connected long sample and then estimate the promotion equation for the period 1998 onward within the connected sample. In the joint estimation of the equation system in (4), we can no longer do that. One option is to estimate the largest connected sample since 1998. However, this will drastically reduce the sample size and make results incompatible to the single-equation. What we will do is to still estimate the connected long sample, and at the same time treat leaders without promotion information as missing data (a total of 201 leaders).

One distinctive feature of the system of equations (4) is that the leader effects have to be estimated jointly from the two equations. It is then clear that a single-equation approach would return biased estimates for both α and the leader effects themselves if simultaneity exists.²¹ A system of equations approach mitigates this shortcoming. We can estimate (4) by the maximum likelihood (ML) method. One caveat is that in deducing the log-likelihood function, we need to add up likelihoods from observations with $p_{i(jt)}$ observed and missing.

Denote $\omega_{1ijt} = X_{i(jt)}\beta + \theta_i + \psi_j + \gamma_t$ and $\omega_{2ijt} = Z_{i(jt)}\delta + \alpha\theta_i + v_k + \eta_t$, the likelihood for each observation where $p_{i(jt)}$ is observed is

$$L_{i(jt)} = \phi_2(y_{i(jt)} - \omega_{1ijt}, p_{i(jt)} - \omega_{2ijt}; \Sigma)$$

where ϕ_2 denotes the density function of the bivariate normal distribution. For observations with $p_{i(jt)}$ missing, only the growth equation is estimated, so the likelihood is thus

$$L_{it} = \phi(y_{ijt} - \omega_{1ijt}; \sigma_{11})$$

where ϕ stands for the density function of the univariate normal distribution. To sum up, the log-likelihood function for the unrestricted model is

$$\ln L = \sum_{p_{i(jt)} \text{observed}} \ln \phi_2(y_{i(jt)} - \omega_{1ijt}, p_{i(jt)} - \omega_{2ijt}; \Sigma) + \sum_{p_{i(jt)} \text{missing}} \ln \phi(y_{i(jt)} - \omega_{1ijt}; \sigma_{11}) \quad (5)$$

To test whether leaders matter, we can estimate a restricted version of (4) where all the leader effects are restricted to equal to zero in the growth equation. As a result, they are also dropped in the promotion equation. We can then perform the likelihood-ratio (LR) test for the null hypothesis imposed on the unrestricted model:

$$H_0 : \theta_i = 0, \text{ for all } i.$$

²¹This means that the estimates provided by Li and Zhou (2005) and other studies for the probabilities of promotion would be biased if the organizational department looks for personal abilities instead of records of economic growth *pe se*.

Then, to examine whether leader effects are important for promotion, we can just test the null hypothesis $\alpha = 0$ in the unrestricted model.

To make it consistent with our baseline estimation, we also estimate a system where the promotion equation is specified as ordered probit (so the model is shorthand as L-OP). The model is

$$\begin{cases} y_{i(jt)} = X_{i(jt)}\beta + \theta_i + \psi_j + \gamma_t + \epsilon_{i(jt)} \\ p_{i(jt)}^* = Z_{i(jt)}\delta + \alpha\theta_i + v_k + \eta_t + u_{i(jt)} \\ p_{i(jt)} = \begin{cases} -1 & p_{i(jt)}^* \leq c_1 \\ 0 & c_1 < p_{i(jt)}^* \leq c_2 \\ 1 & p_{i(jt)}^* > c_2 \end{cases} \end{cases} \quad (6)$$

where $p_{i(jt)} = -1$ denotes retirement, $p_{i(jt)} = 0$ denotes staying as a city-level official, and $p_{i(jt)} = 1$ denotes promotion. The two error terms $(\epsilon_{i(jt)}, u_{i(jt)})$ still follow a bivariate normal distribution specified before. The log-likelihood function can be deduced in the same way as in the L-L model, and is presented in the appendices. The null hypotheses of the two tests are similar to those of the L-L model.

4.4 Results of Joint Estimations

Parallel to the single-equation estimations, we look at two specifications for the system of equation estimations, the L-L model and the L-OP model. Because the θ 's are estimated from personal dummies, convergence of the ML estimation can take a very long time. Instead of estimating the equation system (4) in one shot, we take an iterative approach. In the appendices we present a detailed description for the iterative method, and here we provide a brief description. We first use the growth equation to get an initial estimate of each leader's leader effect $\hat{\theta}_i^0$. Then we impute $\hat{\theta}_i^0$ into the promotion equation and conduct a joint estimation with the θ_i 's in the growth equation treated as unknown using the method developed by Roodman (2011). This round of exercise returns a new estimate for each leader effect, $\hat{\theta}_i^1$ as well as an estimate for α . We then repeat the previous round again. If in round t we find that $\hat{\theta}_i^t$ and $\hat{\theta}_i^{t-1}$ are sufficiently close, we take $\hat{\theta}_i^t$ as the final estimate for θ_i and take the estimate of α in the same round as its true value.

4.4.1 Results of the L-L Model

Table 6 lists the results of three estimations for the unrestricted model when promotion is estimated by the LPM. The results of the restricted model are not reported to save space. The results of the unrestricted model are broadly similar to those returned by the single-equation regressions. In estimation (1), we report the results when the leader effect is interacted with age. The LR statistic equals 2,189.64 and its p-value is less than 0.001, indicating that leaders

do matter for local economic growth.²² The coefficient for the interaction term between the leader effect and age is almost the same as that obtained in the single equation estimation. The only change is that the coefficient of provincial experience is no longer significant in the joint estimation. This means that the positive result found in the single-equation estimations may be caused by the spurious correlation between economic growth and promotion.

[Table 6 inserted here.]

We also use various thresholds of age in place of the continuous variable of age to account for potential nonlinearity. Estimations (3) - (6) in Table 6 present their results. They are also barely different from those provided by the single-equation analysis, and the χ^2 statistics all confirm the contributions of local leaders for economic growth.

4.4.2 Results of the L-OP Model

Table 7 shows the results of the L-OP model where promotion is estimated by the ordered probit model. All the estimations of Table 6 are repeated. The results are broadly similar to those produced by the single-equation estimations. One significant difference is that the impacts of the leader effect among older leaders are shown to become larger. In summary, our single-equation and system of equation estimations provide consistent and robust results that support the thesis that leaders matter for local economic growth, regardless whether we adopt the linear probability model or the ordered probit model to account for promotion.

[Table 7 inserted here.]

5 Concluding Remarks

In this paper we study the role of subnational leaders in local economic growth and their promotion using a unique panel dataset collected on city leaders from 25 Chinese provinces. With the city leaders facing homogeneous national institutional settings, we are able to isolate leaders' personal abilities from institutional factors that may confound the cross-country studies aiming at answering the question whether institutional or human factors contribute more to economic growth. We find that leaders do matter for local economic growth in the sense that their personal abilities estimated from their contributions to local economic growth are sufficiently different from each other. This result is robust to a series of tests trying to address the problem of heteroskedasticity in the data. Our findings lend weight to the thesis that human factors are important for economic growth.

²²As mentioned in the previous section, in the restricted model, leader effects do not appear in either equation. The LR test includes the contribution of the promotion equation.

We also improve on the existing literature on the promotion tournaments in China. Using the leader effect estimated for a leader's contribution to local growth as the predictor for his promotion, we obtain finer results than those offered by the existing studies. In addition, our study sheds some new lights on the promotion tournament. First, we show both theoretically and empirically that shuffling is a way to make the leaders comparable across region and across time. In addition to testing leaders in different cities, shuffling makes cities connected so their leaders can be compared. As a result, the number of shuffling does not need to be large (recall that only 15% of the leaders in our long sample served in more than one city). Second, age is a pivotal factor determining a leader's chances of promotion. A person's chances diminish quickly as he gets older, but in the meantime, personal abilities become a stronger factor determining promotion and reach its peak around the median age of the sample. That is, for the average leader, personal abilities are important for his/her promotion.

References

- [1] Abowd, John M., Francis Kramarz, and David M. Margolis, (1999) "High Wage Workers and High Wage Firms," *Econometrica*, 67 , 251-333.
- [2] Abowd, John M., Robert H. Creedy, and Francis Kramarz, (2006) "Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data," US Census Bureau Technical Report TP-2002-06.
- [3] Abowd, John M., and Francis Kramarz, (2006) "Econometric Analyses of Linked Employer-Employee Data," *Labour Economics*, 6, 53-74.
- [4] Abowd, John M., Francis Kramarz, and Sebastien Roux, (2006) "Wages, Mobility and Firm Performance: Advantages and Insights from Using Matched Worker-Firm Data," *The Economic Journal*, 116, F245-F285.
- [5] Abowd, John M., Kevin McKinney, and Ian M. Schmutte, (2010) "How Important is Endogenous Mobility for Measuring Employer and Employee Heterogeneity?", Manuscript, School of Industrial and Labor Relations, Cornell University.
- [6] Acemoglu, Daron, Simon Johnson, and James A. Robinson, (2005) "Institutions as the Fundamental Cause of Long-Run Growth," in *Handbook of Economic Growth*, Philippe Aghion, and Steven N. Durlauf, eds., Amsterdam: North Holland).
- [7] Ai, Chunrong, and Edward C. Norton, (2003), "Interaction terms in logit and probit models," *Economic Letters*, 80, 123-129.
- [8] Bertrand, Marianne, and Antoinette Schoar, (2003) "Managing with Style: The Effect of Managers on Firm Policies," *Quarterly Journal of Economics*, 118 , 1169-1208.

- [9] Buchinsky, Moshe, Denis Fougere, Francis Kramarz, and Rusty Tchernis, (2010) "Interfirm mobility, wages and the returns to seniority and experience in the United States." *The Review of Economic Studies* 77(3), 972-1001.
- [10] Card, David, Jorg Heining, and Patrick Kline, (2013) "Workplace Heterogeneity and the Rise of West German Wage Inequality," *The Quarterly Journal of Economics*, 128(3), 967-1015.
- [11] Che, Jiahua, Yingyi Qian, and Barry R. Weingast, (2005) "Regional Decentralization and Fiscal Incentives: Federalism, Chinese Style," *Journal of Public Economics*, 89 , 1719-1742.
- [12] Chen, Binkai, and Yang Yao, (2011) "The Cursed Virtue: Government Infrastructural Investment and Household Consumption in Chinese Provinces," *Oxford Bulletin of Economics and Statistics*, 73 , 856-877.
- [13] Cornelissen, Thomas, "The Stata Command felsdvvreg to Fit a Linear Model with Two High-Dimensional Fixed Effects," (2008) *Stata Journal* 8, 170-189.
- [14] Easterly, William, and Steven Pennings. "How Much Do Leaders Explain Growth? An Exercise in Growth Accounting," (2014) *Working paper*.
- [15] Glaeser, Edward L., Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer, 2004, (2004) "Do Institutions Cause Growth?," *Journal of Economic Growth*, 9 , 271-303.
- [16] Graham, John, R, Si Li, and Jiaping Qiu, (2012) "Managerial Attributes and Executive Compensation", *The Review of Financial Studies*, 25 , 144-186.
- [17] Jin, Hehui, Yingyi Qian, Barry R. Weingast, (2005) "Regional decentralization and fiscal incentives: Federalism, Chinese style," *Journal of Public Economics*, 89 (2005), 1719-1742.
- [18] Jones, Benjamin F., and Benjamin A. Olken, (2005) "Do Leaders Matter? National Leadership and Growth since World War II," *Quarterly Journal of Economics*, 120 (2005), 835-864.
- [19] Jones, Benjamin F., and Benjamin A. Olken, (2009) "Hit or Miss? The Effect of Assassinations on Institutions and War," *American Economic Journal: Macroeconomics*, 1 , 55-87.
- [20] Li, Daokui, Linlin Liu, and Hongling Wang, (2009) "GDP zhong laodong fen'e yanbian de U xing guilu" [The U-Shape Law of the Labor Share in GDP], *Jingji yanjiu [Economic Research Journal]* (In Chinese), 13 (2009), 70-82.

- [21] Li, Hongbin, and Li-An Zhou, (2005) "Political Turnover and Economic Performance: The Incentive Role of Personnel Control in China," *Journal of Public Economics* 89 (2005), 1743-1762.
- [22] North, Douglass, and Robert P. Thomas, (1973) *The Rise of the Western World: A New Economic History*, Cambridge: Cambridge University Press.
- [23] Norton, C. Edwards, Hua Wang, and Chunrong Ai, (2004) "Computing interaction effects and standard errors in logit and probit models," *Stata Journal*, 4(2), 154-167.
- [24] Opper, Sonja, and Stefan Brehmm, (2007) "Networks versus Performance: Political Leadership Promotion in China," Manuscript, Department of Economics, Lund University.
- [25] Roodman, David, (2011) "Fitting fully observed recursive mixed-process models with cm-p," *Stata Journal*, 11(2), 159-206.
- [26] Shih, Victor, Christopher Aldolph, and Mingxing Liu, (2012) "Getting Ahead in the Communist Party: Explaining the Advancement of Central Committee Members in China.?" [*American Political Science Review*] 106(1), 166-187.
- [27] Wang, Xianbing, and Xianxiang Xu, (2008) "Difang guanyuan lai yuan, quxiang, renqi yu jingji zengzhang" [Bureaucratic Governance and Performance, Evidence from Chinese Provincial Leaders], *Guanli shijie* [*Management World*] (In Chinese), 2008-3, 16-26.
- [28] Wang, Xianbing, Xianxiang Xu, and Xun Li. (2009). "Difang guanyuan gengti yu jingji zengzhang" [Provincial Governors' Turnovers and Economic Growth: Evidence from China], *Jingjixue (jikan)* [*China Economic Quarterly*] (In Chinese), 8, 1301-1328.
- [29] Xu, Chenggang, (2011) "The Fundamental Institutions of China's Reform and Development," *Journal of Economic Literature*, 49 , 1076-151.
- [30] Xu, Xianxiang, Xianbing Wang, and Yuan Shu, (2007) "Difang guanyuan yu jingji zengzhang: laizi zhongguo shengzhang, shengwei shuji jiaoliu de zhengju" [Local Officials and Economic Growth], *Jingji Yanjiu* [*Economic Research Journal*] (In Chinese), 8 , 18-31.
- [31] Zhang, Jun, and Yuan Gao, (2008) "Term Limits and Rotation of Chinese Governors: Do They Matter to Economic Growth?," *Journal of the Asia Pacific Economy*, 13 , 274-297.

Appendices

A Likelihood Functions

A.1 Linear-Ordered Probit Two-Equation System

The unrestricted model is

$$\begin{cases} y_{j(i)t} = X_{j(i)t}\beta + \theta_i + \psi_j + \gamma_t + \epsilon_{it} \\ p_{it}^* = Z_{it}\delta + \alpha\theta_i + v_k + \eta_t + u_{it} \\ p_{it} = \begin{cases} -1 & p_{it}^* \leq c_1 \\ 0 & c_1 < p_{it}^* \leq c_2 \\ 1 & p_{it}^* > c_2 \end{cases} \end{cases}$$

where $p_{it} = -1$ denotes retirement, $p_{it} = 0$ denotes staying, and $p_{it} = 1$ denotes promotion.

The error term is specified as

$$\begin{pmatrix} \epsilon_{it} \\ u_{it} \end{pmatrix} = \mathcal{N}(0, \Sigma), \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{21} \\ \sigma_{12} & 1 \end{pmatrix}$$

Denote $\omega_{1ijt} = X_{j(i)t}\beta + \theta_i + \psi_j + \gamma_t$ and $\omega_{2ijt} = Z_{it}\delta + \alpha\theta_i + v_k + \eta_t$, we consider three exits:

For $p_{it} = -1$ (retirement),

$$\begin{aligned} L_{it} &= \phi(y_{ijt} - \omega_{1ijt}; \sigma_{11}) \int_{-\infty}^{-\omega_2 + c_1} \phi \left\{ u - \frac{\sigma_{21}}{\sigma_{11}} (y_{it} - \omega_{1ijt}); 1 - \frac{\sigma_{21}\sigma_{12}}{\sigma_{11}} \right\} d\epsilon \\ &= \phi(y_{ijt} - \omega_{1ijt}; \sigma_{11}) \Phi \left\{ c_1 - \omega_2 - \frac{\sigma_{21}}{\sigma_{11}} (y_{it} - \omega_{1ijt}); 1 - \frac{\sigma_{21}\sigma_{12}}{\sigma_{11}} \right\} \end{aligned}$$

For $p_{it} = 0$ (staying at the same position or level),

$$\begin{aligned} L_{it} &= \phi(y_{ijt} - \omega_{1ijt}; \sigma_{11}) \int_{-\omega_2 + c_1}^{-\omega_2 + c_2} \phi \left\{ u - \frac{\sigma_{21}}{\sigma_{11}} (y_{it} - \omega_{1ijt}); 1 - \frac{\sigma_{21}\sigma_{12}}{\sigma_{11}} \right\} d\epsilon \\ &= \phi(y_{ijt} - \omega_{1ijt}; \sigma_{11}) \left[\Phi \left\{ c_2 - \omega_2 - \frac{\sigma_{21}}{\sigma_{11}} (y_{it} - \omega_{1ijt}); 1 - \frac{\sigma_{21}\sigma_{12}}{\sigma_{11}} \right\} \right. \\ &\quad \left. - \Phi \left\{ c_1 - \omega_2 - \frac{\sigma_{21}}{\sigma_{11}} (y_{it} - \omega_{1ijt}); 1 - \frac{\sigma_{21}\sigma_{12}}{\sigma_{11}} \right\} \right] \end{aligned}$$

For $p_{it} = 1$ (promotion),

$$\begin{aligned} L_{it} &= \phi(y_{ijt} - \omega_{1ijt}; \sigma_{11}) \int_{-\omega_2 + c_2}^{+\infty} \phi \left\{ u - \frac{\sigma_{21}}{\sigma_{11}} (y_{it} - \omega_{1ijt}); 1 - \frac{\sigma_{21}\sigma_{12}}{\sigma_{11}} \right\} d\epsilon \\ &= \phi(y_{ijt} - \omega_{1ijt}; \sigma_{11}) \left[1 - \Phi \left\{ c_2 - \omega_2 - \frac{\sigma_{21}}{\sigma_{11}} (y_{it} - \omega_{1ijt}); 1 - \frac{\sigma_{21}\sigma_{12}}{\sigma_{11}} \right\} \right] \end{aligned}$$

For p_{it} = missing,

$$L_{it} = \phi(y_{ijt} - \omega_{1ijt}; \sigma_{11})$$

To sum up, the log-likelihood function is

$$\ln L = \sum_{p_{it}=-1} \ln \left\{ \phi(y_{ijt} - \omega_{1ijt}; \sigma_{11}) \Phi \left\{ c_1 - \omega_2 - \frac{\sigma_{21}}{\sigma_{11}} (y_{it} - \omega_{1ijt}); 1 - \frac{\sigma_{21}\sigma_{12}}{\sigma_{11}} \right\} \right\}$$

$$\begin{aligned}
& + \sum_{p_{it}=0} \ln \left\{ \phi(y_{ijt} - \omega_{1ijt}; \sigma_{11}) \left[\Phi \left\{ c_2 - \omega_2 - \frac{\sigma_{21}}{\sigma_{11}} (y_{it} - \omega_{1ijt}); 1 - \frac{\sigma_{21}\sigma_{12}}{\sigma_{11}} \right\} \right. \right. \\
& \quad \left. \left. - \Phi \left\{ c_1 - \omega_2 - \frac{\sigma_{21}}{\sigma_{11}} (y_{it} - \omega_{1ijt}); 1 - \frac{\sigma_{21}\sigma_{12}}{\sigma_{11}} \right\} \right] \right\} \\
& + \sum_{p_{it}=1} \ln \left\{ \phi(y_{ijt} - \omega_{1ijt}; \sigma_{11}) \left[1 - \Phi \left\{ c_2 - \omega_2 - \frac{\sigma_{21}}{\sigma_{11}} (y_{it} - \omega_{1ijt}); 1 - \frac{\sigma_{21}\sigma_{12}}{\sigma_{11}} \right\} \right] \right\} \\
& + \sum_{p_{it}=\text{missing}} \ln \phi(y_{ijt} - \omega_{1ijt}; \sigma_{11})
\end{aligned}$$

The likelihood function for the restricted model can be deduced similarly, and not presented here.

B Iterative Computation in Estimation

In joint estimation, the growth and promotion equations are estimated simultaneously. However, the leader effect in the promotion equation comes from estimates of the growth equation. When we estimate two equations simultaneously, we do not have the estimates of the growth equation, making estimation implausible.

A direct solution is using dummies of leaders and adding constraints across equation to ensure that the coefficients of dummies are proportional to one another. By this way, the model is ²³

$$\begin{cases} y_{ijt} = X_{ijt}\beta + \Sigma\tau_i d_i + \psi_j + \gamma_t + \epsilon_{ijt} \\ p_{it} = Z_{it}\delta + \Sigma\phi_i d_i + \text{Prov}_{it} + \eta_t + u_{it} \end{cases}$$

s.t. $\frac{\tau_i}{\phi_i} = a$ for all i .

However, as the constraints are nonlinear, coding and convergence are both difficult.

In this paper, we avoid the nonlinear constraints using an iterative method of estimation, executed with following steps.

1. Estimate the growth equation alone in the unrestricted model, save $\hat{\theta}_i$, denoted as $\hat{\theta}_i^0$.
2. Substitute θ_i in the promotion equation with $\hat{\theta}_i^{n=0}$, estimate

$$\begin{cases} y_{ijt} = X_{ijt}\beta + \theta_i + \psi_j + \gamma_t + \epsilon_{ijt} \\ p_{it} = Z_{it}\delta + \alpha\hat{\theta}_i^{n=0} + \text{Prov}_{it} + \eta_t + u_{it} \end{cases}$$

and save $\hat{\theta}_i$, denoted as $\hat{\theta}_i^{n=1}$.

3. Calculate $|\hat{\theta}_i^n - \hat{\theta}_i^{n-1}|$ and
 - If $\max_i |\hat{\theta}_i^n - \hat{\theta}_i^{n-1}| \leq \text{tol}$, convergence achieved.

²³Using L-L model as an example.

- If $\max_i |\hat{\theta}_i^n - \hat{\theta}_i^{n-1}| > \text{tol}$, substitute θ_i in the promotion equation with $\hat{\theta}_i^n$ and start from Step 2 once more.

4. Estimate a restricted model and conduct the LR test.