『国際開発研究フォーラム』27（2004. 8）
Forum of International Development Studies, 27（Aug. 2004）

Intraregional Inequality in China: A Decomposition Analysis

Ulrich REUTER*

Abstract

This paper analyzes the development of intra-provincial regional disparities in China between 1989 and 2001. Although previous studies did find evidence for significant regional disparities on the sub-provincial level, research on this topic remains sparse and mostly piecework. Neither trends nor causalities are sufficiently understood. Therefore, the purpose of this paper is to provide an consistent measurement of level, development and relative importance of intraregional disparities.

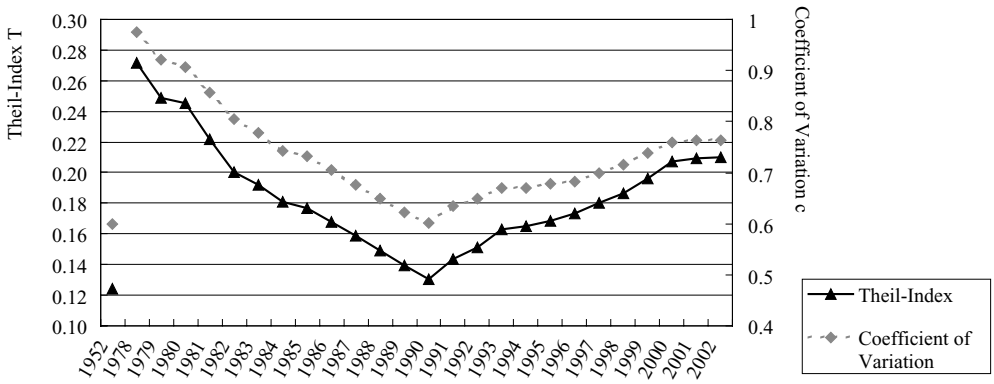
The results show that intraregional disparities contribute significantly to total inequality, and that the relative importance of these disparities is even increasing in recent years. By further decomposing intra-provincial income disparities, it is found that regional components account for about half of the measured intra-provincial inequality. Thus, the spatial component of intraprovincial disparities - compared with a sectoral or rural-urban division - should not be ignored. Additionally, the need for analysis on lower levels of aggregation is highlighted. Generally, the results call for a further analysis of intraregional disparities on the sub-provincial level.

Ⅰ. Introduction

Income disparities in transitional China are a multi-faceted, multi-dimensional phenomenon. They appear between geographical regions, between rural and urban areas, between industries, between ownership forms, and even between people who share all these characteristics. They can be measured, for example, on the national, macro-regional, provincial, prefectural, and household level. It is broadly acknowledged that income disparities for all measures in transitional China decreased in the early era of reform (1978 - mid 1980s), but steadily increased afterwards (see figure 1).

＊Doctoral Student, Graduate School of International Development, Nagoya university
e-mail: ureuterjp@yahoo.co.jp

Figure 1 Trend of Inequality in China, 1978 - 2002



Note: Regional income disparities measured on the provincial level.
Source: Data from NBS 1999a; NBS CSY 1999 - 2003, own calculations.

How serious are these inequalities? On the one hand, a difference in expected incomes - or the incentive to achieve a significantly different income from other people - is one of the most powerful driving forces of a market-coordination based economy. On the other hand, excessive inequality is supposed to have detrimental impacts on incentives, growth, and development. Frequent public and political statements and reports denouncing income disparities in China suggest that rising inequality there already has reached a critical stage. Accordingly, voluminous research on the topic finds alarming levels of disparities. Among all forms of inequality in China, urban-rural disparities are reported to be the biggest contributor; regional disparities are identified to be responsible for a smaller but growing portion of inequality (see World Bank 1997: 15).

No matter how well analyzed *inter*regional disparities are, our current knowledge on the extent and reasons of ***intra***regional inequality is extremely limited. Selective measurement is provided in studies by Khan et al. (1993), Khan and Riskin (2001), Tsui (1993, 1998a, 1998b), Hermann-Pillath, Kirchner and Pan (2002), Gustafsson and Li (2002), Akita (2000, 2001, 2003) and Song, Chu and Chao (2000). The common understanding is that intraregional disparities make a large proportion of total regional disparities. Khan et al. (1993: 66), for example, argue that "a careful analysis of regional differences in sources of inequality could be of much help in devising policies for improving income distribution." Moreover, it is important to understand the degree of intraregional inequality, because excessive inequality in one's close neighborhood is much more evident than a distant one, and shall therefore have much more detrimental impacts on incentives and development. If intraregional disparities preponderate, they should be of more political concern than disparities between macro-regions. However, differences in intraregional disparities "cannot be explained by such simple things as location, the degree of diversification of incomes or the level of development" (Khan et al. 1993: 66).

To facilitate this discussion, this paper attempts to provide a systematic measurement of the regional component of intraregional disparities in China over time and geographic space.

II. Background and Previous Studies

Although voluminous literature on interregional inequalities in China exists, explanations for the pattern of intraregional disparities are - as mentioned above - in short supply. The main reason for this could be the high level of aggregation used in most studies about the subject, which focus on disparities between macro-regions using province level data.

Hermann-Pillath, Kirchert and Pan (2002) analyze how the choice of different levels of aggregation is in the context of analyzing regional inequality in China. Using prefecture level data on economic development for 1993 and 1998, they find the common three-belt approach overly aggregated and inconsistent considering the internal economic structure of the belts. They recommend using provinces as appropriate units for understanding growth and development dynamics. However, they also note the strong disparities between prefectures as well as the continuing importance of the rural-urban gap (compare Hermann-Pillath; Kirchert and Pan 2002: 53). On the other hand, Knight and Song (1993) as well as Peng (1999) argue that counties should be the appropriate unit of analysis, "because every county behaves like a little kingdom."(Peng 1999: 247). By using cities as basic units, this paper follows the second argumentation.

Only a few studies go so far as to disaggregate the data and venture into the analysis of regional disparities or economic development on a sub-provincial level. In the following paragraphs, some major studies that nevertheless did so are introduced and discussed. In general, these studies attempt to measure disparities between regions in China either as regional differences in growth or by calculating direct inequality measures. As data, either regionally aggregated data (for cities or counties) or household income survey data has been used.

Wei (1993) and later Wei and Wu (2001) use city-level data to analyze the growth impact of foreign trade and investment on regional economies. While the first study suffers from a short time period and a very limited sample, the latter is able to establish a strong negative correlation between economic openness and urban-rural disparities (Wei and Wu 2001: 19). From their results, less intra-provincial inequality should be expected in the coastal areas, with highest disparities in the Northwestern regions.

The data set analyzed in Wei and Wu (2001)[1] is very similar to that used here. However, this paper is able to utilize more actual data, up to 2001. Moreover, Wei and Wu (2001: 9, 25) focus on large cities (prefecture- and province-level) that also administer adjacent rural counties, and include data on rural areas;[2] while I restrict my analysis to the urban data so as to obtain a more consistent sample. Occasionally however, I also include county level cities. These different data boundaries are reasonable considering the differing objectives of these papers: urban-rural comparison on the one hand, and cross-provincial comparison over time on the other. In this paper, I assume cities to be the economic gravity centers of their regions, thus representing the surrounding regional economies as

well. Including adjacent rural counties would create new sources of differences between cities depending on the relative importance of those rural areas in the city economy, because the strong rural-urban bias would be introduced into the data.

Khan et al.(1993) and Khan and Riskin (2001) use data from the 1988 and 1995 Household Sample Survey,[3] respectively, to calculate rural and urban Gini ratios for Chinese provinces. Their results show that the variability of intraregional inequality is greater for 1995 than for 1988, and inequality levels increased in almost all of the regions, often very drastically. In both papers, it is repeatedly stressed that no simple source can be identified to explain the pattern of regional inequality (see Khan et al.1993: 56, 58; Khan and Riskin 2001: 48f.). For example, surprisingly, neither rural and urban inequality nor rural and urban incomes appeared to be significantly rank-correlated (Khan et al.1993: 52f.). Therefore, differences in the composition of regional incomes by income source are analyzed, and similar income components are found to cause the interregional inequality to diverge as those observed to affect the national average: the wage income share in rural regions, and (disequalizing) subsidies in urban areas.

Gustafsson and Li (2002) utilize the same household survey data sets to directly address the question: how much income inequality can be measured within and across rural counties? Their explicit approach of decomposing and comparing the regional components of inequality come very close to the objective which this paper pursues. The authors conclude that most of income inequality in rural China has been spatial; however, inter-province and especially inter-belt disparities were more important than disparities between rural counties within provinces (see Gustafsson and Li 2002: 197f.).

Akita (2000; 2001; 2003) applies an analogical decomposition approach, calling it a nested decomposition analysis, for the cases of Indonesia and China. For China, he decomposes sub-provincial data for 1997, and reports that almost two-thirds of the overall inequality is due to within-province disparities. Similar results have also been reported by Tsui (1993) for county and city data in 1982, and Tsui (1998a, 1998b) for rural household survey data for the provinces Sichuan and Guangdong (1985 - 1990), where he finds only a very tiny share of overall inequality to be explained inter-provincially (Tsui 1998a: 794).

Finally, Song, Chu and Cao (2000) examine economic disparities between and within regions in China using city-level data from 1985 (106 cities, Survey of Income and Expenditure of Urban Households) and 1991 (477 cities, China Urban Statistics Yearbook 1992). The results lead Song, Chu and Cao (2000: 259, 252f.) to conclude that - within the widening regional inequality - large but similar intraregional disparities within the eastern, central and western regions exist. However, several serious limitations of their analysis invalidate their conclusions.[4]

Compared with the papers mentioned above, the contribution of this study is to describe the development of different components of inequality - in absolute as well as in relative terms - over time, using a broadly consistent data set on the city level. The decomposition methodology applied is

similar to that of Akita (2000, 2001, 2003) or Gustafsson and Li (2002), and the data used is roughly comparable with that used by Wei and Wu (2001). However, the approach is extended to include the sectoral division of city level data as well, and spans over a longer and more recent time period.

III. Measurement Issues and Data

III.1. The choice of an inequality measure

Various indicators can be used in the analysis of income disparities.[5] Although there is no consensus in the literature on which inequality measure is the most preferable, appropriate indicators could be chosen considering two main evaluation criteria:

- consistency with distributional and welfare axioms;
- practicability considerations.

(i) The most common way to judge the applicability of inequality measures is by comparing their behavior with some *axioms* theoretically derived as preferable properties of such measures. The main issue addressed by this approach is the reasonability of the ordering criterion; which is necessary for any meaningful inequality comparison (see Fields and Fei 1978: 315).

Several systems of axioms have been proposed, differing widely in the number and strength of their required axioms. Usually they include (see Cowell 2000: 97ff.: Fields 2001: 15ff.):

- anonymity (no personal characteristics other than the income determine the ordering principle),
- scale independence or income homogeneity (multiplying all incomes with the same positive scalar does not change inequality),
- population independence or population homogeneity (replicating each income an integral number of times does not change inequality),
- the transfer principle or Pigou-Dalton condition (transfers from a richer to a poorer person do reduce the measured inequality).

Surprisingly, only a few measures can satisfy the commonly accepted axioms. These include the coefficient of variation, the Gini coefficient, the Atkinson class of measures, and finally the generalized entropy family of measures, with the Theil index as the most prominent example.[6] However, only the latter survives very strong versions of the axioms.

(ii) Given an inequality measure fulfils the above stated axiomatic general conditions, a final choice of the appropriate measure can be made considering the practicability of each measure in the specific context. This relates not only to problems like data availability and complexity of calculation, but also to additional properties that an inequality measure can offer, and depends mainly on the specific purpose for which the measure is calculated. In this paper, the importance of sub-provincial regional inequality shall be assessed; therefore an important property of an appropriate inequality measure would be its decomposability (compare Sen 1997: 149ff., esp. 156).

In the context of decomposability, two types of decomposition are addressed: decomposition by

subgroups (e.g.regions, population subgroups, etc.) and by income source (e.g.income from wages, property, subsidies, etc.) (see Cowell 2000: 123). In this paper, decomposition into subgroup components is intended. For this purpose, a desirable measure requires two decomposition properties (see Cowell 2000: 124f.):

- subgroup consistency (which means the positive responsiveness of the overall inequality measure to changes in the inequality levels of constituent groups, see Sen 1997: 157) as a minimum requirement; and
- additive decomposability (overall inequality is the sum of all between-groups and within-groups inequality) as an additional restriction.

These two properties together are only satisfied by the Theil index and the MLD index (see Cowell 2000: 125). Therefore, one of these measures will be applied in the following decomposition analysis.

The Gini coefficient is not decomposable in the sense of subgroup consistency; a decomposition of the index will produce an additional interaction term as long as the subgroups are overlapping (see Cowell 2000: 125 - 127), as is the case with regional incomes. However, given the wide popularity and the otherwise favorable properties of the Gini index, it will be supplied as a measure for total inequality and cross-regional inequality comparisons.

III.2. Data Sources and Selection

In this study, inter- and intraregional disparities are reviewed by analyzing a sample of 215 cities over a 13 year period from 1989 to 2001. The data used in this article is mainly taken from "Fifty Years of Cities in New China" published by the National Bureau of Statistics of China (1999b), a collection of historic data on main economic and social indicators on the city level, mainly based on data from the various issues of the Urban Statistical Yearbooks. The time series data reported in this source cover the period from 1990 till 1998. I also used the Urban Statistics Yearbooks for additional data (volumes 1990, 2000 - 2002) and for crosschecking and verification purposes (various issues).

Generally, the number of cities in China increases over time. While for 1991, there were only data for 479 cities reported in the Urban Statistics Yearbook, this number had risen to 662 cities in 2001.[7] This increase is mainly due to the upgrading of many county-seat towns into county-seat cities.[8] Given the broad and varying coverage of cities within these publications, I had to limit the sample to a consistent set of cities to ensure comparability. Therefore, only cities for which all necessary information was available across the entire period were included. Exceptions were only made in cases where formerly independent cities merged or where simply name changes occurred, as far as these changes became obvious from the data sources. This data selection reduces the total number of observations to only 215 (compared with 662 available observations in 2001).

Nevertheless, even with this cautious and prudent data selection, minor inconsistencies may occur, since some of the time series for identical locations do show relatively high volatility, especially the

case of population data, and may point to significant changes in the size of jurisdictions or the inner-provincial organization of administration, regardless of identity in names. Moreover, keeping consistency over time necessarily comes at the cost of loosing comprehensiveness. Here, a bias may have been introduced into the results, since some of the most dynamically developing cities had to be excluded. These are either previously poorer towns which became vibrant cities during reform; or previous cities which grew very dynamically, thus necessarily changing their organizational structure over time. In the first case, the measurement bias should be expected to overstate (the rise in) regional inequality, because it disregards the catch up process of towns. In the second case, however, even larger inequality might be expected when such cities were to be included into the sample. Thus, although this analysis can shed some light on intraregional inequality development in the sample over time, it may only to a limited extent be regarded as representative for all China. This point as well will be addressed in the following analysis by comparing the city set here with the results for all cities in selected years.

The main data taken from the sources described above includes:

- population data: year-end population (nianmo zong renkou); and annual average population (nian pingyue renkou, only for 1989; all other values calculated as average between two consequent year);

- income data: GDP (guonei shengchan zongzhi) and share of secondary, tertiary sector;

- employment data: year-end total employment (nianmo quanbu congye renyuansu), and percentages for secondary and tertiary sectors.[9]

For decomposition purposes, provinces are grouped into three macro-regions according to the so-called "three economic belts" concept, following the common classification:[10]

- Coastal: Beijing, Tianjin, Hebei, Liaoning, Shanghai, Jiangsu, Zhejiang, Fujian, Shandong, Guangdong, Guangxi, Hainan;

- Central: Shanxi, Inner Mongolia, Jilin, Heilongjiang, Anhui, Jiangxi, Henan, Hubei, Hunan;

- Western: Sichuan, Chongqing, Guizhou, Yunnan, Tibet, Shaanxi, Gansu, Qinghai, Ningxia, Xinjiang.

Of these, the municipalities Beijing, Tianjin, Shanghai and the province Qinghai are excluded from the analysis due to non-available data for the sub-provincial level, and Chongqing and Tibet due to missing data.

III.3. Methodology

For the above described data set, I calculated Gini coefficients and Theil Indices, decomposing the latter into three regional components of inequality: inequality between macro-regions, inequality within macro-regions between provinces, and inequality within provinces. The following section presents the definition of applied measures.

3.1 Overall Inequality Measures:

For measuring aggregate inequality, two measures will be applied: the Gini coefficient and the generalized entropy measure in form of the Theil index.

Following Sen (1997), the Gini index G is often defined as:[11]

(1) $$G = \frac{1}{2n^2\mu_y}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|y_i - y_j\right|$$

with

G = Gini coefficient, n = number of individuals,

μ_y = average income (mean), y_i = income of person i , y_j = income of person j

This formula, however, is designed for individual data. Because the data under consideration here is grouped and, moreover, the partition is not equal, the following formula will be used for calculating the Gini for grouped data:[12]

(2) $$G = \sum_{i=1}^{n-1}\left(\frac{N_i}{N}\right)*\left(\frac{Y_{i+1}}{Y}\right) - \sum_{i=1}^{n-1}\left(\frac{N_{i+1}}{N}\right)*\left(\frac{Y_i}{Y}\right)$$

with

n = number of groups, N_i = cumulative population,

N = total population, Y_i = cumulative income , and Y = total income.

The generalized entropy class of inequality measures as second indicator is defined as (compare Shorrocks 1980: 622; definition of variables as before):

(3) $$I_\alpha = \frac{1}{\alpha(1-\alpha)}\frac{1}{n}\sum_{i=1}^{n}\left[1-\left(\frac{y_i}{\mu_y}\right)\right]$$

for all $0 < \alpha < 1$; and with $\alpha = 1$ as the Theil Index T:

(4) $$T = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{\mu_y}\ln\left(\frac{y_i}{\mu_y}\right)$$

For grouped data, a typical way of writing the Theil-Index T is:[13]

(5) $$T = \sum_i\sum_j\left(\frac{Y_{ij}}{Y}\right)\ln\left(\frac{Y_{ij}/Y}{n_{ij}/N}\right)$$

with

Y_{ij} = total income of the i-j group,

n_{ij} = absolute frequency of population in the i-j group; and

$Y = \sum_i\sum_j Y_{ij}$ as total income over all groups, and $N = \sum_i\sum_j n_{ij}$ as total population.

This Theil index compares the relative income share of each group (Y_{ij}/Y) with its relative share in the total population (n_{ij}/N).

Both indicators will take values between zero and one for perfect equality or concentration of income respectively. Since both indicators satisfy the distributional axioms described above, their results should not differ with regard to the ordering alternative distributions from the same data set. They differ, however, in the weight they attach to a specific income in this distribution, and therefore in their cardinal measurement of inequality. For the Theil index, sensitivity for transfers in different income classes is defined by the parameter $\alpha = 1$; which implies a relative overweighting of lower income groups. In inequality analysis, this statistical property is often regarded as positive, since similar income transfers are considered to matter more for personal utility in the lower end of the distribution. The sensitivity of the Gini coefficient depends on the relative position of the individual in comparison to other individuals (so-called irrelevant alternatives). Therefore, if more people are in the lower end of the income distribution - as is usually the case - these lower incomes will get a stronger weight.

Another possible difference between the indicators is their sensitivity on differing sample sizes. Khan and Riskin (2001: 163) claim that the Theil index is very sensitive to the sample size. To reduce the problems arising from this property, the number of observations has been held constant over time; however, obvisously, a direct comparison between groups is complicated.

3.2 Decomposition Analysis

a) Basic Concept

The Theil index defined above can easily be decomposed into within and between group components for different groups of income receivers, or - as in this case - regions. The common formula for this decomposition is

$$(6)\quad T = T_B + T_W = \left[\sum_i \left(\frac{Y_i}{Y}\right)\ln\left(\frac{Y_i / Y}{n_i / N}\right)\right] + \left[\sum_i \left(\frac{Y_i}{Y}\right)T_i\right]$$

with

$Y_i = \sum_j Y_{ij}$ as the total income of the ith group, $n_i = \sum_j n_{ij}$ as the absolute frequency of population in ith group, and $T_i = \sum_j \left(\frac{Y_{ij}}{Y_i}\right)\ln\left(\frac{Y_{ij} / Y_i}{n_{ij} / n_i}\right)$ as the Theil-Index for the ith group.

In this paper, I go one step further by decomposing the within group inequality. Two ways of decomposing the Theil index for intraregional inequality will be applied; one could call them vertical and horizontal decomposition, depending on whether groups are defined exclusively according to different administrative levels (national, belts, provinces, cities), or whether groups within the same administrative level are defined according to additional characteristics (e.g.a sectoral division).

b) Vertical Decomposition of Intraregional Inequality

First, intercity inequality will be decomposed vertically, distinguishing between three subsequent levels of disaggregation:

- macro-regions (or belts; the coastal, central and western regions);
- provinces and province level municipalities as subgroups of these macro-regions; and
- cities as basic units of which the provinces are composed.

Such nested inequality decomposition into three components exceeds the common decomposition analysis using Theil indices into inequality between and within provinces.[14] It therefore shows disparities below the aggregated provincial level, and at the same time allows a comparison of the importance of each inequality component within the big picture.

For this purpose, the Theil index describing total inequality is redefined as:

$$(7)\quad T = \sum_h \sum_i \sum_j \left(\frac{Y_{hij}}{Y}\right) \ln\left(\frac{Y_{hij}/Y}{n_{hij}/N}\right)$$

with

Y_{hij} and N_{hij} as total GDP and absolute population of city j in province i in macro-region h, $Y = \sum_h \sum_i \sum_j Y_{hij}$ = total income of cities, and $N = \sum_h \sum_i \sum_j n_{hij}$ = total population of all cities.

Overall inequality then can be decomposed into its regional components as follows:

$$(9)\quad T = T_B + T_P + T_C = \left[\sum_h \left(\frac{Y_h}{Y}\right) \ln\left(\frac{Y_h/Y}{n_h/N}\right)\right] + \left[\sum_h \left(\frac{Y_h}{Y}\right) T_h\right] + \left[\sum_h \sum_i \left(\frac{Y_{hi}}{Y}\right) T_{hi}\right],$$

with

T_B = Component of the Theil-Index for 'Per capita GDP'-inequality between macro-regions,

T_P = Component of the Theil-Index for inequality between provinces in macro-regions,

T_C = Component of the Theil-Index for inequality between cities within provinces,

$T_h = \sum_i \left(\frac{Y_{hi}}{Y_h}\right) \ln\left(\frac{Y_{hi}/Y_h}{n_{hi}/n_h}\right)$ = 'Per capita GDP'-inequality between the provinces in macro-region h,

$T_{hi} = \sum_j \left(\frac{Y_{hij}}{Y_{hi}}\right) \ln\left(\frac{Y_{hij}/Y_{hi}}{n_{hij}/n_{hi}}\right)$ = Inequality between cities within province i in region h,

$Y_h = \sum_i \sum_j Y_{hij}$ = total income of region h, $n_h = \sum_i \sum_j n_{hij}$ = total population of region h,

$Y_{hi} = \sum_j Y_{hij}$ = total income of province i in macro-region h,

$n_{hi} = \sum_j n_{hij}$ = total population of province i in macro-region h, and Y_{hij} and n_{hij} as above.

c) Horizontal Decomposition of Intraregional Inequality

In the second step, I address more specifically the nature of disparities within provinces. Therefore, I am going to ask whether it is appropriate to focus the analysis on regional differences between sub-units (in this case: cities) of a province, or whether the real break within provinces rather exists between alternative groups, e.g. urban and rural areas, or different sectors of the local economy. To do so, I will decompose intra-provincial income disparities into income differences between sectors and income differences between cities (or more exactly; between sectors on the provincial level, and within sectors but between cities).

Applying the tripartition into the primary, secondary and tertiary sectors,[15] this classification embraces not only the division between rural (represented mainly by the primary sector) and urban economic activities, but also differences between economic sectors per se. However, this division of urban-rural by sectors is not common in the literature, which usually defines rural and urban activities by geographic areas. Nevertheless, in contemporary China, the so-called *rural* vicinities of large coastal cities might easily appear to be much more urban in the common sense than minor cities in, e.g., the hinterland of the western province of Sichuan; making the previous distinction between rural and urban areas misleading. Moreover, since the data used in this study focuses on urban areas specifically, I prefer the division along economic activities rather than according to an arbitrary rural-urban classification.

Since the derivation of the decomposition formula is similar to the previous case, only the final version is presented below. It is identical down to the provincial level, but redefines the intra-provincial inequality as sum of inter-sectoral inequality T_S and intra-sectoral but inter-city inequality T_{SC}. For this purpose, the new smallest unit of disaggregation becomes the sector within each city (which also leads to a change in the subscript). Another difference is that because population data is not available according to this sectoral division, employment data is used instead. Finally, since this level of disaggregation was not available for all years consistently over all cities in the sample, the time span reduces to the period from 1990 to 1997, and the total number of cities in the sample to 212.

Thus the second decomposition formula becomes:

(10) $$T = T_B + T_P + T_S + T_{SC}$$
$$= \left[\sum_h \left(\frac{Y_h}{Y}\right) \ln\left(\frac{Y_h / Y}{m_h / M}\right)\right] + \left[\sum_h \left(\frac{Y_h}{Y}\right) T_h\right] + \left[\sum_h \sum_u \left(\frac{Y_{hu}}{Y}\right) T_{hu}\right] + \left[\sum_h \sum_u \sum_v \left(\frac{Y_{huv}}{Y}\right) T_{huv}\right]$$

with

T_B = Component of the Theil-Index for inequality in GDP per employee between macro-regions,

T_P = Component of the Theil-Index for inequality between provinces in macro-regions,

T_S = Component of the Theil-Index for inequality between sectors within provinces,

T_{SC} = Component of the Theil-Index for inequality between city sectors within provinces,

$T_h = \sum_u \left(\frac{Y_{hu}}{Y_h} \right) \ln \left(\frac{Y_{hu} / Y_h}{m_{hu} / m_h} \right)$ = 'GDP per employee'-inequality between u provinces within region h,

$T_{hu} = \sum_v \left(\frac{Y_{huv}}{Y_{hu}} \right) \ln \left(\frac{Y_{huv} / Y_{hu}}{m_{huv} / m_{hu}} \right)$ = Inequality between v sectors within province u in region h,

$T_{huv} = \sum_w \left(\frac{Y_{huvw}}{Y_{huv}} \right) \ln \left(\frac{Y_{huvw} / Y_{huv}}{m_{huvw} / m_{huv}} \right)$ = Inequality between w cities within sector v,

$M = \sum_h m_h = \sum_h \sum_u m_{hu} = \sum_h \sum_u \sum_v m_{huv} = \sum_h \sum_u \sum_v \sum_w m_{huvw}$ = total employment in all cities,

m_{hu}, Y_{hu} as total employment and income in province u in region h,

m_{huv}, Y_{huv} as sectoral employment and income in sector v in province u in region h,

m_{huvw}, Y_{huvw} as sectoral employment and income in city w and sector v in province u in region h,

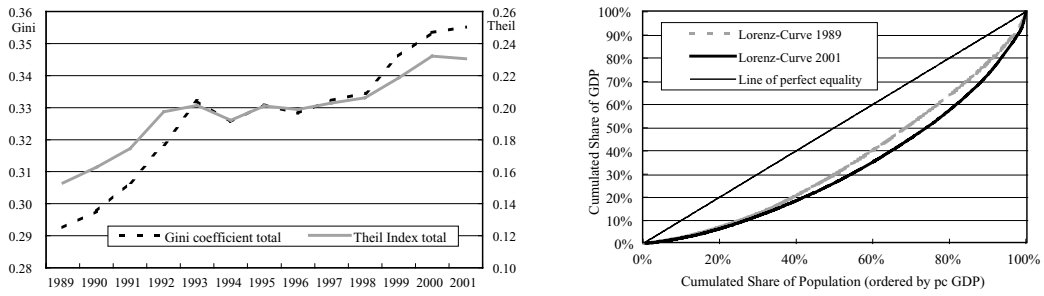
m_h = total employment in region h; and Y_h and Y defined as before.

IV. Results and Interpretation

a) Overall Inequality

As a first result, the following graph shows the development of total inequality for the specific sample on the city level as described by Gini coefficient and Theil index. As expected, a steady increase over the analyzed time span is observed by both measures. However, during the middle of the 1990s, they remain relatively constant; a characteristic that was not detected by the province level data in figure 1.

Figure 2 Overall Income Disparities on the City Level

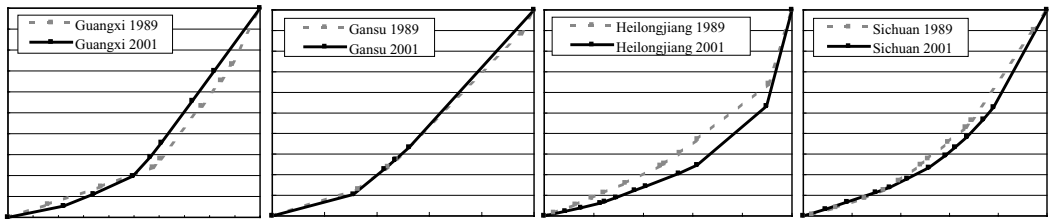


To get a better impression of how to interpret these changes in inequality, the second graph visualizes the distributive changes in the form of two Lorenz-curves. The second curve, describing income distribution in 2001, lies constantly below the curve of 1989; this corresponds to a generally more unequal distribution of regional incomes. However, the distance to the original curve is especially large in the right part of the diagram, which shows the income share of richer regions. Therefore, we can conclude that income distribution did worsen especially at the upper end of the distribution. On this very high level of aggregation, regional income disparities in China appear as a

problem of extreme wealth rather than of extreme poverty.

This fact might lead to the conclusion that the current focus on poverty reduction in development economics might not be the appropriate measure for China. Should China concentrate more on linking the extreme rich cities with the rest of the country, instead of financing poverty reduction programs? The graph, however, depicts only a national average. Consequently, when looking at the local level, a much more diverse picture evolves, as depicted in figure 3. Here, the Lorenz curves for two richer and larger (Heilongjiang, Sichuan) and two smaller and poorer provinces (Guangxi, Gansu) are graphed.

Figure 3 Provincial Lorenz Curves



The picture shows that, while in the richer provinces disparities clearly did widen in the upper part of income distribution, the opposite was true in the poorer provinces; in both Gansu and Guangxi disparities in richer incomes were actually even reduced, as the intersecting Lorenz curves depict. The increased inequality in the poorer income groups in poorer provinces points to the development of poverty pockets in these provinces. Naturally, the increase in these poor and densely populated areas is outmatched when aggregated with the rich and large provinces. Thus, the above stated an interpretation has to be regarded with caution.[16] Both, a better integration of extremely rich cities with the rest of their province, as well as the fight against poverty in backward poverty locks are important policy measures that have to be taken seriously.

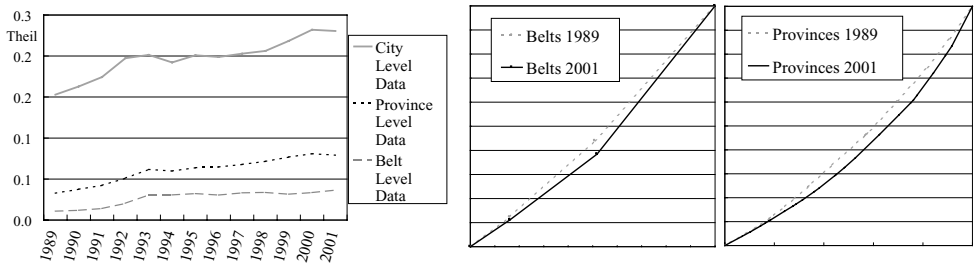
The facts unraveled above do clearly show the importance of incorporating lower level aggregation into the analysis. However, considering these results, what additional insight has been achieved on the aggregate level by using city level data instead of province-level or even macro-level data? To show the differences in results, figure 4 compares Theil indices that result from using cities as basic unit with those derived from provincial or belt-level aggregation.

Besides representing much lower absolute values of inequality - already due to the exclusion of inequality within sub-units and smaller sample size - the higher-level aggregation hides much of the fluctuations in the data. To take one example, the local minimum in 1994 can hardly be identified using aggregated data. One possible explanation might be a surprising one: As in the beginning of reforms, when rapid growth was accompanied by decreasing inequalities, so might the economic boom following Deng Xiaoping's 1992 Southern Journey have had inequality reducing effects, possibly due to a smoother dispersion of growth from until-then isolated growth poles. If periods with accelerated

liberalization and less state intervention turn out to reduce regional disparities in China, aren't these disparities then maybe rather policy induced than a result of the transformation process?

Another way of analyzing the differences is by comparing the corresponding Lorenz-curves, which also are presented below. As before, they show the 2001 curve consistently below the 1989 base line, indicating a higher level of inequality throughout all income levels. However, the stronger increase in the upper end of the income distribution is almost not noticeable.

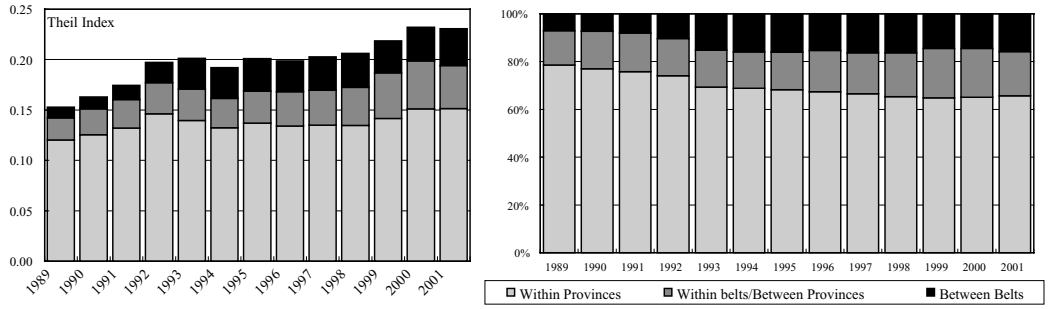
Figure 4 Disparities on Different Levels of Aggregation



b) Vertical Inequality Decomposition

As a next step, overall inequality shall be decomposed into its regional components, applying the methodology described above. The results are graphed in figure 5. This figure shows that the largest share in regional inequality - almost 60 percent - can be attributed to intra-provincial disparities. However, although the absolute value of disparities on the lower level still increased throughout the 1990s, its relative importance declined. Nevertheless, in the last two years its share has stabilized on a high level. Interestingly, except for a steep increase in the beginning of the 1990s, the often reported rise in the relative importance of inter-belt disparities is not supported by the data.

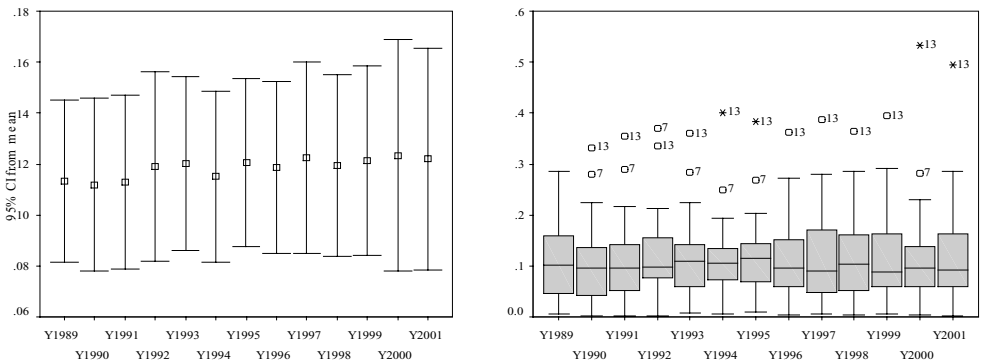
Figure 5 Vertical Decomposition of Theil Index



Given the immense importance of intra-provincial disparities, how much do they vary across China? Or, on the other hand, are they a common phenomenon? In the following, the decomposition results derived above are analyzed further to answer these questions. Therefore, the main measures for location and dispersion of the provincial data on intraregional inequality - mean, standard deviation,

etc. - are calculated (left graph). They reveal a steady increase in the average value as well as in the spread of intra-regional disparities between provinces throughout the 1990s.

Figure 6 Measures of Central Tendency and Dispersion of Intra-Provincial Disparities



Note: 13 = Heilongjiang; 7 = Guangdong

To visualize the results, the error bar and boxplot diagrams are shown in the above figure. The first graph depicts the range of the 95 percent confidence interval for the mean of intraprovincial Theil indices. In this diagram, the rising and dispersing trend of intraregional inequalities is apparent. The second graph presents the median and interquartile range. These are not sensitive to extreme values, thus allowing the identification of outliers and extrema.[17] Thus it becomes obvious that much of the increase in the means and variations of intraregional disparities between provinces is due to outliers and extreme cases in the upper part of the distribution; notably Heilongjiang and Guangdong. This reconfirms one of the earlier results of the analysis: Disparities in China are mainly a problem of differences in the incomes of the rich, rather than a result of extreme poverty. However, the different sources of intra-provincial disparities, as identified before, become not apparent from the graph.

Combining the results of the previous analysis, a rather surprising result appears: Although intraregional inequality in China seems to be on the rise when measured by the mean (a result that many studies on the subject share), the rising trend in inequality largely disappears when outliers are accounted for (e.g.by using measures that are insensitive to outliers like the median). Thus, apart from some very prominent examples, intraregional disparities seem to be rather stable over time. As a first result, the claim of strongly rising inequality in recent years cannot be verified generally.

Nevertheless, a widening spread between relatively equal and relatively unequal provinces is identified by both kinds of measures. Moreover, the extreme increase in inequality in some rich provinces can partly be contrasted with the development in poorer provinces. As seen by comparing provincial Lorenz curves, the trend in these provinces is driven by different forces. While in the richer provinces, the uneven growth of a few economic centers disequalizes the income distribution in its upper part, poorer provinces experienced the development of some poverty locks while generally becoming more equal in their richer regions. Considering the development of disparities over time,

one might suspect that disintegrating policy measures are an important factor in causing intraregional disparities in the richer provinces, while disparities in poorer provinces - usually unnoticed in the aggregated data - might rather be influenced by geographic backwardness.

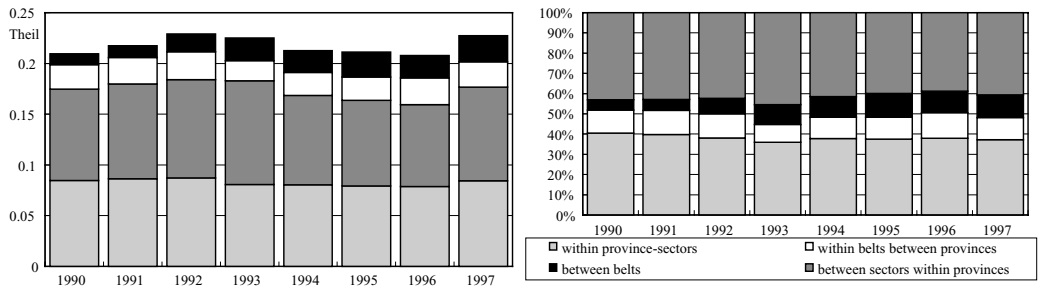
c) Horizontal Inequality Decomposition

Until now, the analysis entirely focused on *unpacking* the aggregated regional data into smaller regional divisions. However, it has been argued that within provinces, characteristics other than location are more important in determining incomes. Especially the division of economic activities into rural and urban is commonly regarded as an important factor. Given these arguments, is it really appropriate to compare sub-provincial jurisdictions to explain intraregional disparities?

To reply to this possible objection, another way of decomposing a less aggregated set of income data is applied. Taking the primary, secondary and tertiary sectors of each city in the sample as the basic units, the contribution of sectoral and locational characteristics in determining intra-provincial disparities are separated. As explained earlier, the sectoral division used here is expected to address not only economic activities, but also to be a proxy for the rural-urban divide within one region. Due to data availability, the time span analyzed is much shorter now and covers only the years 1990 - 1997.

Since the data set used now differs in level of aggregation, sample size and definition of the population variable (year-end employment instead of average population), the following figure shows the new results of the overall inequality measurement and its decomposition. Note that the overall Theil index now is divided into four components: between macro-regions, within macro-regions but between provinces, within provinces but between sectors, and within provincial sectors but between city sectors.

Figure 7 Horizontal Decomposition of Theil Index



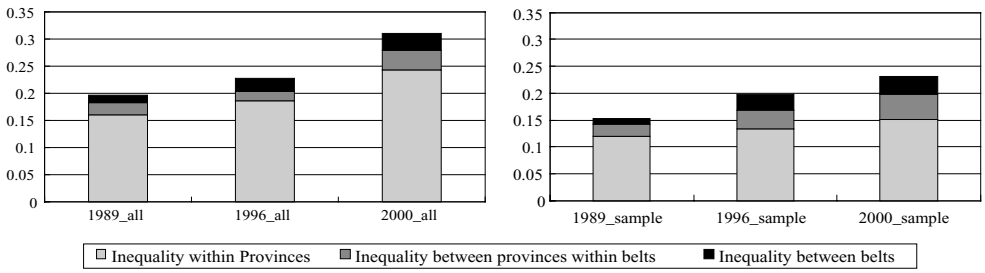
From the second graph, the percentage of contribution of each component to overall sector-regional inequality can be seen. Although up to 1993, the importance of sectoral components (upper part of the bar) increased steadily - mainly at the cost of inter-city disparities -, the trend reversed since then. On average, inter-sectoral and inter-city components explain roughly one half of intra-provincial disparities. It thus can be concluded that both sources - locational characteristics as well as the

different source of local incomes, can explain each a significant proportion of intra-provincial disparities. Therefore, the further analysis of factors affecting intraregional disparities needs to pay attention to both the regional dimension and the material scope of intraregional incomes.

d) Addressing a Possible Selection Bias in the Sample

As already described in section III. 2., the focus on sample consistency in this paper might have compromised the comprehensiveness of the results. The question arising is: Is the sample of 215 cities employed here really representative for (urban) China as a whole?[18] To address this problem, I compared the three-level vertical decomposition results as well as the overall inequality structure described by the Lorenz curve for my sample and for all cities for which data was available in selected years. Beijing, Tianjin, Shanghai and Chongqing[19] remain excluded, since there would have been only one observation available for the entire municipality. The results of these new analyses are shown in figures 8 and 9.

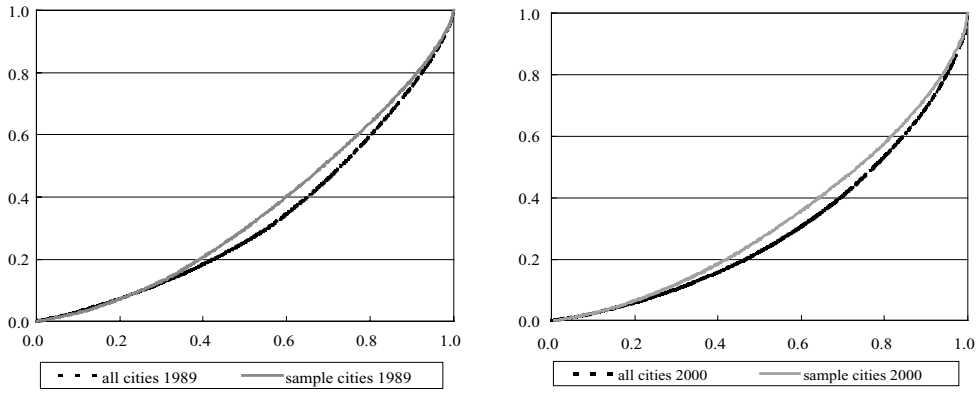
Figure 8 Theil Index Decomposition for All Cities and for the Described Sample, Selected Years



Source: Data from NBS CSY, various issues; own calculations.
Note: Number of cities: 1989 = 440; 1996 = 634; 2000 = 650; for all years: 25 provinces.

Both figures consistently show that the extent of regional inequality was underestimated by the sample used in this paper. However, as it can be seen in figure 8, the trend and relative importance of intraregional disparities identified in both data sets is almost identical. Not surprisingly, the inclusion of more observations on the sub-provincial even increases the absolute and relative share of inequality within provinces. This tendency becomes stronger especially in the later years. Thus, as a first conclusion, the sample used in this study can be regarded to broadly describe the trend and composition of inequality in the whole urban China; however, it tends to underestimate the degree of intraprovincial disparities especially in more recent years.

Figure 9 Lorenz Curve Comparison for All Cities and for the Described Sample, Selected Years



Beside this, figure 9 compares the shapes of total national inequality described by Lorenz curves. In both selected years, 1989 and 2000, the curve for all cities lies significantly below the sample curve for higher incomes. However, in the lower end of the income distribution, the difference in not that large. In 1989, the Lorenz curve for all cities even lies above the sample curve. This indicates that cities that were excluded from the sample were rather rich cities. Thus, if applying the analysis described in this paper to a more complete set of cities, one might expect to find even stronger evidence for the general claim that disparities in China are mainly a problem of differences in the incomes of the rich, rather than a result of extreme poverty.

Conclusion

This paper addressed the development of intra-provincial disparities in China between 1989 and 2001. For this aim, I analyzed city level data a set of more than 200 cities in 25 provinces, which can be regarded as being representative for urban China as a whole. A Theil index decomposition method was applied. The results show that intraregional disparities contribute significantly to total inequality, and that the relative importance of these disparities has even increased in recent years. Thus, the regional component of intraprovincial disparities - compared with a sectoral or rural-urban division - should not be ignored. However, by focusing on specific provinces it could be demonstrated that the pattern of inequality development is not identical within all provinces. The larger provinces, which also dominate the aggregated decomposition results, reconfirm the common perception on inequalities in China to be rather a problem of extreme wealth than of extreme poverty. On the other hand, smaller and poorer provinces do clearly show an increase in inequality especially in the lower end of the income distribution, representing an increase in poverty. Furthermore, some of the results of inequality analysis tend to be very sensitive to the presence of outliers. Thus, the need for careful analysis on lower levels of aggregation is highlighted. Generally, the results call for further study of intraregional disparities on the sub-provincial level.

Endnotes:

1 For about 100 cities over the period 1988 - 1993, taken from NSB USY (various years from 1985 to 1995) and NSB (1999b).

2 In the statistical sources called "regional" (diqu) data or "including counties under jurisdiction" (baokuo shixia xian).

3 For details on the survey in 1988 see Eichen and Zhang (1993), for 1995 Khan and Riskin (2001).

4 The very different nature, size and source of their samples (1985: data on household income and expenditure for 106 cities, Survey of Income and Expenditure of Urban Households; 1991: per capita income and per capita GDP for 477 cities, China Urban Statistics Yearbook 1992) make the comparison between the two years rather difficult. How much the source of data can matter in Chinese statistics has been demonstrated for example by Scharping (2001: 324) for the case of population statistics. Moreover, in analyzing the determinants of regional per capita income differences, the study applies a growth accounting framework on the city level (Song, Chu and Chao 2000: 256ff.); however, without considering any regional aspects (except a coastal province dummy). Therefore, contrary to their claims, their analysis does achieve very little in describing or even explaining intraregional disparities in different regions over time.

5 For an overview and some examples, see Champernowne 1974: 791; Fields 2001: 30; and Sen 1997: 24ff..

6 Strictly speaking, the coefficient of variation can be regarded as the square root of the generalized entropy measure for $\alpha = 2$ (see the exact formula of the generalized entropy family of measures below); and generalized entropy measures with $\alpha < 1$ can be regarded as monotonic transformation of the Atkinson measure. In all cases, α can be interpreted as "inequality aversion" parameter, with lower values of α being more sensitive to changes in the incomes of poorer subgroups. $\alpha = 2$ than represents "transfer neutrality", while generalized entropy measures with $\alpha > 2$ would be especially sensitive to transfers on the upper end of distribution. See Sen (1997: 140f.).

7 See Urban Statistical Yearbook (1990, 2002).

8 See Song, Chu and Chao (2000: 250).

9 The definition of employment statistics, however, changed significantly over time.

10 This "three economic belts" (*sanda jingji didai*) division was officially the first time adopted in the Seventh Five-Year Plan (1986 - 1990), see Fan (1997: 623).

11 There exists, however, a large number of possible definitions of the Gini coefficient. For an overview, see Xu (2004).

12 See Hettige Don (1998: 91). Note that this formula - as well as other formulas used later - assumes equal incomes within each subgroup (in this case: city), and therefore disregards inequality within the subgroups, thus systematically underestimating total inequality. Note also that this computation requires data to be ordered by per capita income.

13 Compare for example Ikemoto (1991: 167) or Hettige Don (2000: 218).

14 So far, a similar approach has been applied to the Chinese case only by Akita (2000, 2001, 2003) and Gustafsson and Li (2002).

15 The sectors are defined as follows: Primary sector: agriculture, forestry, animal husbandry and fishery; Secondary sector: mining and quarrying, manufacturing, production and supply of electricity, water and gas, and construction; Tertiary sector; all economic activities not included in primary or secondary industry.

16 Two more methodological caveats have to be mentioned. First, the data used here nevertheless is aggregated, thus not showing the individual income distribution. Moreover, it focuses on urban areas exclusively, thus, it cannot reflect rural poverty.

17 Outliers are defined as cases with values between 1.5 and 3 box length from the lower or upper edge of the box (which represents the interquartile range). Values with more than 3 box length difference are called extremes.

18 I thank the two anonymous referees for repeatedly pointing out this important qualification.

19 Chongqing, however, is this time included in the 1989 data for "all cities" as part of Sichuan province.

References

Akita, Takahiro. 2003. "Decomposing Regional Inequality in China and Indonesia Using Two-stage Nested Theil Decomposition Method". In: *The Annals of Regional Science*. 37(1): 55 - 77.

Akita, Takahiro. 2001. "Regional Income Inequality in China - A Two-Stage Nested Inequality Decomposition Analysis". International University of Japan Research Institute. *International Development Series Working Paper Nr. 9*.

Akita, Takahiro. 2000. "Decomposing Regional Inequality Using Two-stage Nested Theil Decomposition Method". International University of Japan Research Institute. *International Development Series Working Paper No. 2*.

Atkinson, Anthony B. 1970. "On the Measurement of Inequality". In: *Journal of Economic Theory*. 2(3): 244 - 263.

Cowell, Frank A. 2000. "Measurement of Inequality". In: Atkinson, Anthony B. and Bourguignon, Francois. 2000. *Handbook of Income Distribution*: 87 - 166. Amsterdam: Elsevier Science.

Eichen, Marc and Zhang, Ming. 1993: "Annex: The 1988 Household Survey - Data Description and Availability". In: Griffin, Keith and Zhao, Renwei (eds.). 1993. *The Distribution of Income in China*: 25 - 73. New York, N.Y. : St. Martin's Press.

Fan, Cindy C. 1997. "Uneven development and beyond: Regional development theory in post-Mao China". In: *International Journal of Urban and Regional Research*. 21(4): 620 - 639.

Fields, Gary S. 2001. *Distribution and Development: A New Look at the Developing World*. Cambridge - London: The MIT Press.

Fields, Gary S. and Fei, John C.H. 1978. "On Inequality Comparisons". In: *Econometrica*. 46(2): 303 - 316.

Gustafsson, Björn and Li, Shi. 2002. "Income Inequality Within and Across Counties in Rural China 1988 and 1995". In: *Journal of Development Economics*. 69: 179 - 204.

Herrmann-Pillath, Carsten; Kirchert, Daniel and Pan, Jiancheng. 2002. “Disparities in Chinese Economic Development: Approaches on Different Levels of Aggregation”. In: *Economic Systems*. 26(1): 31 - 54.

Hettige Don, Karunaratne. 1998. “Income Inequality in Sri Lanka: A Disaggregated Analysis by Factor Incomes”. In:*Forum of International Development Studies*. 10: 89 - 111.

Ikemoto, Yukio. 1991. “Income Distribution in Thailand: Its Changes, Causes and Structure”. *I.D.E Occasional Paper Series 25*. Tokyo: Institute of Developing Economies.

Khan, Azizur Rahman; Griffin, Keith; Riskin, Carl and Zhao, Renwei. 1993. “Household Income and its Distribution in China”. In: Griffin, Keith and Zhao, Renwei (eds.). 1993. *The Distribution of Income in China*: 25 - 73. New York, N.Y. : St. Martin’s Press.

Khan, Azizur Rahman and Riskin, Carl. 2001. *Inequality and Poverty in China in the Age of Globalization*. Oxford; New York et al.: Oxford University Press.

Knight, John and Song, Lina. 1993. “ The Spatial Contribution to Income Inequality in Rural China”. In:*Cambridge Journal of Economics*. 17: 159 - 213.

National Bureau of Statistics of China (NBS). 1999a. *Comprehensive Statistical Data and Materials on 50 Years of New China*. Beijing: China Statistics Press.

National Bureau of Statistics of China (NBS). 1999b. F*ifty Years of Cities in New China*. Beijing: China Statistics Press.

National Bureau of Statistics of China (NBS). Various Years. *China Statistical Yearbook* (CSY). Beijing: China Statistics Press.

National Bureau of Statistics of China (NBS). Various Years. *Urban Statistical Yearbook of China* (USY). Beijing: China Statistics Press.

Peng, Yusheng. 1999. “Agricultural and Nonagricultural Growth and Intercounty Inequality in China, 1985 - 1991”. In: *Modern China*. 25(3): 235 - 263.

Scharping, Thomas. 2001. “Hide-and-seek: China’s Elusive Population Data”. *China Economic Review*. 12: 323 - 332.

Sen, Amartya. 1997. *On Economic Inequality*. Expanded edition with a substantial annexe by James Foster and Amartya Sen. Oxford: Clarendon Press.

Shorrocks, Anthony F. 1980. “The Class of Additively Decomposable Inequality Measures”. In: *Econometrica*. 48(3): 613 - 626.

Song, Shunfeng; Chu, George S.-F.; Cao, Rongqing. 2000. “Intercity Regional Disparity in China”. *China Economic Review*. 11: 246 - 261.

Tsui, Kai-yuen. 1998a. “Trends and Inequalities of Rural Welfare in China: Evidence from Rural Households in Guangdong and Sichuan”. In: *Journal of Comparative Economics*. 26(4): 783 - 804.

Tsui, Kai-yuen. 1998b. “Factor Decomposition of Chinese Rural Income Inequality: New Methodology, Empirical Findings, and Policy Implications”. In: *Journal of Comparative Economics*. 26(3): 502 - 528.

Tsui, Kai-yuen. 1993. “Decomposition of China’s Regional Inequality”. In: *Journal of Comparative Economics*.

17(3): 600 - 627.

Wei, Shang-Jin. 1993. "The Open Door Policy and China's Rapid Growth". In: Ito, Takatoshi and Krueger, Anne O. (eds.). 1995. *Growth Theories in the Light of the East Asian Experience*: 73 - 97. Chicago: University of Chicago Press.

Wei, Shang-Jin; Wu, Yi. 2001. *Globalization and inequality: evidence from within China*. NBER Working Paper No. 8611.

World Bank. 1997. *Sharing Rising Incomes: Disparities in China*, Washington, D.C.: World Bank.

Xu, Kuan. 2004. *How has the Literature on Gini's Index evolved in the Past 80 Years?*. Department of Economics at Dalhousie University Working Paper.