Provincial and Local Governments in China: Fiscal Institutions and Government Behavior

Roger H. Gordon, *University of California, San Diego*[*]
Wei Li, *Darden School of Business*

October, 2009

Provincial and local (hereafter "local") governments in China play an important role in the overall economy. To begin with, local tax revenue in recent years is around 7% of GDP, extra-budgetary revenues (largely income from land) come to another 4% of GDP, and expenditures are close to 13% of GDP. This scale of activity is broadly comparable to that of state and local governments in the U.S., where own tax revenue is around 13% of GDP and expenditures are 16% of GDP. Local governments in both settings have primary responsibility for education, local infrastructure, and local public services.

The similarity largely stops there, however. In the U.S. local governments are mainly financed by a property tax and user fees, and state governments by a combination of personal income taxes and sales taxes. In China, until 1994 local governments were mainly financed through a tax on the profits and sales of non-state firms. Since then they receive a fraction of the VAT and corporate profits tax collected in their jurisdiction, and all the revenue from personal income taxes, business taxes, and (until recently) taxes on agriculture.

Chinese local governments also play a much more central role in the local economy than do local governments in the U.S., controlling the allocation of land and in the past exercising substantial controls over the allocation of bank credit. The initial growth in China at the beginning of the economic reforms in fact is largely

[*]This paper was prepared for an NBER Conference on "Capitalizing China," to be held during December 15-16 at the Chinese University of Hong Kong. We would like to thank Hua Li and Alex Gelber for comments on an earlier draft of this paper.

attributed to the initiative of local governments in setting up TVE's and other non-state firms.

Oversight over local government in China is also far different than in the U.S. In the U.S., oversight occurs through both voice and exit. Local officials are elected by residents, so can be removed from office if residents are not satisfied with outcomes. In addition, residents can vote with their feet and leave a poorly performing district, putting pressure on local officials to keep current residents satisfied. In China, in contrast, these officials are appointed by higher levels of government, and their promotion again depends on the judgment of higher officials. The pressure to prevent people from leaving is also much weaker, since Chinese are subject to the hukou system, at least on paper tying them to their current location. In addition, farmers are tied to their land, since there is not a well functioning market for selling or leasing this land, due to their lack of legal ownership of the land.

The objective of this paper is to examine the incentives faced by local governments in China, given that they face oversight neither from voice nor from exit of current residents. Incentives instead must come through officials' dependence on particular sources of finance, through their controls over the allocation of land and (in the past) credit, and through the preferences of higher officials who control their job retention and promotion. What economic choices would local officials then be expected to make, given the incentives they face? What behavior do we see? When incentives have changed over time, do we see the expected changes in behavior? What are the key sources of inefficiency in the allocation of resources by local governments, given these incentives? What institutional changes might help to alleviate these misallocations?

The organization of the paper is as follows. Section 1 summarizes briefly the role of voice and exit in affecting the incentives faced by local officials. Section 2 develops an alternative model in which the incentives faced by officials depends on the tax revenue they receive minus whatever they need to spend on local public services. We develop this model using the institutions that have existed in China during the reform period. In section 3, we then examine how these incentives changed over the course of the reform period, as the tax law changed and as market reforms were introduced. Section 4 then examines the additional incentives faced by officials due to their implications for the likelihood an official will be promoted or at least retained in their position. Section 5 considers a range of policy reforms aimed at improving the incentives faced by local officials, while sectioni 6 provides a brief summary.

1 Traditional models of government oversight

We begin with a brief summary of the U.S. literature on the forms of oversight of local officials, and why on paper we would then expect to see officials making choices that are largely in the best interests of residents.

One source of oversight is the election process. Each potential official proposes a platform to voters. Voters choose that candidate whose proposed platform provides them the highest utility. If candidates simply care about being elected and voters are homogeneous, then in equilibrium each candidate's platform will maximize the utility of voters. Inefficiencies can arise due to differences in the preferences of the median voter compared with the overall costs vs. benefits of a project, as emphasized by Buchanan and Tullock (1962). Candidates may come in with their own preferences, or can be influenced by special interests, and are not obliged to follow through on their campaign promises. Voters also face a free rider problem, having no personal incentive to vote or to be informed about the candidates. The quality of oversight through the voting process is therefore uncertain on net. Banerjee and Duflo (2006), for example, find in India that voters provide surprisingly poor oversight over officials.

More central to the literature on fiscal federalism is the (Tiebout, 1956) model. Under this model, officials propose a tax structure and spending package, and gain utility from any tax revenue left after financing promised expenditures. Residents then choose where to live, with land prices adjusting to generate an equilibrium residential allocation. Competition among communities forces down the net profits available to officials, induces them to provide the package of public services residents are willing to pay for, and forces them to finance these expenditures with user fees (or head taxes if all residents benefit equally from the spending) and to provide the services at minimum cost. This competition is most intense if residents are costlessly mobile, if they can carry their income with them, and if there are many competing communities. In equilibrium, when competition is intense, outcomes should be efficient.

Officials in China, though, are not subject to either voting pressures or much pressure from the mobility of potential residents, given the hukou system. In order to understand the incentives faced by these officials, we need to turn elsewhere.

2 Incentives created by the source of tax revenue: general model

In this section, we focus on how the available sources of revenue affect the incentives faced by Chinese officials. Our set-up follows the structure of the Tiebout model in that the utility of officials depends on their net "fiscal profits": tax revenue, profits from firms owned by the local government, plus income generated by land rents minus expenditures on public services.

We begin by laying out a general model with the following stylized institutional features. With the hukou system in place we assume that labor is not mobile across jurisdictions (Wang and Zuo, 1999), but can freely move between jobs within the jurisdiction. Due to capital controls, we assume that the capital market in China is segmented (Gordon and Li, 2003): foreign capital can enter and exit with relative ease, but domestic capital is landlocked. We further assume for now that domestic capital does not move freely across regions within China. Land is owned by the government, which allocates use rights for a finite number of years. Farmers have use rights for their plots. In the next section, we then relate these more general assumptions to the particular institutions that existed during various time periods under the reforms in China.

Firms can be privately owned, local-government owned, or subsidiaries of foreign multinationals.[1] Privately-owned firms in each industry i located in the jurisdiction face a tax at rate τ_i on their profits, denoted π_i, an excise tax at rate s_i on their sales, plus an implicit tax denoted by κ_i on their capital.[2] Here, $\pi_i = p_i(1-s_i)Q^i - wL_i - (r+\kappa_i)K_i - n_iA_i - u_iG$, where p_i is the output price (which the jurisdiction takes as given), $Q^i = Q^i(L_i, K_i, A_i; G, R)$ is local output produced in industry i using effective labor (L_i), domestic capital (K_i), and land (A_i), with local infrastructure G aiding production and local regulations R affecting productivity. Effective labor is measured in efficiency units: If an unskilled worker's efficiency unit of labor supply is normalized to 1 and industry i hires skilled workers, we must have $L_i > N_i$, where N_i is the number of workers hired. Here, w is the local wage rate for effective labor, r is the interest rate charged by banks on loans to the firm (set nationally), and n_i is the implicit rent the government charges industry for use of land, while u_i is a fee (if any) charged for use of G which at most equals the marginal product of G. Depending on the time period, officials may control the allocation of domestic

[1] In this analysis, we ignore state-owned firms, since allocations to these firms are largely controlled by the national government.

[2] These taxes comes in part from the fraction of capital not financed with debt, and also from use of depreciation allowances rather than expensing for new investment.

capital across local firms through their oversight of the local banks. Not only do tax rates differ by industry but they also may differ by size of firm, with the national government at times receiving all of the revenue from state-owned firms, but local governments at times receiving all of the tax revenue from both private firms and firms set up by the local government.

For government-owned firms, the government receives not only the above tax revenue from the firms but also the after-tax profits, $(1-\tau)\pi_i$. In total, it therefore simply receives the entire pre-tax profits.

For subsidiaries of multinationals, various things change. First, the tax rates differ. We can capture this difference in tax rates by adding a superscript $*$ to each tax rate. In addition, foreign firms are not dependent on the local government for capital, though they do depend on it for land. The opportunity cost of capital for them also differs. We denote their opportunity cost by r^*. An additional consideration, though, is that foreign subsidiaries are mobile, whereas local entrepreneurs are not due to the hukou system. Jurisdictions then compete with each other to attract foreign subsidiaries. If this competition is sufficiently intense, then at the equilibrium agreement between a jurisdiction and a foreign subsidiary, the jurisdiction will just break even from attracting the subsidiary, and no other jurisdiction can profitably offer the subsidiary a more attractive arrangement, leading to Pareto efficient outcomes. At the margin officials then act as if they care about the profits of the subsidiary before any local taxes (though after any national taxes).

Depending on the time period, officials may also receive revenue from agriculture, both through explicit taxes and through requiring farmers to sell output to the government at a below-market price. Assume for simplicity that the tax revenue from agriculture equals $\sigma p_f F$, where p_f is the market price for agricultural output and F is the quantity produced, with $F = F(L_f, K_f, A_f; G_f, R_f)$. Here, G_f is another set of public services aimed at agriculture, provided at a user charge of u_f, while R_f are regulations affecting agriculture.

Farmers have a use right for an area of land A_f^0 without paying explicit rent. If officials reallocate some of this land for industrial or residential uses, they must compensate farmers by paying them the marginal product of land used in agriculture, an amount we denote by $c \equiv (1-\sigma)p_f F_A$, where $F_A = \partial F/\partial A_f$ is the marginal product of land.[3] Similarly, the wage rate firms must pay to attract local workers satisfies $w \equiv (1-\sigma)p_f F_L$. To simplify the subsequent notation, assume that all units of output are redefined so that $p_i = p_f = 1$.

The government also provides services to each household, G_h, which the household in part pays for through a user fee u_h. Let n_h denote the rent received per unit

[3]Throughout, we use subscripts of a function to denote partial derivatives.

of land $A_h$allocated to housing. In general, n_h equals residents' willingness to pay for use of a unit of residential land, $n_h = q(A_h, \sum_i L, w, G_h)$, where q is decreasing in A_h, increasing in the number of workers and their wage rate, and increasing in G_h.

For simplicity, we start by assuming a fixed total supply of each factor to the jurisdiction, e.g. $\sum_i A_i + A_f + A_h = A^T$, where the superscript T signifies the total amount of a factor available in the jurisdiction. Given the lack of mobility, total factor supplies are clearly fixed for labor and land. For the moment, we assume that the supply of domestic capital is fixed as well, based on the deposits under the control of local banks. We also assume that factors are fully employed, so for any given allocation of factors to industries, the agricultural output is simply, $F(L^T - \sum_i L_i, K^T - \sum_i K_i, A^T - \sum_i A_i - A_h; G_f, R_f)$.

Assume that firms with $i \in I_P$ are privately owned, firms with $j \in I_G$ are owned by the government, while those with $k \in I_F$ are foreign owned. The objective of officials is to maximize the sum of tax revenue from privately-owned firms, profits from government-owned and foreign-owned firms, agricultural taxes, and land rents, minus compensation to farmers and minus the net cost (net of user fees) of public expenditures on local infrastructure and minus the effort expended on regulations:

$$\begin{aligned}\sum_{i\in I_P}(\tau_i\pi_i + \kappa_i K_i + s_i Q^i + n_i A_i) + \sum_{j\in I_G}(Q^j - wL_j - rK_j) \\ + \sum_{k\in I_F}(Q^k - wL_k - r^* K_k) + \sigma F + qA_h - c(A_f^0 - A_f) \\ - G(1 - \sum_{i\in I_P} u_i) - G_f(1-u_f) - N^T G_h(1-u_h) - e(R) - e_f(R_f) \quad (1)\end{aligned}$$

Here, N^T is the size of the local population. Officials then allocate land and capital and choose how much to spend on each form of public services to maximize expression (1). Local wage rates and labor allocation are determined by the local labor market. While we assume that officials must take output prices as given, they will take into account the effects of their actions on the local wage rate, labor allocations, and equilibrium rents in each sector.

Consider first the allocation of domestic capital to government-owned firms in industry j. The first order condition satisfies

$$Q_K^j = r + \sigma F_K - (\frac{s_j}{1-s_j} - \frac{\sigma}{1-\sigma}) w dL_j - (1-\sigma)(F_{AK} + F_{AL} dL_j)(A_f^0 - A_f) - q_L A_h dL_j, \quad (2)$$

where the subscripts in Q^j and F denote first order and second order partial derivatives with respect to capital, labor and/or land.[4] Given the local resource constraint, the extra capital allocated to industries has to be taken from agriculture. Now r is paid by the government rather than by farmers, introducing one cost. The next term reflects the foregone tax revenue from agriculture due to the drop in capital there. The third term captures the net change in tax revenue due to the market-induced reallocation of labor, dL_j, from agriculture to industry in response to the shift in capital, where $dL_j = -Q_{LK}/Q_{LL}$. Tax revenue falls in agriculture, but also rises in the government-owned firms on our assumption that the firm hires workers until $(1 - s_j)Q^j_L = w$. The fourth term measures the gain due to the fall in land values in agriculture from the drop in use of capital and labor there, resulting in less compensation being paid to farmers for any land shifted out of agriculture to industry. The final term measures the increase in rents paid for residential land in response to the increased demand from the extra industrial workers. On net, we conclude that $Q^j_K < F_K$, implying too much investment in industry than in agriculture on efficiency grounds, as long as r is small, $\sigma < 1$ and $s_j \geq \sigma$, as we expect.

Consider next the allocation of domestic capital to private firms. This first-order condition satisfies

$$Q^i_K = \frac{1}{T_i}\{\tau_i r - (1-\tau_i)\kappa_i - (1-T_i)Q^i_{AK}A_i + \sigma F_K - (\frac{s_i}{1-s_i} - \frac{\sigma}{1-\sigma})wdL_i \\ - (1-\sigma)(F_{AK} + F_{AL}dL_i)(A^0_f - A_f) - q_L A_h dL_i\} \quad (3)$$

There are a variety of differences here compared to the condition for government-owned firms. First, the government receives only the fraction $T_i = s_i + \tau_i(1 - s_i)$ of the resulting extra profits. Offsetting this, it bears only the fraction τ_i of the interest charges paid by the firm for the capital, and also receives extra tax revenue from the implicit taxes on the firm's capital. After-tax rental payments by the firm, $(1 - \tau_i)n_iA_i = (1 - T_i)Q^i_AA_i$, also increase to the degree that land and capital are complements in production. The remaining terms carry over unchanged. As long as the firm can reject the capital if it loses as a result, then the benefits from allocating capital to a private firm are less than from allocating it to a government-owned firm, while the opportunity costs are the same. The equilibrium marginal product of capital in private firms will then be above that in government-owned firms, implying that private firms receive less investment. The marginal product of

[4] In general, wage rates can change, leading to addition terms. However, if $Q_{LK}/Q_{LL} = F_{LK}/F_{LL}$ for all industries, as we assume for simplicity, then wage rates are unaffected by reallocations of capital across industries.

capital will be lower in private firms than in agriculture, however, as long as r is small, and $s_j \geq \sigma$, as we expect.

Note that investment in either private firms or agriculture increases as tax rates on the sector increase. With a higher tax rate, the benefits to the government of investment in the sector rise, leading to additional investment. This counter-intuitive result arises because the government is making allocation decisions based on the implications for tax revenue, rather than having firms make the decision based on implications for after-tax profits.

Turn now to the allocation of land for industrial and residential uses. The first order condition for land allocated to government-owned firms or foreign subsidiaries is

$$Q_A^j = F_A - (\frac{s_j}{1-s_j} - \frac{\sigma}{1-\sigma})wdL_j - q_L A_h dL_j - (1-\sigma)(F_{AA} + F_{AL}dL_j)(A_f^0 - A_f) \quad (4)$$

By shifting an extra unit of agricultural land to industrial use, the government pays $(1-\sigma)F_A$ to farmers as compensation and bears a fall in agricultural revenue by an amount σF_A, for a combined opportunity cost of F_A. In addition, since less land in agriculture raises its marginal product in agriculture, the government needs to provide more compensation to farmers by an amount $-(1-\sigma)F_{AA}(A_f^0 - A_f)$. The resulting movement of labor from agriculture to industry raises tax revenue directly, raises residential rents, and lowers the value of agricultural land, all providing additional benefits. In general, it is unclear whether the equilibrium marginal return to land in industry is higher or lower than that in agriculture.

One consideration we have assumed away in these expressions is the impact of the land reallocation on equilibrium wage rates. However, land in industry can potentially support many more jobs than the same amount of land in agriculture. The reallocation then can be expected to generate an increase in market-clearing wage rates. From the perspective of officials, this increase in wage rates is an added cost to consider when reallocating land from agriculture to industry.

Consider next the first-order condition that arises when officials consider reallocating land from government-owned firms to private firms. Here, we find that $Q_A^i + (1-T_i)Q_{AA}^i A_i = Q_A^j$. The government, being a monopoly supplier of land restricts land allocations to private firms in order to drive up rents.

For residential land, the government also acts as a monopoly supplier. It compares the marginal revenue it receives from rents to the same types of terms as above measuring the opportunity cost of the land.

One implicit assumption in the above derivation is that officials compare the flows of rent in each use. In fact, they need to pay a lump-sum compensation to

farmers reflecting the present value of the land in agriculture when land is taken from agriculture. If the trade-off that officials face is between this lump-sum payment to farmers and an increased flow of rents from industry during the limited time period the official remains in power, they would favor leaving land in agriculture. The land use policies and practice since the late 1990s have instead been that officials sell rights to the land when it is reallocated to industry. Officials therefore compare present values. As long as the discount rate is the same in each market, the above results then carry through unchanged, except that the respective rents are divided by a discount rate, giving them much more weight. Farmers face a harder time acquiring funds, though, since farmland cannot be used as collateral, unlike industrial or residential land, so their discount rate should be higher, leading to a further factor favoring a reallocation of land from agriculture to industry.

What about the choice of expenditures on public services? The choices made by government officials, based on their own self-interest, would be efficient only if local firms and individuals together are left unaffected on net by a marginal change in G.[5] Any benefits to government-owned firms already go in their entirety to the government, as do all benefits to foreign subsidiaries through adjustments in the amount the government needs to pay them to attract them to the jurisdiction. The remaining net benefits equal zero, leading to efficient choice, only if

$$\sum_{i \in I_p} [(1 - T_i)(Q^i_G - Q^i_{AG} A_i) - (1 - \tau_i) L_i dw_G] + (L^T - L_f) dw_G$$
$$- (1 - \sigma) F_{AL} dL_G (A^0_f - A_f) - q_L A_h dL_G = \sum_{i \in I_p} (1 - \tau_i) u_i. \quad (5)$$

Here dw_G and dL_G are the marginal increase in industrial wage rates in the local labor market and the associated reallocation of agricultural labor into industry ($dw_G = -(1 - \sigma) F_{LL} dL_G$) due to the marginal increase in G, which, we presume, raises the marginal product of industrial labor.

To achieve an efficient result, user fees must fully reflect the direct benefits to private firms minus any losses they incur due to changes in either commercial rents or wages. Changes in the price of goods and factors traded among residents provides no net benefits to residents as a group. However, any net increase in wage rates due to the provision of better services aids the private sector as a whole, while any net

[5] This condition on marginal incentives is sufficient if the second-order conditions are satisfied, e.g. decreasing returns to scale. If there are increasing returns to scale, however, then the marginal conditions do not assure that the project as a whole will be attractive to officials. They will ignore any consumer surplus, generated by the project as a whole, so tend to underinvest.

decrease in the compensation being paid to farmers for land shifted out of agriculture to industry and any increase in residential rents hurt the private sector as a whole. At least for roads with tolls, a firm makes use of these roads to the point where $(1-T_i)Q^i_G = (1-\tau_i)u_i$. Efficiency then requires that the remaining terms on the left hand side of Equation (5) equal zero. Among the remaining terms, the only one reflecting a net benefit to the private sector is the increase in wage rates. If wage rates hardly change, due to the overhang of surplus labor in agriculture, then the private sector loses at the margin from more public services. Governments therefore have too strong an incentive to provide these services, and would be expected to provide subsidies to the private firms that undertake these infrastructure investments.

Similarly, the choice of G_f yields an efficient outcome if the private sector is left indifferent at the margin to any marginal change in provision of public services. This condition holds if

$$(1-\sigma)F_G + (1-\sigma)(F_{AG} - F_{AL}dL_{G_f})(A^0_f - A_f) - \sum_{i\in I_P}(1-T_i)Q^i_{AL}A_i dL^i_{G_f}$$
$$- q_L A_h dL_{G_f} + [L^T - L_f - \sum_{i\in I_P}(1-\tau_i)L_i]dw_{G_f} = u_f. \quad (6)$$

Here dw_{G_f} is the marginal increase in industrial wage rates in the local labor market due to the marginal increase in G_f, which, we presume, raises the marginal product of agricultural labor and causes a marginal reallocation of industrial workers back into agriculture. The marginal reallocation of labor from industry back to agriculture,[6] $-dL_{G_f} > 0$, satisfies $dw_{G_f} = (1-\sigma)(F_{LG_f} - F_{LL}dL_{G_f})$.

In order for incentives on government officials to be efficient, user fees must fully reflect the net-of-tax benefits to farmers from extra public services to agriculture, minus any net benefits farmers receive through increased compensation for land transferred out of agriculture but also plus the net benefits the non-agricultural sector receives from lower rents on commercial and residential property due to a migration of people back to agriculture. If farmers make use of public services until marginal benefits and marginal costs are equal, so that $(1-\sigma)F_G = u_f$, then efficiency again requires that the remaining terms on the right hand side of Equation (6) sum to zero. All of these terms reflect a net benefit to the private sector except for the change in wage rates. If wage rates hardly change due to surplus labor in agriculture, then the private sector benefits on net from additional expenditures on public services to agriculture. The government then provides too few such services.

[6] The term $-dL^i_{G_f}$ in equation (6) denotes the marginal reallocation of labor from agriculture to industry i.

Expenditures on G_H are efficient only if the dollar benefits per household equal their required user fee. If use of the public services is voluntary, e.g. tolls on a road, then individuals make use of the service until the marginal benefit equals the user fee, yielding efficient incentives on government officials.

Finally, what can we say about regulatory policies? Again, decisions by government officials are efficient only if the private sector is left indifferent at the margin to any changes in regulation. As with public services, the private sector benefits from any increase in after-tax profits, and is affected by any changes in land rents that arise (directly or indirectly) in response to these extra profits. With no extra user fees, though, there is no offsetting price that can adjust so that the private sector can be left indifferent on net. As a result, officials face inadequate incentives to put effort into industrial regulations benefiting private firms, though they would face efficient incentives if there were separate policies for government-owned firms. For similar reasons, there are inadequate incentives to regulate agriculture well.

While local governments in China control the allocation of land and did in past years control the allocation of capital, they don't control the allocation of labor. From their perspective, too much labor ends up migrating to lightly-taxed industries away from more heavily-taxed industries. To that extent, production is excessive in these more lightly-taxed industries. As a result, local governments can potentially gain through making use of any further instruments to shift production from lightly-taxed to heavily-taxed industries. One such instrument is controls over trade between their jurisdiction and the rest of China. In particular, each local government has an incentive to restrict imports in heavily taxed industries and restrict exports in the most lightly-taxed industries. By shifting the composition of local production towards goods that are more heavily taxed, government revenue in the jurisdiction increases. Largely, this increase comes at the expense of government revenue in other jurisdictions, who lose export markets for their most highly taxed goods and have a harder time buying elsewhere the most lightly-taxed commodities. National prices then fall for the most heavily taxed goods, and rise for the more lightly-taxed goods, weakening any further incentive to intervene to restrict trade. These negative fiscal externalities result in an inefficient choice of government policies from the joint perspective of local governments, providing a motivation for the national government to intervene to lessen these trade distortions. Consistent with these forecasts, Young (2000) reports evidence that local governments restricted trade patterns, leading to too many firms of too small scale in the heavily-taxed industries.

3 Application of model to different time periods

We next use this general model to forecast the behavior of government officials during particular sub-periods under the reforms in China, and examine how behavior should have changed over time. .

3.1 1979-1994

Institutions in fact were changing frequently during this time period, but our sense is that the following description fits this period reasonably well.

To begin with, patterns of resource allocation in 1979 were largely fashioned by years of intermittent central planning punctured by political turmoil (Li, 2008). Under central planning, heavy industries had been favored at the expense of agriculture and industries that catered to consumer demands. The government collected revenue not with explicit taxes but by directly accessing state-owned firms' profits and by taking grain from peasants and leaving them just enough for subsistence. To economize on the cost of revenue collection, the government used price scissors to channel profits to a few industries located in large cities. Agricultural goods were priced the lowest, followed by raw materials, energy, industrial goods, consumer necessities and then consumer durables. To capture price scissors, we assume

$$p_1^0 \geq p_2^0 \geq ... \geq p_I^0 \geq p_f^0 \equiv 1, \tag{7}$$

so the lower numbered industries are higher profit-margin consumer durables industries. Here the total number of industries is I and the superscript 0 denotes planned prices. In order to maintain consistency with the structure of the above model, we assume that these price distortions were implicitly implemented with differential sales taxes, with the highest tax rate on consumer durables and the lowest on agriculture. In response to these price distortions, the implicit factor prices would have adjusted to favor those factors used in agriculture, presumably land and labor, and penalize those factors primarily used in the more heavily taxed industries, presumably capital. The national government then used its control over the allocation of factors to produce those goods demanded at these prices, limiting allocations to the highly-taxed industries.[7]

Since the use of household responsibility contracts in agriculture in the late 1970's generated dramatic improvements in agricultural productivity (Lin, 1990), the government started to experiment with offering managerial responsibility contracts in state-owned enterprises using financial incentives in the early 1980's. In

[7]Some consumer goods were rationed, though, limiting consumer demand by fiat.

part to make measurements of financial performance more informative, the government implemented the "taxes-for-profits" reform between 1982 and 1984 aimed at roughly equalizing after-tax profit margins across industries. The reform introduced excise taxes and profit taxes (s_i, τ_i and κ_i) with provisions for levying additional adjustment profit taxes and differential treatment of depreciation allowance and interest expense deduction so that the after-tax profit margins were roughly equalized across industries at the existing allocations.

Control rights and cash flow rights of small and medium-sided SOE's and even some large SOE's were officially transferred to local governments, in addition to new firms that they set up or sponsored. Tax collection was also officially delegated to local governments, which in turn shared the revenues with the central government using responsibility contracts. The local government not only received the tax payments but also controlled use of the remaining after-tax profits. In part due to the connivance of local officials, locally controlled firms paid very little in taxes to the national government at the margin. As a result, the objective of officials is to maximize the sum of tax revenue, land rents, and after-tax profits, minus the cost of public expenditures on local infrastructure and minus the effort expended on regulations:[8]

$$\sum_i (p_i^0 Q^i - wL_i - rK_i) + \sigma F - G - G_f - G_h - e(R) - e_f(R_f) \tag{8}$$

This objective is a special case of equation (1), but with no private firms and no control over the allocation of land for residential housing. The key point, though, is that the pre-existing allocation was designed to produce the goods demanded at the distorted prices. With the reforms, local governments faced undistorted incentives but prices that differed sharply from marginal costs. Local officials had effective control over the allocation of existing bank credit among different firms. Labor contracts inherited from the period of central planning, however, set industrial wage rates that exceeded the opportunity cost of labor in agriculture. What developed instead was a dual-labor market, with new "contract" workers being hired at a different (lower) wage rate than was received by existing workers. In addition, we assume that so little land was yet in use in industry that $A_f^0 \approx A_f$.

What do these conditions imply for the allocation decisions? The first order

[8] At least initially, user fees were unusual.

conditions with respect to capital, labor and land satify

$$Q_K^j = (r + \sigma F_K)/p_j^0 \tag{9}$$

$$Q_L^j = (w + \sigma F_L)/p_j^0 \tag{10}$$

$$Q_A^j = \sigma F_A/p_j^0 \tag{11}$$

Conditional on these initial prices, production should have been efficient. However, the initial prices were not market-clearing prices, leading to surplus output in the industries with artificially high prices and shortages in the industries with artificially low prices. The resulting competition among local governments to gain market shares in high-margin industries led to over-capacity and inefficiently small scales in those industries, exposing the incompatibility of pricing under the plan with decentralized decision-making.

These growing surpluses and shortages, and the resulting interventions by local governments to preserve their market share in the overpriced goods, quickly forced the national government to introduce a dual-track pricing system, whereby a fixed quantity, rationed among firms, must be sold at the original prices and all further output must be sold at market prices (Li, 1999). With undistorted incentives on local governments and market prices for all marginal transactions, allocation decisions within industry should indeed have been efficient. Empirical studies by Gordon and Li (1995), Groves et al. (1994) and Li (1997) confirm the efficiency enhancing impact of the reform in the 1980's. It was also documented in Li (1997) that between 1980 and 1989 more investment did flow to industries that had higher combined taxes and (after-tax) profits per yuan of sales and that product market competition among enterprises did bring about marked improvements in total factor productivity.

However, given that $\sigma < 1$, there would be too little land allocated to agriculture. With $r \approx 0$, there would also be too little capital investment in agriculture, as is apparent in the data given the lack of any farm machinery investment during this period.

In spite of an unclear legal environment, some private firms did enter during this period. However, since local governments could keep only the tax revenue from private firms, there would be underinvestment and potentially even no land allocated to private firms.

What about public expenditures? First, without (much of) a private sector and no land rents, there are no marginal effects on the private sector of any changes in public services to industry, so that government incentives lead to efficient outcomes as long as changes in wage rates can be ignored. Agriculture would benefit directly from extra services, but without user fees officials have no reason to take these

benefits into account, leading to inadequate incentives. Without user fees, there are no incentives to provide services to households.

Finally, what about regulatory policies? With full control over the entire return to improvements in industrial productivity, officials should have invested the efficient level of effort in designing effective regulations for industry. Sharing less in productivity gains in agriculture, they would have invested less effort there.

The reforms starting in the early 1980's offered local officials strong incentives for industrial development. These incentives encouraged officials to pour resources into sectors that the planners had previously restricted. Relative prices started to change, undoing price scissors and moderating the incentives to invest in high-margin industries. Given the direct link between price scissors and tax rates, the introduction of the dual-track system was tantamount to a tax cut at the margin. While government revenue continued to grow, it fell relative to GDP during this period. The central government's share of revenue also declined sharply during this period.

The fiscal decentralization undermined macroeconomic stability of the economy. During the period, local officials had effective control over the allocation of bank credit among different firms. In principle, they controlled only the balances available in the local banks, though in practice governments often were able to lend more funds than they had available, with the national government covering the short-fall. The decentralized money creation led to inflation, forcing the national government to freeze bank lending on occasion, resulting in boom and bust cycles during the period (Brandt and Zhu, 2000).

3.2 Post-1994

In response to the distortions introduced by the earlier reforms, the Chinese government implemented extensive economic reforms around 1994, with many further gradual changes since then. Our stylized summary of the institutions since 1994 are as follows:

The dual track system was phased out by the mid-1990s. Planned prices were largely eliminated, and market prices dominated markets for goods. This shift in infra-marginal rents, though, didn't change marginal incentives so shouldn't have affected market allocations.

The formal tax structure changed dramatically in 1994. To begin with, the national government separated national and local tax administration. Excise taxes with rates that varied by industry were replaced by: 1) VAT on mining and manu-

facturing industries,[9] administered by the national government, at a uniform rate of 17%; 2) business tax, a streamlined excise tax on service industries,[10] administered by local governments; and 3) an excise tax on luxury goods and goods with consumption externalities administered by the national government. In addition, the statutory corporate income tax rate fell from 55% to 33%, and its administration was split between national and local governments. The national government collected corporate income taxes from financial institutions and firms controlled directly by the national government, while local governments collected corporate taxes from local firms.[11] Local governments kept all business taxes and corporate taxes from local firms and 25% of VAT from firms located in their jurisdiction. There was no obvious change in the degree to which capital was part of the corporate profits tax base.

If officials still controlled the after-tax earnings of these firms, these changes in statutory tax rates simply change parameter values in the above equations.[12] However, there was a steady push towards selling off control over non-state firms. After this occurs, local governments simply receive their share of the taxes collected from these firms. On net, these changes lowered dramatically the incentives officials faced to improve the performance of local firms.

Another major shift, though, was to reduce the government's control over the allocation of credit from the banking system, so that loans would be made based on commercial principles. With a commercial market for credit, we then must presume that r becomes a market clearing price.

[9] When first introduced, expenditures on fixed assets were not permitted as a deduction under the VAT. In addition, the VAT was production-based rather than consumption-based. China switched from a production-based VAT to a consumption-based VAT effective on January 1 2009, which was expected to reduce tax revenue by RMB 120 billion in 2009. Source: Ministry of Finance.

[10] The business tax rates range from 3% for transportation, construction, post and telecommunications, and culture and sports, 8% for finance and insurance, 5-20% for entertainment, and 5% for other services.

[11] Since 2002, the national government has assumed administration of corporation income taxes on all new firms. But this new rule had a grandfather clause. It allowed local governments to continue to collect and keep corporate taxes from existing local firms. The national government has kept most of the corporate taxes from new firms and used it to fund revenue equalizing grants for less developed regions.

[12] In particular, for existing government-owned firms, the only change is that the national government now collects some VAT revenue from these firms, so that the local government does not receive quite all of the pre-tax profits. For existing private firms, statutory tax rates change, but the expressions themselves remain unchanged. New private firms, though, generate no corporate tax revenue, unlike existing private firms, though they still generate some VAT revenue for local governments.

What can we then say about the allocation of capital after these reforms? These decisions are now made by firms rather than by local governments, subject to a market clearing interest rate. Firms would choose to invest until $(1-s_j)f_K^j = r+\kappa_j$. Taxes now discourage investment, and can distort the allocation of capital across sectors due to variation in VAT coverage or in implicit tax rates on capital.[13]

Farmers initially still faced agricultural taxes, so would invest until $g_K = r + \sigma r/(1-\sigma)$. While some distortions to the allocation of capital remain due to differential tax rates by sector, distortions are likely much less than before these reforms. However, taxes on agriculture were eliminated recently, at least in principle, a change that should lead to a shift towards over-investment in agriculture relative to industry. Insecure use rights to the land, though, may still inhibit investments in agriculture.

When firms are left indifferent to adding more capital, however, local governments gain from further capital investment due to the resulting taxes. They can add to the local capital stock by favoring capital-intensive over labor-intensive industries. One way to do this is to continue restricting imports to the jurisdiction of more capital-intensive products, in order to increase demand for local production in these industries.

With the loss of control over the allocation of capital, the remaining control over the allocation of land took on greater importance. Due to the rapid rate of growth in industry and the pressure for large reallocations of land, land allocations became an important issue.

Rather than allocating land specifically to one firm or another, the practice instead was to auction the land to the highest bidder. The key question is then the amount of land to remove from agriculture and make available for industrial or residential use. Given the institutions prevailing since 1994, all of the terms in equation (4) become relevant. As described above, most of these terms push for too much land left in agriculture relative to industry, particularly if $s_i < \sigma$ after the shift from excise taxes to a VAT. The excess land in agriculture keeps the auction price high, and means that the required compensation to farmers remains low due to the resulting low marginal product of land in agriculture.

Due to the restrictions on land available for industry, industrial rents are artificially high. As a result, land rents have become a major source of finance for local governments in China. This institution is very much reminiscent of the role of land controls in place in Hong Kong, where again the government limited rights to land development.

[13]Note that the market-clearing interest rate will be below the marginal product of capital due to the taxes on capital.

The economic reforms also substantially changed the incentives to finance public services. Part of the change was the growing use of user fees to finance infrastructure as well as services to households. With the growth in the fraction of output coming from private firms, public service levels remain efficient only if there are sufficient user fees. Note, though, that user fees are sufficient to induce an efficient level of public services even if they cover considerably less than the marginal cost of these services, given that the resulting improvements in productivity lead to an increase in equilibrium land rents. With many services (e.g. highways) provided by private firms with user fees fully covering the costs to these firms, we forecast that the private sector as a whole loses at the margin from increased services, due to the resulting increases in land rents, leading to excessive spending on these services.

What can we say about the efficiency of spending on G_F? User fees such as road tolls now equal the after-tax benefit to farmers from the use of public services, e.g. $(1-\sigma)F_G = u_F$. As seen from equation (6), though, allocations are then efficient only if there are no net effects of the extra public services on land rents or on wage rates. However, any extra services to agriculture benefit farmers due to the increase in compensation paid for land removed from agriculture, and also benefit the non-agricultural sector through a fall in land rents there. To that extent, government incentives to provide services to agriculture are insufficient.

Services to households will be provided only if they are fully financed with user fees. We have heard this referred to as the "commoditization" of education and health care.

One recent change was to eliminate any taxes on agriculture, setting $\sigma = 0$. This change improves the private incentives of farmers to invest capital in agriculture. It raises the value of land in agriculture, making it more expensive for officials to shift land to other uses. However, it also lowers the incentives on officials to provide public services to agriculture, since this extra source of incentives is eliminated.

3.3 Labor mobility

Contrary to our assumptions above, there is some labor mobility in China, particularly between rural areas but also between rural and urban residence. Officially, individuals need to change their hukou in order to move, requiring approval of both the new jurisdiction and the old jurisdiction. Mobility therefore requires the worker as well as both jurisdictions to benefit. For a move to benefit both jurisdictions will normally require side payments between the two jurisdictions. Such side payments inevitably will be time consuming to negotiate. Perhaps due to these costs, we often see jurisdictions allowing migrant workers to enter without granting them

official residence, but also without providing them public services. Our aim in this section is to understand the implications of both forms of labor movement, and their implications for government policies more broadly.

To begin with, what are the determinants of the benefits to a jurisdiction from having more labor enter, or the loss to a jurisdiction from having workers leave? Making use of expression (1), we can calculate the impact on a jurisdiction from a marginal increase in the number of workers. The net benefits/costs of having an extra worker consist of several components. First, sales tax revenue increases on the extra output. Second, the extra labor force will generate further capital investment, leading to extra tax revenue both directly due to the implicit tax on capital and indirectly through further increases in sales tax revenue. Third, land rents change: industrial land rents go up due to the increases in both capital and labor, residential land rents increase due to the larger industrial labor force, but the compensation that must be paid to farmers for land taken out of agriculture also goes up since the value of agricultural land increases. Fourth, wage rates fall, in itself raising tax revenue. Fifth, the new worker must be provided public services, but pays any associated user fees – these two effects cancel out if user fees fully finance any public services.

The size of these net gains will vary by jurisdiction for a variety of reasons. The gain in sales tax revenue depends on the industrial composition of the jurisdiction. For example, if the jurisdiction is mainly agricultural, then there are little gains given that agriculture is no longer taxed. This gain is also larger in jurisdictions with a high local wage rate, since then the marginal product of labor is higher, leading to larger increases in sales tax revenue. The increase in capital investment would normally be larger the more capital intensive are the key industries in the jurisdiction. Capital intensity should be higher in part when the local wage rate is higher. The increase in industrial land rents would normally be greater the larger the increase in capital and labor, while the higher compensation to farmers for their land is less important in more urban jurisdictions.

What migration do we then expect to see? Workers will want to move to jurisdictions that provide them higher utility. We assume that their utility can be expressed by $U(w, q, G_h, u)$: utility should be increasing in the wage rate, falling due to higher land prices, and higher when the package of public services and user fees is more attractive. The old and new jurisdictions will together sanction this move if the new jurisdiction gains more than the old jurisdiction loses, with enough of a difference to justify paying the fixed costs of negotiating an agreement. All the terms are larger for more skilled workers, so their moves are more likely to be sanctioned. In addition, we expect that moves will more likely be approved when

they involve migration from an agricultural area to an industrial area, because of the differences in relative tax rates.

Unskilled workers are less likely to have their moves approved, even when all parties gain from the move, since the required transactions costs may be too high. When a jurisdiction benefits from having extra unskilled workers, though, they should still allow unskilled workers to enter, even if they do not bother to negotiate a shift in their hukou.

We have heard anecdotes, though, of jurisdictions trying to prevent entry of unskilled workers? Why might this be? We have assumed that all workers are perfect substitutes in production. More realistically, unskilled and skilled workers are not perfect substitutes, while industries vary in their relative demands for skilled vs. unskilled workers. The industrial composition of the jurisdiction should then in equilibrium adjust so that demands for different skill levels match supplies. Given that some industries pay more in taxes than others, a jurisdiction would then want to adjust the skill composition of its labor force to match the desired skill composition of the more heavily taxed industries. If skill-intensive industries are more heavily taxed, as seems plausible, then jurisdictions have an added reason to try to increase the ratio of skilled to unskilled workers, so to prevent the in-migration of unskilled workers.

Jurisdictions gain also from attracting workers, per se. To attract more workers, a jurisdiction needs to provide these workers higher utility. Wage rates and residential rents are both market-clearing prices, so not directly under the control of local governments. Officials do control, though, the level and price of public services. They can attract more workers by supplying the set of public services households desire. The government can make these services yet more attractive by offering them at a fee below their marginal cost, paying the residual costs out of the extra tax revenue that results from the in-migration of additional workers. If communities need to compete for workers and workers are highly mobile, then user fees will fall to the point that jurisdictions just break even from attracting an extra worker.

When jurisdictions benefit more from skilled workers than unskilled workers, there is an incentive to induce students to get more education, requiring a reduction in user fees. As long as these workers get enough education that any moves will be sanctioned, providing revenue to the original jurisdiction, then the jurisdiction is protected from possible mobility. Among less skilled workers, though, if exit becomes more likely as their education increases, this increased threat of exit may discourage providing them more education, if having the extra worker outweighs having a less educated labor force.

4 Incentives generated by promotion and retention procedures

Another important source of incentives for local officials in China arises from the implications of their performance for possible promotion to higher positions, or possible demotion (or worse). The national government also attempts to regulate directly the behavior of local officials.

Current intervention by the national government takes many forms. For one, the national government has ruled that local governments are obliged to provide free education through grade nine. We found above, though, that providing such a free allocation is against the interests of local officials. Given the difficulties faced by the national government in keeping tabs over a huge country, any such attempts at oversight will likely have minimal effect. For example, education can be free of tuition but there can be many fees imposed (for uniforms, books, etc.).[14]

The national government has made explicit its mechanism for judging the qualifications of local officials for possible promotion: judgments will be based in good part on the rate of growth in national tax revenue from the jurisdiction. How do these added incentives affect our prior results?

Previously, the objectives of officials reflected local tax revenue plus rental income, minus the net costs of supplying public services. National promotion policies induce officials to give some weight as well to national tax revenue. The national government collects 75% of the overall VAT payments, increasing the importance of sales tax revenue. They collect all of the corporate tax revenue from new firms and from state firms, tending to equalize the incentives local officials face to aid one category of firm compared to another. The national government, though, does not collect any revenue from agriculture, so that these incentives increase the weight placed on industry compared with agriculture. The national government also does not share in land rents, strengthening the incentives to allocate land to industry where the extra land increases national VAT revenue. Since the VAT does not allow deductions for capital, promotion incentives create yet more of an incentive to favor capital-intensive industries over other sectors. As previously, this can be done in part through trade restrictions protecting these sectors. Providing complementary inputs to capital-intensive firms can also help, e.g. better supplies of electricity and infrastructure may favor capital-intensive firms.

The national government also affects the incentives faced by local governments

[14] This is the strategy pursued, for example, by the University of California in response to a decision by the legislature that students from California owe no tuition for college.

through its control over the allocation of intergovenmental transfers of funds. Funds are in part allocated based on geography, going particularly to inland provinces, and to this extent do not change marginal incentives. Any allocations that can be affected by local choices, though, do implicitly alter local incentives, and local behavior. The process of allocating transfers is not explicit, though, preventing us from assessing the resulting incentives.

5 Alternative policies that can improve incentives for local officials

Under existing incentives faced by officials, outcomes are inefficient for a wide variety of reasons. Governments make use of their controls over the allocation of land to keep prices high for industrial and residential uses, and low in agriculture. Since any movement of resources from agriculture to industry generates more tax revenue, officials will make use of their control over public services and regulations to aid industry relative to agriculture. They may try to interfere with inter-jurisdictional trade to protect more highly taxed local industries. Migration of workers is limited given the high costs of negotiating the shift in a worker's hukou from one jurisdiction to another, leading to an inefficient allocation of labor across jurisdictions as reflected in the sharp differences in wage rates between rural and urban areas and among urban areas. What policy changes at the national level might lead to better incentives for local officials and a more efficient allocation?

There are many possible directions for reform, and we make no pretense of knowing which steps might be considered seriously. Our aim has instead been to provide a framework for the analysis of alternative institutions.

That said, let us examine a few possible directions for reform. One reform that has recently been taken is to shift from a production-based VAT to a consumption-based VAT. Suggested by this change, consider a reform under which corporate tax revenue and any implicit taxes on capital go solely to the national government, while local governments rely instead on an increased share of this consumption-based VAT plus business taxes at the same rate on those goods (largely services) not covered by the VAT. The individual's budget constraint equals $wL = C + qA_h + u_h$. A uniform VAT on consumption, housing, and user fees is then equivalent to a labor income tax. With tax rates fixed, raising tax revenue is then equivalent to raising labor income, algining the interests of officials with those of residents. Higher labor incomes also help attract new residents.Under particular assumptions, laid out in Atkinson and Stiglitz (1976), any differences in consumption tax rates across goods

should be avoided.

What about the misallocations of land described above? Since officials are compensated heavily based on the price differential of land in alternative uses, they face strong incentives to shift land to higher-value uses. Misallocations occur primarily due to the goverment's market power. This market power is limited, though, to the extent that firms and individuals are mobile.

Introducing policy changes that increase mobility is the mechanism emphasized in the Tiebout model to induce officials to provide the efficient level and composition of services. To the extent people are mobile, officials are pressed to adopt policies that attract potential residents and induce existing residents from stay. With intense competion, policies end up maximizing the utility of residents, so are efficient.

There are a variety of policy changes that can affect mobility. One would be to maintain the discretion of jurisdictions over whether to grant residence to migrants, but to eliminate the need to compensate the original jurisdictions for the.loss of their residents. By eliminating these transactions costs, individuals can move to jurisdictions offering higher utility as long as the jurisdiction itself gains. The jurisdiction can limit entry in part by limiting the supply of residential housing, forcing rents up to the point that only the more skilled may enter. This change alone then can help push towards equalization of utility across locations, and increase the pressure on officials to maximize this utility.

One factor inhibiting mobility is the lack of a market for farmland. Those working in agriculture then face the potential loss of much of the value of this use-right to the land if they migrate. Only if the utility gain is large enough, given these hurdles, will people move. Increasing the ease of selling use rights to agricultural land may then be a key step in increasing the potential mobility of farmers, putting more presure on officials to find policies that induce them to stay.

6 Conclusions

What pressures affect the allocation decisions made by local officials? How do changes in institutions under the control of the national government affect the decisions made by local officials?

In this paper, we started with the assumption that local officials act to maximize the tax revenue they receive plus income from land minus the costs of public services, and explored its implications. The behavior of local officials forecast by the model to our mind corresponds closely to the stylized facts we see in the data. If we accept this model as a valid characterization for how officials behave, then the model provides a mechanism to help guide the redesign of these incentives in order to induce officials

to behave as desired by the national government. Potential reforms were discusssed briefly, though many others may also improve allocations.

References

Banerjee, Abhijit, and Esther Duflo (2006) 'Addressing absence.' *The Journal of Economic Perspectives* 20(1), 117–132

Brandt, Loren, and Ziaodong Zhu (2000) 'Redistribution in a decentralized economy: Growth and inflation in china under reform.' *The Journal of Political Economy* 108(2), 422–439

Buchanan, James M., and Gordon Tullock (1962) *The Calculus of Consent* (Ann Arbor: University of Michigan Press)

Gordon, Roger H., and Wei Li (1995) 'The change in productivity of chinese state enterprises, 1983–1987.' *Journal of Productivity Analysis* 6, 5–26

Gordon, Roger H., and Wei Li (2003) 'Government as a discriminating monopolist in the financial market: The caseof China.' *Journal of Public Economics* 87, 283–312

Groves, Theodore, Yongmiao Hong, John McMillan, and Barry Naughton (1994) 'Autonomy and incentives in chinese state enterprises.' *Quarterly Journal of Economics* 109, 183–209

Li, Wei (1997) 'The impact of economic reform on the performance of Chinese state enterprises,1980–1989.' *Journal of Political Economy* 105(5), 1080–1106

Li, Wei (1999) 'A tale of two reforms.' *Rand Journal of Economics* 30, 120–136

Li, Wei (2008) 'Maoist economics.' In *The New Palgrave: A Dictionary of Economics,* ed. Steven N. Durlauf and Lawrence E. Blume, second ed. (Palgrave Macmillan)

Tiebout, Charles M. (1956) 'A pure theory of local expenditures.' *The Journal of Political Economy* 64(5), 416–424

Wang, Feng, and Xuejin Zuo (1999) 'Inside china's cities: Institutional barriers and opportunities for urban migrants.' *The American Economic Review* 89(2), 276–280

Young, Alwyn (2000) 'The razor's edge: Distortions and incremental reform in the people's republic of china.' *The Quarterly Journal of Economics* 115(4), 1091–1135