Accounting for China's Urbanization

Qin Chen* Zheng (Michael) Song†

This version: June 2014

Abstract

It is widely acknowledged that large-scale urbanization plays a pivotal role in China's miraculous economic growth over the past two decades. Yet many of the basic statistics and facts remain disputable. The contribution of this paper is two-fold. First, based on the publicly available 2000 and 2010 census data, plus some auxiliary information from other sources, we develop an accounting method to back out the scale and composition of China's urbanization. We find that urbanization accounts for 80.4 percent of the total urban population growth of 211 million in the 2000s. Moreover, more than half of the urbanized population, about 85.6 million, is due to rural-urban migration. Our findings suggest that rural-urban migration increased by two thirds from the 1990s to 2000s, while the population urbanized by land reclassification is roughly the same across the two periods.

JEL Classification: C82, J11, J61, J82

Keywords: Census, China's Urbanization, Population Dynamics, Rural-Urban Migration

*Fudan University, School of Economics. Email: chenqin@fudan.edu.cn

†University of Chicago, Booth School of Business. Email: zheng.song@chicagobooth.edu.

1 Introduction

A wave of urbanization has swept over China in the past two decades. The urban population ratio skyrocketed from 26 percent in 1990 to 50 percent in 2010 in a country with more than one billion people. It is widely acknowledged that urbanization plays a pivotal role in China's miraculous economic growth. Moreover, there is little doubt that urbanization will maintain its momentum and continue to be an important driving force for economic growth in China in the next one decade or two.[1] The obvious importance of the issue has sparked a growing literature devoted to characterizing the basic patterns of China's urbanization (e.g., Li et al. (2012); Meng (2012)).

Yet our knowledge remains thin for the following reason. Urbanization is a rich dynamic process involving two basic components: (i) rural-urban migration; and (ii) urbanization by urban land expansion. A first-order issue is to identify the scale and composition of urbanization. Since urbanization is dynamic *per se*, direct identification requires both the current and historical residential status at the individual level. To this end, the literature often resorts to the surveys that provide such information. These surveys, however, are designed for various purposes and tend to have various degrees of representativeness for the urbanized population. It is not surprising to see large disparities in the results based on different surveys. For instance, the literature provides a range of the estimated migration that seems too wide to be taken seriously. According to Knight, Deng and Li (2011), migrant workers would account for about 40 percent of the total urban labor force. By contrast, Cai and Du (2011) suggests a ratio of barely 13 percent.

This paper proposes a novel way to account for China's urbanization. The idea is to back out population flows from population stocks in the publicly

[1] At the current speed, China's urban population would account for 68 percent of the total population by 2030 (see Song et al. (2014)).

available census data. The challenge is obvious: The 2010 census does not ask about historical residential status, except for those without local Hukou.[2] Worse, the rural-urban population dynamics are difficult to construct without access to the individual-level data that are not publicly available. To tackle the issue, we develop an accounting framework that characterizes population dynamics. To the best of our knowledge, this is the first attempt in the literature to build a mapping from population stocks in census to population flows by residential and Hukou status. The model allows us to identify the two channels of urbanization in a structural way. We find that urbanization accounts for 80.4 percent of the total urban population growth of 211 million from 2000 to 2010. The rest of the growth is due to the natural growth of the urban population. Among the urbanized population, rural-urban migration accounts for more than half, about 85.6 million. The rest of the urbanized population, about 84.2 million, comes from the expansion of urban land.

It is worthing emphasizing that without the guidance of the model, there is no obvious way of measuring urbanization from the census data. No papers in the literature construct an explicit mapping between the population flow from rural to urban areas (i.e., urbanization) and the population stocks available in the census data. Recent work on China's urbanization has exploited three other major datasets: (i) NongMinGong Survey (see, e.g., Knight, Deng and Li (2011) and Meng (2012)); (ii) China National Rural Survey (see, e.g., Rozelle et al. (2008) and Wang et al. (2011)); and (iii) Rural Urban Migration in China and Indonesia (see, e.g., Kong, Meng and Zhang (2010) and Meng (2012)). Some other survey data are also used, such as China Family Panel Studies (see, e.g., Xu and Xie (2010)). Each dataset has its own strengths and limitations. For instance, compared with censuses, survey data often provide more direct information on rural-urban migration. The downside of these surveys is that their samples are typically drawn from population databases in local administrative

[2]By contrast, the 2000 census asks about historical residential status for everyone. See below for an detailed explanation of Hukou status.

offices, which underrepresent migrants.

The rest of paper is organized as follows. Section 2 discusses the data. We present the accounting framework in Section 3 and, then, use it to back out the scale and composition of China's urbanization in the 2000s. Section 4 conducts several robustness checks and Section 5 concludes.

2 Data

This paper mainly exploits the 2000/2010 census data for variables on population by residential status (rural/urban) and by Hukou status (local/non-local and agricultural/non-agricultural). We also use the 2005 one-percent population survey (henceforth, the mini-census) and the 2008 China General Social Survey to complement the census data. The China General Social Survey was conducted jointly by the Department of Sociology at Renmin University of China and the Survey Research Centre of Hong Kong University of Science and Technology. As will be shown below, the 2005 mini-census and the 2008 China General Social Survey provide information on Hukou status change that is an important identification device in our accounting exercise.

The rural/urban classification and the multi-dimensional Hukou status deserve discussion. Following the definition given by China's National Bureau of Statistics in 2000, we classify rural/urban by city (urban), town (urban) and township (rural).[3] The urbanized population in a certain period is, thus, referred to as the urban population that used to live in a rural area prior to that period. Similarly, rural-urban migrants are those who physically relocate from a rural to an urban area in a given period.

[3]The rural/urban definition has changed a bit in the 2000s, but the difference is too small to have any major effect on our results. Specifically, according to the 2010 definition, urban areas should be located in or contiguous to the area where the local government is located. The only difference in the 2000 definition is that if a region is NOT located in or contiguous to the area where the local government is located, but located in a municipal district with a population density above 1500 people per square km, it would be classified as urban, while it is rural by the 2010 definition. The difference is negligible since it is hard to find a discontinuous area in a high-density municipal district. See Chan and Hu (2003) and Chan (2007) for more details on the definitions.

Hukou status is multi-dimensional: agricultural/non-agricultural and local/non-local. The agricultural/non-agricultural classification is often based on the rural/urban status of a person's original Hukou registration place, though the agricultural Hukou can be converted under circumstances that will be discussed below. In the 2000 census, 24.7 percent of the total population had non-agricultural Hukou. The number increased to 29.1 percent in the 2010 census.[4] The local/non-local classification shows whether a person's current place of residence is his Hukou registration place. In the 2010 census, 261 million people (19.6 percent of the total population) did not have local Hukou. The number had increased by 81 percent since 2000. As will be shown below, the changing Hukou compositions will be essential for identifying rural-urban migration.

Although census adopts the same rural/urban definition for the current residential status, it classifies the Hukou registration place into four jurisdictive categories that are not entirely consistent with the standard rural/urban definition: (i) Street; (ii) Residents' Committee of Town; (iii) Villagers' Committee of Town; and (iv) Township. To deal with potential inconsistencies, the literature often uses "Street" + "Residents' Committee of Town" and "Villagers' Committee of Town" + "Township" as proxies for urban and rural areas, respectively.[5] This approach, which is exactly how the 1990 census defines rural/urban area, tends to deliver a reasonably good approximation. Admittedly, it would bias the results from the more recent censuses since some places under the jurisdiction of "Villagers' Committee of Town" have been classified as urban areas by the above new rural/urban definition. We will follow the literature in our benchmark analysis and address the issue of potential bias in Section 4.1.

[4] The Ministry of Public Security also provides information on agricultural/non-agricultural hukou. According to the reason detailed in Appendix A, the Ministry of Public Security data tend to overestimate the size of the population with agricultural Hukou.

[5] See, for example, Wang (2004) and Cai and Wang (2008).

3 An Accounting Framework

We now present the accounting framework. It is useful to begin with the general structure of China's urban population growth. We decompose the growth from 2000 to 2010 into three parts: (i) the natural growth by birth and death; (ii) rural-urban migration; and (iii) urbanization by land reclassification. Figure 1 illustrates the composition. The first part is self-evident. We use the methodology in Song et al. (2014) to simulate the natural urban population growth, and the rest of the paper will focus on how to back out the urbanized population – i.e., the second and third parts. The second part, rural-urban migration, can further be divided into two categories: (i) rural-urban Hukou migrants whose Hukou registration place is identical to their current place of residence; and (ii) rural-urban non-Hukou migrants whose Hukou registration place is inconsistent with current place of residence. The third part, urbanization by land reclassification, refers to those whose place of residence was rural but has been reclassified as urban.

[Insert Figure 1]

It is worth mentioning that our approach of quantifying China's urbanization is fundamentally different from that in the literature focusing on off-farm migrant workers with agricultural Hukou, known as "NongMinGong" (see, e.g., Knight, Deng and Li (2011), Meng (2012)). Although "NongMinGong" play a key role in China's urbanization, they are not identical to the urbanized population defined above for the following three reasons. First, "NongMinGong" may work in either rural or urban areas. Many migrants actually work in rural areas and, thus, do not contribute to urbanization. Second, the urban "NongMinGong" are a subset of the rural-urban non-Hukou migrants in Figure 1. According to Cai, Park and Zhao (2008), the 2000 census suggests that 34.9 percent of rural-urban migrants obtained local Hukou. Our paper finds a similar proportion of 31 percent. In other words, the urban "NongMinGong" account for, at best, two thirds of the rural-urban migrants. Finally, "NongMinGong"

do not include the population undergoing land reclassification, which accounts for about half of China's urbanization, as we will show below.

3.1 Urbanization in the 2000s

We first back out the overall scale of urbanization – i.e., the sum of rural-urban migration and urbanization by land reclassification. The 2×2 matrix in Figure 2 defines four states by residential and Hukou status, on which the accounting framework is based. The horizontal line shows the rural/urban status of residence. The vertical line represents a person's agricultural/non-agricultural Hukou status. Urbanization occurs when a person moves from the left to the right. More precisely, the size of urbanization between 2000 and 2010 follows:

$$Urbanization_{2000-2010} = P_{12} + P_{14} + P_{32} + P_{34}, \tag{1}$$

where P_{ij} denotes the population of those who are in block i in 2000 and switch to block j in 2010. We know from the census data the population size in each block, P_i^t, where the superscript t denotes the census year. However, census does not provide any direct information on any of the population flows, P_{ij}, in (1). Therefore, our first step is to infer from the population stocks, P_i^t, the scale of urbanization in the 2000s.

[Insert Figure 2]

A person in block 2 in 2000 would end up in any of the four blocks in 2010:

$$P_2^{2000} = P_{21} + P_{22} + P_{23} + P_{24} + d_2, \tag{2}$$

where d_i denotes the population of those who are in block i in 2000 and dead in 2010. Symmetrically, by population inflows, we have an analogous formula for population composition in block 2 in 2010:

$$P_2^{2010} = P_{12} + P_{22} + P_{32} + P_{42} + b_2, \tag{3}$$

where b_i denotes the population of those who are born in 2000-2010 and in block i in 2010. In other words, except for the newborns, a person in block 2 in 2010

must come from any of the four blocks in 2000. A combination of (2) and (3) leads to

$$P_2^{2010} - P_2^{2000} = P_{12} + P_{32} + b_2 - P_{21} - P_{23} - P_{24} - d_2. \tag{4}$$

Here, we drop P_{42} since non-agricultural Hukou is generally not allowed to be converted to agricultural Hukou.[6] A similar procedure decomposes the population change in block 4 as follows.

$$P_4^{2010} - P_4^{2000} = P_{14} + P_{24} + b_4 + P_{34} - P_{43} - d_4, \tag{5}$$

where P_{41} is dropped for the reason explained above.

We make two important assumptions.

Assumption 1: Those who live in rural areas always have agricultural Hukou – i.e., $P_3^t = 0$.

Assumption 2: Those who live in urban areas and have agricultural Hukou in 2000 continue to live in urban areas in 2010 – i.e., $P_{21} = 0$.

The first assumption is primarily for simplicity. Quantitatively, P_3^t appears to be small relative to the population size of the other blocks. In the 2000 census, for instance, only 4.7 percent of the rural population had non-agricultural Hukou. The share dropped to 4.3 percent in the 2010 census, accounting for only two percent of the total population. Moreover, as we show below, what really matters for our results are $P_{32} - P_{23}$ and $P_{34} - P_{43}$. To get a sense of how large they would be, let us assume proportional population flows such that $P_{3i} = \left(P_i^{2010} / \sum_j P_j^{2010}\right) P_3^{2000}$ and $P_{i3} = \left(P_3^{2010} / \sum_j P_j^{2010}\right) P_i^{2000}$.[7] Then, we would have $P_{32} - P_{23}$ and $P_{34} - P_{43}$ of 4.5 million and 4.0 million, respectively,

[6] Only under one strict condition can people with non-agricultural hukou change to a agricultural hukou. Agricultural Hukou of college/university students will automatically be converted into non-agricultural hukou after entering college/university. When college/university students who had agricultural hukou before fail to find a job in city for two years after graduation, they could apply to restore their agricultural Hukou. The restoration also needs approval from the village committee of the student's place of origin.

[7] This assumption implies that 97.2 percent of the population in block 3 in 2000 would move to the other blocks in 2010, and 97.9 percent of the population in block 3 in 2010 would come from the other blocks in 2000. It certainly inflates the population flows during the 2000s. We use the assumption to shoot upper bound estimates of the population flows.

about 2.5 and 2.2 percent of our estimated urbanized population. Therefore, dropping $P_{32} - P_{23}$ and $P_{34} - P_{43}$ does not seem to have any major effect.

The second assumption has two implications. On the one hand, we can argue that the population of urban-rural migrants – i.e., those moving from urban to rural areas – is quantitatively small. On the other hand, Assumption 2 is conceptually important. By assuming away the reversed migration, we may interpret, in a more precise way, our estimated urbanization as a net population flow.

Under Assumptions 1 and 2, (4) and (5) establish the first result of this paper:

$$\begin{aligned} Urbanization_{2000-2010} &= P_{12} + P_{14} \\ &= P_2^{2010} - P_2^{2000} + P_4^{2010} - P_4^{2000} - \sum_{i=2,4} (b_i - d_i)\,. \end{aligned} \tag{6}$$

We know from the 2000 and 2010 censuses that $P_2^{2010} - P_2^{2000}$ and $P_4^{2010} - P_4^{2000}$ equal 123.8 and 87.2 million, respectively. Following Song et al. (2014), we can back out the natural population change, the last term on the right-hand side of (6).[8] Specifically, $b_2 + b_4$ and $d_2 + d_4$ equal 61.5 and 20.3 million, respectively. Therefore, (6) implies the total urbanization of 169.8 million.

3.2 Rural-Urban Migration in the 2000s

$P_{12} + P_{14}$ consist of both rural-urban migrants who have physically relocated and those whose place of residence was reclassified as urban. The latter refers to urbanization by land reclassification, as shown in Figure 1. To figure out the structure of urbanization, we need to separate the two components. For notational ease, let us define "$*$" as the population that physically relocated from rural to urban areas in the 2000s. The sum of P_{12}^* and P_{14}^* would, thus, be equal to the population of rural-urban migrants.

To back out $P_{12}^* + P_{14}^*$, we turn to a finer definition of P_2, the urban residents with agricultural Hukou. Define P_{2_n} and P_{2_l} as urban residents with *non-local*

[8] See Appendix B for details.

and *local* agricultural Hukou, respectively. Analogous to (1) and (2), we can decompose $P_{2_n}^{2000}$ and $P_{2_n}^{2010}$ by population flows:

$$\begin{aligned} P_{2_n}^{2000} &= P_{2_n 2} + P_{2_n 4} + d_{2_n} \\ &= P_{2_n 2_l} + P_{2_n 2_n} + P_{2_n 4} + d_{2_n}, \end{aligned} \tag{7}$$

$$\begin{aligned} P_{2_n}^{2010} &= P_{12_n} + P_{22_n} + b_{2_n} \\ &= P_{12_n} + P_{2_l 2_n} + P_{2_n 2_n} + b_{2_n}, \end{aligned} \tag{8}$$

where d_{2_n} denotes the population of those who are in block 2_n in 2000 but dead in 2010; b_{2_n} denotes the population of newborns in 2000-2010 who are in block 2_n in 2010. Notice that Assumptions 1 and 2 allow us to drop $P_{2_n 1}$, $P_{2_n 3}$, P_{32_n} and P_{42_n} in the above two equations.

We make two additional assumptions.

Assumption 3: Those in P_{12_l} are urbanized by land reclassification – i.e., $P_{12_l} = P_{12} - P_{12}^*$.

Assumption 4: $P_{2_n 2_l} - P_{2_l 2_n} = 0$

P_{12_l} refer to those satisfying the following three conditions: (i) Their place of residence switches from rural to urban in 2000-2010; (ii) Their Hukou is "agricultural" in 2010; and (iii) Their Hukou is "local" in 2010. It is the combination of the last two conditions that justify Assumption 3. To see this, first notice that, by definition, P_{12_l} is exclusive to non-Hukou rural-urban migrants in Figure 1: P_{12_l} have local Hukou, while non-Hukou migrants do not. P_{12_l} is also exclusive to Hukou rural-urban migrants in Figure 1. This is because the latter have local and non-agricultural Hukou, while those in P_{12_l} have local and agricultural Hukou. So, we can exclude rural-urban migrants from P_{12_l}. Together with the fact that everyone in P_{12_l} is an urbanized individual, this implies that P_{12_l} has to represent the urbanized population through land reclassification.

An alternative way of interpreting the assumption is to use the definition that $P_{12} = P_{12_l} + P_{12_n}$, which leads to

$$P_{12_n} = P_{12}^*. \tag{9}$$

In other words, we assume that those in P_{12} with non-local agricultural Hukou, P_{12_n}, coincide with those in P_{12} who migrate from rural to urban areas.

Although Assumption 3 or (9) looks reasonable, there are certainly exceptions in reality. To gauge the quantitative importance of the assumption, look at the two possibilities in which $P_{12_n} \neq P_{12}^*$. First, some people in P_{12_n} may be urbanized by land reclassification and, hence, are not in P_{12}^*. The 2000 census shows that only 2.6 percent of the rural population have non-local agricultural Hukou. Unless the ratio is a lot higher among the rural population that is urbanized by land reclassification – i.e., $P_{12} - P_{12_n}$ – the first possibility cannot have any quantitatively large effect on our results. The second possibility is that some people in P_{12}^* may have local agricultural Hukou. According to a subsample of the 2000 census in which individual-level data are available, among those who moved from rural to urban areas in 1995-2000, only 3.7 percent changed their Hukou registration place while maintaining their agricultural Hukou status.[9] Therefore, the effect of the second possibility must also be quantitatively small.

Like Assumption 1, Assumption 4 is also made for simplicity. In words, $P_{2_n 2_l}$ or $P_{2_l 2_n}$ refer to those who (i) live in urban areas; (ii) have agricultural Hukou; and (iii) move back to or out of their Hukou registration place, respectively. According to the 2005 mini-census, the respective sizes of $P_{2_n 2_l}$ and $P_{2_l 2_n}$ are 1.15 million and 1.43 million. Assume that $P_{2_n 2_l}$ and $P_{2_l 2_n}$ in 2000-2010 are twice as large as those in 2000-2005. $P_{2_n 2_l} - P_{2_l 2_n}$ would account for less than one percent of the inferred rural-urban migration, which will be shown below. Assumption 4, therefore, should have little effect on our main results.

The two assumptions allow us to derive the following equation from (7) and

[9]Such information is no longer available from the 2010 census even if there is access to individual-level data. Recall that the 2010 census asks about historical residential status for those with non-local Hukou only.

(8):

$$\underbrace{P_{12}^{*}+P_{14}^{*}}_{\text{rural-urban migration}} = P_{2_n}^{2010}-P_{2_n}^{2000}-\underbrace{(b_{2_n}-d_{2_n})}_{\substack{\text{natural population change of}\\ \text{urban residents with non-local hukou}}}+\underbrace{P_{2_n4}+P_{14}^{*}}_{\text{YiDiNongZhuanFei}}. \tag{10}$$

Recall that P_{14}^{*} refers to the rural-urban migrants whose agricultural Hukou was converted into non-agricultural Hukou between 2000 and 2010; P_{2_n4} is the urban population that changed their non-local agricultural Hukou to local non-agricultural Hukou between 2000 and 2010. Therefore, the union of P_{14}^{*} and P_{2_n4} corresponds to those who satisfy the following two conditions: (i) They change their agricultural Hukou to non-agricultural Hukou; and (ii) they change their Hukou registration place. This coincides exactly with the definition of so-called "YiDiNongZhuanFei" – i.e., converting agricultural to non-agricultural Hukou in a place other than where the Hukou was registered.

We next calculate the three components on the right-hand side of (10) in order.

3.2.1 $P_{2_n}^{2010}-P_{2_n}^{2000}$

With some abuse of notation, let us refer to P_n^t as the population of residents with non-local Hukou at period t. Although $P_{2_n}^{2010}$ and $P_{2_n}^{2000}$ are not directly available from the census data, we can back them out by

$$P_{2_n}^{t} = P_n^t \times \frac{P_{2_n}^t}{P_n^t}, \tag{11}$$

where P_n^t is observable in census and $P_{2_n}^t/P_n^t$ can be proxied by its counterpart in a ten-percent subsample of census reported by the *Long Table*.

[Insert Table 1]

According to Tables 1-2a and 1-2b in the 2000/2010 census tabulation, P_n^{2000} and P_n^{2010} equal 144.39 and 261.38 million, respectively. Regarding $P_{2_n}^t/P_n^t$, Table 1 reports the population of those with non-local Hukou by their current residence (the first column) and Hukou registration place (the first row) in the

subsamples of the two censuses. $P^t_{2_n}$ corresponds to the framed cells in each panel – i.e., those who live in urban areas (City or Town) but have their Hukou registered in rural areas (Township or "Villagers' Committee of Town").[10] The counterpart of $P^t_{2_n}/P^t_n$ is, thus, the share of the sum of all the framed cells to the sum of all cells in each panel. Assume that the ratios in the ten-percent subsamples are identical to those in the censuses. We find $P^t_{2_n}/P^t_n$ of 48.8 and 52.7 percent in 2000 and 2010, respectively. Substituting the ratios back into (11) leads to $P^{2010}_{2_n}$ and $P^{2000}_{2_n}$ of 137.75 and 70.46 million, respectively, and $P^{2010}_{2_n} - P^{2000}_{2_n}$ of 67.28 million.

3.2.2 $b_{2_n} - d_{2_n}$

$b_{2_n} - d_{2_n}$ is the natural population change of P_{2n}. As shown in Appendix B, the age- and gender-specific fertility/mortality rates are needed to calculate natural population change. The data for P_{2_n} are not available in census. So, we assume the natural population growth of P_{2_n} to be the same as that of $P_2 + P_4$, which is the total urban population:

$$\frac{b_{2_n} - d_{2_n}}{P^{2000}_{2_n}} = \frac{\sum_{i=2,4} b_i - d_i}{P^{2000}_2 + P^{2000}_4}. \tag{12}$$

We know that $\sum_{i=2,4} b_i - d_i$, which sits in (6), and $P^{2000}_{2_n}$ from (11). (12) gives $b_{2_n} - d_{2_n}$ of 8.5 million. Given the size of $P^{2000}_{2_n}$, alternative assumptions on the natural population growth would have minor effects on our main results below.

3.2.3 YiDiNongZhuanFei

Two sources contribute to the growth of the population with non-agricultural Hukou: (i) natural population growth; and (ii) the population that changed their agricultural Hukou to non-agricultural Hukou – i.e., the so-called "NongZhuanFei" population. "YiDiNongZhuanFei" is a subset of "NongZhuanFei." The former refers to those who also changed their Hukou registration place. According to the 2000 and 2010 censuses, the population with non-agricultural Hukou grew

[10] Here, we identify the rural status of Hukou registration by Township or Villagers' Committee of Town. Section 4.1 will use the NBS definition to check the accuracy of the result.

from 305.1 million to 384.34 million. After excluding the natural growth of the non-agricultural Hukou population, we find a "NongZhuanFei" population of 59.9 million in 2000-2010.[11] We next look at the ratio of the "YiDiNongZhuanFei" population to the "NongZhuanFei" population in the 2008 China General Social Survey, which averaged 45 percent between 2000 and 2008. Assuming the average ratio to be identical to the nationwide ratio in 2000-2010, we can back out the "YiDiNongZhuanFei" population of 26.9 million. By adding up all elements on the right-hand side of (10), the scale of rural-urban migration in the 2000s is 85.6 million.

3.3 Results

As shown in Figure 1, the urbanized population includes both rural-urban migrants and those experiencing land reclassification. Subtracting 85.6 million rural-urban migrants from the urbanized population of 169.8 million, we can infer an urbanized population of 84.2 million due to land reclassification.

The results are summarized in Table 2. In the 2000s, China had a total urbanized population of 169.8 million, which is substantially larger than the 139 million in the 1990s, estimated in Wang (2004). It would be instructive to know the extent to which rural-urban migration contributed to this acceleration of urbanization, a key driving force of China's growth. Compared with the estimates from Chan (2012), rural-urban migration tended to slow down in the 2000s, while the results in Wang (2004) would suggest the opposite. Methodologically, our approach is more in line with Wang (2004), which first backs out rural-urban migration and, then, takes the residual as urbanization by land reclassification.[12] According to the rural-urban migration of 51.7 million in the 1990s in Wang (2004), our result would imply a 66-percent increase

[11] To calculate the natural population growth, we take the assumption that is analogous to (12).

[12] Chan (2012) adopts the opposite procedure: First, back out the urbanization by land reclassification. Specifically, Chan assumes that all the people who lived in the 8,439 newly designated towns in the 1990s were the reclassification population. Moreover, these towns have an average population of 3,500. The two assumptions lead to a reclassification population of 30 million, which accounts for 22 percent of the total urban population growth in the 1990s.

in rural-urban migration in the 2000s. This provides support to the view that rural-urban migration has also accelerated and plays an even more important role in the Chinese economy.

[Insert Table 2]

Incidentally, our quantitative results turn out to be close to those in Dragonomics (2011) and Chan (2012). However, it is unclear how the two studies come up with their results. By contrast, our accounting framework makes the whole procedure highly transparent and can, thus, be used to back out the scale and structure of urbanization from any past and future censuses.

4 Robustness Check

4.1 The Alternative Rural/Urban Classification

As discussed in Section 2, census reports original Hukou residential status by administrative units, which cannot perfectly be matched with rural/urban classification. Table 3 reports the numbers in the four basic administrative units by the 2010 NBS rural/urban classification. It turns out that 19.9 and 3.6 percent of the "Villagers' Committee of Town" and "Township," respectively, are classified as urban by the NBS definition. Using the administrative categories to identify the rural/urban status of a person's original residency would, thus, lead to biased results. As a robustness check, we assume that the original residency of rural-urban migrants is evenly distributed within each type of basic administrative unit. We make this seemingly unreasonable assumption is taken to get a sense of how large the bias could be. Then, the shares in Table 3 would imply that 80.1 percent of those moving from "Villagers' Committee of Town" to urban areas are rural-urban migrants by the NBS definition. Symmetrically, 21.5 percent of those moving from "Street" to urban areas would be counted as rural-urban migrants. We then recalculate $P^t_{2_n}/P^t_n$ and, thus, $P^{2010}_{2_n} - P^{2000}_{2_n}$. The adjusted $P^{2010}_{2_n} - P^{2000}_{2_n}$ is 64.36 million, about 3 million less than the bench-

mark result. The difference is apparently too small to make any major change to our main findings.

[Insert Table 3]

4.2 Newborns

So far, urbanization in this paper has not involved newborns. However, one may argue that the children of the urbanized population, if born in the period of urbanization, should also be counted as part of the urbanized population. To check the quantitative importance of the alternative classification, we rewrite (6) as follows:

$$\begin{aligned} Urbanization_{2000-2010} &= P_{12} + P_{14} + \sum_{i=2,4} b_i^u \qquad (13) \\ &= P_2^{2010} - P_2^{2000} + P_4^{2010} - P_4^{2000} - \sum_{i=2,4} \left(b_i - b_i^u - d_i\right), \end{aligned}$$

where b_i^u refers to the newborns in 2000-2010 who satisfy the following two conditions: (i) They were in block i in 2010; and (ii) their parents were urbanized in 2000-2010 and ended up in block i in 2010. Accordingly, $b_i - b_i^u$ and $b_i - b_i^u - d_i$ refer to the newborns whose parents were residents in block i in 2000 and the natural population change in block i during 2000-2010, respectively. Following the same method in Appendix B, we find the natural population change, $\sum_{i=2,4}(b_i - b_i^u - d_i)$, to be 28.7 million. The adjusted urbanization by (13) is, thus, 182.3 million,12.5 million more than the benchmark result.

We may change (10) in an analogous way:

$$\underbrace{P_{12}^* + P_{14}^* + b_{2_n}^u}_{\text{rural-urban migration}} = P_{2_n}^{2010} - P_{2_n}^{2000} - \underbrace{\left(b_{2_n} - b_{2_n}^u - d_{2_n}\right)}_{\substack{\text{natural population change of} \\ \text{urban residents with non-local hukou}}} + \underbrace{P_{2_n 4} + P_{14}^*}_{\text{YiDiNongZhuanFei}}, \qquad (14)$$

To back out $b_{2_n} - b_{2_n}^u - d_{2_n}$, we adjust (12) accordingly:

$$\frac{b_{2_n} - b_{2_n}^u - d_{2_n}}{P_{2_n}^{2000}} = \frac{\sum_{i=2,4}\left(b_i - b_i^u - d_i\right)}{P_2^{2000} + P_4^{2000}}.$$

This gives $b_{2_n} - b_{2_n}^u - d_{2_n}$ of 4.4 million. The adjusted size and composition of China's urbanization is reported in the second row of Table 2.

4.3 A Dual Approach

Our accounting exercise is based primarily on the population growth in P_2 and P_4 – i.e., the urban population growth. There is an alternative way of backing out $P_{12} + P_{14}$ using the rural population change. Symmetrically, the same assumptions would yield

$$P_{12} + P_{14} = -\left(P_1^{2010} - P_1^{2000}\right) + b_1 - d_1, \tag{15}$$

where b_1 refers to those that were born in 2000-2010 and were in block 1 in 2010. The implied urbanization is 168.7 million, very close to the size of 169.8 million implied by (6). To deal with the issue in the previous section, let us denote as b_1^u the newborns of the urbanized who were in block 1 in 2000. Theoretically speaking, b_1^u should be identical to $b_2^u + b_4^u$ since they refer to the same set of newborns. Then, (15) can be rewritten as:

$$P_{12} + P_{14} + b_1^u = -\left(P_1^{2010} - P_1^{2000}\right) + b_1 + b_1^u - d_1. \tag{16}$$

Here, $b_1 + b_1^u - d_1$ stands for the natural population change in P_1, which is found to be 47.9 million. This implies an urbanized population of 181.2 million. Once again, it is very close to the size of 182.3 million implied by (13).

5 Conclusion

Urbanization is set to be one of the most important driving forces for China's future economic growth. The reforms regarding urbanization, such as lifting Hukou restrictions and tightening land regulations, have been among top agenda items in the policy circle. Any serious policy discussion has to be based on reliable statistics. When it comes to China's urbanization, however, the existing studies often disagree with each other on some of the basic statistics. To address this issue, we develop an accounting framework which establishes a mapping from population stocks to dynamics. We then apply the method to measure China's urbanization in the 2000s by exploiting the census data. Our

approach has two advantages: (i) the representativeness of the census data; and (ii) the highly transparent accounting procedure. While the results certainly have policy implications on China's urbanization, the methodological contribution is not limited to this specific topic. Generally speaking, the accounting framework allows us to back out population flows across multidimensional status over time by using repeated cross-sectional data. Therefore, the model can easily be extended to study population dynamics across regions, industries and occupations. This is an important direction for future research.

Reference

Cai, Fang, Albert Park, and Yaohui Zhao. 2008. "The Chinese Labor Market in the Reform Era." In *China's Great Economic Transformation.* , ed. Loren Brandt and Thomas G. Rawski, 167–214. Cambridge University Press: New York.

Cai, Fang, and Meiyan Wang. 2008. "A Counterfactual Analysis on Unlimited Surplus Labor in Rural China." *China and World Economy*, 16(1): 51–65.

Cai, Fang, and Yang Du. 2011. "Wage Increases, Wage Convergence, and the Lewis Turning Point in China." *China Economic Review*, 22(4): 601–610.

Chan, Kam Wing. 2007. "Misconceptions and Complexities in the Study of China's Cities: Definitions, Statistics, and Implications." *Eurasian Geography and Economics*, 48(4): 383–412.

Chan, Kam Wing. 2012. "Crossing the 50 Percent Population Rubicon: Can China Urbanize to Prosperity?" *Eurasian Geography and Economics*, 53(1): 63–86.

Chan, Kam Wing, and Ying Hu. 2003. "Urbanization in China in the 1990s: New Definition, Different Series, and Revised Trends." *The China Review*, 3(2): 49–71.

Dragonomics. 2011. *Urbanization and Housing Markets in China.* Beijing, China: Dragonomics.

Knight, John, Quheng Deng, and Shi Li. 2011. "The Puzzle of Migrant Labour Shortage and Rural Labour Surplus in China." *China Economic Review*, 22(4): 585–600.

Kong, Sherry Tao, Xin Meng, and Dandan Zhang. 2010. "The Global Financial Crisis and Rural–Urban Migration." In *China: The Next Twenty Years of Reform and Development.* 241–265. Beijing, China:The Australian National University E-Press, Canberra and Social Sciences Academic Press.

Li, Hongbin, Lei Li, Binzhen Wu, and Yanyan Xiong. 2012. "The End of Cheap Chinese Labor." *Journal of Economic Perspectives*, 26(4): 57–74.

Meng, Xin. 2012. "Labor Market Outcomes and Reforms in China." *Journal of Economic Perspectives*, 26(4): 75–101.

Rozelle, Scott, Jikun Huang, Linxiu Zhang, and Xiaofei Li. 2008. "Chinas Labor Transition and Future Educational Challenges." *Unpublished paper, Stanford University. http://scid. stanford. edu/group/siepr/cgi-bin/scid.*

Song, Zheng, Kjetil Storesletten, Yikai Wang, and Fabrizio Zilibotti. 2014. "Sharing High Growth Across Generations: Pensions and Demographic Transition in China." *American Economic Journal: Macroeconomics*, forthcoming.

Wang, Fang. 2004. "The Composition of Urban Population Growth between 4th Census and 5th Census (in Chinese)." *Population Research*, 28(3): 60–67.

Wang, Xiaobing, Jikun Huang, Linxiu Zhang, and Scott Rozelle. 2011. "The Rise of Migration and the Fall of Self Employment in Rural China's Labor Market." *China Economic Review*, 22(4): 573–584.

Xu, Hongwei, and Yu Xie. 2010. "The Causal Effects of Rural-to-Urban Migration on Childrens Wellbeing in China." In *Annual Meeting of the Population Association of America*. 241–265. New Orleans.

A Data Sources of Agricultural Hukou and Non-Agricultural Hukou

The information on the population with agricultural/non-agricultural Hukou is available from two different datasets. The 2000 and 2010 censuses show an increase of 79 million in the non-agricultural Hukou population, while the Ministry of Public Security data suggests an increase of 138 million.

Excluding the natural growth of the non-agricultural population in the 2000s reduces the increase in the "NongZhuanFei" population (i.e., those whose agricultural Hukou was converted into non-agricultural) to 59.85 million and 117.7 million in the census data and the Ministry of Public Security data, respectively.[13] Therefore, the ratio of "NongZhuanFei" in the 2000s to the total population with non-agricultural Hukou in 2010 is 15.5 percent in the census data. The ratio is 25.6 percent in the Ministry of Public Security data.

We next look at the ratio of "NongZhuanFei" to the total population with non-agricultural Hukou in the 2008 China General Social Survey data, which has an average of 10.2 percent between 2000 and 2008. Assume that the "NongZhuanFei" process in the 2008 China General Social Survey data has a speed identical to that of the nationwide process in 2000-2010, the ratio of "NongZhuanFei" to the total population with non-agricultural Hukou would be 13.5 percent, which is close to the ratio implied by the census data, and only half of the ratio suggested by the Ministry of Public Security data. Thus, we believe that the Ministry of Public Security data overestimates the non-agricultural Hukou population.

[13] To calculate the natural population growth, we take the assumption that is analogous to 12.

B Projecting the "Natural" Population Growth

We follow the methodology used in Song et al. (2014) to project the natural population growth. We first simulate the "natural" population. We separate the total population into four blocks, male/female and 101 age specific cells, and adopt the following notations: $i \in \{1,2,3,4\}$, $j \in \{f,m\}$ and $k \in \{0,1,2,\ldots,100\}$ stand for the four blocks defined in the text, gender and age, respectively. $p_t^{i,j,k}$ represents the population with type i, j and k in the period-t census. For instance, $p_{2000}^{1,f,10}$ stands for the population of females of age 10 in block 1 in the 2000 census. The total population in block i at time t can be written as:

$$P_i^t = \sum_{j,k} p_t^{i,j,k}$$

Let $\hat{p}_t^{i,j,k}$ represent the simulated natural population in period t according to:

$$\hat{p}_t^{i,j,0} = \sum_i \sum_k \theta_{t-1}^{i,j} br_{t-1}^{i,k-1} \hat{p}_{t-1}^{i,f,k-1}, \tag{17}$$

$$\hat{p}_t^{i,j,k} = \sum_i (1 - dr_{t-1}^{i,j,k-1}) \hat{p}_{t-1}^{i,j,k-1}, k > 0, \tag{18}$$

where $br_t^{i,k}$, $dr_t^{i,j,k}$ and $\theta_{t-1}^{i,j}$ denote the fertility rates, mortality rates and sex ratios at birth in period t, respectively. These demographic parameters in 2000 and 2010 are calibrated by their counterparts in the census data. The missing parameters between the two census years are generated by linear interpolation. We use the parameters calibrated by the Township data in census to simulate the population in blocks 1 and 3. Symmetrically, the City and Town data in census are used for simulating blocks 2 and 4. By construction, the simulated population at period t would be identical to the population in the census data – i.e., $p_t^{i,j,k} = \hat{p}_t^{i,j,k}$, if t is a census year. Following the assumption in Song et al. (2014), fertility rates are adjusted upwards to correct for the potentially under-reported birth rates in the 2000 census. We multiply $br_{2000}^{i,k}$ from the 2000 census tabulation by 1.2.

Second, we calculate the natural population growth, $\sum_{j,k}\left(\hat{p}^{i,j,k}_{2010}-p^{i,j,k}_{2010}\right)$. Rewrite the natural population growth as

$$\sum_{j,k}\left(\hat{p}^{i,j,k}_{2010}-p^{i,j,k}_{2010}\right)=b_i-b^u_i-d_i,$$

where

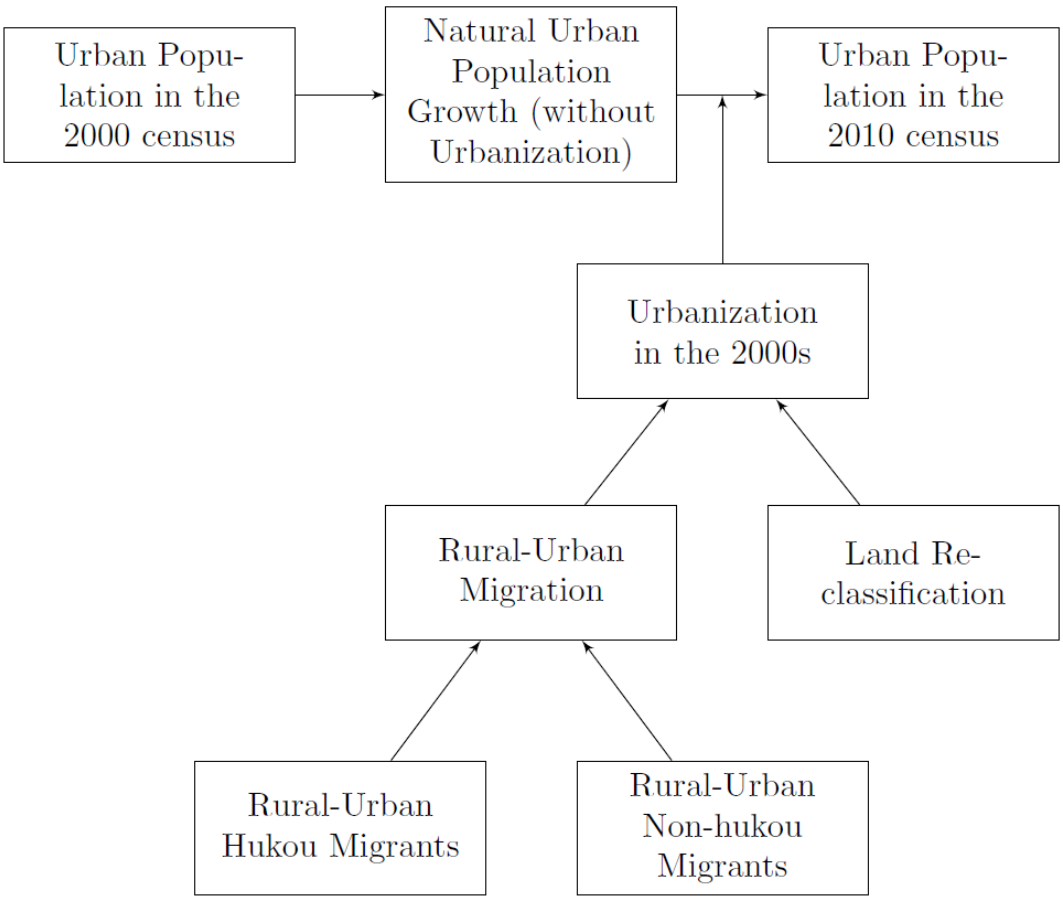
$$\begin{aligned}
b_i &= \sum_{j;k\leq 10} p^{i,j,k}_{2010},\\
d_i &= \sum_{j,k} p^{i,j,k}_{2000}-\sum_{j;k>10}\hat{p}^{i,j,k}_{2000},\\
b^u_i &= \sum_{j;k\leq 10} p^{i,j,k}_{2010}-\hat{p}^{i,j,k}_{2010}.
\end{aligned}$$

The natural population growth rate in block i follows

$$\frac{\sum_{j,k}\left(\hat{p}^{i,j,k}_{2010}-p^{i,j,k}_{2010}\right)}{\sum_{j,k}p^{i,j,k}_{2000}}=\frac{b_i-b^u_i-d_i}{P^{2000}_i}.$$

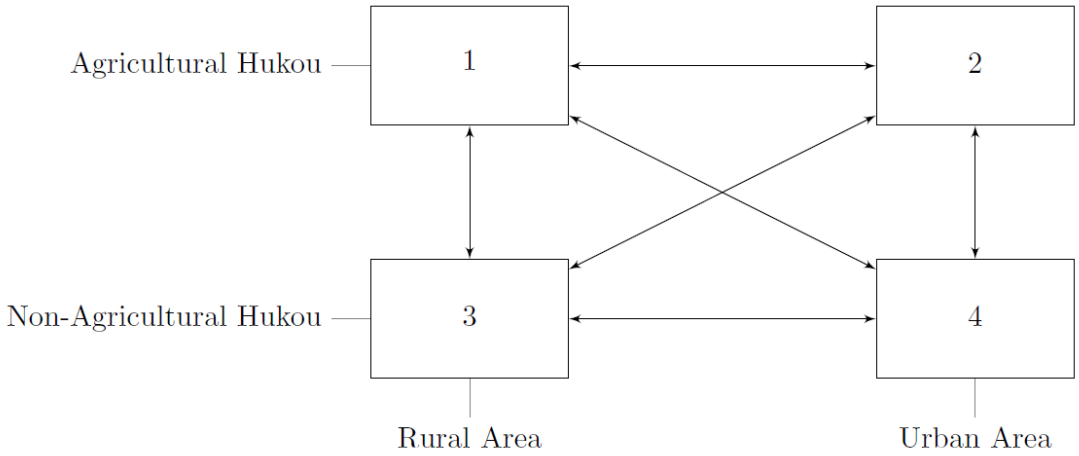
The above formula is used to calculate the natural population growth reported in the text.

Figure 1



This figure illustrates the composition of urban population growth.

Figure 2



This figure presents a 2×2 population matrix. The urbanized population is the sum of the population flows from the left two blocks to the right two blocks.

Table 1: Population without Local Hukou

Year	Current Place of Residence	Hukou Registration Place			
		Township	RCT	VCT	Street
2000	City	1570156	2551633	1999862	3414135
	Town	806405	1123832	851960	238122
	Township	1206845	1291268	1117164	262016
2010	City	3273927	1679495	5821197	5450000
	Town	1355337	767106	2389724	488573
	Township	853872	286985	1645633	343525

RCT and VCT represent “Residents’ Committee of Town” and “Villagers’ Committee of Town,” respectively. Source: Table L-7-3a/3b/3c in the Tabulation, the 2000 census, and Table L-7-1 in the Tabulation, the 2010 census.

Table 2: China's Urban Population Growth and Urbanization (Millions)

Sources	Period	Urban Population Growth				
			Natural Population Growth	Urbanization		
					Rural-Urban Migration	Land Reclassification
This paper	the 2000s	211	41.3	169.8	85.6	84.2
This paper*	the 2000s	211	28.7	182.3	89.7	92.6
Chan and Hu (2003)	the 1990s	157	31.4	125.6	91.1	34.5
Wang (2004)	the 1990s	167.8	28.5	139.3	51.7	87.6
Dragonomics (2011)	the 2000s	207	39.3	167.7	78.7	89
Chan (2012)	the 2000s	207	30	172	90	82

* stands for the results by the alternative definition of urbanization, which includes the newborns of the urbanized population in the period of urbanization.

Table 3: Basic Administrative Units by the NBS Rural/Urban Classification

	Street	RCT	VCT	Township
All	83974	29124	372662	185430
Urban	65879	27754	74311	6687
	(78.5%)	(95.3%)	(19.9%)	(3.6%)
Rural	18095	1370	298351	178743
	(21.5%)	(4.7%)	(80.1%)	(96.4%)

RCT and VCT represent “Residents’ Committee of Town” and “Villagers’ Committee of Town”, respectively. Source: Code of Urban/Rural Classification of Basic Administrative Units at http://www.stats.gov.cn/zjtj/tjbz/tjyqhdmhcxhfdm/2010/index.html