How Migration Restrictions Limit Agglomeration and Productivity in China*

Chun-Chung Au and Vernon Henderson

Brown University, August, 2004

Abstract: China strongly restricts rural-urban migration, resulting in a well acknowledged surplus of labor in agriculture. But migration is also restricted within sectors. This paper argues that these intra-sector restrictions lead to insufficient agglomeration of economic activity in both the rural industrial and urban sectors, with resulting first order losses in GDP. For urban areas the paper estimates a city productivity relationship, based on city GDP numbers. The effects of access, educational attainment, FDI, and public infrastructure on productivity are estimated. Given these, worker productivity is shown to be an inverted *U*-shape function of city employment, with the peak point shifting out as industrial composition moves from manufacturing to services, as predicted by urban theory. As far as we know this is the first paper to actually estimate the relationship between output per worker and city scale for any country. The majority of Chinese cities are shown to be potentially undersized -- below the lower bound on the 95%confidence interval about the size where their output per worker peaks. The paper calculates the large gains from increased agglomeration in both the rural industrial and urban sectors.

key words: China, migration, urbanization, scale economies, productivity. **JEL** Reference Nos. J600, O100, O400, P200, R100, R200, R300.

* We gratefully acknowledge support for this work from a grant from the Research Committee of the World Bank and by a Mellon Foundation grant to the Population Studies and Training Center at Brown University. The opinions expressed in the paper do not necessarily represent those of these organizations. Remy Prud'homme provided early inspiration and insight in formulating some of the issues in the paper. Comments by two anonymous referees were very helpful in revising the paper.

China restricts internal migration of its population between urban and rural areas, within the rural sector, between big and small cities, and between regions. The prime instrument of control is the household registration, or hukou system described below. The accepted view is that the primary effect of the restrictions is to enforce a surplus of labor in the rural sector. As China's rate of economic growth has fluctuated in recent years, a policy issue is whether permitting greater migration to cities would encourage growth by moving workers from low marginal product employment in the rural sector to high marginal product employment in the urban sector (see Prud'homme (2000)). A related policy concern is national income inequality, a serious domestic social and political issue. In 1978 China's Gini coefficient was similar to India's and urban real incomes were at least 2.0 times those in the rural sector. Since 1978, China's rural-urban income disparity has increased (Johnson (2001)); and the national Gini coefficient has risen and now far exceeds countries such as India, Bangladesh and Pakistan, not to mention Korea and Japan.[1] Allowing greater migration might reduce inequality.

Evaluating the gains from lifting restrictions to migration is a complex exercise, beyond simple comparisons of labor productivity in the rural versus urban sectors. This paper focuses on one aspect of resource misallocation from migration restrictions, and examines a second. The primary idea is that there would be huge productivity gains just from internal reorganization of the urban and rural sectors, without changing the allocation of labor between these sectors. In both sectors there are unexploited economies of scale, from spatial agglomeration of resources. We will show empirically that migration restrictions across cities, as well as rural-urban migration restrictions, have left many cities significantly undersized, with unexploited economies of scale, resulting in large productivity losses. Similarly, migration restrictions on rural to rural migration and restrictions on the movement of rural firms strongly inhibit spatial agglomeration of rural industry, also resulting in large productivity losses.

The second point developed in the paper is based on the fact that the rural sector itself consists of two parts, agriculture and rural industry called town and village enterprises [TVE's]. While there is a distinct surplus of labor in agriculture, the rest of the picture is cloudy. At least in the 1990's, the marginal product of labor in the urban sector exceeds that in the TVE sector, but for capital the reverse is the case: the return to capital in the urban sector is distinctly lower than in the TVE sector. Capital allocations in China are subject to strong controls, with investment directed towards urban industry through quantity controls and discriminatory taxation. As a result, Jefferson and Singhe (1999, Table 1.9) find in 1995 and 1996 that the pre-tax return on

[1] In 1995 the World Bank estimated China's Gini to be .42 compared to about .30 for the noted South Asian countries (WDR (1998)). Korea and Japan are viewed as having among the lowest Gini's in the world.

capital for rural industry is 25% higher than in the state owned urban sector. Moreover education is very unequally spread among sectors, with migration policies which appear to bias the supply of more educated people towards the largest cities.

In terms of the primary focus of the paper on within sector spatial agglomeration, spatial agglomeration involves the clustering of people and resources into cities and towns (in the rural sector), so as to exploit local scale externalities and market linkages in the production and distribution of goods (Henderson (1974), Fujita, Krugman, and Venables (2000), Duranton and Puga (2001)). However, the size of any city is limited by diseconomies in commuting and, urban spatial structure that also arise with agglomeration, as well as the spatial extent of a city or town's market. Systems of cities models postulate an inverted *U*-shape function to utility, or to real output per worker in city population or employment size. There is a city size at which output per worker is maximized; but the size where that peak occurs will vary with city industrial composition.

To assess whether there is too little (or too much) spatial agglomeration in a country, we need to assess whether cities are consistently under- (over)-sized. To do that assessment, we need to estimate how real output per worker varies with city size and how the inverted *U* -shape shifts with industrial composition. China presents a unique opportunity to do this because it collects GDP figures for geographic units corresponding to the urbanized area of a city. China's migration restrictions turn out also to be fortuitous from a research perspective. Since cities are found to both the left and right of the peak of the inverted curve, we can identify its shape. As such this is a fundamental research contribution of the work, in addition to the analysis of the costs of migration restrictions. This type of analysis has yet to be conducted for any other country.

1. Migration Restrictions and Urbanization Policy in China

Migration in China is restricted through the hukou system. Here we describe its application in the period of data analysis 1995-1997, and then comment on the current situation. The hukou system in China is similar to an internal passport system (see Chan (1994) for a detailed description). A person's local "citizenship" and residence is defined as a birth right, until recently by the mother's place of legal residence. The entitlements and details of the system differ for urban and rural residents. Legal residence in a city entitles one to local access to permanent jobs, regular housing, public schooling, and public health care (where almost all health care is public). Until the early 1990's, it also entitled urban people to "grain rations" -- rations of essentials such as grain and kerosene.

Legal residence in a village or rural township entitles citizens to land for farming, township housing, job opportunities in rural industrial enterprises, and access to local health and schooling facilities. Residents also have some degree of "ownership" in local enterprises, although distributed profits all go to the local public budget, which may be used to finance township housing and infrastructure. Again, until recent years, legal residence in a township also entitled a "peasant" to some share in locally produced (or allocated from the outside) grain and other essentials.

How does a person change their local citizenship? One mechanism is education. A smart rural youth may persist through the competitive, discriminatory education system to get a spot in a college; upon college graduation, the rural youth will be hired into an urban job, with an urban hukou. It is this aspect of migration policy which potentially skews the allocation of educated people towards larger cities. Second, the state at times has opened the gates, permitting factories to hire permanent workers from rural areas, permitting family reunification, or permitting legal migration from rural areas to nearby small cities. Permanent migration is not without costs. Permanently leaving a village means abandoning ownership claims without compensation to agricultural land that one's family may have farmed for decades and to the profits of local rural industries which are distributed in-kind, as for example with township housing.

Apart from citizenship changes, people can migrate to an area without local hukou, either illegally ("unregistered") or legally as a temporary worker, or as a "permanent resident" on a long-term permit. For example, a rural person may be hired legally as a "contract worker" in industry or services, for a term of three years. Or people may move illegally, and work in the informal sector for low pay, under poor conditions, with risk of deportation. Temporary migrants to larger cities typically have no, or very high priced access to health care and schooling facilities and regular, "legal" housing. Cities have strict national guidelines on conversion of agriculture to urban land; and institutional difficulties in housing markets in expanding supply makes it particularly difficult for migrants to find decent housing. Legal temporary migration requires a permit from the city of in-migration and cities can impose various hurdles to getting a permit --- permission from the home location, proof of a guaranteed job and specific housing, and the like. Cities also publish job lists, citing jobs for which migrants are not eligible; in 2000, Beijing listed over 100 occupations as non-eligible ones. Migrants may still have to pay taxes to their rural home village for services they don't consume and on land left fallow. Finally migrants typically have faced direct fees (Cai (2000)). There is a license fee to work outside the home township paid to the township that can be equivalent to several months wages. At the destination there can be fees for city management, for being a "foreign" worker, for city construction, for crime fighting, for temporary residence, and even for family planning if the migrant is female.[2] All these restrictions sharply reduce the benefits and raise the costs of migration, particularly into large cities. Migration is limited and most migration is short-term, or “ return” migration. These restrictions on migration were loosened for incoming migrants in the late 1990’s in parts of southeastern China and have been relaxed since 2000 more generally. Yet the basic provisions remain in place even today.

The application of migration restrictions in part reflects the urbanization provisions in the Sixth Five Year Plan (1981-1985) (Chan (1994), Lin (1992b), and Henderson (1988)). Several features of national urban planning in the 1980's strongly influence the urban sector today. For most of the 20 years prior to 1982, urbanization in China was strongly restricted, with the urban sector's share of national population declining, reflecting in part the bias against urbanization and its association with capitalism and Western moral pollution. The Sixth Five Year Plan intended to constrain the growth of larger cities, although allowing population growth of smaller cities, through flexibility in transferring full or limited hukou to small cities from rural areas. But given overall urbanization was still to be restricted, a key issue was what to do with the surplus rural population. A policy of “leave the land but not the village” promoted development of rural industry. Rural industrial development became a key engine of Chinese growth through to the mid-1990's. But the restriction of “not [leaving] the village” makes rural agglomeration of industry difficult, limiting its overall ability to be competitive.

A final aspect of planning in the 1980's was the strict spatial hierarchy that is still in place. The large (“sophisticated”) lead the small. So only the largest coastal cities were to have initial access to new technologies and FDI; and in turn were supposed to transfer their older technologies to smaller and hinterland cities. Capital allocations, tax policies and governance structure favor the large over the small; and this hierarchical structure may influence productivity differentials across cities. We need to account for aspects of these factors in estimation.

2. The Urban Sector

This section starts with an overview of the urban sector, looking at its key features. Then we present a model of productivity of cities for econometric implementation, analyze its error structure and identification, and discuss data issues. We present results for prefecture level cities at the top of the urban hierarchy first and then for county level cities.

[2] Officially, most of these fees were abolished in 2001. Unofficially a number persist.

2.1 China's Urban System

We have data for 1990-1997 on about 225 prefecture and provincial level cities, the larger cities at the top of the urban hierarchy. Post 1997, numbers on effective total labor force of cities are no longer published, making an update of the analysis difficult, since our key size measure will be employment. In addition we suspect the ability of city statistical bureaus (see below) to measure value added in the rapidly growing, more informal service sector starts to deteriorate in the late 1990's. For 1990, 1991 and 1994-1996 we have data on about 435 below-prefecture level cities, generally urban places over 100,000 population. Those latter cities are either traditional county cities or new cities; we call them county cities. In 1990, at all levels, there are about 465 cities; by 1996 that number grows to about 665 cities, so there is entry of an additional 200 cities. Virtually all new cities up through 1997 are county ones.

Table 1 presents some basic facts about the urban system in China. Columns in Table 1 give figures for all cities present in 1990, for prefecture level cities present in 1990, for county cities present in 1990, and for new county cities after 1990. The first four rows show high overall non-agricultural employment growth from 1990 to 1996, over 4.3% per year for all cities, but much slower population growth, about 1.9% per year. The differential reflects two things. In 1990 some cities contained significant rural-agricultural populations which during the decade moved into non-agricultural employment. This is urbanization from within, although overall local absolute agricultural employment in cities grew modestly from 1990 to 1996. Second and more critically, population numbers may exclude certain types of immigrants while employment figures better reflect them. These are shorter-term or longer-term migrants working in the city, who either live in the city but are not counted in population enumeration or who can't obtain housing within the city and reside just beyond the boundaries of urban districts (in often "illegal" rural housing). As a result, in the paper, our city size measure will always be in terms of non-agricultural employment.

Table 1 also shows that real output per worker grew at an incredible rate during the period; for all cities, the average annual rate was about 7.8% a year. Finally over time the manufacturing to service ratio declines. The decline involves some redefinition of manufacturing activity to be service activity in around 1993-1994. In the data, over the 1990-93 period the ratio declines modestly by about 1% a year; between 1993 and 1994 it declines by 24% with economic redefinitions based in key reforms in 1993-1994; and from 1994 on it declines by 4-5% a year, as restrictions on private service sector development are loosened with the economic reforms.

Looking across the urban system, prefecture level and above cities are larger and have much greater human capital. Among county cities, there are many new cities, which are larger than traditional county towns, have more industry, and are nearer the coast. These cities reflect rapid urbanization and industrialization of former rural towns and townships.

The period 1990-1996 is one of rapid industrial reforms, removing most of the remaining props under state owned industry and exposing them to increasing competition. Mostly heavily hit were interior and northern heavy industry cities, especially under the reforms in 1993-94, which moved most remaining planning functions to a market basis and represent a break point in the data in terms of how outputs are evaluated. Along with the rapid growth of business and financial service activity, the result, in this very short period of time, is to dramatically shake up the urban system. Later in estimation to help achieve identification, we will utilize this '93-'94 split, viewing economic data from the early 1990's as more heavily influenced by planning and the history of planning, and economic data from 1996 and 1997 as primarily driven by market forces.

2.2 Productivity in the Urban Sector

We model productivity in the urban sector, using a standard urban model, incorporating developments from the new economic geography. Suppose countries are composed of a variety of different types of cities, where each type produces a particular good for export to the rest of the country from the city, in an urban hierarchy with an endogenous number of cities of each type. In any city type, final city output is produced with local inputs from manufacturers and from service producers. So at the bottom of the hierarchy we might have heavy industry with, say, steel type hinterland cities that combine rolled steel production with local finance, personnel, and market services to sell rolled steel to other cities. At the top in a coastal city we might have headquarters activity for manufacturers whose services (to hinterland plants) are produced mostly with local export, R&D, and management inputs.

More generally for city i of type j, value-added of the city Y_j^i, is a function of production amenities in city i, A^i, local manufacturers' output M_j^i, and a function with intermediate non-traded service inputs, $\bar{g}_j(x_1^i, x_2^i \ldots x_s^i)$. x_ℓ^i is output of service firm ℓ and there are s^i service firms in city i. Usually symmetry is imposed (i.e., x_ℓ^i firms have identical technology and enter $\bar{g}$ identically, so $\bar{g}_j(\cdot)$ is written as $g_j\ (s^i, x^i)$, where x^i is the output of the representative service firm in city i. The $\bar{g}_j(\cdot)$ function represents the Dixit-Stiglitz-Ethier notion that diversity, s^i, plays a role in enhancing productivity in Y_j^i production. Thus value added of city i in the simple case of Hicks' neutrality of A^i is

$$Y_j^i = A^i f_j \ (M_j^i, \ g_j \ (s^i, \ x^i)) \qquad (1)$$

Each of the x^i is produced according to technology

$$x^i = x(k^i, n^i) \qquad (2)$$

where k^i and n^i are per firm inputs of capital and labor in the service sector in city i . Manufactures, M_j^i , are produced with inputs of capital and labor, or

$$M_j^i = M_j \ (K^i . N^i) \qquad (3)$$

City endowments of capital and effective labor are $\overline{K^i}$ and $\overline{N^i}$ where effective labor is defined below, so

$$K^i + s^i k^i = \overline{K^i} \qquad (4a)$$

$$N^i + s^i n^i = \overline{N^i} \ . \qquad (4b)$$

Capital and labor are observed but not their allocations between the final output sector and the intermediate input service sector. Accordingly, if we impose internal allocation rules, we can reformulate eq. (1) as a 'meta-production' function in terms of $\overline{K^i}$ and $\overline{N^i}$. For example, there could be optimal allocation rules, where after substituting (2)-(4) into (1) and optimizing with respect to K, N and s (dropping sub and superscripts) we have

$$sf_M M_K = f_g g_x x_k \qquad (5a)$$

$$sf_M M_N = f_g g_x x_n \qquad (5b)$$

$$g_s = g_x \ (x_n n + \ x_x k) / s \qquad (5c)$$

Note the last represents the usual monopolistic competition trade-off from increasing the number of service sector firms and hence varieties of local service sector products against the reduction in resources for any one existing service sector firm. Utilizing these to solve out $k, n,$ and $s,$ allows us to write (1) as

$$Y_j^i = A^i \tilde{F}_j \ (\bar{K}^i, \ \bar{N}^i) \qquad (6)$$

There are two points in evaluating results based on (6). First the allocation rules used to derive a meta-production function need not be optimal. So, for example, we could assume marginal products of capital or labor are only equalized across sectors within cities up to a factor of proportionality, ϕ_K or ϕ_N, where the ϕ's might depend on human capital in the city or the extent of economic liberalization. Second, contained in (6) are local scale externalities in cities which determine the shape of the inverted-*U* relationships between output per worker and city sizes that we seek to identify later. One source of local scale externalities, the role of diversity,

s^i, has been discussed. Another is that manufacturers can also experience scale externalities, dependent on local manufacture's scale (measured by, say, either M_j^i or N_j^i and reflected in the A^i term in (8) as discussed below). Such scale externalities are based on local information spillovers and intra-industry specialization (Marshall (1980); see Duranton and Puga (2004) for a review of the literature on micro-foundations). Allocation rules of cities reflect the extent to which these externalities are successfully locally internalized. Our results are conditional upon allocation rules in place, although as we point out later, some basic switches in allocation rules only shift the inverted-U's we seek to estimate up or down, without changing their shape.

2.2.1 The Inverted U-Shape to Output Per Worker and City Types

So far we have discussed the existence of scale economies in cities which are the basis of local agglomeration. However with undiminished scale economies, we would have no limit to the benefits of agglomeration and city sizes. Apart from limits on a city's market potential (Fujita et al (1999)) discussed later, the key ingredient in most urban models involves the notion of "effective labor". In a city, the down-side to local agglomeration which limits city sizes is represented by commuting/congestion costs (Henderson (1974)). People live in residential areas surrounding a central business district; as city population, or employment size increases the per person time devoted to commuting (to claim more land for housing at the expanding city edge) increases on average.[3] The net disadvantages of city size increases can be represented as lost working time. So if L is total city population (= labor force), the fraction available for work is $\tau(L),$ where

$$\overline{N} = \tau(L)L, \ \ \tau < 1, \tau^{'} < 0 \tag{7}$$

so effective labor, $\overline{N},$ is a fraction of the labor force, $L,$ where that fraction declines with $L.$ Incorporating commuting costs, we rewrite (6) as

$$Y_j^i = A^i F_j(\overline{K}^i, L^i). \tag{8}$$

Empirically we will estimate a version of (8) where value-added per worker will be viewed as an inverted U-shaped function of scale, L. As city scale increases from low levels, initially scale economies dominate, but then commuting cost considerations kick-in, so there is a city scale which maximizes value-added per worker. One issue will be how to capture these non-

[3] The sum of rents plus commuting costs increases for everyone; but rents are redistributed, say, equally back to residents as Arrow-Debreu shareholders in their city. Even allowing for polycentric cities, per person commuting costs relative to agglomeration benefits rise with city size.

linear effects. A second issue is how to incorporate notions of hierarchy so that the peak point to the inverted U -shape of value-added per worker shifts with industrial composition.

As we move up the urban hierarchy, empirical evidence for a market economy, the USA, indicates the manufacturing to service ratio declines monotonically as city size increases. Kolko (1999) looking at six city size categories has an employment manufacturing to business service ratio of .67 at the top of the size hierarchy which rises continuously to 2.95 at the bottom. In China, an economy in transition, we do not see such a neat monotonic relationship; but the manufacturing to overall service ratio and city size are negatively correlated, with distinctly lower relative manufacturing activity in the largest cities. Thus we postulate that cities should become increasingly service oriented as size increases, in terms of the composition of local observed production of M_j versus $\overline{g}_j(\cdot)$ based on the shape of $f_j(\cdot)$ in equation (1). For this to come out of a theoretical model we must account for the juxtaposition of parameters as they vary by type j, that describe the intensity of service versus manufacturing inputs in Y_j. Au and Henderson (2004) have a fully specified, specific functional form model, that illustrates the interactions. There the share of $\overline{g}_j(\cdot)$ in Y_j^i in eq. (1) whose measure is a function of observed manufacturing to service ratios is specified to rise as we move up the urban hierarchy. But in practice scale economies and the importance of diversity also vary across the urban hierarchy and in the next section we allow for general interactions, capturing in a meta-function approach how inverted-U's shift with industrial composition across the urban hierarchy.

How does the literature put together different city types and inverted U-shape curves in a general equilibrium context? In standard urban models (see Duranton and Puga (2004) or Anas and Abdel-Rahman (2004) for reviews), in a market economy with perfect migration, free capital markets, and developers and/or local government involved in formation of new cities, any city type j operates near its peak point to value-added per worker, which will be proportional to real compensation per worker with log-linear technology, as we will see below. All cities face the same horizontal national supply curve of labor (as viewed by an individual city). As we move up the urban hierarchy, bigger cities have their peak points to real compensation per worker shifted right, cresting near the supply curve. In particular, with perfect divisibility of cities, many cities of each type, and all cities having identical amenities, A^i, each inverted U -shape curve for real compensation per worker is tangent to the supply curve at its peak point. If A^i's vary, say, then those with high A^i's operate to the right of their peak points in stable equilibria.

2.3 Econometric Implementation

How do we estimate eq. (8)? The key issues concern how to represent city types and scale effects, how economic variables in (8) are measured, and what is the error structure. We start by noting the role of transport costs and presenting the basic estimating equation. Then we turn to issues of error structure and finally to issues concerning the data, in particular the measurement of capital.

For transport costs, in eq. (8), if $\hat{Y}^i$ is value added evaluated in national coastal markets at uniform world prices, what we observe in the data is the local evaluation Y^i. The link is transport costs; purchasers of products of cities further from national coastal export markets will pay higher transport costs and pay producers in those cities less for any amount of output. Initially we model this as $\ln \hat{Y}^i = \ln Y^i + T^i\phi$ where ϕ is a vector of parameters and T^i a set of variables, in particular distance from city i to the coast. Later we experiment with adding measures of local market potential.

In the formulation of (8), we need to account for both scale economies and diseconomies in producing an inverted U-shape to real output per worker and we need to allow for urban specialization in different types of cities with differing equilibrium or efficient sizes. We use the ratio of secondary to tertiary sector activity, which we call manufacturing and services to indicate how the constellation of scale economy effects varies as we move up the urban hierarchy. The estimating equation is

$$\begin{aligned}\ln (Y_t^i / L_t^i) = {} & A_t^i\gamma - T^i\phi + \alpha \ln (K_t^i / L_t^i) \\ & + \beta_0 MS_t^i + \beta_1 L_t^i + \beta_2 (L_t^i)^2 + \beta_3 (L_t^i MS_t^i) \\ & + \delta_t + \varepsilon_t^i\end{aligned} \qquad (9)$$

where MS is the manufacturing-service ratio. We use the ratio of a value-added in the two sectors based on the specific functional for model in Au and Henderson (2004), where that ratio exactly represents how the share of $g_j(\cdot)$ in (1) rises across the urban hierarchy and accentuates the role of scale economies in diversity of local intermediate inputs. The A_t^i term captures technology measures discussed below. In (9), scale effects are expressed as a quadratic in an expotential function (explaining Y/L), where the linear term varies with the manufacturing to service ratio, MS. Output per worker peaks at an employment of

$$\overset{*}{L}{}_t^i = \frac{\beta_1 + \beta_3 MS_t^i}{-2\beta_2}, \ \beta_2 < 0, \ \beta_1 + \beta_3 MS > 0 \qquad (10)$$

We tried more complex general forms to (9) and (10) where there are a variety of other interactive terms. None of these are significant, including interaction with potential indicators of sophistication of technology. Nor is interaction of MS with L^2 significant nor a general Taylor series expansion beyond the listed terms. We expect $\beta_3 < 0$ under the notion that city size where Y/L peaks will decline monotonically as MS rises. We will explore that monotonicity with refinements of (9), discussed later. In (9) while the MS ratio affects scale economies, the allocation of factors within the city between manufacturing and services is hidden. However, most critically, α does not vary with MS, L, or controls for technology, empirically. And in terms of how the specification of (9) is affected by allocation rules, for simple regimes switches (e.g., first best planning to market outcomes), as shown in Au and Henderson (2004) the shape of the inverted *U*-curve is unaffected; only its height changes.

In (9), as noted, α is capital's share in output technology. In (10), by focusing on cities operating at the peak to inverted U -shape curves, we are seeking to maximize the residual return in cities to workers, or $(Y - rK/L)$ where r is the cost of capital assessed a city. If either K/L is fixed to a city, or capital is allocated to the city so $\alpha Y/K = r$, then by maximizing Y/L, the residual return is also maximized. Labor may get this residual either directly from firms in the form of $(1-\alpha)\ Y/L$, or some directly from firms and the rest in in-kind compensation in the form of housing and services provided by the city. Of course that utility of workers may differ from real income per worker. While the formulation of $Y(\cdot)$ incorporates resources losses from agglomeration such as increasing commuting times, we haven't accounted for city public services and amenities. However there is no evidence that these vary consistently as a group by city size or, in particular (see later), that they affect the shape to the real income per worker curve. Thus we assume maximizing real output per worker maximizes worker utility.

2.3.1 Error Structure and Identification

We estimate equation (9), pooling data on prefecture level cities in 1996 and 1997. We restrict ourselves to these two years because we want to look at productivity after the regime switch in 1993-4, to try to represent outcomes in an (almost) market economy. We exclude 1995 since it is too close to the regime switch; and, if that year is included, specification tests on IV estimation (see later) fail. As we will see momentarily, in IV estimation instruments from 1990 are used.

In the error structure in (9), δ_t are national technology shocks, as well as a control for inflation and national price re-evaluations. The issue concerns the ε_t^i term. It may contain elements representing current shocks to city productivity such as a recent import or adaptation of a new technology to the city, which affect factor allocations and covariates. It may contain time persistent shocks to do with unmeasured location-geographic features or local political and institutional environments that again affect both productivity and factor allocations. Finally variables may be measured with error, resulting in attenuation bias.

Our strategy to deal with these problems affecting identification is to instrument for all contemporaneous time-varying covariates with historical characteristics of the city. Because current magnitudes present accumulation processes (see below), our instruments are strong, with first stage F's and R^2's averaging .57 and .66 respectively.[4] The issue is their validity, or exogeneity to current shocks affecting current productivity. In this section we articulate an economic rational for choices of specific instruments; and then we turn to the practical aspect – tests for such exogeneity. Much of that discussion is delayed to section 2.5.2. The economic rationale for choice of instruments has two parts, each relating to a specific set of historical variables.

Planning Variables.

The 1990 capital to labor ratio, manufacturing to service ratio, spatial area of the central business district, agriculture to other sector ratio, FDI to capital stock, sales of independent to all enterprises (see data section) and whether a city had FDI or not are treated as variables set by planning, rather than market forces in the late 1980's leading into 1990. The key argument is that provincial planners in the late 1980's in making allocations to cities ignored unmeasured (to us) aspects of the local environment in 1996-97 which affect both productivity and covariates in 1996-97. The argument is based on two assumptions. First, in the late 1980's, for prefecture level cities, these variables listed above were still largely determined based on decisions made by planners and government officials, unlike, say, the rural sector which was more market driven. In the "twice-up, twice-down" planning process, provincial along with city planners decided what lines of what products in prefecture level cities should be produced and what their investment allocations should be, not to mention the center determining where FDI would be permitted. Second, either planners and officials' objectives were not to make allocation decisions to maximize output per worker, but rather to satisfy certain planning and political objectives; or, even if maximizing output per worker was part of the objective, decisions by planners were based on the same observables we have access to, and not on the unobservables. Planners' objectives varied by province and their formulation was constrained by political considerations, by notions

[4] The minimum is for roads per capita, where the F and R^2 are respectively 17.5 and .43.

of maximizing the amount of physical production, by notions of use of input-output and planning coefficients, and by lack of proper calculation of firm profits. But again the key is whatever unobservables affect both productivity and covariates in 1996-97, allocation decisions did not account for these in the late 1980's.

As we move into the 1990's, reforms in the state-owned urban sector are less cosmetic and more cutting. In particular in 1993-94, the state-owned sector is moved in a dramatic fashion to a market basis and the service sector, particularly business services, as noted earlier is freed up with vast expansion of private services. We perceive this as a regime switch, where many of our cities between 1993 and 1994 stretching into 1995 have enormous changes (up or down) in $Y/L,\ K/L$ and MS for roughly the same urban scale, indicating dramatic shifts in the way quantities were evaluated. By 1996 and 1997 we perceive economic magnitudes much more driven by market forces.

In summary the first rational for instrumenting is that there are a set of planning variables from 1990 which are strong instruments but are unaffected by unobservables affecting productivity in 1996-97. The economic rationale I have given for this is obviously debatable, especially the notion that there was somewhat of a sudden regime switch, but on the practical side the use of 1990 planning variables as instruments is supported by statistical tests discussed below. However, in that testing, it seems that the education level and size of labor force in 1990 are not exogenous. Planning *ratios*, or planning coefficients like capital per worker are, but not *absolute scale*.[5] That could reflect either (1) even in 1990 to the extent possible migrants responded to unobservables affecting productivity and effecting earnings or (2) planners did account for certain unobserved (to us) geographic and locational advantages in determining scale but used fixed planning ratios. For labor force currently and historically we have a different instrumenting rational, which is based on migration decisions and parallels the classic case of using demand variables to instrument for price in estimating supply curves. To see this we briefly model how populations are determined in 1997.

<u>Amenity Variables</u>.

Each city offers a utility level to a resident which is a function of its real wage and local quality of life, Q_t^i. Real wages are related to the city's allocation of labor and capital and scale (dis)economies, determining labor marginal products. Thus city utility that can be offered to migrants is some function

$$U_t^i = U\ (\bar{K}_t^i,\ L_t^i\ A_t^i\ Q_t^i)\ . \tag{11}$$

[5] Of course unobservables could affect scale of all variables in the same proportion and hence not affect their ratio.

In (11) we expect U_K, U_A, $U_Q > 0$ but the sign of U_L is ambiguous (given the inverted U-shape to real wages). Focusing on the determinants of city population, L_t^i, in China growth comes primarily rural counties surrounding a city.

To model migration frictions, suppose city i draws from a surrounding rural population, where utility in the rural sector is given by

$$R_t^i = R\ (K_t^{Ri},\ \overline{L}_t^i - L_t^i,\ A_t^{Ri},\ Q_t^{Ri})\ . \tag{12}$$

$\overline{L}_t^i$ is the fixed population of the whole region of i, encompassing the urban and rural sectors, so the rural sector population is this total less the urban population, L_t^i. As noted below and in the data Appendix for China, $\overline{L}_t^i$ is fairly well defined for the 225 largest cities. K_t^{Ri}, A_t^{Ri}, and Q_t^{Ri} are respectively the capital, production amenities (e.g., soil quality, rainfall) and quality life measures in the rural sector. $\partial R_t^i / \partial L_t^i$ may be greater than zero, implying diminishing real wages in the rural sector, so R_r^i rises with outmigration. If there were no migration costs or restrictions L_t^i would adjust to equalize R_t^i and U_t^i in (11) and (12). In China, we presume $U_t^i > R_t^i$, a differential sustained by migration restrictions, which operate as frictions that have the per person cost of in-migration rising as the rate of net in-migration to the city, $\dot{L}_t^i \equiv (dL_t^i / dt) / L_t^i$, rises. At any instant the gap between urban and rural utility equals the cost of migration $m(\cdot)$, or

$$U_t^i - R_t^i = m(\dot{L}_t^i),\ m,\ m' > 0\,. \tag{13}$$

Eq. (13) is the specification of migration frictions in the USA (due to rising costs of housing with in-migration in the short run), used in Mueser and Graves (1995) and Rappaport (2000).

In estimation of an urban production relation, if L_t^i is viewed as endogenous, then Q_t^i, $\overline{L}_t^i$, A_t^{Ri}, could be instruments in IV estimation, since they influence L_t^i through (13) but arguably don't directly affect city productivity. The differential (13) also provides a link between current values of city employment and historical variables. In 1990 there is a rural-urban utility gap based on historical allocations within the municipality in question. So we take 1990 population of the rural area as exogenous and the base for much of the migration into the nearby city determining 1997 population. And we have measures of consumer urban amenities in 1990 which we presume are related to 1997 amenities. These are books per capita, doctors per capita, telephones per capita, and roads per capita. These amenities along with the measure of surrounding rural population, we call amenity instruments. An issue might be whether these also reflect production amenities in 1997, but we will test for their exogeneity.

IV Estimation

The characterization and rational we have given for planning and amenity instruments are obviously debatable. In a country where experimentation in market reforms, the strength of the party, the strength of the center and politics vary enormously locally, descriptions of what China is like in 1997 vs. 1990 are highly varied. We have given a rationale but the test lies with statistical results. For IV estimation, we do 2SLS, 3SLS, and what is labeled as GMM. The last adjusts coefficients relative to 3SLS for within year conditional heteroskedasticity (see for example Arelleno and Bond (1991)).[6] 2SLS (where standard errors are adjusted for clustering by city) and 3SLS results are virtually identical, so we will report only 2SLS and GMM. The adjustment for heteroskedasticity affects one or two coefficients strongly, as we will see.

In evaluating exogeneity of instruments, formally we rely on Sargan test results, reported in Section 2.5.2. In this context, these are very sharp in terms of sorting among instruments. We also conducted more informal analyses. Early on, for 1995-97 we calculated a "back of the envelope" city "fixed effect" for average total factor productivity.[7] We then looked at the simple correlation matrix between that fixed effect and our 1990 instruments as an initial guide. Almost all simple correlation coefficients between the fixed effect and the instruments we use are under .14 and the maximum is under .2. Later once we settled on a formulation, to an OLS version of our basic specification we added as covariates all instruments, as a further guide to choice of instruments. If coefficients on model covariates are unaffected by the addition of the instruments, that indicates instruments are not correlated with unobservables that influence values of model covariates. We report (the supportive) results on this in Section 2.5.2.

We note that we do not utilize fixed effect methods; nor do we time difference the equations (to eliminate a fixed effect) and use predetermined level variables as instruments (which from (13) would potentially be strong instruments). Because of measurement error, fixed effect and simple time differenced estimates suffer from magnified attenuation bias. In time differenced equations, instrumental variable estimation suffers from weak instruments. For example first stage regressions of $d\ln(K/L)$ for 1996 and for 1997 on predetermined level covariates have F's of .58 and 4.64 respectively in our best specification. Moreover fixed effects

---

6 Specifically we use the DPD98 Gauss program in estimation. In 2SLS for each city-year observation we use the same 1990 instruments, but in 2SLS we allow the form of the first stage regression to vary by year (as it should from eq. (13)).

7 This is the average residual of the regression ($(\ln(Y_t^i / L_t^i) - \alpha \ \ln(K_t^i / L_t^i) = a_0 + b_1$ time dummy + b_2 distance to coast, where α is fixed at an acceptable number (we used .3 and .4).

impose a very strong structure on the error terms, which may be implausible, rather than just allowing for simple time correlation (as in 3SLS or GMM).

2.4 Data

The description of the data and associated caveats are in the Appendix. Here we note some critical features. First, cities are defined as the urbanized area (called the "city proper") of the prefecture region. The definition of the metropolitan urbanized area generally defines what we want -- all contiguous agglomerated urban activity around the city center -- and is regularly updated. Second, city level data are collected locally by the city statistical bureaus, where in the planning era almost all data used nationally was collected by local statistical bureaus; data on value-added in particular are reputed to be very good.[8] While we have value-added by the three sectors -- primary (which we exclude), secondary ("manufacturing") and tertiary ("services"), we do not have data on sectoral breakdown of the capital stock. And the sectoral breakdown of labor is noisy – workers effectively laid off by manufacturing firms are often carried on their books (and receive stipends) but may work in the service sector on the side. For these reasons, we do not attempt separate production function estimates for services and manufactures, but use the meta-function approach in eq. (8).

In eq. (9), for elements of A_t^i we have variables related to effective technology: accumulated FDI from 1990 to 1997 as a fraction of the capital stock in 1997 and education of the adult population in 1990 as measured by the percent of the population (over age 6 (!)) completing high school ("senior middle school"). We also have a control for public capital – surface area of roads per capita within the city. For capital stock we use 1996 and 1997 original book value. We had no way to construct a perpetual inventory series capital stock, given problems with published investment figures (separation of different forms of capital: housing, working capital, completed investment). Book value of stock is the only alternative. In using original book value, one hopes that inflation rates are approximately offset by depreciation rates.

An issue might be that in slower growing cities, relative to faster ones, capital stock is underrepresented because inflation was higher than the depreciation rate. Another might be that in faster growing cities, employment is relatively more undercounted, inflating output per worker or capital per worker values. In the estimations below we experimented with adding the 1990-95 employment growth rate as a covariate. Treating that covariate as either exogenous or endogenous results in a completely insignificant coefficient for that covariate, with coefficients on other covariates unchanged.

---

[8] We don't wish to implicate people, but we corresponded with several long-term China experts about data quality.

There is one key issue about the capital stock measure. The capital stock data we have is for “independent accounting units”. In the mid-1990's independent accounting unit numbers cover virtually the whole state owned sector but less than half of urban collectives and private firms. The ratio of total activity to that of independent accounting units varies substantially across cities and is negatively correlated with observed capital by construction. Thus we should control for the factor of proportionality. The numbers to do so are less than perfect. For 1996 (and earlier) we know the fraction of total output of independent accounting units to “all industrial enterprises”, which is about .78. What comprises “all industrial enterprises” is not precisely clear. For 1997 we know total value added and are given the value-added of independent accounting units, but post 1994 the numbers for independent accounting units appear low.[9] Nevertheless, we use these ratios as cross-city controls to assess whether independent accounting units, which include most state owned firms, may face a lower opportunity cost of capital than non-independent accounting units, as suggested in Jefferson and Singhe (1999).

For interpretation consider the following simple framework. Assume two sectors in a city, I for independent and N for non-independent, where total city output is $A\left[K_N^{\alpha}L_N^{1-\alpha}+K_I^{\alpha}L_I^{1-\alpha}\right]$. Both sectors have identical technologies drawing on underlying manufacturing and service production, so we have meta-functions for each sector. Inside the square brackets are technologies for producing Y_N and Y_I. The variation in Y_I/Y_N by city reflects the history of local policy adjustments from the planning regime. If labor is internally efficiently allocated in the city (equalized private marginal products) then we can solve $L_N = L\ (Y_N/Y)$ and $L_I = L\ (Y_I/Y)$. For capital, assume nationally sector I faces a cost r_I and sector N a cost of r_N, where we expect $r_I/r_N < 1$. Then markets will yield $K_N = (r_I/r_N)\ K_I(Y_N/Y_I)$. Substituting for L_N, L_I and K_N back in the basic production function we get

$$Y = AK_I^{\alpha}L^{1-\alpha}(Y/Y_I)^{\alpha}\ [(\frac{r_I}{r_N})^{\alpha} + (\frac{Y_I}{Y})\ (1-(\frac{r_I}{r_N})^{\alpha}] \qquad (14)$$

If $r_I/r_N = 1$, then eq. (14) reduces to

$$Y = A\ (K_I(Y/Y_I))^{\alpha}\ L^{1-\alpha} \qquad (14a)$$

---

[9] Recorded value-added of independent accounting units falls by 35% between 1994 and 1995 and stagnates. Capital stock of the same units grows by 38% between 1994 and 1995 and, in fact, overall grows smoothly over the years.

That is, if $r_I / r_N = 1$ then total capital is simply K_I inflated by Y / Y_I. In estimation we will generally adopt a form of (14a) where α is the coefficient for K_I but Y / Y_I is a assigned a coefficient of σ. If (14a) is correct, then in estimation α will equal σ. More generally differentiating (14) we see that the coefficient by for $d(\ln Y) / d\ln(Y / Y_I)$ is

$$\alpha - (1 - (r_I / r_N)^{\alpha}) (1 + (r_I / r_N)^{\alpha} Y_N / Y_I)^{-1} \quad (15)$$

If $r_I / r_N < 1$ then the coefficient (σ) is less than α, whereas if $r_I / r_N > 1$, the opposite is the case. Finally, we can and will directly estimate eq. (14) by non-linear methods.

2.5 Results for Prefecture Level Cities

In this section, we present results for prefecture level cities in Table 2. We divide the discussion into several parts. First we focus on the capital to labor coefficient and scale effects, devoting considerable space to an analysis of scale effects in Section 2.5.1. Section 2.5.2 turns to sensitivity of these scale results to choice of covariates and instruments, and discusses specification test results. Section 2.5.3 discusses results on other covariates, such as access, FDI, public capital, and non-independent versus independent sector capital allocation.

In Table 2, the coefficient on capital in prefecture level cities is high, over .4 and approaching .5 under GMM estimation. For rural industry, as we will see later, the coefficients are much lower, in the traditional .25-.35 range. The only explanation we have is that these prefecture level cities have an industrial base emanating from Soviet style capital-intensive planned production; perhaps these historical extreme capital intensive methods of production persist in making these cities capital intensive in their choice and development of technology, relative to the rest of the country with restricted access to capital. Interacting the K / L ratio with other variables such as the manufacturing to service ratio or city employment scale produces insignificant capital intensity differences across city types.

2.5.1 Scale Effects

Before turning to scale effect results in Table 2, we note that in the raw data and in estimation that does not allow for there to be different city types, it looks like scale effects don't exist. Figure 1 plots the relationship between real output per worker, Y / L, and city scale, L, for cities for 1995-97 data. Figure 1 shows the data points and then draws a curve through the median values of Y / L for ten intervals of L (which is measured in 10,000's). Whether we look at the entire range of L or truncated versions to see more detail, the curve itself is rather flat, although there is modest, visual evidence in the 0-200 range of L of an inverted U-shape. This might be expected. For example, in a market context with free migration and competitive city formation in national land development markets, we would expect to find a flat line. As discussed earlier, with different kinds of cities, the utility, or real income curve each city can offer varies with the manufacturing to service ratio $[MS]$, so that with an inverted U-shape, the peak point to Y/L (and hence to real compensation per worker, $(1-\alpha)\ Y/L)$ shifts as MS changes. Then in equilibrium under free migration, a typical city of each type would operate near the peak point for that type, to offer roughly the same real wage. So a typical city of each type would have an inverted U-shape that peaks near a horizontal line, representing the going national real wage clearing national labor markets. China does not have free migration. But there is no particular reason to expect a specific shape to the graphed lines in Figure 1. The point is that to find inverted U-shapes to Y/L as a function of city scale, we need to control for city type by controlling for industrial composition. If scale effects in eq. (9) are just represented by either $\ln L$ or by a simple quadratic (no $MS*L$ term), the coefficients on scale variables in OLS or IV estimation are zero (Au and Henderson, 2002).

Scale and City Type

The basic scale effects are the coefficients on city employment, employment squared, and employment interacted with the manufacturing to service ratio. While we focus on the IV specifications, OLS results are consistent in pattern, although they give much larger efficient city sizes. We are going to argue Chinese cities are generally undersized; and we do so from the IV results which predict smaller efficient sizes than OLS. In column (4) for the GMM estimates, the employment variable has a strong quadratic effect $\beta_1 > 0,\ \beta_2 < 0$ in eq. (9), yielding an inverted-U shape to $\ln(Y/L)$ as a function of L. And the interactive term with the manufacturing to service ratio, MS has a strong negative coefficient $(\beta_3 < 0)$, so that in the efficient city size expression $((\beta_1 + \beta_3\ MS)/(-2\beta_2)$, as MS rises, efficient size declines.

In Table 3 for the GMM estimates, we present for a range of MS values the corresponding efficient city sizes and the 95% confidence interval on those based on applying the delta-method.[10] We start with MS at its lowest typical value .4 and raise it to MS = 2, where manufacturing activity is double services. Note at MS = 2.16, the numerator in eq. (10) becomes negative and global diseconomies set in. This likely derives from the restriction of the simple functional form, although at MS = 2.16, by most standards in the world for a city the overall service sector is anemic. Efficient city size declines monotonically as relative manufacturing

[10] 2SLS results have efficient size at 3.1m, 1.8m, 1.0m and .17m for MS =. 4, 1.0, 1.4 and 1.8 respectively, generally lower size values. Confidence intervals are also a little wider given large standard under the clustering calculations. Regular standard errors for 2SLS are smaller than the GMM ones.

content rises, from 3.1 million workers at $MS = .4$ to .27 million workers at $MS = 2$. Note the employment variable is measured in 10,000's, but Table 3 is written in 1,000's.

The Table also reports the 95% confidence intervals. At the lower tail the intervals are wide, $\pm$ 1.0 million workers for $MS = .4$. They narrow in absolute terms although not relative terms as MS approaches typical values of 1.2 to 1.4 and then widen absolutely and relatively as MS rises further. Error bands are appropriately wide, given the crude control for industrial composition, as well as possibly the inability to adjust value-added, or GDP numbers for purchasing power parity discrepancies, although these seem to be minimal in China (Johnson (2001)).

Where do Chinese cities stand relative to efficient sizes? About 11% of cities lie beyond the peak points, but almost none significantly so. Another 13% have no inverted-U shape (MS is too high). Overall then, 76% of cities are to the left of their peak points and potentially undersized. Of these, 56% of all cities are below the lower band on the 95% confidence interval -- that is 56% of cities are statistically significantly undersized. This is startling and indicates that indeed migration restrictions are inefficiently limiting city sizes. Later we will examine the welfare losses from this, but here we explore this result.

Alternative Approaches

The calculation of peak points may be sensitive to specification and estimation method. For example 2SLS results in column (3) in Table 2 suggest modestly smaller efficient sizes. -- 1.9m at MS 1.0 (vs. 2.0m), 1.1m at MS = \$1.4 (vs. 1.3m), and .70m at MS = \$ 1.6 (vs. .97m). But the basic conclusion is unaltered: only 11% of cities lie to the right of their peak points. The rest are undersized, or have too high a MS to calculate a peak point. The real issue however is whether the functional form is too strong a restriction, masking possible non-monotonicities in how the peak shifts as MS changes or more subtle changes in shape. We tried higher order terms in L and MS, getting insignificant coefficients. Another way to proceed would be to break MS into intervals and to estimate a separate quadratic in city scale for each. We did that for OLS which always yields larger efficient sizes than IV estimation. Under OLS for septiles of MS, efficient size declines monotonically with MS, except for the lowest septile.[11]

To test for non-monotonicity of declining efficient size with MS under IV estimation, we split the sample around the MS value of 1.3 in 1990, an "exogenous split", separating off the

[11] For the first septile (MS < .844), efficient size employment is 2.3m. For the second (.845 \$ 1.10), it jumps to 4.3m; and then it declines monotonically thereafter as we move from the second to third and on to the seventh septile (MS > 2.18), taking values of 2.4m, 1.4m, 1.3m, .60m and .28m.

lower 25% of the sample. Then for 75% of the sample where MS in 1990 $\geq$ \$ 1.3, we estimated one form to the quadratic ($\beta_1,\ \beta_2,\ \beta_3$) in (9) and (10) for 1996-97. For the 25% in the lower tail we estimated its own peak (just β_1 and β_2). Estimation is by 2SLS, separately for each sub-sample, with considerable loss of efficiency. For the lower tail, coefficients (and standard errors) are β_1 = .00505 (.00297) and β_2 = -.0000172 (.0000107), so the peak is at 1.5m (recalling the variable is in 10,000's). A typical MS value in 1997 for this sample is .9. For the upper tail, precision is poor; but β_1 = .00425(.00398), β_2 = -.00000245 (.00000688), and β_3 = -.00362 (.00178). The peak at MS = 1.0 (in 1997) is 1.3m. Thus the results under a split sample with 2SLS are consistent with peak employment declining monotonically as MS rises, as in Table 3.

2.5.2 Sensitivity and Specification Tests

On sensitivity we noted two results already. Adding the 1990-95 city labor force growth rate as a (either exogenous or endogenous) covariate leaves the results in column (4), Table 2 unchanged, with a zero coefficient on the added covariate. Second, dropping the variable identifying effects of city type on scale, $L * MS$, yields zero scale effects and raises the Sargan test statistic so much that the p-value drops from .98 to .27. Controlling for city type, as expected is critical to identifying scale effects.

How sensitive are the scale effects to dropping other covariates? Dropping the controls for Y_I / Y has little effect. In fact dropping all variables except $\ln\ (K / L),\ L,\ L^2,\ L * MS$, distance to the coast and regional and time dummies, results in the following coefficients (all significant) for these variables: .54, .014, -.000017, and -.0062. These are to be compared to column (4), Table 2 coefficients of .49, .013, -.000017 and -.0058. But the change in calculation of peak points for different MS values is under 10%, (given the slight declines in both the L and $MS * L$ absolute value of coefficients). Dropping all these other covariates of course is not desirable. It raises the Sargan test statistic and drops the p-value from .98 to .36.

What about the choice of instruments? The procedure we follow is to look at the effect on OLS values of covariates of adding instruments as covariates, as a rough guide as to problematic instruments. Then we estimate the full model and rely on Sargan test results. The Sargan test is a joint test on model specification and orthogonality of instruments to error terms in the moment condition. We have already seen the test statistic deteriorates badly (i.e. rises) if we drop various variables in column (4), Table 2. Here we see how choice of instruments affects it.

The full instrument list of planning and amenity variables yields a very low Sargan value, with a p-value of .98, indicating an excellent specification and choice of instruments. Given the full list of amenity and planning variables, if we add as instruments the excluded absolute labor force and educational attainment in 1990, the Sargan test statistic rises so much the test fails (p-value of .023). For instruments, in terms of experimenting with adding instruments as covariates to OLS regressions the most troublesome are the planning variables, K/L in 1990 and *MS* in 1990 * business district area (in 1990). They have fairly significant coefficients and most importantly in OLS have strong impacts on the $MS * L$ coefficient (making it zero). However the effect on GMM coefficients of excluding these variables as instruments is absolutely minimal; and the Sargan test statistic actually deteriorates slightly with their exclusion. Compared to the column (4), Table 2 coefficients on $\ln(K/L)$, L, L^2, and $MS * L$ of .49, .013, -.000017, and -.0058, the coefficients with the dropped instruments are .54, .012, -.000016, -.0058. But the drop in both the L and L^2 coefficients leaves peak points again changing by less than 10%. We felt the column (4), Table 2 results were better, especially for the capital coefficient.[12]

2.5.3 Other Factors Directly Affecting Productivity.

In Table 2 a variety of other factors are considered which might affect productivity. First, time and regional dummies are insignificant, indicating for the latter that we have already captured regional differences through measured covariates. Second the manufacturing to service ratio does not and in our mind should not affect output per worker. Compositional effects in the meta-function approach affect scale economies (by city type), but have no other effect on productivity (or relative GDP evaluation) per se in a market context.

We then consider two factors which should affect productivity directly, education of the population and the relative extent of FDI. Education is measured by the percent of people six years of age and over who have completed high school (a similar college measure has no impact), in 1990. In the '90-'92 OLS regression that variable is significant but by 1996 or 1997 it is not because it may poorly measure education that has evolved by migration and age cohort turnover of the relevant population. While coefficients are all positive they are not significant. We do note for the record that a one-standard deviation increase in education for the GMM estimate raises output per worker by 5%, a noticeable effect.

For FDI, the idea is that local FDI involves technology transfer, or knowledge spillovers. We choose to measure FDI intensity by the accumulated (from 1990) sum of FDI investments relative to the capital stock -- approximately the fraction of the capital stock from FDI.

[12] Finally we note if we use just the amenity list plus FDI or not and 1990 area of the business district (a historical planning variable) as instruments, while we don't have enough instruments to estimate the full model, in the stripped version discussed above where the only covariates are K/L, L, L^2, and $MS * L$ as well as distance to the coast and regional and time dummies, with full instruments the coefficients are .54, .014, -.000017 and -.0062, while with just amenity instruments the coefficients are .61, .013, -.000014, and -.0082.

Coefficients move around and the coefficient under 2SLS is insignificant. However adjusting coefficients for heteroskedasticity makes a big difference. Under GMM a one-standard deviation increase in the FDI ratio raises output per worker by 8.5%. This would suggest cities receiving FDI have an advantage. The FDI result also perhaps lends support to the practice of Chinese subsidization of FDI, since it brings knowledge spillovers. The Chinese originally tried to limit FDI investment to certain coastal cities and clearly that may have advantaged those cities relative to hinterland ones.

The public infrastructure result is classic. Roads per capita has a strong effect in OLS estimation; but an insignificant effect in IV estimation, with a zero coefficient under GMM. So public infrastructure investments and productivity are positively correlated; but investments don't cause higher productivity. That is the usual interpretation in the infrastructure literature. However the (tax) costs to the city of local road construction are not represented per se. In the public finance literature, a zero coefficient in a meta-production function is consistent with there being an optimal level of infrastructure where cities have optimally traded off their costs and benefits of changing infrastructure and further investments would not raise city net output. (A positive coefficient would represent underinvestment, meaning that infrastructure is too low, so given its costs and benefits in net it should be raised on average in cities.)

Access.

Finally, before turning to capital market issues, is access. We constructed several pieces of information on access -- straight-line distance to the nearest coastal point, straight-line distance to the provincial capital, whether a city is on a major highway (on a late 1990's map), whether a city is on or near navigable waterways, and whether a city is on a rail. Essentially all prefecture level cities are on some rail system. Being on a late 1990's highway doesn't indicate much for 1996 and "major" does not mean divided highway. The only variable with a consistent, stable effect is the first -- distance to the coast.

In Table 2 column (4) a one-standard deviation improvement in access (reduction in distance to the coast) raises Y/L by 15%. This huge effect in China with its very poor inter-regional trucking systems could reflect just pure transport, but it could also potentially reflect technology diffusion from abroad (after controlling for FDI). Since the effect is so large, it seems important to look more closely. First, the effect is specified as a simple iceberg where $Y\ (1-\phi T)$ in logs is approximated by $\ln\ Y - \phi T$. Small movements away from the coast could be the most costly, where for long distance hauls additional kilometers are cheap. So we tried a quadratic form, which we will report on momentarily.

Second, work in international trade (Evenett and Keller (2001)) as well as on cities in the USA (Black and Henderson (2003)) suggests domestic market potential is a key measure of market demand and access, as well as access to the coast and international markets. The issue is important here because population in China is historically heavily concentrated in the fertile coastal regions, compared to the more arid hinterland regions. So access to the coast and domestic market potential are positively correlated for historical-geographic reasons. To try to separate the two, we constructed a traditional measure for each city of market potential where for city i

$$MP_i(t) = \sum_{j=1, i \neq j} \frac{VA_j(t)}{d_{ij}} \qquad (16)$$

$VA_j(t)$ is value-added in all other prefecture and county city locations to city , covering almost the whole country. d_{ij} is the straight-line distance from i to j, discounting market (income) demand by remoteness of locations. For the own prefecture city, its rural VA (in the surrounding rural part of the prefecture) is included with no discounting, although that has little effect on the results. In IV estimation we instrument with 1990 market potential.

Results are in Table 4, based on the column (4), Table 2 specification for all other variables. First a quadratic form to access seems not unreasonable. In column (1) access effects die out, although at two-standard deviations above mean distance, the marginal effect is still -.03 (recall in Table 2 the constant marginal effect is -.025). At maximal distance around 45, the marginal effect is -.012. In column (2), adding market potential to a linear distance term cuts the access coefficient in half relative to Table 2. Now a one-standard deviation improvement in access raises Y/L by 7%; while a one-standard deviation improvement in market potential raises Y/L by 15%. Similarly in column (3), adding market potential to the quadratic access specification weakens the role of access.

These are important results. Simple economic geography claims on the importance of being on the coast (Rappaport and Sachs (2002)) tend to ignore the role of historical factors, such as population accumulation near the coast. Any dominance of coast regions derives from two factors. First is self-sustaining market demand/potential for coastal products, deriving from history. Second is on-going benefits of pure coastal access. For regions of China such as Szechuan where there are high inland population concentrations, that domestic market demand could have strong self-sustaining potential.

Capital Allocation

Finally, we turn to the issue of capital allocation in Table 2. Recall the capital stock measure is for independent accounting units, where we control for Y/Y_L, the ratio of total to independent accounting unit activity. As noted earlier, for 1996 we control for a sales ratio; for 1997 a value-added ratio. We focus on the 1997 control, given the output measure is a value-added one. If independent and non-independent accounting units face the same opportunity cost, the coefficient on this control, σ, should equal α, the coefficient for $\ln(K/L)$ (based on eq. (14a) discussion). In Table 2 for OLS and GMM, σ is about .16, while for 2SLS it is .25. For GMM, σ is significantly less than α, indicating in eq. (16a) that $r_I / r_N < 1$, or independent accounting units face a lower opportunity cost of capital. If we use eq. (15) (as the value σ) and solve it for an approximate value of r_I / r_N (for Y_n / Y_I = 1.5 in 1997) we get r_I / r_N = .18. That indicates a pretty extreme opportunity cost differential, but is not inconsistent with results in the last section on the marginal product of capital in different uses in China. Two-stage least squares results suggest r_I / r_N = .37, still a huge differential. We also estimated eq. (14) by non-linear ordinary least squares for just 1997 to directly estimate r_I / r_N and get a value of .16 (with a standard error of .09). However, generally, these non-linear two-stage least squares results are not robust.

2.6 County Cities

We now turn to county cities where we have more limited data. There is less information on base characteristics of cities especially for instruments from 1990. There are no data on these cities for 1992 and 1997. Areas of county cities do not distinguish the urbanized area from the rural area also governed by the county city. Finally for these cities there is extensive entry, so for many cities we do not have 1990 variables to use as instruments. Our before and after regime switch in OLS estimation is based on just two years each, '90-'91 and '95 to '96, . IV estimation is based on just 1996 for the 236 (of 432) cities for which we have 1990 variables as instruments. As for prefecture level cities, if 1995 observations are added to the IV estimation, Sargan specification tests fail. Given the data limitations, we focus less on results for county cities; but what we find is of interest.

Results are in Table 5. As in Tables 2, OLS results for '90-'91 and '95-'96 are quite different; and IV results differ from OLS ones. Here GMM and 2SLS results are almost the same. The discussion is focused on column (4), the GMM results. In Table 5, the main result concerns scale economies. We were unable to estimate the non-linear scale-composition version of the model in eq. (9), in the sense that all the relevant coefficients are completely insignificant. However the results are instructive. With these insignificant coefficients, the shape is for output per worker to decline initially, bottom out (at a point well below mean sample size), and then rise unbounded.[13] So we turn to a specification, with unbounded scale economies and no composition effects, entering scale in the traditional fashion as $\ln L$, whose coefficient is an elasticity. Here that elasticity is large, almost .15, and significant. Doubling the typical non-prefecture level cities non-agricultural employment would cause city output per worker to rise by 11%, a big gain.

In general the suggestion is that county level cities are a cut at the bottom end of the city size distribution (a cut made by Chinese political hierarchy notions in cities, separating prefecture and county cities). There are not enough large county cities to define the inverted U -shape function, where some cities operate near and beyond peak points. That is, as a group county cities are all to the left of their peak points, the sizes where output per worker is maximized. As a result, we estimate unbounded scale effects. The same issue will arise in the rural TVE sector; and we will discuss the implications of this finding in the last sector.

In terms of other covariates, the coefficient on the control Y/Y_I is not significantly different from α; and the suspicion is that in these cities no firms have great access to capital markets. Favored firms are in the higher order prefecture level cities. For access and technology variables, education and now surprisingly access to the coast are insignificant. The access result may suggest that county cities simply serve very local, markets, so coastal access is unimportant. Market potential for non-prefecture level cities has a large coefficient (.00106), if that variable is added to the specification, although it is statistically weak (standard error of .000841). As with prefecture level cities, public capital under IV estimation has no effect, perhaps indicating in general it is satisfactorily provided. Only the FDI ratio has a big impact. A one-standard deviation increase in the FDI ratio for non-prefecture level cities (.041) results in a 15% increase in Y/L.

3. Productivity in the Rural Sector

The rural sector in China has key features noted in the introduction. Overall it appears that there is a surplus of labor within the sector due to restrictive migration policies. While under the “leave the land but not the village" policy, workers have moved out of farming to rural industrial production, there remains a surplus of labor in agriculture, probably felt most strongly in more remote regions where rural industrialization is not very competitive and residents can't migrate to richer rural and agricultural areas. Second, as we will discuss momentarily, the rural

[13] Coefficients (and standard errors) for β_1, β_2 and β_3 in eq. (9) are -.0137 (.0234), .000161 (.000239), and .00950 (.0129).

industrial, or town and village enterprise [TVE] sector faces constraints that restrict TVE development.

The shift to the TVE sector has been dramatic. In 1981 TVE's accounted for 6% of all industrial gross output, while the urban state owned and collective sector accounted for the balance. By 1997, TVE's accounted for 28% of all gross industrial production with the urban state owned and collective share dropping from 93% in 1981 to 40%. Of course apart from the TVE sector's share growing from 6 to 28%, the absolute growth is astronomical. However, the rural industrial sector faces two major constraints. First, as noted earlier, the sector is capital constrained. Second, the TVE sector is constrained by forced factor immobility within the rural sector, not only across provinces, but also, in particular, within provinces.

Rural industrial activities can't relocate to appropriately spatially agglomerate to exploit the local scale externalities inherent in agglomeration, whether they are local information spillovers or firm and intra-industry local specialization in production and exchange of parts, components, and services. Even today, under so-called privatization, “former” TVE's cannot readily be bought and sold, nor shares transferred among individuals. People in one location cannot readily buy and relocate to their location, a TVE from another location. Similarly if workers leave their township permanently they lose their rights to agricultural land and the in-kind (e.g., housing) distribution of profits of local TVE's, with a limited ability to garner such rights elsewhere. While some provinces such as Zhejiang argue have they rural industrial cluster policies, little geographical clustering and relocation has yet been permitted.

If TVE's could spatially relocate and cluster, of course, that would transform the rural sector. With agglomeration the “rural” sector would become two distinct parts, a rural agricultural sector with supporting services and industry and an urbanized sector with some of the many townships transformed into cities. In fact, as noted earlier, the rapid entry of new county cities noted above consists in part of rural townships that have grown into cities, so although constrained, some transformation has occurred.

Following Yang and An's (2002) insightful work, we analyze the rural sector as it largely operated in the 1990's -- local townships with integrated agricultural and non-agricultural sectors. A town's function in part is to allocate resources to farm activity (farming), to non-farm agricultural activity which is mostly animal husbandry and fisheries, and to TVE and related activities covering manufacturing, services, construction, transportation and so on. Our data are limited. We have provincial level data for three years, 1995, 1997, and 1998, on value-added in the agricultural and in the TVE sector. We have specific TVE employment numbers, but agricultural employment is difficult to measure, being simply defined as persons “primarily engaged” in agriculture, where at harvest, agricultural employment may be augmented by TVE workers. While we have capital figures (original value of assets, with similar results for horsepower) for TVE and agricultural production, some TVE capital equipment (e.g. TVE trucks) may also serve agriculture. Thus we rely on a meta-production function approach for value-added in the rural sector, as in the urban sector. This also follows Yang and An (2002). However we also examine the TVE sector separately, especially since scale effects there are better defined.

Our work adds two components to the literature. First they are focused on township level scale effects, not a focus of other studies. Second, estimates of factor intensity coefficients even for labor for agriculture and the overall rural sector in China are all over the place (Lin (1992a) and (Putterman and Chiacu (1994)). While our estimates of labor share will correspond to those of Yang and An (2002), they generally find zero returns to land and almost zero returns to capital. We believe our results and the formulation of scale effects, to be more plausible.

For the overall rural sector, value-added in province p, Z^{Rp}, is postulated to be a function of capital K^{Rp}, sown land S^{Rp}, labor N^{Rp}, and education E^{Rp} in the rural sector. We estimate

$$\ln(Z_t^{Rp} / N_t^{Rp}) = \beta_0 \ln(K_t^{Rp} / N_t^{Rp}) + \beta_1 \ln(S_t^{Rp} / N_t^{Rp}) + \beta_2 E^{Rp} + \beta_3 \ln N_t^{Rp} + \delta_t + \delta^{Rp} + \varepsilon_t^{Rp} \quad (17)$$

In (17), given a meta-production function with underlying labor allocation rules, the “ social return” to labor is $[(1-\beta_0-\beta_1)+\beta_3]\, Z^{Rp} / N^{Rp}$, where scale effects derive from adding a worker to the rural sector in a province. While adding labor to the provincial rural sector with a fixed number of townships indicates the scale benefits of adding a worker to some township somewhere, it is a little vague. For TVE's, it will be possible to be more specific.

In estimation of (17), the key issues are error structure and data quality. Input choices are not exogenous to provincial effects, δ^{Rp} representing long term weather conditions, soil fertility, local political climate and extent of economic and social liberalization. We do have regional dummies, distinguishing the coast, central and western regions. However at a minimum, we would like to employ provincial fixed effects, but annual variations in all inputs (especially land) are poorly measured and fixed effect results are insignificant with a zero return to land. We rely on two types of estimates. First are time and region fixed effect estimates, with provincial random effects for the three years of data. Second are IV estimation results using instruments from 1981. Instruments are values of 1981 variables that meet two criteria. They are determined by political and historical forces under Russian style industrialization, planning, and the political chaos from the mid-1960's on, which are independent of current unobserved influences on rural sector productivity. Second they have a strong impact on influencing TVE development and agriculture populations today.[14] First stage regressions indicate a strong set of instruments.

The Overall Rural Sector.

The results for the meta-production function for the rural sector are in Table 6. In the random effect results, the residual labor share coefficient is .53 (=1 -.379 -.094) and the labor scale effect is .10, although the latter is not statistically strong. The two sum to .63, which is within the realm of Yang and An's results on the social marginal product of labor (.52 to .61). The main difference between this result and Yang and An's is that we have significant and reasonable coefficient on capital. The IV results are stronger, and yield an overall labor coefficient of .66, with very significant scale effects of .16. In both cases, education affects productivity very strongly where a one standard deviation in percent in the rural population completing high school (2.5) results in a 27% increase in output for IV results. This result supports Yang and An's hypothesis that education improves overall productivity, by improving allocation decisions between the urban and rural sectors. In Table 6 the interpretation of scale effects is difficult since they are at the provincial level. Attempts to refine the scale measure to average township size in a province produced insignificant coefficients for the overall rural sector. For TVE's, however we can be more precise.

TVE's

TVE output which includes both manufacturing and services is modeled as in (17) after dropping the land variable. Random effects and IV results on TVE's are in Table 7. There are two sets of results. The first is for a generalized scale effect – adding labor to the TVE sector in a province. The second is more precise – the effect of increasing average local township TVE sector size in a province, or the degree of local industrial agglomeration. In general, as we move from random effect to IV results, capital's coefficient rises, but the IV results vary by scale formulation. Results are consistent with Jefferson and Singhe (1999, Table 7.4) who find a capital coefficient of .29, mid-way between the two IV numbers in Table 7. For scale effects, the generalized IV formulation has a scale elasticity of .13 and a social marginal product of labor coefficient of .77 to be compared with Jefferson and Singhe's of .97. Under the more specific

---

[14] Instruments are (1) the number of communes, and their population prior to any free migration reflecting the base from which populations are adjusting today; (2) provincial industrial output and the share of light industry in that outside the rural sector, that formed an early basis for local demand for TVE industrial parts and components; (3) enrollments in regular secondary schools and institutions of higher learning that influence adult educational attainment today; and (4) arable land defined under Mao's "grain everywhere" policy that defines land that can be planted with grain but not necessarily economically viable today (again a base from which China is adjusting in determining what is productive land today).

township scale effect, IV results suggest a scale elasticity of .30 and a labor social marginal product coefficient of 1.04. What is generally apparent is the very high level of scale effects in our results, as well as in Jefferson and Singhe (who focus on plant size effects). This would suggest that the costs of constraining agglomeration in the TVE sector are very high; and that we have "unbounded" scale economies in the relevant range of average local (township) industrial employment.

Finally we note that rural education has no direct positive impact on TVE productivity. That combined with the positive overall effect of education on rural productivity suggests education's effect works through improved allocation decisions for rural sector activities (Yang and An (2002)).

4. Productivity in the Urban and Rural Sectors

This summary section considers the allocation of resources across sectors in China. The analysis is divided into two parts. First we look at the issue of agglomeration within sectors, holding relative capital allocations (the capital to labor ratio in each sector) constant. The issue is: what are the magnitudes of costs of under-agglomeration within sectors and what are the implications of permitting increased agglomeration? Second we will examine the more static, traditional issue of sectoral allocation, where the sectors are prefecture level cities, TVE's and the overall rural sector.

4.1 Agglomeration Gains Within the Spatial Hierarchy

The different sections in this paper argue that spatial agglomeration in China is insufficient. TVE industry and county cities are all undersized, with as yet unexhausted scale economies. Table 8a shows the gains from doubling local scale such as the size of a county city or total TVE employment in a township, according to scale elasticities from Tables 5 and 7. For TVE's the effects are enormous -- a 23% increase in output per worker, holding input ratios fixed -- but even for county cities the 11% number is very large. It is not clear what the local scale of “rural industry” or county cities should ultimately be. In particular allowing spatial agglomeration of TVE's would transform townships into cities. It would seem reasonable to presume that, if county cities or townships were permitted to grow with population reclustering, they would become subject to the same econometric representation for scale effects as prefecture level cities. In the process a number of less competitive townships and cities would "die out". The whole population geography of the country would be changed.

For prefecture level cities where there is a huge size range, we are able to estimate the size where output per worker peaks as industrial composition varies. Based on IV results in Table 2, Table 8b tells us the percentage gain in output per worker (holding input ratios and other city characteristics fixed) of moving from a size that is 50, 40, 30, and then 20% below the size where Y/L peaks, to the size where it peaks. The gains are independent of industrial composition (although the peak sizes differ). If a city is half its peak size, the gain is 35%. But the gains drop off quickly as the gap shrinks. So if a city is only 20% below its peak size the gain to moving to peak size is only 4%. However the majority of Chinese prefecture cities are at least 50% below their estimated peak size.

In summary, results suggest that the losses within sectors from under agglomeration are enormous. Doubling sizes of many current rural industrial or urban agglomerations would substantially increase national output. Having said that, there is a timing problem. One wouldn't want to double city sizes overnight. First there is the issue of timely adjustment of public infrastructure and housing capital and up-grading managerial expertise in cities. Moreover there would not be instantaneous massive migration, given behavior of migrants, where most moves are local and involve youth soon after high school graduation age. The idea, looking also to the next section, would be to allow free migration across cities and to introduce freer rural-urban migration so as to increase the annual rate of labor force growth of successful cities. The latter could be done in stages—first free intra-province migration and then free inter-province migration. While there is a view that the general Chinese gradualist approach has served the country well, for migration it has simply been vastly overdone. The next section looks at inter-sectoral allocations.

4.2 Allocation Gains across the Spatial Hierarchy

A focus of Chinese policy has been the gain from permitting increased rural to urban migration. The idea is that rural-urban migration restrictions leave a gap between the marginal product of labor in rural versus urban areas. We believe this focus misses the main point. Free migration would redefine the rural sector with some townships becoming cities and many others dying out. In the urban sector similarly some county cities would become very large and a number of less competitive cities would diminish over time.

Nevertheless looking at the sectoral allocations of factors given current sizes does raise important issues. We will compare the marginal product of labor and capital in prefecture level cities, in the overall rural sector, and in TVE's, at the time period of our data. With scale economies, we have private and social market products. For the rural and TVE sectors, we use a national weighted average of marginal products: the private and social marginal products of labor are respectively $(1-\beta_0-\beta_1)\ \bar{Y}/\bar{L}$ and $(1-\beta_0-\beta_1+\beta_3)\ \bar{Y}/\bar{L}$ where $\bar{Y}$ and $\bar{L}$ are the national value-added and labor for the relevant sector ($\beta_1 = 0$ for TVE's). These are listed in Table 9a for the overall rural and TVE sectors using the coefficients from column (2) in Table 6 and column (4) in Table 7 respectively, for 1997 data.

For prefecture level cities, accounting for scale non-linearities in defining "average" sector marginal products is difficult: most cities are grossly undersized but there is also a tail of very large oversized cities with low or negative social marginal products. We give two calculations. First is the private marginal product of labor under Table 2 column (4) estimates, which would equal the social marginal product if all cities were of efficient size. This is quite low. We then calculate the social marginal product for a representative city with an MS = 1.4 set a size which is half efficient size. These results are also in Table 9a.

Table 9a indicates the overall rural surplus of labor, with a social marginal product in the rural sector about a third of that in the urban sector. For TVE's versus the urban sector, the picture is less striking. Private marginal products in the TVE sector are less than in the urban sector and the social marginal product of labor in the typical undersized city exceeds that of TVE's. But the gaps are surprisingly modest.

The real issue are the caveats. The first involves education. If the rural sector had the national average education (12.5% high school) vs. its current level (7.4%), the estimates say output would increase by 70%. The second involves capital. Given current sizes of cities and the provincial rural and TVE sectors, we calculate the marginal products of capital. For the rural and TVE sectors it is easy: $\beta_0\ \bar{Y}/\bar{K},$ where $\bar{Y}$ and $\bar{K}$ are the national value-added and capital in the relevant sectors. The calculations of capital's marginal product in Table 9b suggest very high numbers for the rural and TVE sector, where we believe the TVE number are based on more reliable point estimates of coefficients.

For cities, we have two sets of results. First we assume capital is allocated to equalize marginal products $(r_I = r_N)$ across independent and non-independent accounting units. Then the marginal product of capital is $\beta_0\ \bar{Y}/\bar{K},$ where $\bar{Y}$ and $\bar{K}$ are national value-added and estimated national capital in the prefecture city sector. $\bar{K} = \sum_i (K^i Y^i / Y_I^i)$ where each city's K is inflated by Y/Y_I based on eq. (14a). In this case, Table 9b suggests a relatively low return to capital of .14 in the urban sector.

However the results in Table 2 indicate $r_I \neq r_N$. For that case when an additional unit of capital is added to a city we assume it is added and split across sectors so Y/Y_I is unchanged. Then one can show that a weighted average of marginal products of capital equals

$$\beta_0 \overline{Y} / (\sum_i (K^i \ (1 + \frac{r_I}{r_N} \frac{Y_N^i}{Y_I^i})) .$$

As noted earlier to solve for (the national) r_I / r_N (=.176), we set eq. (15) equal to the estimated coefficient of $\ln \ (Y/Y_I)$ in Table 2, using a Y_N / Y_I in that equation equal to the national ratio in 1997 for prefecture level cities. Now the marginal product of capital in the urban sector in Table 9b rises to .30. But still the estimated marginal products of capital in the rural and TVE sector far exceed that in the urban sector. For the TVE versus urban sector, the gap is similar to that of Jefferson and Singhe (1999): urban returns are 25% less the TVE returns.

These intersectoral comparisons in Table 9 are suggestive. But an overall assessment of what would happen if labor and capital markets were completely freed up is beyond our reach, because so much reagglomeration would occur within the current sectors as well as across. The Table 8 and 9 numbers relate to marginal changes.

References

Arellano, M. and S. Bond (1991), "Some Tests of Specification in Panel Data: Monte Carlo Evidence and Application," Review of Economic Studies, 58, 537-559.

Au, C.C. and J.V. Henderson (2004), "Estimating Net Urban Agglomeration Economies with an Application to China", Brown University.

Au, C.C. and J. V. Henderson (2002), "How Migration Restrictions Limit Agglomeration and Productivity In China", NBER Working Paper #8707.

Black, D. and V. Henderson (2003), "Urban Evolution in the USA," Journal of Economic Geography.

Cai, Fang (2000), Zongguo Liudong Renkou Wenti (The Mobile Population Problem in China), Henan People's Publishing House: Zhengzhou.

Chan, H.W. (1994), Cities With Invisible Walls, Oxford University Press: Hong Kong.

Duranton, G. and D. Puga (2001), "Nursery Cities," American Economic Review, 91, 1454-1463.

Duranton, G. and D. Puga (2004), "Micro-Foundations of Urban Agglomeration Economies", J.V. Henderson and J-F. Thisse (eds.) Handbook of Urban and Regional Economics, Vol. 4, Amsterdam, North Holland.

Evenett, S and W. Keller (2001), "On Theories Explaining the Success of the Gravity Equation", Journal of Political Economy..

Fujita, M. P. Kurgman, and A. Venables (1999), The Spatial Economy, MIT Press.

Hansen, L. (1982), "Large Sample Properties of Generalized Methods of Moments," Econometrica, 50.

Henderson, J.V. (1974), "The Size and Types of Cities," American Economic Review, 64, 640-656.

Henderson, J.V. (1988), Urban Development: Theory, Fact and Illusion, Oxford University Press.

Jefferson, G. and I. Singhe (1999), Enterprise Reform in China: Ownership Transition and Performance, Oxford University Press: New York.

Johnson D., Gale (2001), "Have The Urban-Rural Disparities Increased Since 1978 in China?" University of Chicago mimeo.

Kolko, J. (1999), "Can I Get Some Service Here? Information Technology, Service Industries and the Future of Cities," Harvard University mimeo.

Lin, J.Y. (1992a), "Rural Reforms and Agricultural Growth in China," American Economic Review, 82, 34-51.

Lin, J.Y. (1992b), The China Miracle: Development Strategy and Economic Reform, Hong Kong: Chinese University Press.

Prud'homme, Remy (2000), "Estimating the Benefits of Urbanization in China," Report to World Bank on China's Urbanization Policy.

Putterman, L. (1993), Continuity and Change in China's Rural Development: Collective and Reform Eras in Perspective, Oxford University Press: New York.

Putterman, L. and A. Chiacu (1994), “Elasticities and Factor Weights for Agricultural Growth Accounting: A Look at Data for China,” China Economic Review, 5, 191-204.

Marshall, A. (1890), Principles of Economics, London: MacMillan.

Mueser, P. and P. Graves (1995), “Examining the Role of Economic Opportunity and Amenities in Explaining Population Redistribution,” Journal of Urban Economics, 37, 176-2000.

Rappaport, J. (2000), “Why Are Population Flows So Persistent?,” Federal Reserve Bank of Kansas City mimeo.

Rappaport, J. and J. Sacks (2001), "The US as a Coastal Nation", Federal Reserve Bank of Kansas City, Working Paper No. RWP 01-11.

Yang, D.T. and M.Y. An (2002), “ Human Capital, Entrepreneurship and Farm Household Earnings,” Journal of Development Economics, 68, 65-88.

Table 1. The Chinese Urban System

	All cities present in 1990	Prefecture level cities present in 1990	County (non-pref.) cities present in 1990	New county cities
Avg. pop. 1996 (thousands)	811	1050	589	706
Pop. growth 1990-96	12%	15%	7.9%	n.a.
Avg. non-agri. employ. 1996 (thousands)	327	511	158	183
Non-agri. employ. growth 1990-96	29%	26%	40%	n.a.
% pop (over 6) completed high school in 1990	14.6%	21.7%	12.8%	8.8%
Distance to coast	7.1	6.3	9.4	5.4
Avg. output per worker in 1996 (1990 yuan)	8942	10198	7802	9229
Growth in real output per worker 1990-96	57%	53%	62%	n.a.
Avg. manu. service ratio 1996	1.52	1.59	1.47	1.59
Change in manu. serv. ratio, 1990-1996	-25%	-32%	-8.2%	n.a.
N	459	220	240	193

Table 2. Modeling City Scale Effects
(dependent variable $\ln\ (Y/L)$)

	(1)	(2)	(3)	(4)
	OLS	**OLS**	**IV Estimation '96-'97**	**IV Estimation '96 – '97**
	('90 – '92)	**('96-'97)**	**2SLS**	**GMM**
ln (capital/employ.)	.324**	.452**	.428**	.492**
	(.0645)	(.0398)	(.108)	(.0759)
employ.	.00332**	.00345**	.0165**	.0126**
	(.00152)	(.00108)	(.00565)	(.00405)
employ. sq.	-.000000393	-.00000327*	-.0000214**	-.0000166**
	(.00000184)	(.00000147)	(.00000970)	(.00000762)
(MS) * employ.	-.00182**	-.00151**	-.00845**	-.00583**
	(.000667)	(.0007078)	(.00278)	(.00172)
MS	.0596	.0548	.233	.0650
	(.0510)	(.0344)	(.173)	(.0884)
% high school	.00761**	.00475*	.00603	.00621
	(.00298)	(.00255)	(.00561)	(.00492)
distance to coast	-.0131	-.0227**	-.0282**	-.0246**
	(.00935)	(.00579)	(.00894)	(.00705)
ln (output/output of	.00428	.386**	.470**	.291**
ind. acct. units) 1990-96_	(.115)	(.101)	(.213)	(.0970)

Table 2. Modeling City Scale Effects – (Continued)
(dependent variable $\ln\ (Y/L)$)

	(1)	(2)	(3)	(4)
	OLS	**OLS**	**IV Estimation ’96 – ’97**	**IV Estimation ’96 – ’97**
	(’90 – ’92)	**(’96-’97)**	**2SLS**	**GMM**
ln (VA/VA of ind.	n.a.	.167**	.304**	.163**
acct. units 1997		(.0629)	(.151)	(.0593)
FDI/Capital	1.66**	.910**	.491	1.254**
	(.743)	(.279)	(.629)	(.353)
ln (roads per capita)	.0541*	.108**	.126	.0204
	(.0296)	(.0301)	(.138)	(.0924)
time effects	yes	yes	yes	yes
region effects	yes	yes	yes	yes
R^2	.565	.537	.278	n.a.
Sar p-value	n.a.	n.a.	n.a.	.981
N [cities]	663 [223]	434 [221]	417 [212]	417 [212]

** Significant at 5% level. * Significant at 10% level.

Table 3. City Employment at Peak Point of Value-Added Per Worker

	manufacturing to service ratio								
peak point	**.60**	**.80**	**1.0**	**1.2**	**1.4**	**1.6**	**1.8**	**2.0**	**2.6**
employment (1000’s)	2730	2380	2030	1670	1320	970	620	270	
IV ’96 – ‘97									
95% conf. interval - lower	1880	1680	1420	1090	670	180			
- upper	3590	3080	2630	2260	1980	1760	1580	1430	

Table 4. Market Potential and Coastal Access

(Results from IV estimation)

	(1)	(2)	(3)
distance to coast	-.0415**	-.0121*	-.0249**
	(.0128)	(.00685)	(.0107)
distance to coast squared	.000644**		.000474*
	(.000319)		(.000270)
Market Potential (sum of		.000890**	.000881**
distance discounted GDP)		(.000253)	(.000244)
Sargan p-value	.991	.677	.702

** Significant at 5% level. * Significant at 10% level.

Table 5. Non-Prefecture Level Cities

(dependent variable $\ln(Y/L)$)

	(1)	(2)	(3)	(4)
	OLS	**OLS**	**IV Estimation (’96)**	**IV Estimation (’96)**
	(’90, ’91)	**(’95, ’96)**	**2SLS**	**GMM**
ln (capital/employ)	.220**	.417**	.385**	.351**
	(.0452)	(.0354)	(.0289)	(.0812)
ln (employ.)	-.0782**	-.00898	.132**	.146**
	(.0288)	(.0284)	(.0563)	(.0600)
% high school	-.00113	-.0105**	.0166	.0155
	(.00380)	(.00393)	(.0203)	(.0201)
distance to coast	.000299	.000537	.00225	.00223
	(.00291)	(.00304)	(.00480)	(.00429)
ln (output/output of	.273**	.451**	.689**	.604**
ind. acct. units	(.0749)	(.0587)	(.211)	(.180)
FDI/capital	.192*	1.47**	3.93**	3.59**
	(.112)	(.475)	(1.65)	(1.50)
ln (roads per capita)	.0323	.0960**	-.0487	-.0441
	(.0216)	(.0213)	(.169)	(.170)
time effects	yes	yes	n.a.	n.a.
regional effects	yes	yes	yes	yes
R^2	.272	.505	.296	n.a.
Sargan p-value	n.a.	n.a.	n.a.	.346
N [cities]	486 [250]	835 [432]	236	236

** Significant at 5% level. * Significant at 10% level.

Table 6. Productivity in the Rural Sector

(dependent variable: $\ln (Y/L)$)

	random effects	**IV estimation**
ln (capital/labor)	.379**	.396**
	(.0884)	(.0544)
ln (sown land/labor)	.0942	.102
	(.133)	(.0937)
ln (labor)	.101	.156**
	(.0662)	(.0179)
% completing high school in rural sector	.0648**	.106**
	(.0303)	(.0250)
central region	-.169	-.0767
	(.134)	(.0476)
west region	-.560**	-.392**
	(.155)	(.0458)
time dummies	yes	yes
adj R^2	.849	n.a.
Sargan value	n.a.	.605
N	87	87

** Significant at 5% level.

Table 7. Productivity in TVE's
(dependent variable: $\ln\ (Y/L)$)

	Generalized Scale Effect		Township Scale Effect	
	random effects	**IV estimation**	**random effects**	**IV Estimation**
ln (capital/labor)	.229**	.355**	.211**	.256**
	(.0672)	(.0412)	(.0664)	(.0271)
ln (labor)	.119**	.128**	n.a.	n.a.
	(.0410)	(.0158)		
ln (avg. township TVE size)	n.a.	n.a.	.158**	.300**
			(.058)	(.0346)
% completing high school	.011	-.00494	-.0180	-.0120**
	(.0077)	(.0118)	(.0173)	(.00582)
central region	-.0625	.120**	-.0102	.057*
	(.105)	(.043)	(.111)	(.033)
west region	-.515**	-.335**	-.504**	-.258**
	(.119)	(.172)	(.114)	(.0623)
time dummies	yes	yes	yes	n.a.
R^2	.786	n.a.	.781	n.a.
Sargan value	n.a.	.760	n.a.	.846
N	87	87	87	87

** Significant at 5% level. * Significant at 10% level.

Table 8. Gains from Increased Within-Sector Agglomeration

A. Gain in output per worker from doubling size of local employment agglomeration

Non-prefecture level cities	11%
TVE's (township TVE employment)	23%

B. Gain from moving to size where output per worker peaks for prefecture level cities

Percent size is below peak:	**50**	**40**	**30**	**20**
Percent gain in output in output per worker	35%	20%	9.5%	4.1%

Table 9. Marginal Products

A. Marginal Product of Labor (1997)

	Rural Sector	**TVE Sector**
Private:	3875	11824
Social:	5115	16528

	Urban Sector
Private (also social if cities efficient sized)	13841
Social: at 50% of efficient size for $MS = 1.4$	17640

Part B. Marginal Product of Capital (1997)

Rural Sector	**TVE Sector**	**Urban Sector** $r_I = r_N$	**Urban Sector** $r_I < r_N$
.66	.42	.14	.30

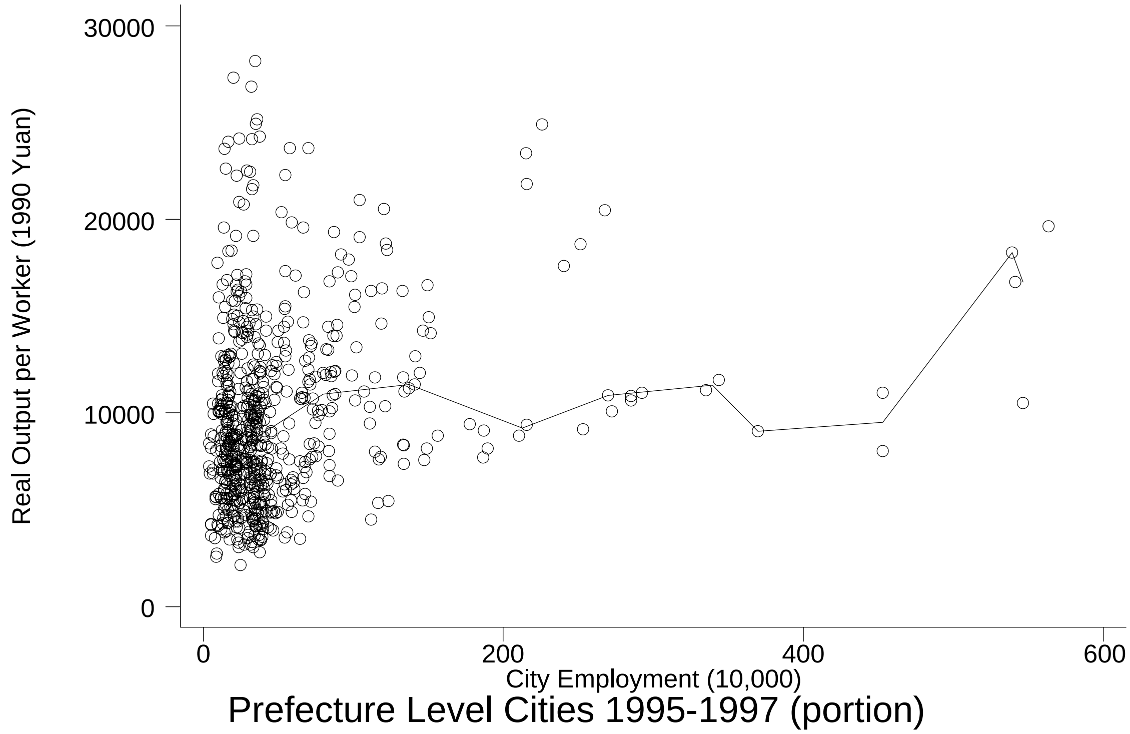

FIGURE 1a. OUTPUT PER WORKER AS A FUNCTION OF CITY SIZE (WHOLE SAMPLE)

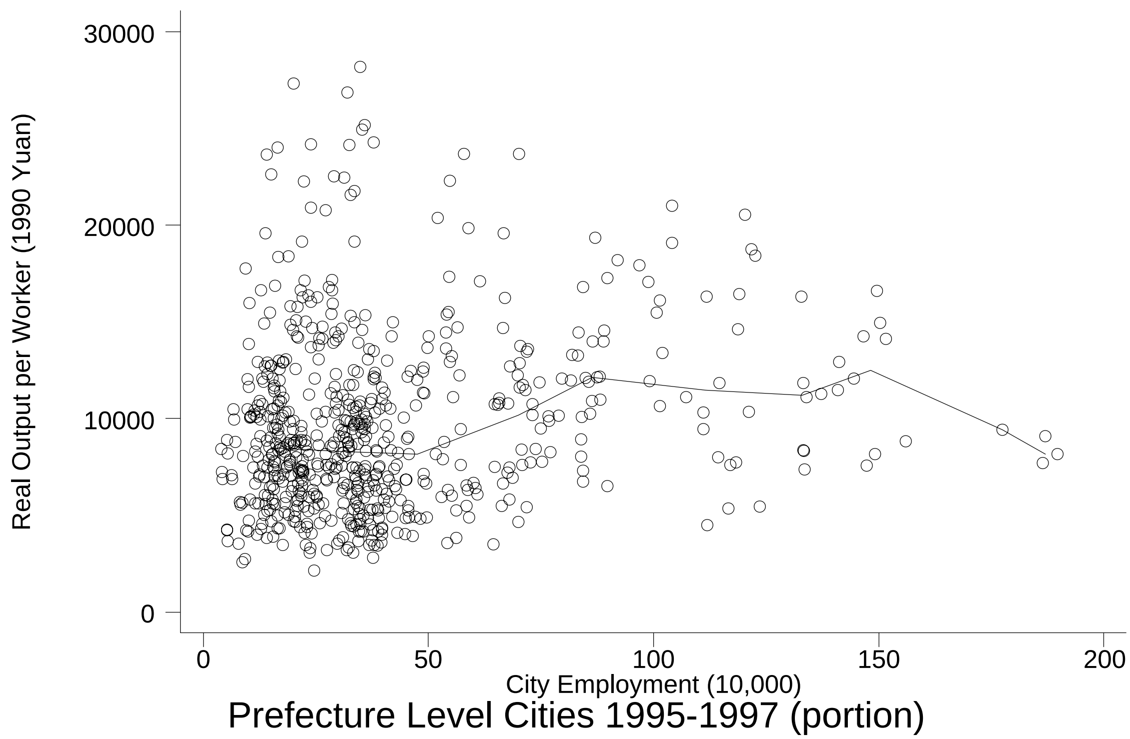

FIGURE 1B. OUTPUT PER WORKER AND CITY SIZE (SIZE < 2M EMPL0YMENT)

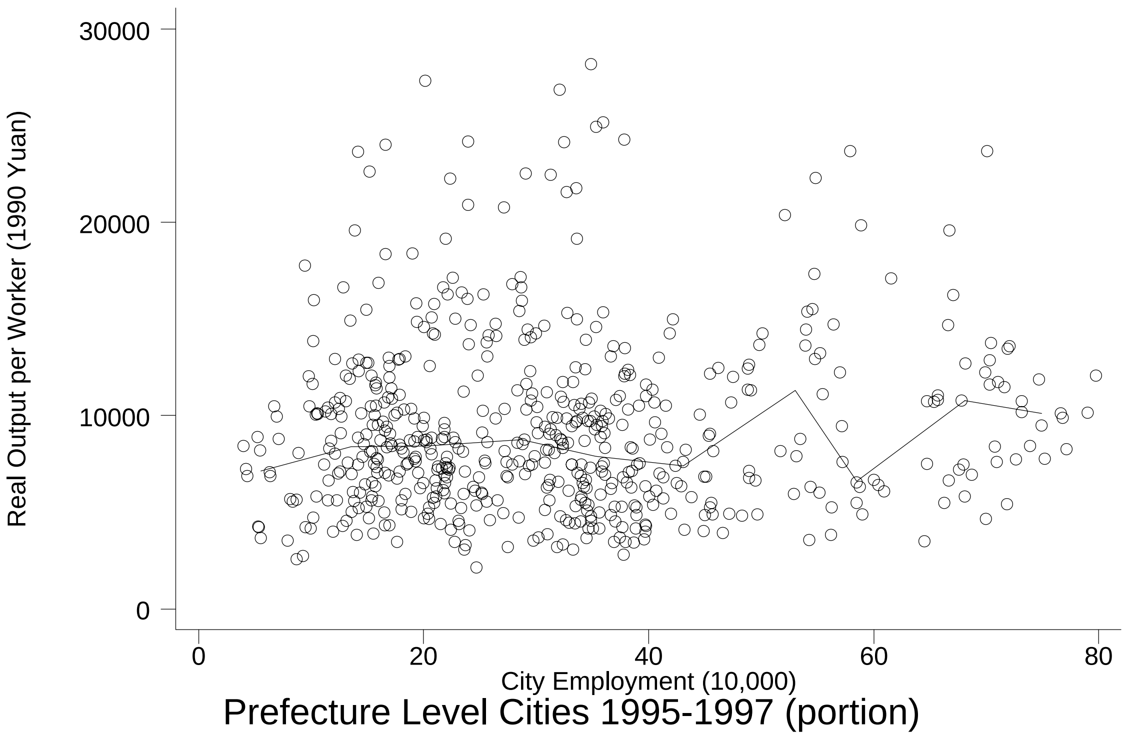

FIGURE 1C. OUTPUT PER WORKER AND CITY SIZE (EMPLOYMENT < .75 M)

Data Appendix

The Urban Sector

City level data used in our analysis come from several sources. Most economic and amenity variables were taken from the 1991 to 1998 annual volumes (for data years 1990 to 1997) of the *Urban Statistical Yearbook of China* (hereafter *Yearbook*)[1], and *Cities China 1949-1998*. The former contains data for both prefecture level and county level cities. The latter includes a compilation of selected data in 1990 to 1997 for prefecture level cities from the *Yearbook* volumes and a complete history of new city establishment and changes in administrative area of all cities during the period. Distance proxies are measured with a ruler from *Map of China*. Highway access is read directly from the same map (occasionally with help from a more detailed map). Educational attainment is aggregated from the *China County-Level Data on Population (Census) and Agriculture, Keyed to 1:1M GIS Map, 1990.*

GDP figures are not available for 1992 and 1993 from the *Yearbook*, but are documented for prefecture level cities in *Cities China 1949-1998*. County level cities are no longer included in the *Yearbook* starting from the 1998 volume (containing 1997 data). A change in definition of labor force (number of employed person) was effected starting from data year 1998. Therefore, the panel data set that we use in our analysis consists of prefecture level cities each year between 1990 to 1997, and county level cities in 1990, 1991 and 1994 to 1996. Table A shows the number of cities with available data in each year. It should be noted that all city level data that we use are those of the more confined city proper (shi qu) rather than the municipal district (di qu). The city proper corresponds to an "urbanized area" in the USA, or the urbanized portion of a metropolitan statistical area. In our analysis of prefecture level cities, we have excluded three oil-dominant cities[2], and a minimal set of outlying city-years[3] based on extraordinary year-to-year change in output per capita, capital-labor ratio or manufacturing to service ratio which are likely the results of misdocumentation. The identified outliners constitute less than 2.5% of the available sample of prefecture level city-years. Brief descriptions of the variables used in our analysis are in Table B and C.

Three issues should be noted here. First, capital is original book value of capital of industrial enterprises with independent accounting systems. Constructed capital data are used for prefecture level cities in 1992 and 1993. These are the only two years in which actual data for capital are not available in the *Yearbooks*. In these years we observe the ratio of gross output to capital for independent accounting units, but we don't observe gross output for these independent accounting units. However the gross output value of industrial enterprises with independent accounting system has a very close correspondence with the gross output value at township and higher levels for prefecture level cities in 1990, 1991, 1995 and 1996 (a simple correlation of 0.997 with sample size 902)[4]. We assume that the ratio of the two gross output values remains stable in 1992 and 1993, and is equal to the average of the ratio over the four years of 1990, 1991, 1995 and 1996 for each prefecture level city. We have data for gross industrial output value at township and higher levels in 1992 and 1993 and based on the assumed ratio, we estimate the gross industrial output value of independent accounting units in each city. Dividing this estimate by the industrial output value realized per 100 yuan of fixed assets at book value of industrial enterprises with

[1] A combined volume was published for 1993 and 1994.

[2] Daqing, Dongying and Karamay in all years.

[3] Chuzhou in 1992, Maanshan in 19193, Mianyang in 1993, Mundanjiang in 1996, Yingtan in 1992, Haikou in 1994, Wanxian in 1992, Jining of Shandong province in 1994 and 1997, Fangchenggang in 1995 and Songyuan in 1992-1997.

[4] The 1994 figures for industrial output value realized per 100 yuan of fixed assets at book value (from which the gross output values are calculated) are abnormally high as compared to that of the other years. In 1994, the mean value among prefecture level cities is 245. The mean value in the other years are: 136 (1990), 134 (1991), 113 (1995), 108 (1996).

independent accounting systems, we arrive at an estimate of the capital for each prefecture level city in 1992 and 1993. There is no significant change in our regression results for prefecture level cities with the inclusion of these estimated capital stocks.

Second, since only capital of independent accounting units is measured, we need to control for the share of independent accounting units among production units in a city. Ideally we would want to use the share of value-added of independent accounting units in the city economy. The data needed to derive this share is available only in 1997. For 1990-1996, a proxy based on the share of industrial output value (i.e. sales value) of independent accounting units among the industrial output value of a city is used. We allow the effects of the actual share and the proxy share to be different in all regressions of prefecture level cities. For county level cities, only the proxy share is available in the relevant years. As noted in footnote 4, the data required to derive the industrial output value is probably not very reliable.

Third, for comparison of real growth of output (GDP), we use the provincial level urban resident consumer price index to deflate nominal GDP's. The index is taken from the Price Indices section of the annual *China Statistical Yearbook* in the relevant period. To compare the real output across cities, we have to assume comparability based on nominal prices in a certain year (1990 in our case).

The Rural Sector

Provincial level data used in our analysis of the rural sector come mainly from two sources: annual volumes of the *China Statistical Yearbook* in 1996, 1998 and 1999 (data year 1995, 1997 and 1998). Provincial educational attainment in rural areas is aggregated from relevant variables of counties not included in urban areas from the *China County-Level Data on Population (Census) and Agriculture, Keyed to 1:1M GIS Map, 1990, 1990*. Data year of 1995, 1997, and 1998 are chosen because these are the data years when value-added (instead of gross output value) of TVEs is available. Table D shows brief description of the variables of the rural sector.

Table A. Number of Cities with Available Data

Year	Number of Prefecture Level Cities	Number of County Level Cities	Total Number of All Cities
1990	225	239	464
1991	225	251	476
1992	227	/	
1993	225	/	
1994	226	393	619
1995	226	411	637
1996	225	433	658
1997	223	/	

Table B. Description of Variables (Urban Sector)

Variable	Description	Source(s)
population	- population at the end of the year	S1, S2
output of a city	- GDP of city in 2nd and 3rd sectors at current prices	S1, S2
manufacturing to service ratio (MS)	- ratio of GDP in 2nd sector to GDP in 3rd sector	S1, S2
employment	- number of persons employed in 2nd and 3rd sectors	S1, S2
capital	- original value of capital of industrial enterprises with independent accounting system	S1
output (value-added) of all units	- gross industrial output value (GDP in 2nd and 3rd sector of city) at current prices	S1, S2
output (value-added) of independent accounting units	- gross industrial output value (value-added of industry) of industrial enterprises with independent accounting system at current prices[5]	S1
FDI	- accumulated sum of foreign direct investment (foreign capital actually used) since 1990	S1, S2
roads per capita	- paved area of all roads with width greater than 3.5 meters	S1, S2
% high school	- percentage of population aged 6+ that has completed senior middle school or above	S3
distance to coast	- shortest horizontal distance from coast, measured in centimeters from map S4	S4, S5
distance to provincial capital	- horizontal distance from capital of province in which a city is located, measured in centimeters from map S4	S4, S5
on highway	- dummy for cities with access to highway (the highest category of all roads on map)	S4, S5
area (1990)	- built-up area in city proper	S2
doctors per capita (1990)	- number of medical doctors per capita	S2
books per capita (1990)	- number of books in public library per capita	S2
telephone per 100 persons	- number of telephones per 100 persons	S2
ratio of municipal agriculture to city value-added	- ratio of total GDP in 1st sector in municipal area to total non-agricultural GDP in city proper	S1, S2
ln (ex-urban population)	- log of (1+ population living in municipal district excluding city proper)	S1, S2

[5] Calculated from industrial output value realized per 100 yuan of fixed assets at book value (value-added realized per 100 yuan of fixed Assets at book value) and fixed assets at book value of industrial enterprises with independent accounting system

Table C. Urban Variable Means and Standard Deviations

	Prefecture Level Cities in 1997		Non-Prefecture Level Cities 1996	
	mean	standard deviation	mean	standard deviation
output per worker	23079	11077	16836	9429
capital per worker	30335	18197	15030	13794
employment (1000's)	510	661	158	104
total value added/value added of independent acct. units	3.33	2.05	n.a.	n.a.
total value sales/sales of independent acct. units 1996	1.42	.522	2.07	1.30
roads per capita	4.72	3.56	2.75	3.38
ln (roads per capita)	1.36	.675	.668	.771
% high school	21.9	8.48	12.8	6.41
distance to coast	6.09	6.06	9.36	10.5
on highway (highest road category)	.66	.47	.271	.445
manufacturing to service ratio (GDP)	1.46	.739	1.47	.809
accumulated FDI since 1990 ($)/capital (yuan)	.0374	.0685	.0160	.0411

Table D. Description of Variables (Rural Sector)

Variable	Description	Source(s)
output of TVE	- added-value of township and village enterprises at current price	S6
labor of TVE	- total number of staff and workers employed in TVE	S6
capital of TVE	- original value of fixed assets (year-end)	S6
average township TVE size	- labor of TVE divided by number of township and town governments	S6
output of rural sector	- output of TVE plus GDP in 1st sector, deflated to 1990 constant prices[6]	S6+S7
labor of rural sector	- labor of TVE plus rural labor force in farming, forestry, animal husbandry and fishery	S6
capital of rural sector	- capital of TVE plus total original value of productive fixed assets of rural households[7]	S6
sown land	- total sown areas of farm crops	S6
% completing high school	- percentage of population aged 6+ that has completed senior middle school or above in rural areas	S3
persons in the communes (1982)	- number of persons in the communes	S6
number of communes (1982)	- number of people's communes	S6
provincial industrial output(1982)	- gross industrial output value of industrial enterprises in province	S6
share of light industry (1982)	- share of light industry in gross industrial output value of industrial enterprises	S6
arable land (1979)	- arable land area	S6
enrollment in higher institutions (1981)	- student enrollment in institutions of higher learning	S6
enrollment in secondary schools (1981)	- student enrollment in regular secondary schools or above	S6
rainfall	- annual total precipitation in capital city of province	S6
hp (1981)	- total horse power of farm machinery (year-end)	S6

[6] Deflation is done with the ratio of provincial total output value of agriculture, forestry, animal husbandry and fishery calculated at current prices to that calculated at 1990 constant prices.

[7] Calculated from original value of productive fixed assets per rural household times number of rural households.

Sources of Data

S1. State Statistical Bureau, Urban Social and Economic Survey Team [Guojia Tongjiju Chengshi Shehui Jingji Diaocha Zongdui], *Urban Statistical Yearbook of China* [*Zhongguo Chengshi Tongji Nianjian*], Beijing: China Statistics Press, 1991 to 1998 (annual volumes).

S2. State Statistical Bureau, Urban Social and Economic Survey Team [Guojia Tongjiju Chengshi Shehui Jingji Diaocha Zongdui], *Cities China 1949-1998* [*Xin Zhongguo Chengshi Wushi Nian*], Beijing: Xinhua Press, 1998.

S3. China in Time and Space (CITAS), *The China County-Level Data on Population (Census) and Agriculture, Keyed to 1:1M GIS Map, 1990*. Data can be downloaded from: ftp://ftpserver.ciesin.org/pub/data/China/CITAS/census_agr/

S4. *Map of China* [*Zhongguo Quantu*], Haerbin Map Press, 3rd ed., February 1999. #1280529-158

S5. *Transportation Map of China* [*Zhongguo Jiaotong Yingyun Licheng Tuji*], Beijing: People's Communication Press, 2000. ISBN 7-114-03553-5.

S6. State Statistical Bureau, *China Statistical Yearbook* [*Zhongguo Tongji Nianjian*], Beijing: China Statistical Publishing House, 1996, 1998 and 1999 (annual volumes) and other relevant years.

S7. State Statistical Bureau, Department of Rural Social and Economic Statistics [Guojia Tongjiju Nongcun Shehui Jingji Tongji Si], *Rural Statistical Yearbook of China* [*Zhongguo Nongcun Tongji Nianjian*], Beijing: China Statistics Press, 1996, 1998 and 1999 (annual volumes).