Spatial Price Differences – Implications for Poverty in China

Åshild Auglænd Johnsen

Master Thesis for the Master of Philosophy in Economics degree

University of Oslo

August 31th, 2009

Preface

This thesis marks the end of my studies, and it has been an exciting process from which I have learned a lot. In a sense it summarizes the various steps I have taken throughout my years at the university – from an initial interest in development studies to a broader economic perspective, followed by a fascination with China that led to Mandarin studies, next the return to economics and finally, spatial price differences, poverty and well-being. I owe gratitude to many.

I wish to thank my supervisor Ingvild Almås. Her insightful feedback and enthusiasm towards the thesis was an inspirational support throughout the process. She has offered me a lot of her time, shared thoughts and ideas, and reminded me of the importance of asking questions. For this I am grateful.

I would like to thank Arne Melchior for introducing me to the exciting world of international trade, economic geography, PPPs and much more. The same goes for the rest of economics department at Nupi, providing me with food for thought by the coffee machine. I also want to thank Nupi for offering me a scholarship, unlimited access to coffee, a nice desk and good people for interesting lunch breaks with daily updates on international relations. Gratitude goes to Marc Lanteigne for interesting discussions on China and helpful comments on my draft, to Kam Wing Chan for clearing up issues on Chinese population statistics, and to the National Bureau of Statistics China for providing additional information on various subjects. I also would like to thank ESOP for my time there as a research assistant, the professors that I assisted for introducing me to the world of academics, and the other research assistants for an enjoyable life at the office.

My thanks go to Anders Nordraak Aasheim for spellchecking and for putting up with me (and all home chores), as my focus has been elsewhere. Finally, I want to thank my family for supporting me through a special time for all of us.

Of course, all flaws and mistakes are mine, and mine alone.

Summary

Since reforms were initiated in 1978, the economic development of China has been tremendous. The World Bank reports that the poverty reduction has been significant in this period (The World Bank 2009). The poverty measures are however subject to debate and uncertainty (Chen and Ravallion 2008; The World Bank 2009). Correcting for the cost of living is essential to poverty measurement, making prices a central part of the poverty reduction discussion (Gong & Meng, 2008). Considering the geography and large population of China, regional price levels are likely to differ. It has been suggested in the literature that urban prices are systematically higher than those in rural areas (Brandt and Holz 2006). Alwyn Young (2000) documents that provincial protectionism evolved throughout the reform process, which contributed to increasing price level differences. It is thus reasonable to expect that failing to adjust for regional price levels has a significant impact on poverty estimates. More specifically, as we expect prices to be relatively lower in rural areas, this has the potential to lead to an overestimation of rural poverty relative to urban poverty. However, identifying price levels that allow for comparisons across provinces, i.e. Spatial Price Indexes (SPI) is challenging methodologically as well as empirically, as lack of sufficient data is often an issue.

In this thesis we make a first attempt to identify Chinese SPIs by applying a simple, but empirically robust, economic theory – namely Engel's law – on Chinese household data. Incomes are then adjusted using the new price estimates given by the SPI, providing new estimates of real income. National poverty lines using the 1 dollar a day definition are constructed, and poverty estimates based on nominal and real incomes are then compared in order to investigate the effect of adjusting for spatial price level differences.

We estimate Engel curves for food based on Chinese household data on consumption expenditures. The resulting price level estimates are used to derive the SPIs. This method is similar to that of Hamilton (2001). Hamilton uses Engel's law to estimate bias in the consumer price index. Hamilton suggests that the Engel curve approach could be extended and used in the estimation of movements in the cost of living. By acknowledging the analogy between the SPI and the CPI we are able to deal with the problems related to the construction of the SPI directly– by applying the method proposed by Hamilton to estimate spatial price levels for Chinese provinces.

This allows us to investigate whether provinces have different price levels, and furthermore whether the price levels differ according to whether a household is located in the urban or rural part of the

provinces. Engel's law provides the theoretical background, and the method is based on the same principles as Hamilton's method. Consequently, the idea is that if two identical households located in different provinces have the same budget share for food but different nominal income; this reveals a price level difference. Adjusting incomes using these price level estimates, we get new estimates of real income, which allows us to investigate how adjusting for spatial price level differences affect poverty estimates.

Two findings for relative prices are now presented. First, we find that there are large differences between rural and urban price levels in both 1995 and 2002. Second, the SPI estimates show less national price dispersion in 2002 compared to 1995. Subsequently, adjusting incomes for spatial price differences has a large effect on poverty estimates, from which six poverty findings appear. First, national poverty rates are significantly lower when adjusting for spatial price differences. Second, nominal incomes on the one hand indicate that there was a reduction in poverty from 1995 to 2002 – real incomes, on the other hand, indicate the opposite. Third, the urban poverty incidence is higher when correcting for local price levels. Fourth, nominal incomes indicate that urban poverty is increasing in the period from 1995 to 2002, while urban poverty is reduced in this period according to real income measures. Fifth, adjusting for different price levels greatly reduces rural poverty incidence, nevertheless poverty is still a rural issue. Sixth, the urban/ rural poverty gap is reduced throughout 1995-2002 in nominal terms, but this pattern is not evident when examining poverty estimates based on real incomes.

The chosen approach in this thesis has two clear advantages. For one, even in cases where regional price data actually exists, the construction of a SPI is a time-consuming and a complex procedure (see Brandt and Holz (2006) for a thorough explanation). The Engel curve approach however, is much more straightforward and less tedious approach. Second and perhaps the most important argument, the strength of the Engel curve approach is that the cost of living is inferred directly from consumer behaviour (Hamilton 2001). Moreover, the strength of this analysis is the inclusion of a large number of *rural* as well as *urban* households covering several provinces in all of China's regions. When considering poverty rates, inclusion of rural areas is of utmost importance. This allows for the investigation of the relative price levels of the less advanced economic regions compared to urban areas.

The statistical software used in this thesis is Intercooled Stata 9.0. The household data are provided by the Inter-university Consortium for Political and Social Research (ICPSR), and the results are derived from data from the two following studies:

ICPSR Study 3012[1]

Title: Chinese Household Income Project, 1995

Principal Investigator(s): Riskin, Carl, Zhao Renwei, and Li Shi

ICPSR Study 21741

Title: Chinese Household Income Project, 2002

Principal Investigator(s): Shi, Li

[1] Riskin, Carl, Zhao Renwei, and Li Shi. CHINESE HOUSEHOLD INCOME PROJECT, 1995 [Computer file]. ICPSR version. Amherst, MA: University of Massachusetts, Political Economy Research Institute [producer], 2000. Ann Arbor, MI:Inter-university Consortium for Political and Social Research [distributor], 2000.

Contents

List of tables and figures

1 Introduction

Since reforms were initiated in 1978, the economic development of China has been tremendous. The World Bank reports that the poverty reduction has been significant in this period (The World Bank 2009). The poverty measures are however subject to debate and uncertainty (Chen and Ravallion 2008; The World Bank 2009). Correcting for the cost of living is essential to poverty measurement, making prices a central part of the poverty reduction discussion (Gong & Meng, 2008). Considering the geography and large population of China, regional price levels are likely to differ. It has been suggested in the literature that urban prices are systematically higher than those in rural areas (Brandt and Holz 2006). Alwyn Young (2000) documents that provincial protectionism evolved throughout the reform process, which contributed to increasing price level differences. It is thus reasonable to expect that failing to adjust for regional price levels has a significant impact on poverty estimates. More specifically, as we expect prices to be relatively lower in rural areas, this has the potential to lead to an overestimation of rural poverty relative to urban poverty. However, identifying price levels that allow for comparisons across provinces, i.e. Spatial Price Indexes (SPI) is challenging methodologically as well as empirically, as a lack of sufficient data is often an issue.

In this thesis we make a first attempt to identify Chinese SPIs by applying a simple, but empirically robust, economic theory – namely Engel's law - on household data. Incomes are then adjusted using the new price estimates given by the SPI, providing new estimates of real income. National poverty lines using the 1 dollar a day definition are constructed, and poverty estimates based on nominal and real incomes are then compared in order to investigate the effect of adjusting for spatial price level differences.

First we provide two findings for relative prices. We find that there are large differences between rural and urban price levels in both 1995 and 2002. Second, the SPI estimates show less national price dispersion in 2002 compared to 1995. Subsequently, adjusting incomes for spatial price differences has a large effect on poverty estimates, from which six poverty findings appear. First, national poverty rates are significantly lower when adjusting for spatial price differences. Second, nominal incomes on the one hand indicate that there was a reduction in poverty from 1995 to 2002– real incomes on the other hand indicates the opposite. Third, the urban poverty incidence is higher when correcting for local price levels. Fourth, nominal incomes indicate that urban poverty is increasing in the period from 1995 to 2002, while urban poverty is reduced in this period according to real income measures. Fifth, adjusting for different price levels greatly reduces rural poverty incidence, nevertheless poverty is still a rural issue. Sixth, the urban/ rural poverty gap is reduced throughout 1995-2002 in nominal terms, but this pattern is not evident when examining poverty estimates based on real incomes.

The most commonly used measure of prices in the economic literature is the consumer price index (CPI). More precisely, the CPI is constructed with the intention to measure *changes* in the cost of living. Hence, the CPI measures only inter-temporal changes in prices across provinces; it does not provide spatial price *level* differences for provinces (Brandt and Holz 2006). We need a Spatial Price Index (SPI) in order to provide price level differences across provinces, and this thesis measures and uses such indices. In this sense, the SPI is analogous to the CPI. The construction of both the CPI and SPI rely on detailed and extensive price data. Furthermore, the procedure usually means compromising between data availability and the consistency with consumer preferences, leading to well-known problems such as the quality, substitution, outlet and weighting biases (Moulton 1996; Costa 2001; Hamilton 2001; Brandt and Holz 2006; Almås 2007)[1]. The construction of a SPI requires highly detailed price data at a regional level, which is usually not available. Without such data, income can not be adjusted for systematic differences in spatial price levels (Brandt & Holz 2006).

The problems related to the construction of the CPI are dealt with by Hamilton (2001). Hamilton (2001) uses Engel's law to estimate bias in the consumer price index. Engel's law states that a household's budget share for food is inversely related to household real income (ibid.). This theory implies that there is a unique relationship between the budget share for food and total expenditures[2]. Hamilton's main idea is to see the potential in applying Engel's law to measure the cost of living. If two households with identical characteristics, observed in different periods, have the same budget share for food, they should also have the same real income. As real incomes are produced by deflating nominal income by the CPI, a difference in their measured real incomes reveals a CPI bias.

Hamilton suggests that the Engel curve approach could be extended and used in the estimation of movements in the cost of living (ibid.). By acknowledging the analogy between the SPI and the CPI we

[1] More specifically, the CPI is a measure of the change in costs of obtaining a fixed basket of goods. The fixed basket is based on a selection of goods, and the CPI follows the price changes of these. The fixed basket approach do not take account of consumer's behaviour in the sense that consumers tend to substitute their consumption away from items becoming relatively more expensive towards goods that have become relatively less expensive. This is called the substitution bias. Ideally, all items included in the fixed basket should be identical across regions. In reality the quality of goods vary across space and time, and this generates the quality bias. The outlet bias refers to rapidly expanding outlet stores being underrepresented in the sample of outlets from which the CPI prices are collected. Finally, the weighting bias is due to expenditure weights in the index being outdated/ incorrect.

[2] Ernst Engel observed that poorer households spent a larger share of total expenditures on food than richer households did (Deaton and Muellbauer 1980a). The same applied to large households relative to smaller households for the same level of total expenditures. Based on this, he suggested that household budget share for food could be used as an indicator of indirect welfare (Deaton and Muellbauer 1980a).

are able to deal with the problems related to the construction of the SPI and CPI directly – by applying the method proposed by Hamilton to estimate spatial price levels for Chinese provinces. This allows us to investigate whether provinces have different price levels, and furthermore whether the price levels differ according to whether a household is located in the urban or rural part of the provinces. Engel's law provides the theoretical background, and the method is based on the same principles as Hamilton's method. Consequently, the idea is that if two identical households located in different provinces have the same budget share for food but different nominal income; this reveals a price level difference.

The chosen approach in this thesis has two clear advantages. For one, even in cases where regional price data actually exists, the construction of a SPI is a time-consuming and a complex procedure (see Brandt and Holz (2006) for a thorough explanation). The Engel curve approach however, is much more straightforward and less tedious approach. Second, and perhaps the most important argument, the strength of the Engel curve approach is that the cost of living is inferred directly from consumer behaviour (Hamilton 2001).

In the literature, other methods to identify SPIs have been proposed. First, nominal values could be used as an approximation to real income, thus ignoring spatial difference s[1]. This approach contradicts the basic premise on which this thesis is based, namely that prices matter. As we expect that there will be considerable spatial price differences, in particular with regards to rural/ urban price levels, this approach is far from ideal for a country such as China. Second, we could assume that prices were the same in all regions in a specific base year and then use the regional CPIs to lead us from this base year to comparable cross-regional price levels for the year that we study. Brandt and Holz (2006) follow the second solution and construct spatial deflators based on this method for 1990[2]. It is possible to argue that this method is attractive in the case of transitional economies with former centralized pricing systems, such as China. However, this method has to two disadvantages. First, prices can differ in the base year. Second, the method relies on the CPI, which is a biased measure of price changes. Gluschenko (2006) compares such a CPI proxied price level with a SPI constructed for Russian regions and he concludes that this method fails to provide precise estimates

[1] Gluschenko (2006) defines real income strictly as spatial price adjusted income. Examplifying this, he refers to a study on regions in Japan, USA and Europe by Barro and Sala-i-Martin (2003) , where incomes by region are deflated using national values of the consumer price index. See Gluschenko (2006), page 7 and Barro and Sala-i-Martin (2003).

[2] At that point in time, some prices were still administered centrally and thus uniformly defined across the country (Brandt and Holz 2006). Based on these price data available on the province level, they construct comparable provincial price levels which are adjusted for all other years using regional CPIs.

of cross-regional price variation[1]. Thus, neither of these proposed methods proves themselves to be ideal for identifying regional price levels.

Gong and Meng (2008) apply the Hamilton method to identify SPIs for households in the urban parts of different provinces for the period 1986-2001. The approach in this thesis is similar to theirs, but with one major difference. The strength of this analysis is the inclusion of a large number of *rural* as well as *urban* households covering several provinces in all of China's regions, whereas they cover only urban households. When considering poverty rates, inclusion of rural areas is of utmost importance. This allows for the investigation of the relative price levels of the less advanced economic regions compared to urban areas.

This thesis is organized as follows. Chapter 2 gives a brief introduction to Chinese economic development in the past thirty years, and discuss why we would expect regional price differences to be significant. Chapter 3 presents the method used to derive the SPIs, the household data and construction of the variables. Furthermore, two sets of poverty lines are derived, and these are compared to those of other poverty studies on China. In chapter 4 the results and analysis of the main specification are presented. Tables on regression results, the derived SPI and the poverty estimates are presented. In chapter 5 we include several robustness checks. Conclusions are given in chapter 6.

[1] Furthermore, Gong and Meng (2008) point out three additional weaknesses to the method of Brandt and Holz. First, the price data used were collected for the use of CPI, which involves issues of quality adjustment of goods across provinces. Second, prices were collected from the capital cities only. Third, manufacturing wages were used as proxies for prices of non-tradable goods (ibid.).

2 Background

Before we turn to the method and findings of this thesis, we briefly discuss historical reasons for why we would expect spatial price level differences in China. Finally, two forces inherent in the liberalization process are emphasized as relevant for what results to expect.

There is significant price variation across provinces in China, and possible explanations for this can be linked to geographical matters and elements of the past thirty years of economic reform (Gong and Meng 2008; Brandt and Holz 2006). There are large differences between provinces, and natural resources are distributed differently across the country, leaving the local population different possibilities for economic development (Gong & Meng 2008). Prior to economic reform, China could be characterized as an egalitarian country (Riskin, Zhao, and Li 2001). Riskin argues however, that there existed gaps between rural and urban areas already then, and an important contributing factor to this was special policies concerning prices for agricultural products. In order to ensure accumulation of funds for industrialization, purchasing prices for these products were kept artificially low. Migration to the cities was not possible due to a strict residence permit system (ibid.). The Hukou household registration system required people to live and work only where their residency permits allowed them to. Reforms of this very restrictive migration control system started in the beginning of 2001 (Liu 2005), though the system continues to restrict labour market integration nationwide, leading to potentially segmented markets (Brandt and Holz 2006).

Reforms were concentraded on traditional institutions such as collective farming, state owned enterprises, the central planning system, the labour employment and remuneration system (Liu 2005). A defining element to this was what is known as the dual-track approach (Qian 2002). Until 1984, prices were centrally administered. Gradual price reforms were initiated by allowing for the co-existence of centrally determined and market determined prices (Marangos 2006). Price setting was finally decentralized by 1992, and 95 percent of the prices on consumer goods and 80 percent of investment goods were fully determined by the market by 1999. Regional prices diverged significantly throughout the whole period of reform because of the different regional industrial structures (Gong & Meng, 2008).

Concerning the history of economic reforms, location has also been a crucial factor. As reforms started in 1978, state investments were concentrated in the eastern regions (Riskin, Zhao, and Li 2001). Special economic zones were set up, receiving preferential treatment by the state while opening up to international trade (ibid.). According to Riskin et al. (2001) this lead to major differences between the eastern economy on the one side and the western and central regions on the other, concerning industrial development, infrastructure, levels of education and the development of a commodity market. Restrictions on factor mobility were reduced after reforms, which could be expected to lead to price convergence (Gong and Meng 2008). As the economic zones opened up to international trade, it attracted foreign direct investment. This further increased the capital/labour ratio of these zones, leading to even larger differences between the coast and the rest of the country (ibid.). Based on this, Gong and Meng suggest that regional prices may have converged more slowly than expected. Another aspect of regional differences in economic development concerns proximity to markets. Distance to major markets can affect regional price differences (ibid.). Regional protectionism is a recognized issue in China, as investigated by Young (2000). According to Young, incremental economic reforms led to new distortions in economy, such as interregional barriers to trade. Other adverse effects could involve local monopolies and bureaucratic difficulties (Gong & Meng, 2008).

As the economic gap between the eastern and the western parts of China increased throughout the 1980s and 1990s, a development plan known as the "Go West" plan was initiated in 1997. Inland development was set forth as an important priority for 1996 to 2010 (Chen and Lin 2004). This strategy is similar to the urbanization strategy implemented in the coastal areas, with the aim of promoting rural-urban convergence (The World Bank 2008). As the data covers only the years 1995 and 2002, this is one of many elements that might influence the estimated spatial price levels – which will be discussed in chapter 4.

As all prices had been officially decentralized by 1992 in a process that was started twelve years earlier, the following period constitutes the next stage in a gradualist approach to become a market economy. In this regard, two features of the liberalization process could be of particular relevance to the SPIs: one of national character concerning price decentralization, the other concerning exposure to international trade. As internal market integration could be expected to be positively linked to price convergence, increased integration could lead to spatial price level convergence, everything else given. Exposure to international trade being concentrated in specific areas is on the other hand an argument in favour of price divergence, in particular for urban and rural areas. Young (2000) found in his empirical study that regional prices in China diverged throughout the period 1990-1999, with increasingly fragmented markets. Fan and Wei (2006) challenge this by finding empirical support for convergence to the law of one price[1] for prices in different regions in China throughout 1993-

[1] The law of one price implies that an identical good sold in two separate competitive markets free of transportation costs and other barriers to trade should sell for the same price when converted to the same currency (Krugman and Obstfeld 2003).

2003. They argue that this implies increased regional market integration in China and a successful transition to a market economy[1]. As both studies rely on data from 36 cities, it is interesting to compare these strikingly opposing results to the urban estimated SPI. Our findings indicate that urban price variation is larger in 2002 compared to 1995, which is in support of urban price divergence, as argued by Young (2000). The national price variation however, which is the "correct" variable to inspect, indicates that there was an overall reduction in price variation for China. The main point here is that we have two opposing effects concerning spatial price levels. The estimated SPIs will be analysed in conjunction with these two features in chapter 4.

[1]As they also find an upward trend in the overall price dispersion analogous to Young, they link this to an aggregation bias. Splitting the product categories in sub-samples , they argue that the price dispersion estimates show a downwards trend (Fan and Wei 2006).

3 Methodology

In this chapter, we present the econometric specification and how to derive the SPIs. The household survey data is presented along with a description of the construction of the included variables in the analysis. After settling on the definition of income, the next step is to do the same for poverty. In chapter 3.3.1 the concept of poverty lines will be introduced. Purchasing Power Parity (PPP) exchange rates are of importance in this context, and we construct two poverty lines based on two different PPP exchange rates. Finally, we briefly discuss adjusting incomes for economies of scale, and the chosen equivalence scale is presented.

3.1 Econometric specification

Following the approach of Hamilton (2001), cross-provincial Engel curves for food for the years 1995 and 2002 are estimated by using the Almost Ideal Demand Structure (AIDS) proposed by Deaton and Muellbauer (1980b). Household data for several provinces and municipalities in China for 1995 and 2002 are used to estimate the relationship between the budget share for food and household income. Based on the assumptions that the demand function is correctly specified, that consumer's preferences are stable throughout the period, and that the variables contain no systematic errors, a set of urban and rural dummy variables reveal a set of price levels. Based on the dummy coefficient estimates, the SPIs are constructed.

Why food

According to Hamilton (2001) and Costa (2001), food is an ideal indicator good for measuring real income for the following reasons. First, the indicator good should be sensitive to variation in income, which is the case for food as the income elasticity of food is substantially different from unity. Second, food can be characterized as a nondurable good. Expenditures and consumption of food in one period are nearly identical, as opposed to a durable good, which is bought in one period but consumed throughout several periods of time. Third, the definition of food is straightforward, as opposed to other goods such as leisure (Hamilton 2001). Finally, empirical studies show that AIDS is a suitable functional form for estimating demand for food (Banks, Blundell and Lewbel 1997; Leser 1963). The AIDS system is given by:

(1)
$$m_{h,p,u} = a + b(\ln y_{h,p,u} - \ln P_{p,u}) + \gamma(\, lnP_{f,p,u} - \, lnP_{n,p,u}) + \theta X_{h,p,u} + \varepsilon_{h,p,u}$$

Where $m_{h,p.u}$ is the budget share for food for household *h* in province *p* in rural/ urban area *u*, $y_{h,p,u}$ is nominal income for household *h* in province *p* and rural/ urban area *u*. $P_{p,u}$ is the variable for the cost of living, and it is a composite price for consumption in province *p* and rural/ urban area *u*. Relative prices $P_{f,p,u}$ and $P_{n,p,u}$ are composite prices for food and non-food in province *p* and rural/ urban area *u*, respectively. $X_{h,p,u}$ is a vector of demographic control variables for household *h* in province p and rural/ urban area *u*. Finally, $\varepsilon_{h,p,u}$ is the residual.

Food and non-food prices are not observable, so relative prices are not included in the main estimation, which is the same as implicitly assuming that the budget share for food is not influenced by relative prices. This is not ideal, but several studies using this method have shown that the estimated prices are very robust to the variation in relative prices (see e.g., Almås (2007), Costa (2001) and Hamilton (2001)). The AIDS system then reduces to:

(2) $$m_{h,p,u} = a + b\left(\ln y_{h,p,u} - \ln P_{p,u}\right) + \theta X_{h,p,u} + \varepsilon_{h,p,u}$$

The budget share for food in (2) is specified as a function of real income and the control variables. The survey provides the information needed in order to construct the variables for budget share for food, income and the demographic control variables. However, the overall price level in (2) is not observable. The identification strategy is the following: $P_{p,u}$ is the only variable that is specific for each province *p* and area *u*, and hence by including dummy variables corresponding to these areas *p* and *u*, $d_{p,u}$, we thus have a method to identify the local price level differences. The AIDS specification given by (2) can then be estimated by:

(3) $$m_{h,p,u} = a + b\left(\ln y_{h,p,u}\right) + \theta X_{h,p,u} + \sum_{p=1}^{N} d_{p,u} D_{p,u} + \varepsilon_{h,p,u}$$

The budget share for food in (3) is specified as a function of nominal income, the control variables and dummy variables for areas, $d_{p,u}$. The data set for 1995 includes information on rural and urban households for 11 provinces. For additional 8 provinces we only have data on rural households. Thus, we have 8 plus 11 rural dummy variables and 11 urban dummy variables. Therefore, the estimated equation includes 30 dummy variables of which 19 represent rural provinces and 11 urban provinces in 1995. Analogously, 34 dummy variables (22 rural, 12 urban) are estimated for 2002. Simplifying the terms used for geographical locations slightly, we will refer to these dummy variables as *provincial* from here on. When necessary, we distinguish between rural and urban areas by referring to them as *rural/ urban provinces.* Beijing (urban) is the base province, and is therefore left out of the estimation function. Hence, the dummy coefficient for urban Beijing is then by definition set equal to 0. Using (2) and (3) we are able to identify the price for province *p* and area *u* relative to urban Beijing, given by the following[1]:

(4) $$d_{p,u} = -b \ln P_{p,u} \quad \Leftrightarrow \quad P_{p,u} = e^{-\frac{d_{p,u}}{b}}$$

A positive dummy variable for province *p* in urban/ rural area *u* implies that the budget share for food for this specific province is relatively higher than that of Beijing (urban) for a given level of income. The budget share for food is decreasing in income; the coefficient for nominal income, *b*, is negative. Hence, if the provincial dummy is positive, the price level exceeds unity, implying that the price level of this province exceeds that of urban Beijing. A negative estimate yields a price estimate lower than that of urban Beijing. Based on the price estimate, a spatial price index (SPI) is calculated. China as a whole is taken as the reference location, given by the average price level. The average price level is a product of the number of households in a province and the estimated relative price level of that province divided by the total number of households in the *P* provinces[2], where *U*=1 if we only have rural households for this province, and *U*=2 if we have both rural and urban:

(5) $$\bar{P} = \frac{1}{\sum_{u=1}^{U}\sum_{p=1}^{P} Households_{p,u}} \sum_{u=1}^{U}\sum_{p=1}^{P} Households_{p,u} P_{p,u}$$

Thus, the SPI for each province relative to the national average in the year *t* is given by the estimated price level over the average price level:

(6) $$P_{p,u}^{rel} = \frac{P_{P,u}}{\bar{P}}$$

Real income is then given by nominal income divided by the SPI:

(7) $$y_{p,u}^{real} = \frac{y_{p,u}^{nominal}}{P_{p,u}^{rel}}$$

[1] Gong and Meng derives the same expression, but somehow they find that the relative price level can be expressed by $e^{\frac{d_j}{b}}$, which has the opposite sign (2008).
[2] This calculation relies on the assumption that the survey is representitative at the national, as well as the provincial and also on an urban/ rural level.

3.2 Household survey data

Household data used in the estimation are provided by the “Chinese Household Income Project”, collected in 1995 and 2002. Both data sets are separated into a rural and an urban part, based on information provided by rural and urban households. These households were selected from larger samples consisting of approximately 65 000 rural households and 35 000 urban households in 1995 drawn by the State Statistical Bureau[1].

Two sets of household survey questionnaires accompany the data sets. Number of individuals and households for 1995 and 2002 are presented in . The household composition is, as already discussed, restricted to two adults and one child in the main analysis. As we can see from table 1, the average household size for rural households is larger than the urban average for both years.

Table 1 : Comparison of the surveys	**1995**			**2002**		
	Individuals	**Households**	**Average household size**	**Individuals**	**Households**	**Average household size**
Rural	34 739	7 998	4.35	37 969	9 200	4.14
Urban	21 687	6 931	3.13	20 632	6 835	3.02
Total	56 426	14 929	3.79	58 601	16 035	3.66

The total number of provinces, autonomous regions (AR) and Direct-control municipality (DM) in China is 30 in 1995 and 31 in 2002 (excluding Taiwan), and data covers all provinces but four. The missing provinces, autonomous regions and municipalities are listed in table 2. Chongqing, the largest of the four municipalities, was not established until 1997. Prior to that it was a part of Sichuan, and hence included in the 1995 sample, since Sichuan is included in both years. In 2002 Chongqing is included as a separate entity. In 1997 Hong Kong came under China’s rule, while Macau was handed over in 1999. Neither of these are included in the samples of 1995 and 2002. Large economic centres such as Shanghai and Tianjin are not included, and more remote areas such as Tibet and Inner Mongolia are excluded as well.

Ideally, the data set would contain information on all provinces, municipalities and autonomous regions in China at that particular time. But based on own calculations, we can state that population numbers for regions in 1995 (NBS 1996) show that the included provinces cover roughly 83 percent

[1] The data set for 2002 does not provide information on this. Rural and urban household surveys in China were started in 1956 in order to keep track of the national economic deveolopment, and this number reflects the sample selected by the National Bureau of Statistics (Cramer, Fang, and Wailes 1998).

of the national population (excluding Taiwan). For 2002, the number is 88 percent (excluding Taiwan, Macao and Hong Kong) (NSB 2003).

Table 2: Regions	Included regions		Missing regions			
	1995	2002	1995		2002	
			Provinces	AR/ DM	Provinces	AR/ DM
			Fujian	Guangxi	Fujian	-
Rural provinces	19	22	Hainan	Inner Mongolia	Hainan	Inner Mongolia
			Heilongjiang	Ningxia	Heilongjiang	Ningxia
Urban provinces	11	12	Qinghai	Shanghai (DM)	Qinghai	Shanghai (DM)
				Tianjin (DM)		Tianjin (DM)
				Tibet		Tibet
Both rural and urban area covered	11	12		Xinjiang		-
			Special administered regions/ other			
			Hong Kong, Macao, Taiwan			

The main specification includes the control variables elders, age and gender of head of household. Furthermore, it is restricted to households of two adults and one child, in order to make these households comparable. Hamilton (2001) argues that if the cost of living varies across groups, there is no point in including all observations in order to obtain a random sample of the national population. Rather, data quality and comparability should be the selection criteria when using Engel curves to estimate the cost of living. Restricting household composition, the sample is reduced from the initial 14578 to 2935 for 1995 and from 15708 to 3548 in 2002. By restricting the analysis, we run the risk of a selection bias. If poorer households tend to have more members, we base our analysis on the relatively richer rural households. As having more than one child in China is illegal and punishable by fines that are significant for poor households, underreporting of number of children is a potential issue. These fines are small for more wealthy households, thus restricting household composition can lead to a possible exclusion of larger urban households that are relatively well off (Branigan 2008).There is a trade-off between inaccuracies due to possible aggregation bias and a selection problem. In chapter 5 we include all households as a robustness check. As will be shown, the results from the robustness check support the findings from the main specification.

3.2.1 Defining the variables

Income

In the following chapter a detailed explanation on how the variables are constructed is provided. There is one reoccurring issue, namely the trade-off between ensuring comparability of the variables and inclusion of all potentially relevant information when constructing the variables. The questionnaires differ for urban and rural households, and between 1995 and 2002. They do not always include the same variables and in some cases the variables are differently defined. Based on this, the variables have been constructed with these things in mind, namely to include as much relevant information as the household data allow for, given that comparability of the variables is maintained. This is especially a problem for the value of self-production for rural households, and in-kind income for urban households, but unfortunately an unavoidable issue.

Reported income versus expenditures

Income is measured by household consumption expenditures. Deaton argues that measuring income by consumption is the most appropriate for developing countries where household expenditure surveys are available (Deaton and Zaidi 2002). First, consumption is a more satisfactory measure of well-being. Second, income can be erratic, especially in agricultural societies. Self-employment can involve several sources of income, which can lead to large variations in annual income. Expenditures however, capture consumption-smoothing. Third, consumption data can be cheaper to collect relative to income data in developing countries compared to more formalized industrial economies, and on occasion consumption data can be the only available information. Consumption is smoother over the period of a year, and more reliable in the sense that it reflects actual behaviour. Fourth, there are no obvious reasons to underreport consumption expenditures as compared to income. With income data, the survey reporters might underreport income if they e.g. suspect that these data could become available to the tax authorities (Deaton and Zaidi 2002). Thus, measuring income in terms of consumption has clear advantages in the case of China. From here on, we refer to expenditures as income.

Rural income

The rural household questionnaire provides several options for constructing the income variable. Two income aggregates are reported, namely *Total Household Expenditures* and *Total Cash Expenditures on Consumption*. While the former seems to incorporate expenditures on purchasing fixed capital for production and expenditures on taxes and fees, the latter reflects expenditures on various consumption goods. According to the questionnaire, total cash expenditures can be decomposed into all cash expenditures on food and non-food industry products, clothing, transport and communications, daily consumption goods, durable goods, medical care, educational costs, housing, expenditures related to support of family, fines, insurance and gifts.

A third approach to defining the rural income variable involves construction of the income aggregate by simply summing up the value of the various consumption expenditures, leaving out capital investments and taxes. This is the only option that ensures comparability over the years and across urban and rural households. This income aggregate does not match any of the readily defined income aggregates. This mismatch could be due to household reporting in-kind income or expenditures related to production of food for self-consumption in the measure of total consumption. But since we do not know how this question has been interpreted by participating households, we can not use it as our income variable. As already mentioned, there is a trade-off between comparability and inclusion of information. We believe that the third approach is the best option in this situation, and hence we apply this.

Urban income

Defining urban income poses similar challenges concerning inclusion of information versus comparable variables. The same approach is selected, i.e. the income aggregate is constructed based on the components of consumption expenditures, similar to those in the rural section. The questionnaire differs slightly between 1995 and 2002 concerning the definition of durable goods. A robustness check (not included) show that the exclusion / inclusion of these do not change the results; hence durable goods are included in the main estimation.

The urban household expenditures also includes expenditures on electricity, telephone, child care and labour and other services in addition to what is included in the rural questionnaire. Particularly, the rural and urban questionnaires differ with regards to definitions of public services. As the rural survey does not include these elements, we run the risk of underestimating rural expenditures. On the other hand, the surveys have been designed to match the living conditions facing the respondents. As these expenditures are of importance for urban citizens, they must be included in the income aggregate.

The 2002 urban questionnaire includes a section on household non-monetary income (in-kind and services) measured in Yuan, and a less detailed section is provided by the 1995 survey. In-kind income can be food and clothes received at the work unit or at school, and this represents consumption and should hence be included in a measure of total expenditures. Urban in-kind income for both years is decomposable into food and non-food so that this can be added to total expenditures and the budget share for food. This is not the case for rural in-kind income, which is restricted to in-kind income from the work unit. As no consistent way of including in-kind income into expenditures exist in this situation, these are not included in the main specification.

The budget share for food

For rural households, the food aggregate is composed by cash expenditure on food industry products, staple and non-staple food. In "Guidelines for Constructing Consumption Aggregates for Welfare Analysis", Deaton and Zaidi (2002) argues that including food consumption from all sources is of great importance, especially when examining poverty. Leaving out food consumption from other sources has the potential to weaken the results when it comes to poor rural households. But adding food from other sources is imbued with issues related to comparability and value measurement. For one, the 1995 questionnaire reveals information on food received from relief, from collective distribution, gifts, bought at subsidized prices for food grain such as corn, rice and "other". Relying on the household's evaluation of market prices for food grain, the value of this can be estimated. Prices are however left out in the 2002 questionnaire, thus food grain can not be included. Moreover, the rural surveys allow for construction of the value of food consumed from self-production. Here the complicating factors involve the prices estimated by households, as broad categories such as "meat" are deemed to produce rough approximations to the actual value. In-kind income was excluded from urban households' total expenditures, and the 2002 data includes limited information on self-production. To ensure comparability across years and for rural and urban households, the value of self-production is excluded.

The budget share for food is other food industry products, staple and non-staple food divided by income. Analogous, the budget share for food in the urban case is simply expenditures on other food industry products, staple and non-staple food as share of income. The questionnaire for 1995 does not separate between food consumed at home and other, so all food consumption is included in the constructed food aggregate.

Public goods

Consumption of public goods has been included in the income aggregate. However, Deaton (2002) recommends that a household's valuation of public goods should not be included. There are several reasons for this. Provision of public goods varies across space, and it is difficult to estimate the value of consumption for services as health and education.

Amount of medical services consumed by the rural households is composed by private, public and insurance-financed spending. For urban households, total consumption expenditures captures self-financed medical care only. Furthermore, the level of state provision of medical services varies between urban and rural areas. The issues concerning health service consumption is a possible matter when it comes to expenditure on education as well. In spite of the obvious complicating factors characterizing consumption of public goods, education and health service have been included in the income aggregate. The design of the questionnaire is chosen in order to match the actual expenditures of the recipients as close as possible. Consumption of these services make up an important part of expenditures, and leaving them out weakens the result possibly as much as excluding them.

Durable goods and housing

For rural households, purchase, maintenance and construction costs are included in expenditures on housing for both 1995 and 2002. In 2002, housing costs are defined as rent on leasehold/ owned housing. The urban questionnaire does not include explicit questions about housing expenditures in 1995, so the average rent for 1995 is used as an approximation to annual costs. The urban questionnaire includes questions on electricity, water and telephone expenditures, while there's only information on fuel in the rural part for both 1995 and 2002.

Control variables

The age and sex of head of household is included in the main regression. Head of household is restricted to individuals older than 15. In the rural dataset for 1995 all but 328 (352 in 2002) individuals are male head of households, while 2289 (2220 in 2002) out of the urban heads of household are female.

A variable denoting total members of household is constructed. Average number of members in a household included in the analysis is 3.1 (largest 8) for urban households and 4.3 for rural (largest 10) for 1995. The variable for number of adults was constructed by subtracting number of children from total members of household. Children are defined as being younger than 16. Elders are defined by the official retirement age in China, which is 60 for men, and 55 for women.

Education of head of household is included in a robustness check in chapter 5. Education is defined by seven categories in the urban section. The rural questionnaire also includes an additional question regarding illiteracy. I construct three dummy variables on the basis of these categories, defined as higher, middle and lower education for head of household, provided that he or she is older than fifteen. The questionnaire for 2002 included nine categories for level of education. These were divided in order to match the definition for 1995.

Other variables were also considered, such as temperature (capturing difference in tastes), but finding the corresponding temperature for each village and city was not feasible. Minority was considered. Besides the majority group Han, 56 ethnic minority groups are recognized in China, which represent approximately 10% of the total population. However, this variable is highly correlated with income, and was consequently left out. In addition to the gender of head of household, a variable for the ratio of female to household was also tested.

3.3 Measuring poverty

In order to measure poverty, five choices have to be made. First, we have to decide on which indicator of well-being to use. Second, we have to choose between absolute versus relative poverty lines. Third, we must settle on a poverty line of 1 or 2 dollars a day. Fourth, we need to select which PPP exchange rates to use. Finally, in order to adjust incomes for economies of scale, we must pick an equivalence scale. In the following we present the chosen measures and in brief the arguments in favour of doing so.

3.3.1 The poverty line

Poverty is multidimensional concept. *A Sourcebook for Poverty Reduction Strategies* defines poverty as not to have enough of some dimension of well-being (Klugman 2002). Here the chosen measure of household well-being is defined in monetary terms, given by household income. There are three arguments in favour of choosing income as an indicator of well-being. First, the availability of data is better. Second, we need to choose an index in order to measure well-being. Adding several dimensions creates an aggregation challenge. Third, we believe that income is correlated with - although not perfectly – with the opportunity for health, schooling and other potential dimensions of well-being. An example of a non-monetary poverty line is to have a certain level of education. Another is a specified minimum of calories that an individual must consume each day in order not to be defined as poor (ibid.).A monetary poverty line could be set at a certain expenditure level. A household, whose expenditures are below this line, would be then defined as poor (ibid.). The poverty lines developed in this chapter will be monetary.

Next, a poverty line can refer to absolute measures, as well as being defined relative to another measure. This is better illustrated by examples. An absolute measure could be the cost of a basket of food satisfying a defined standard of nutrition. Then the corresponding absolute poverty line would define an individual with a budget less than the cost of this basket to be poor (ibid.). Poverty estimates for Norway usually refer to relative poverty lines. The relative poverty line could be defined as 50% of the median income for the population as a whole for a year. Then an individual with an income less than 50% of the median income would then be defined as poor (Normann 2009). The relative poverty line is an inequality measure which focuses on the lower tail of the distribution, as the median influences the size of the poverty line. A higher median produces a higher poverty line. In the case of increased inequality, this leads to a higher number of poor – the poverty rate depends on the income distribution. The absolute line, on the other hand, is a defined standard of what we perceive as an extremely low income, and it is based on the individual's power to consume. This constitutes an advantage when we want to measure poverty.

Having decided that we are going to use the absolute poverty line, we now need to define this line. For international comparisons of poverty, the one-dollar-a-day measure is often used (Ravallion, Chen and Sangraula 2008). This poverty line is absolute in the sense that it remains fixed over time, only adjusted for inflation (Klugman 2002). The history of the one-dollar-a-day poverty line goes back to 1990, when the World Bank suggested that poverty measures should be based on the standards of the poorest countries (Ravallion, Chen and Sangraula 2008). The poverty lines used by the low-income countries themselves were adjusted for differences in purchasing power and compared. The lines fell in the range of $275 and $370 per year in PPP terms at 1985 prices. The latter, being fairly close to $365 dollars, gave rise to what is known as "the dollar-a-day-line". New Purchasing Power Parity (PPP) data including more developing countries became available in 2000, and the poverty line was revised on this basis. The new line was set equal to the median poverty line of the ten poorest countries, which was $1.08 a day measured at purchasing power parity based on 1993 prices (USAID 2008)

The poverty line is converted to Chinese currency (Yuan) using two different Purchasing Power Parity (PPP) exchange rates (Chen and Ravallion 2007). Several institutions provide PPP measures, such as the EU, OECD, the Penn World Table (PWT), the World Bank and the IMF, but the two main sources are the PWT produced by the University of Pennsylvania and the World Bank PPP estimates. The PWT and the World Bank use the same basic data, but their methods differ and in particular with regards to aggregation methods used when constructing the PPPs [1](World Bank 2006). In this analysis, we construct one poverty line based on the PWT and one based on the World Bank data. The poverty results provided by these lines differ significantly, and a comparison of the results will be given in chapter 4.

[1] For a brief introduction to the construction of and difference between these PPPs, see: http://siteresources.worldbank.org/ICPINT/Resources/About_the_ICP_and_PPPs_FAQ_11.2006.doc

We now need a specific poverty measure to evaluate the results of applying the poverty line on income. The chosen measure is the headcount index, which is a measure of the incidence of poverty for a specified group (Klugman 2002). The headcount index is the percentage of the sample population with income per person (or equivalence scale adjusted income) below the poverty line, which is set to be $1.08/day measured in 1993 prices. The headcount index is the ratio of this number to the population – the share of people being poor, according to these definitions (ibid.). In addition to this, measures exploring the depth and severity of poverty exist (ibid), but these will not be considered here.

Thus, the poverty line in this thesis is set to be equal to $1.08/ day measured in 1993 prices. The position of the poverty line is arbitrary in a sense; the main issue concerns the effect of adjusting for prices on the poverty estimates. We could also have based the analysis on the $ 2/ day poverty line, but we deliberately chose a more conservative measure. The following procedure for constructing the poverty line is similar to that of Fang et al (2002). In order to construct a poverty line measured in Chinese currency, information on the consumption purchasing power parity PPP exchange rate in 1993 for China is needed. The poverty line can then be converted to Yuan in 1993 prices by multiplying the PPP by 1.08. Then we can use the Chinese CPI to adjust the poverty line for price increases from 1993 to 1995. This is done using the Chinese Consumer Price Index for the years 1993 and 1995, both with base year 1978[1]. Finally, the poverty line should reflect annual consumption expenditures, which is found by multiplying by 365 days. The general formula for the poverty line is:

$$P^{1995} = 1.08 * \frac{PPP_{CHN}^{1993}}{PPP_{US}^{1993}} * \frac{CPI^{1995}}{CPI^{1993}} * 365 \tag{8}$$

The 2002 line analogously:

$$P^{2002} = 1.08 * \frac{PPP_{CHN}^{1993}}{PPP_{US}^{1993}} * \frac{CPI^{2002}}{CPI^{1993}} * 365 \tag{9}$$

[1]Fixed-base consumer price indices are provided online by the China Statistical Yearbook, table 8-2 (NBS 2008).

The Penn World Table line

The Penn World Table[1] provides purchasing power parity exchange rates for 188 countries for some or all of the years between 1950 and 2004 (Heston, Summers, and Aten 2006). The poverty line is finally calculated from (8) and (9), and the PPP ratio refers to purchasing power parity over GDP (unit: US=1 in general variables). Then, according to the Penn World Table, the nominal poverty line of $1.08/day equals 845.09 Yuan in 1995:

(10) $$P_{PWT}^{1995} = 1.08 * 1.48 * \frac{396.9}{273.1} * 365 = 845.09$$

If the same poverty line is adjusted using CPI for 2002 (base year 1978), the nominal poverty line equals 923.02 Yuan in 2002:

(11) $$P_{PWT}^{2002} = 1.08 * 1.48 * \frac{433.5}{273.1} * 365 = 923.02$$

The 2005 ICP line

The International Comparison Program (ICP) provided new regional estimates of purchasing power parities (PPPs) in 2008, which received a great deal of attention. The 2005 PPP estimates for China were significantly higher than the corresponding 1993 PPP (which was based on data not collected by the ICP), and led to a 40% reduction in the estimated size of China's economy measured in PPP terms (Chen and Ravallion 2008). This estimated poverty line is therefore from here on referred to as the ICP-line. The PPP conversion factor (US$=1) for China equals 3.45 in 2005 (ICP 2008). According to Chen and Ravallion (2008), the 2005 ICP gives the most complete assessment of how the cost of living varies across countries in the world. The implied 1993 PPP conversion rate of the 2005 PPP can be found by deflating the PPP conversion rate by inflation in China and the US[2], as done by Reddy and Minoiu (2006). First, finding the corresponding value of one dollar in 1993, $1\text{USD}^{05} * \frac{CPI_{US}^{93}}{CPI_{US}^{05}} = 0.74\ \text{USD}^{93}$, and then the value of one Yuan in 1993, $1\ \text{Yuan}^{05} * \frac{CPI_{CHN}^{93}}{CPI_{CHN}^{05}} = 0.59\ \text{Yuan}^{93}$ the 2005 PPP conversion rate can be found by the following formula:

(12) $$\frac{PPP_{CHN}^{1993}}{PPP_{US}^{1993}} = \frac{PPP_{CHN}^{2005}}{PPP_{US}^{2005}} * \frac{1\text{Yuan}^{05} * \frac{CPI_{CHN}^{93}}{CPI_{CHN}^{05}}}{1\text{USD}^{05} * \frac{CPI_{US}^{93}}{CPI_{US}^{05}}} = 3.45 * \frac{\frac{273.1}{464}}{\frac{144.5}{195.3}} \frac{\text{Yuan}^{93}}{\text{USD}^{93}}$$

$$= 2.74\ {}^{\text{Yuan}^{93}}/_{\text{USD}^{93}}$$

[1]Alan Heston, Robert Summers and Bettina Aten, Penn World Table Version 6.2, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania, September 2006

[2] Fixed-base consumer price indices for the US are provided online by the Consumer Price Index History Table (Bureu of Labor Statistics 2009).

In order to calculate the poverty line, the dollar must be converted to Chinese currency (Yuan) and then adjusted for inflation for the period between 1993 and 1995, as when deriving the PWT line:

(13) $$P_{WB}^{1995} = 1.08 * 2.74 * \frac{396.9}{273.1} * 365 = 1571.19\, Yuan$$

Adjusting the same poverty line using the CPI (base year 1978) for 2002, the nominal poverty line equals 1716.08 Yuan in 2002:

(14) $$P_{WB}^{2002} = 1.08 * 2.74 * \frac{433.5}{273.1} * 365 = 1716.08\, Yuan$$

Poverty lines for rural and urban China, in addition to the country as a whole are presented in appendix A. Comparing the PWT and ICP poverty lines to these, we see that our poverty lines are located between the lowest poverty lines for the rural provinces and the highest lines for the urban parts. Compared to other poverty lines, the PWT and the ICP line seem to correspond well with what is perceived to be a high and a low poverty threshold, respectively.

3.3.2 Equivalence scales

One potential way to measure poverty would be to compare per capita income to the poverty lines derived in chapter 3.3.1. But by simply dividing household spending on the number of household members, we implicitly assume that a three-person household with a 9000 Yuan budget is equally well off as a one person with a budget of 3000 Yuan (Klugman 2002). The demographic composition of a household and number of members affect the level of expenditures on goods and services[1]. In large households the members might benefit from sharing commodities and housing, and large purchases might result in lower expenditures per member due to lower unit costs relative to that of smaller households. In order to correct for these economies of scale in the households, we follow the proposals of Klugman (2002), and adjust expenditures using an equivalence scale. The idea behind equivalence scales is that the needs of a household do not grow proportionally with each new member of the family.

The OECD Social Policy Division defines three equivalence scales, presented in appendices B (OECD 2009). The OECD-scale implies some, but not extensive, presence of economics of scale. This scale

[1] Khan and Riskin (2001) and Fang et al (2002) among others, use only per capita expenditures when measuring poverty.

produces the best estimate compared to other studies on economies of scale in China; the main analysis will thus be based on the OECD-scale (OECD 2009):

$$OECD^{ES} = 1 + (adults - 1) * 0.7 + children * 0.5 \qquad (15)$$

The ICP poverty line is equal to 1716.08 Yuan a year in 2002, hence if annual equivalence scale adjusted income is less than this, an individual will be defined as poor.

3.3.3 Population data

Before we turn to the results of the estimation results, one issue remain. In order to calculate the actual number of poor individuals based on the estimated headcount rates, we need population data. Moreover, providing headcount poverty rates for China requires population data that separates between urban and rural areas on a province level. Unfortunately, such data are released only from 2005 and onwards. The calculations therefore have to rely on national population data, accompanied by the following caveats reported by Chan (2009). First, many Chinese cities include large agricultural populations. Second, due to migrants and the hukou issue, the actual urban population is larger than the registered population for most large cities[1]. Third, new definitions on urban areas in 2006 have lead to major revisions in population estimates for cities. These caveats should be kept in mind when evaluating headcount estimates based on urban and rural shares of the population resulting from this analysis.

[1] See appendice C for a brief introduction to internal migration in China.

4 Results

In the following, the results from the regression function are presented. After commenting on the coefficient estimates, the SPIs are derived and subsequently the resulting headcount poverty rates are presented.

4.1 Regression results

The main regression given in equation (3) is estimated for the years 1995 and 2002. As a robustness check, equation (3) is estimated using the whole sample in chapter 5. The results from equation (3) are presented in table 3, table 4 and table 5. Expenditures are, as predicted by Engel's law, negatively related to the budget share for food. For the 1995 survey data, an increase in expenditure of 1% reduces the budget share for food by 0.16. The predicted effect from the 2002 data is -0.17. Both coefficients of the logarithm of real income are statistically significant at 1%.

These results are similar to those found by Gong and Meng (2008)[1]. Three control variables are included. Age of head of household has a small and positive effect on the budget share for food, and the variable is statistically significant at a 1% level for both years. A female head of household has a small negative, but insignificant effect on the budget share for food for both years. Number of elders has a positive as well as a significant effect on the budget share for 2002, the coefficient estimate is however not significant in 1995.

The dummy coefficients for the provinces are presented in two separate tables: table 4 for the urban dummy variables, and rural estimates in table 5. Out of the 19 rural dummy variables for 1995, 18 are significant at a 5% significance level. 9 out of the 11 urban variables are significant at a 5% level for the same year. 20 out of 22 rural and all urban dummy variables are significant at a 5% level for 2002. Thus, a large majority of the estimated province coefficients are significant.

[1] The price elasticities can be calculated using the following formula from Costa (2001) and Hamilton (2001), where the expenditure elasticity of the budget share for food η is a function of the coefficient of the logarithm of income, and the average budget share ω:

$$\eta_{p.u} = 1 + \frac{\beta}{\omega}$$

Average budget share for food for 1995 is 0.45 and 0.47 in 2002 for the whole sample. This is large compared to the estimates of Gong and Meng, but sensible since they base their analysis on urban households only. This yields elasticity of 0.65 in 1995 and 0.66 in 2002 compared to 0.56 and 0.54 for 1995 and 2001, respectively (Gong and Meng 2008). These estimates imply that a 1% increase in expenditures in 1995 lead to a 0.65% increase in the budget share for food. Seven years later, an increase in expenditures would lead to the same effect of the budget share. The coefficients correspond to price elasticities which seems reasonable, and of a similar value to that of Hamilton (2001) and Costa (2001).

Budget share for food	1995	2002
Log(Income)	-0.16	-0.17
	(36.48)**	(33.49)**
Constant	1.96	2.00
	(40.21)**	(40.31)**
Age head of household	0.00	0.00
	(4.21)**	(3.83)**
Sex head of household	0.00	0.00
	(0.74)	(0.14)
Elders	0.01	0.04
	(0.99)	(3.92)**
Observations	2935	3548
R-squared	0.45	0.31

Table 3: Regression results main specification

The regression results are presented in three tables. Table 3 include the explanatory and control variable estimates, table 4 contain the dummy variable estimates for urban areas and table 5 is for rural dummy variables. This table reports two sets of estimates, one for each year. Robust t-statistics are in parentheses. The preferred specification is based on a subset of households consisting of two adults and one child. * Significant at 5%; ** significant at 1%.

Urban dummy variables	1995	2002
Shanxi	-0.10	-0.16
	(8.64)**	(12.58)**
Liaoning	-0.03	-0.08
	(2.62)**	(5.97)**
Jiangsu	-0.02	-0.08
	(2.24)*	(6.32)**
Anhui	-0.03	-0.06
	(2.55)*	(4.62)**
Henan	-0.08	-0.15
	(6.76)**	(11.87)**
Hubei	-0.03	-0.09
	(3.18)**	(7.14)**
Guangdong	0.06	0.03
	(5.35)**	(2.38)*
Chongqing	-	-0.09
		(5.88)**
Sichuan	-0.02	-0.06
	(2.23)*	(4.48)**
Yunnan	-0.02	-0.04
	(1.92)	(2.85)**
Gansu	-0.04	-0.13
	(3.54)**	(9.07)**

Table 4: Regression results - urban dummy variables, main specification

This table reports two sets of dummy variable estimates, one for each year. Chongqing is included only in 2002. Urban Beijing is set equal to 0. Robust t-statistics are in parentheses. * Significant at 5%; ** significant at 1%. The regression results are presented in three tables. Table 3 include the explanatory and control variable estimates, table 4 contain the dummy variable estimates for urban areas and table 5 is for rural dummy variables.

Rural dummy variables	1995	2002
Beijing	-0.14	-0.18
	(5.31)**	(6.04)**
Hebei	-0.36	-0.26
	(10.71)**	(10.94)**
Shanxi	-0.36	-0.25
	(7.02)**	(7.90)**
Liaoning	-0.27	-0.22
	(11.17)**	(9.99)**
Jilin	-0.23	-0.23
	(10.64)**	(12.90)**
Jiangsu	-0.25	-0.18
	(13.39)**	(9.73)**
Zhejiang	-0.10	-0.14
	(4.90)**	(7.99)**
Anhui	-0.27	-0.12
	(9.62)**	(6.09)**
Jiangxi	-0.23	-0.15
	(5.87)**	(4.82)**
Shandong	-0.26	-0.16
	(12.81)**	(8.61)**
Henan	-0.40	-0.21
	(15.34)**	(9.36)**
Hubei	-0.39	-0.20
	(14.14)**	(9.07)**
Hunan	-0.30	-0.12
	(13.15)**	(5.25)**
Guangdong	-0.06	-0.06
	(1.15)	(1.15)
Guangxi	-	-0.10
		(2.59)**
Chongqing	-	-0.13
		(5.80)**
Sichuan	-0.35	-0.11
	(18.46)**	(5.93)**
Guizhou	-0.41	-0.14
	(9.73)**	(4.24)**
Yunnan	-0.29	-0.01
	(9.68)**	(0.20)
Shaanxi	-0.46	-0.24
	(10.83)**	(8.76)**
Gansu	-0.49	-0.16
	(14.89)**	(6.08)**
Xinjiang	-	-0.18
		(7.59)**

Table 5: Regression results - rural dummy variables, main specification

The table reports two sets of dummy variable estimates, one for each year. Guangxi, Chongqing, Xinjiang is included only in 2002. Robust t-statistics are in parentheses. * significant at 5%; ** significant at 1%

4.2 Spatial price differences

Based on the estimated coefficients from chapter 4.1, we can identify the regional price levels relative to urban Beijing by equation (4). In this chapter we use these relative prices to identify the spatial price indexes, which are measured relative to a national average, i.e., we use equation (5) to normalize to the weighted national mean.

4.2.1 SPI results

The SPI gives us measures of regional price level differences. As the weighted national mean of the SPIs has been normalized to one for each year, it is not possible to compare between years, only across provinces. The resulting SPIs are presented in table 6, which also includes summary statistics on average spatial price levels for urban and rural areas separately; table 6 reveals the first two findings.

First, there are large differences between rural and urban price levels, according to the estimated SPIs for both years. In 1995, the urban price level is on average four times larger than the rural according to the SPI estimates. In 2002, this ratio is reduced – urban prices are approximately 1.6 times larger than the rural average price level. According to these estimates, there was a large gap in price levels between rural and urban areas in 1995. In 2002 this gap is considerably smaller, though still existing. Second, there are clear differences in price variation. The coefficient of variation (CV)[1] can be used as a measure of the spread of the estimated price levels. The CV for the whole of China is approximately 80 in 1995 and 60 in 2002, indicating less price dispersion in 2002 than in 1995.

Moreover, the SPI estimates also allow for investigation of variation in prices across rural and urban areas. Inspecting the values for price variation in table 6, we see that the variability of rural prices exceed that of urban prices in both 1995 and 2002. This goes in particular for 1995, where the rural CV equals 81, as compared to 26 for urban SPIs. In 2002 this difference is reduced, however. The rural price variation is markedly lower, while urban price variation has become larger. Rural price variation still exceeds urban price variation, but the difference is noticeably smaller in 2002 compared to 1995.

[1] $CV = \frac{\sigma}{\mu} * 100$, where μ is the average and σ the standard deviation.

Urban	SPI 95	SPI 02
Anhui	1.65	1.35
Beijing	1.94	1.99
Chongqing	-	1.16
Gansu	1.50	0.90
Guangdong	2.77	2.41
Henan	1.21	0.77
Hubei	1.59	1.15
Jiangsu	1.70	1.18
Liaoning	1.66	1.21
Shanxi	1.09	0.71
Sichuan	1.71	1.37
Yunnan	1.72	1.58
Rural		
Anhui	0.38	0.93
Beijing	0.83	0.65
Chongqing	-	0.88
Gansu	0.10	0.72
Guangdong	1.39	1.38
Guangxi	-	1.07
Guizhou	0.16	0.84
Hebei	0.22	0.39
Henan	0.18	0.54
Hubei	0.19	0.58
Hunan	0.33	0.96
Jiangsu	0.42	0.63
Jiangxi	0.48	0.78
Jilin	0.50	0.48
Liaoning	0.39	0.51
Shaanxi	0.12	0.45
Shandong	0.40	0.72
Shanxi	0.22	0.42
Sichuan	0.23	1.02
Xinjiang	-	0.63
Yunnan	0.35	1.89
Zhejiang	1.10	0.85
Mean (national)	0.88	0.97
Mean (urban)	1.68	1.32
Mean (rural)	0.42	0.79
Coefficient of variation (national)	81.64	48.42
Coefficient of variation (urban)	25.62	37.29
Coefficient of variation (rural)	80.81	43.96

Table 6: Spatial Price Index rural/ urban provinces

The table provides two sets of spatial price level estimates, one for 1995 and one for 2002. The price level estimates have furthermore been split in an urban and a rural section. The province price level estimates are presented alphabetically. A summary of the average price level and variation for the rural and urban SPIs is provided below. There is no SPI estimate for rural Chongqing in 1995, and no SPI estimate for urban Chongqing, Guangxi and Xinjiang in 2002.

Evaluation of the SPI

Before we turn to the presentation of the poverty estimates resulting from the real income estimates based on the SPI estimates, two questions are briefly discussed. First, can we relate the development of relative prices to Chinese economic development introduced in chapter 2? Second, does it follow the so-called Balassa-Samuelson hypothesis developed for cross-country comparisons, stating that the richer a country (in our case province), the higher the price level?

First, the findings indicate increasing market integration, following the line of reasoning presented in chapter 2. The overall and rural price variation is reduced from 1995 to 2002, which could be a result of rural areas being more integrated in the overall economy. Second, the results reveal increasing urban price dispersion and less urban-rural price level differences in 2002. Linking this to international trade, this could be due to uneven exposure to international markets and hence economic development for urban areas.

The latter is related to the Balassa-Samuelson hypothesis, which is explained in appendices A. It should be noted that we focus solely on the relationship between the price level and real income, and not the mechanisms described by this hypothesis. Investigating the empirical relationship between the estimated SPIs and real incoes, we do find evidence in favour of the Balassa-Samuelson hypothesis for 1995 (see graphs presented in the appendices D). We can see that there is a positive relationship between the average real income in an urban/rural province and the local price level[1]. For 2002 however, the results are mixed. The relationship between the price level and real income is negative when basing the analysis on households with the same composition. Extending the analysis to all households however, we do find evidence in favour of the Balassa-Samuelson hypothesis, see appendices A. As real incomes are produced by adjusting nominal income for different price levels, the price adjusted income will give a different picture with regards to poverty levels than nominal incomes do. Basically, the results from this test indicate that poorer areas have relatively lower prices, while richer areas face relatively higher prices. How will this affect the poverty estimates? Consider a poor household with a nominal income below the defined poverty threshold. This household is furthermore located in a low-price area. Real income then, is relatively higher than what is indicated by nominal measures. This could result in a situation where income measured in nominal terms place the household below the defined poverty line, while real income indicates that the household should not be considered as poor. The opposite effect is likely to affect a low-income

[1] The data material is based on average provincial income and province price levels estimates for rural and urban areas separate. The relation is significant on a 5 percent level.

urban household located in a high-price area as well. We investigate the effect of this (and more) on poverty estimates for China in the next chapter.

4.3 Poverty measurement

Poverty estimates are sensitive to price adjustment, and this allows for an investigation of the trend in poverty rates when adjusting for spatial price differences. When having estimated the SPIs, and subsequently new household real incomes, new poverty measures can be calculated. In this chapter we provide these new poverty measures. The poverty lines given by the PWT and ICP PPP exchange rates were derived in chapter 3.3.1. The method for deriving price level estimates and adjusting income was presented in 3.1. The main results are presented in two tables; table 7 presents the headcount ratio, and table 8 the actual number of people whose expenditures fall below the defined poverty lines.

Nominal consumption means income measured with no price adjustment; real consumption is nominal consumption deflated by the SPIs. The headcount provides the percentage poor for individuals from urban, rural and the national level, respectively.

Adjusting incomes for spatial price differences has a large effect on poverty estimates.
Six poverty findings appear. First, national poverty rates are significantly lower when adjusting for spatial price differences. Second, nominal incomes indicate a reduction in poverty from 1995 to 2002– real incomes does not. Third, the urban poverty incidence is higher when correcting for local price levels. Fourth, nominal incomes indicate that urban poverty is increasing in the period from 1995 to 2002, while urban poverty is reduced when we apply real income. Fifth, even though adjusting for different price levels greatly reduces rural poverty incidence, poverty is still a rural issue. Sixth, the urban/ rural poverty gap is reduced throughout 1995-2002 in nominal terms, but this pattern is not evident when examining poverty estimates based on price adjusted income.

Headcount rate (percent)		ICP 1995		ICP 2002		PWT 1995		PWT 2002	
		Nominal income	Real income	Nominal income	Real income	Nominal income	Real income	Nominal income	Real income
Equivalence scale (OECD) adjusted income	Rural	77.60	19.96	44.08	24.79	47.78	3.93	8.04	3.56
	Urban	0.80	4.45	1.06	2.04	0.04	0.23	0.05	0.11
	Total	47.74	13.93	28.75	16.68	29.22	2.49	5.19	2.33

Table 7: Headcount Poverty Estimates
This table reports the percentage of poor individuals of the sample for the two poverty lines for both years. Urban poverty rates refers to the percentage of poor urban individuals out of the urban sample population, rural to the rural sample and finally total poverty rates as share of the total sample population. Real income is nominal income adjusted for prices using the estimated SPI.

Number of poor (in 1000)		ICP 1995		ICP 2002		PWT 1995		PWT 2002	
		Nominal income	Real income	Nominal income	Real income	Nominal income	Real income	Nominal income	Real income
Equivalence scale (OECD) adjusted income	Rural	666 953	171 553	344 910	193 997	410 656	33 698	62 878	27 889
	Urban	2 807	15 641	5 328	10 238	148	804	246	565
	Total	669 759	187 194	350 238	204 235	410 804	34 503	63 124	28 454

Table 8: Number of Poor (1000 persons)
This table reports the absolute number of poor based on the headcount ratios (1000 persons). The urban and rural headcount ratios have been applied to the actual rural and urban population numbers for the years 1995 and 2002 for both poverty lines. Real income is nominal income adjusted for prices using the SPI estimates.

National poverty

The first finding is based on a comparison of national poverty rates from nominal and adjusted incomes, respectively. On a national level, the effect of price adjustment is clear. When incomes are adjusted for price levels, national poverty estimates are lower than what is indicated by nominal incomes. This is the case for both 1995 and 2002, using both the ICP and the PWT-line. This could be due to the Balassa-Samuelson effect that was discussed earlier. Adjusting for differences in spatial price levels has the largest effect in 1995 for both the ICP and the PWT poverty line.

Investigating how poverty rates changes from 1995 to 2002, table 7 shows two different trends for the two measures of income. This is the second finding. Poverty estimates based on nominal incomes indicate that there has been a large reduction in overall poverty from 1995 to 2002, while real income indicate little or no sign of a poverty reduction. The poverty rates based on the ICP line and nominal incomes are extremely high for 1995, and though the number of poor is reduced by more than 300 million people, the poverty rate for 2002 is still close to 30 percent. The ICP line based on real income however, indicates that poverty increased by roughly 2.5 percent from 1995 to 2002, totalling 204 million people.

The PWT line applied to nominal incomes gives a poverty incidence of 30 percent in 1995, while being reduced to roughly 5 percent in 2002. This corresponds to a total number of 63 million people. The headcount ratio provided by the PWT line and real incomes however, implies that the poverty reduction in this period was minor. According to these estimates, 2.49 percent of the national population in China survived on less than one dollar a day in 1995. The corresponding result for 2002 is 2.33 percent, which is minimally smaller (though equivalent to 6 million fewer poor).

Thus, according to the poverty estimates based on real incomes, there is little sign of a poverty reduction in China from 1995 to 2002. This result is strikingly different from the poverty estimates based on nominal incomes. The poverty estimates for the sample as a whole based on nominal incomes seem to overestimate poverty. This result can be due to lower price levels in poorer regions relative to those in richer areas. Taking lower provincial price levels in rural areas into consideration then leads to lower poverty rates for the rural population. The opposite effect should then affect the urban poverty rates. This is explored in the next paragraph.

Urban poverty

If adjusting for relatively lower rural price levels leads to lower rural poverty estimates, adjusting urban household income for relatively higher urban price levels could potentially lead to higher urban poverty incidence. This is exactly what the figures indicate. Adjusting incomes for spatial price differences produces higher urban poverty rates for both poverty lines, which is the third finding. Investigating the ICP line, we see that the poverty rate is 4.45 percent when using price adjusted income, versus 0.8o percent implied by nominal income. This is the case for both poverty lines, in both 1995 and in 2002.

As for the trend for national poverty estimates, the urban poverty rates based on real and nominal incomes give rise to two opposing effects. Urban poverty incidences measured in nominal terms indicate that urban poverty increased slightly from 1995 to 2002, from 0.8 percent to 1.06 percent when measured using the ICP line, and by 0.01 percent by the PWT line. In real terms however, urban poverty rates were halved from 1995 to 2002. Thus, urban poverty is under-estimated when using nominal incomes, according to the estimated SPI – this is the fourth finding.

Rural poverty

Table 7 and table 8 reveal the fifth finding, that poverty is mainly a rural phenomenon. Adjusting for prices reduces rural poverty incidence greatly, but the large majority of poor individuals are still located in rural areas. Rural individuals make out 61 percent of the total sample, but but between 92 to 98 percent of all individuals in the sample defined as poor are from rural areas (measured in real terms). The share is roughly 99 percent when we use nominal income - by and large poverty is a rural issue.

Inspecting the poverty estimates based on nominal values, the ICP line indicates that poverty was reduced from 78 to 44 percent from 1995 to 2002, while the real values indicate the opposite, namely that the rural poverty incidence increased by almost 5 percent.. The PWT line indicates a massive reduction in rural poverty for this period for nominal incomes, reducing the sheer number of people living for less than a dollar a day by 350 million people. The real income poverty estimate suggests that the reduction was considerably less, and roughly equal to 5.5 million people.

Adjusting for spatial price differences reduces the absolute number of rural poor extensively, but on the other hand we see little evidence of reduced poverty from 1995 to 2002.

Urban/ rural poverty gap

Finding six concerns the development in the rural / urban poverty gap in the period between 1995 and 2002. In nominal terms, the urban-rural poverty gap is reduced by a large fraction from 1995 to 2002, as rural poverty is greatly reduced while urban poverty incidence increases. The poverty estimates based on real incomes and the ICP line actually imply a widened urban-rural poverty gap. There is no change when applying the PWT line. This is due to increasing rural poverty rates from 1995 to 2002 in combination with decreasing urban poverty incidence.

5 Extended models

In this chapter we present the results from two extended regression analyses, neither of which change the results chapter 4.3. Equation (3) is estimated using the whole sample. This is to check whether we have a selecton bias in our main results. Second, we test whether the inclusion of more control variables change the results.

5.1 Including all households

As a robustness check, we run a test including all households. Instead of restricting household composition to two adults and one child, the number of children and adults in the households are controlled for by including them as control variables. In short, the results support the findings presented in chapter 4.3 the poverty results will be presented first, followed by a short comment on the coefficient estimates and the SPI.

The regression results are based on the full sample, roughly 15 000 observations for each year, and they are presented in appendix E. The coefficient of the logarithm of income is -0.13, which is lower than that of the main regression. Including all households or not for the 2002 data does not change the coefficients much. The number of adults in the household has positive and significant effect on the budget share for food. The coefficient estimate for the number of children is positive in 1995 and negative in 2002, significant on a 5 percent level both years. Age of head of household has a positive coefficient estimate, significant for both years, while gender has a negative but not significant effect on the budget share. The coefficients for number of elders in the household are both significant on a 5 percent level for both years, and are estimated to have a positive effect.

The new SPI confirm the results derived from the main analysis. Concerning the first finding, there are large differences when it comes to average urban/ rural price levels, even larger than what is indicated by the main SPIs for 1995. Furthermore, according to the new SPIs, the rural/urban price level gap is reduced by even more than what the main specification indicates in 2002. When it comes to finding 2 - price variation - the new SPI strengthens the results concerning price integration. Comparing the new SPI to the main, price variation is relatively higher in 1995, while relatively lower in 2002. Finally, including all households reduces rural price variation in both years. In 2002 the CV for rural and urban areas are almost identical. The results based on all households thus strengthen the findings presented in chapter 4.3.

The poverty estimates based on nominal incomes are by definition unchanged, while the headcount rates based on real incomes are slightly changed due to new price level estimates. First of all, none of the six findings presented in chapter 4.3 are rejected. On the contrary, the estimates based on all households strengthen the findings from the main analysis. Investigating the wheteher the first finding holds, we see that the total poverty incidence is even lower after incomes have been adjusted for local price levels. On finding two, the robustness check provides the same results - nominal incomes indicate poverty reduction, while real incomes either indicate a minor reduction, or an increase in the poverty incidence. On finding three, the robustness check indicate that urban poverty is somewhat higher than what is indicated by the main analysis. Finding four is also supported, we see that real incomes indicate decreasing urban poverty from 1995 to 2002, while nominal incomes indicate the opposite trend. Fifth, poverty is mainly a rural issue in both the main specification and in the robustness check. On the sixth and final finding, we see that the urban / rural poverty gap is widened by the inclusion of all households.

Headcount rate (percent)		ICP 1995		ICP 2002		PWT 1995		PWT 2002	
		Nominal income	Real income	Nominal income	Real income	Nominal income	Real income	Nominal income	Real income
Equivalence scale (OECD) adjusted income	Rural	77.60	12.99	44.08	19.63	47.78	1.93	8.04	1.33
	Urban	0.80	5.75	1.06	2.47	0.04	0.43	0.05	0.16
	Total	47.74	10.18	28.75	13.52	29.22	1.35	5.19	0.91

Table 9: Headcount Poverty Estimates for all households, controlling for number of children and adults
This table reports the percentage of poor individuals of the sample for the two poverty lines for both years. Urban poverty rates refers to the percentage of poor urban individuals out of the urban sample population, rural to the rural sample and finally total poverty rates as share of the total sample population. Real income is nominal income adjusted for prices using the estimated SPI.

5.2 More control variables

The main specification only includes three control variables. In this chapter the specification is extended , including several control variables, namely the number of children, elders and adults in the household, the age of the spouse (of head of household) , the educational level of the head of household and number of females. Comparing these estimates to those given by the main specification, we find strong support in table 9 for the six findings presented in chapter 4.3.

The coefficient estimates are presented in appendix F. In short, the head of household's age, the number of elders and adults have a positive effect on the budget share for food. Only elders have a significant (5%) coefficient estimate for both 1995 and 2002. Share of females in the household and the level of education have a negative and significant (5%) impact of the food's share. Whether the head of household is male/ female and the age of the spouse does not have a significant effect. The numbers of children have a negative effect in 1995, but a positive coefficient estimate in 2002, both significant on a 5 percent level. All of the rural coefficient estimates are significant for both 1995 and 2002, and so are the urban dummy estimates for 2002. But only four of the urban dummy coefficient estimates are significant on a 5 percent level in 1995.

The SPI results are reassuring. Finding one on urban/rural price level difference is confirmed - variation is large in 1995, and reduced in 2002. Concerning the second finding,the CV indicates that price dispersion was even larger in 1995, and moreover that the price variation in 2002 is even lower than the other estimates.

Headcount rate (percent)		ICP 1995		ICP 2002		PWT 1995		PWT 2002	
		Nominal income	Real income	Nominal income	Real income	Nominal income	Real income	Nominal income	Real income
Equivalence scale (OECD) adjusted income	Rural	77.60	10.79	44.08	18.42	47.78	1.45	8.04	1.24
	Urban	0.80	6.22	1.06	2.52	0.04	0.43	0.05	0.18
	Total	47.74	9.01	28.75	12.75	29.22	1.05	5.19	0.86

Table 10: Headcount Poverty Estimates for all households, including more control variables
This table reports the percentage of poor individuals of the sample for the two poverty lines for both years. Urban poverty rates refers to the percentage of poor urban individuals out of the urban sample population, rural to the rural sample and finally total poverty rates as share of the total sample population. Real income is nominal income adjusted for prices using the estimated SPI.

In the following, the results will be compared to the six findings presented in chapter 4.3, as done for the previous robustness check. First, the total poverty incidence is even lower than the poverty estimates in chapter 5.1 after incomes have been adjusted for local price levels. Second, nominal incomes indicate that poverty was reduced from 1995 to 2002, real incomes do not. Third, urban poverty increases when prices are corrected for, and by more than what is indicated by the main analysis. Fourth, real incomes and the ICP line indicate that the urban poverty reduction was even larger compared to the main analysis. Fifth, including more control variables and all households produces the highest urban poverty rates, but poverty is still mainly a rural issue. What is noticeable concerning finding five is the large increase in rural poverty as indicated by the ICP poverty line and real incomes, from 11 percent in 1995 to 18 percent in 2002. Sixth, we see no sign of the urban / rural gap decreasing when inspecting poverty estimates based on real income.

6 Conclusion

In this thesis, we have investigated the extent to which spatial price levels differ in China in 1995 and 2002. Detailed household data allows for provincial price level identification, while also separating between rural and urban areas. Urban-rural differences are well-documented and by no doubt an aspect that must be considered when measuring poverty, but access to data is always a challenge. The inclusion of a large number of rural and urban households from a national representative sample of China's population is a desirable quality and strength of this analysis. We apply the estimated SPIs on incomes and produce new estimates of real income. Based on these, we calculate poverty estimates.

The price level estimates reveal some interesting findings, and so does the poverty estimates based on price adjusted income. There are two results concerning the relative price levels that are particularly interesting. The average urban and rural price levels are relatively different for both years. This tells us what we would expect, namely that prices differ between rural and urban areas. However, the difference between the urban and rural price level is less in 2002 than in 1995. This leads us to the second finding, namely that there is less national price dispersion in 2002 compared to 1995. Adjusting incomes using these SPIs, the first observation is that adjusting for local price differences has a large effect on the poverty estimates. More specifically, the poverty estimates can be summarized by six poverty findings. First, national poverty rates are significantly lower when adjusting for spatial price differences. Second, nominal incomes indicate that there was a reduction in poverty from 1995 to 2002– real incomes does not. The higher poverty line actually indicates that poverty increased in the period in question. Third, the urban poverty incidence is higher when correcting for local price levels. Fourth, nominal incomes indicate that urban poverty is increasing in the period from 1995 to 2002, while urban poverty is reduced when we apply real income. Fifth, even though adjusting for different price levels greatly reduces rural poverty incidence, poverty is still mostly a rural phenomenon. Sixth, the urban/ rural poverty gap is reduced throughout 1995-2002 in nominal terms, but this pattern is not evident when examining poverty estimates based on price adjusted income.

As the main topic in this thesis is price adjustment and the effect of this, the analysis has been based on a comparison of the results from using nominal and real incomes. It would however be interesting to draw a comparison with other broad studies. The World Bank (2009) has published an assessment report on poverty reduction in China for the past twenty years. The period of time being analyzed in this thesis is comparable to what is referred to as the *9th Plan period*, 1996-2001. According to this report, the period in question is characterized by a marginal decline in poverty. As the method applied in their report and in this thesis differs, the poverty estimates in themselves are not comparable. Ravallion and Chen (2007) also find that poverty reduction stalled in the late 1990s. To make a complete assessment of poverty estimates and methods for China is a daunting task which would require a thesis in itself. A more thorough comparison of poverty rates is therefore not included. The main point still, is that it is reassuring to know that other studies do not find support for a large poverty reduction in this period.

The choice of poverty line influences the results. The poverty line given by the PWT is approximately half the size of the line indicated by the ICP exchange rates. Needless to say, this produces two different sets of poverty estimates. The lower poverty line indicates little or no change in poverty from 1995 to 2002. The higher line implies that poverty increased from 1995 to 2002. It is tempting to connect these results to the debate on whether inequality has increased in China throughout the years or not. The lowest poverty corresponds to other poverty lines for rural households[1], capturing household at the bottom end of the income distribution. The higher poverty line is similar to those defined for urban households, including households with an income slightly above the absolute minimum required for rural households. If inequality is increasing, more households will be located in the tails of the income distribution. A higher poverty line will capture this effect, naturally leading to increased poverty estimates. The literature and views on inequality and its characteristics is extensive (Khan and Riskin 2001; Riskin et. al. 2001; Fang et. al. 2002; Chen and Lin 2004; Sicular et al. 2007; Wan 2007; The World Bank 2009). A comprehensive discussion and evaluation of these results in this regard then is beyond the scope of this thesis. This argument is nonetheless interesting, and adds to the credibility of the results.

As the "true" poverty rates in China are unknown, a discussion on what constitutes reasonable and not so reasonable estimates of poverty has been based on the results from nominal and real incomes. Income not adjusted for prices is in any case not a proper measure to base poverty estimation on, but it reveals important information on what difference adjusting for local price levels make. Our findings are robust to specifications including all households and more control variables, which is reassuring. As the poverty estimates do not support the perceived "fact", namely that the poverty reduction has been large and ongoing for a long time in China, there is clearly a need for further analysis. One potential approach is to use a higher poverty line, as the estimates here are

[1] See appendix A.

based on a conservative estimate of one dollar a day. Another issue concerns urban poverty. As poverty is perceived to be mainly a rural phenomenon, much work has been oriented towards this. As our estimates indicate that urban poverty is higher than what is suggested by nominal incomes, this calls for further analysis of urban poverty. Including migrants in an analysis of urban households has the potential to reveal more information on this, and as we have argued throughout this thesis, the Engel curve approach is a valuable tool in this context.

References

Almås, I. (2007): "International Income Inequality: Measuring PPP bias by estimating Engel curves for food", LIS Working Paper Series 473.

Banks, J., R. Blundell, and A. Lewbel. (1997): "Quadratic Engel curves and consumer demand" , Review of Economics and Statistics 79, No. 4, 527-539.

Barro, R. J., and X. Sala-i-Martin (2003): Economic Growth 2nd edition. The MIT Press.

Brandt, L., and C. A. Holz (2006): "Spatial price differences in China: Estimates and implications" , Economic Development and Cultural Change 55, No. 1, 43-86.

Branigan, T. (2008): "China considers ending one-child policy", The Guardian, Feb. 28th 2008.

Bureau of Labor Statistics (2009): "Consumer Price Index" , ftp://ftp.bls.gov/pub/special.requests/cpi/cpiai.txt, accessed 30/08-2009.

Chan, K. W. (2009): "Measuring the urban millions" , China Economic Quarterly March, 21-26.

Chang, G. H., and Q. Shao (2004): "How much is the Chinese currency undervalued? A quantitative estimation." , China Economic Review 15, No. 3: 366-371.

Chen, S. and Ravallion, M. (2008): "China is poorer than we thought, but no less successful in the fight against poverty." , Policy Research Working Paper 4621.

Chen, S. and M. Ravallion (2007): "Absolute poverty measures for the developing world 1981–2004" , *Proceedings of the National Academy of Sciences of the United States of America* 104, no. 43, 16757-16762.

Costa, D. (2001): "Estimating real income in the United States from 1888 to 1994: Correcting CPI bias using Engel curves" , Journal of political economy 109, No. 6, 1288-1310.

Deaton, A. and J. Muellbauer (1980): "Economics and consumer behavior". Cambridge University Press.

Deaton, A. and J. Muellbauer (1980): "An Almost Ideal Demand System" , The American Economic Review 70, no. 3, 312-326.

Deaton, A. and S. Zaidi (2002): "Guidelines for constructing consumption aggregates for welfare analysis". World Bank Publications.

Dunaway, S. and X. Li (2005): "Estimating China's' equilibrium'real exchange rate" , *IMF Working Paper* 05, No. 202.

Fan, C. S. and X. Wei (2006): "The Law of One Price: Evidence from the Transitional Economy of China" , Review of Economics and Statistics 88, no. 4, 682-697.

Fang, C., X. Zhang, and S. Fan (2002): "Emergence of urban poverty and inequality in China: evidence from household survey" , China Economic Review 13, No. 4, 430-443.

Fang, Z., E. Wailes, and G. Cramer (1998): "China's rural and urban household survey data: collection, availability, and problems" , vol. 202, CARD Working Paper 98.

Frankel, J. (2006): "On the yuan: The choice between adjustment under a fixed exchange rate and adjustment under a flexible rate" , CESifo Economic Studies 52, No. 2, 246–275.

Gente, K. (2006): "The Balassa-Samuelson Effect in a Developing Country" , Review of Development Economics 10, No. 4, 683-699.

Gong, C. H. and X. Meng (2008): "Regional Price Differences in Urban China 1986-2001: Estimation and Implication." IZA Discussion Paper 3621.

Guillaumont, S. J. and P. Hua (2002): "The Balassa-Samuelson effect and inflation in the Chinese provinces" , China Economic Review 13, No. 2-3, 134-160.

Hamilton, B. W. (2001): "Using Engel's law to estimate CPI bias." , American Economic Review 91, No. 3, 619-630.

Heston, A., R. Summers and B. Aten (2006): "Penn World Table 6.2." Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania (CICUP), Philadelphia.

Huikang, J. (2003): "History of Guangdong", News Guangdong March 20th 2003.

Khan, A. R, and C. Riskin (2001): "Inequality and Poverty in China in the Age of Globalization" , Oxford University Press, USA.

Klugman, J. (2002): "A Sourcebook for Poverty Reduction Strategies." , World Bank Publications.

Krugman, P. R. and M. Obstfeld (2003): International Economics: Theory and Policy, Sixth Edition. Boston, Ma: addison Wesley.

Leser, C. E. V. (1963): "Forms of Engel Functions." , Econometrica 31, No. 4, 694-703.

Lin, W. L, and T. P Chen (2004): "China's widening economic disparities and its 'Go West Program'." , Journal of Contemporary China 13, No. 41, 663–686.

Liu, Z. (2005): "Institution and inequality: the hukou system in China." , Journal of Comparative Economics 33, No. 1, 133-157.

Lu, Y. (2009): "China's Floating Population." , ISERP Quarterly Newsletter 5, No. 3.

Marangos, J. (2006): "Were the Chinese reforms a feasible alternative for transition economies?" , International Journal of Social Economics 33, No. 3, 221 - 240.

Moulton, B. R. (1996): "Bias in the Consumer Price Index: What is the Evidence?" , The Journal of Economic Perspectives 10, No. 4, 159-177.

National Bureau of Statistics of China (2005): "Communiqué on 2004 Rural Poverty Monitoring of China." , http://www.stats.gov.cn/english/newsandcomingevents/t20050513_402249491.htm, accessed 30/8-2009.

National Bureau of Statistics of China (2008): China Statistical Yearbook 1996, 2003, 2008. Beijing: China Statistics Press.

OECD Social Policy Division (2009): "What are Equivalence Scales?" Directorate for Employment, Labour and Social Affairs, http://www.oecd.org/LongAbstract/0,3425,en_2649_33933_35411112_1_1_1_1,00.html, accessed 30/8-2009

Qian, Y. (2002): "How Reform Worked in China." , William Davidson Institute Working Paper 473.

Ravallion, M. and Chen, S. (2007): "China's (uneven) progress against poverty." , Journal of Development Economics 82, No. 1, 1-42.

Ravallion M., Chen, S. and P. Sangraula (2008): "Dollar a day revisited." , Policy Research Working Paper 4620, 1-35.

Reddy, S. G., and C. Minoiu (2008): "Chinese poverty: assessing the impact of alternative assumptions." , Review of Income and Wealth (forthcoming), http://ssrn.com/abstract=799844, accessed 30/8-2009.

Riskin, C., R. Zhao, and S. Li. (2001): China's retreat from equality: income distribution and economic transition. East Gate Book.

Saunders, P. (2007): "Comparing Poverty Among Older People in Urban China Internationally." , The China Quarterly 190, No. 1, 451-465.

Sicular, T., X. Yue, B. Gustafsson, and S. Li. (2007): "The urban-rural gap and income inequality in China." , Review of Income and Wealth 53, No. 1, 93-126.

USAID (2008): "Poverty Assessment Tools", *Frequently Asked Questions*. http://www.povertytools.org/faq/faq.html, accessed 30/8-09.

Wan, G. (2007): "Understanding regional poverty and inequality trends in China: methodological issues and empirical findings." , Review of Income and Wealth 53, No. 1, 25–34.

World Bank (2006): "About the ICP and PPPs - Frequently Asked Questions." , http://siteresources.worldbank.org/ICPINT/Resources/About_the_ICP_and_PPPs_FAQ_11.2006.doc , accessed 30/8-2009

World Bank (2008): Global purchasing power parities and real expenditures - 2005 International Comparison Program. Washington DC: World Bank.

World Bank (2008): Reshaping Economic Geography. Washington, DC: The World Bank.

World Bank (2009): From poor areas to poor people: China's evolving poverty reduction agenda - An assessment of poverty and inequality in China. Poverty Reduction and Economic Management Department, East Asia and Pacific Region, World Bank, Beijing.

Young, Alwyn (2000): "The Razor's Edge: Distortions and Incremental Reform in the People's Republic of China." Quarterly Journal of Economics 115, No. 4, 1091-1135.

Zhang, K. H. and S. Song (2003): "Rural-urban migration and urbanization in China: Evidence from time-series and cross-section analyses." China Economic Review 14, No. 4, 386-400

Appendices

A Compared with other poverty lines

The PWT and ICP derived in chapter 3.3.1 correspond well to other poverty lines. In this chapter we present poverty lines calculated by other authors.

Poverty thresholds derived by other researchers can be used to evaluate the sensitivity of the measured lines. Several authors use the $1/ day in 1993 prices as a measure of poverty, for one Reddy and Minoiu (2006). They use PPP exchange rates provided by the PWT 5.5 and World Economic Outlook, which is somewhat lower than those provided by the PWT 6.2 and ICP used in chapter 3.3.1[1]. Adjusting for price increases using the CPI for 1993 and 1995, the poverty lines are presented in 1995 and 2002 prices in table 10 below. As can be seen from the table 10, the lower PPP estimates leads to slightly lower poverty lines compared the poverty lines derived here.

Khan and Riskin (2001) derives sets of high and low poverty lines for urban and rural areas respectively, where the low threshold is set to be 810 Yuan (rural) and 1604 Yuan (urban), and the high threshold 1159 Yuan (rural) and 2291 (urban), nominal values. This is based on a more complex procedure. First they multiply a defined minimum of food energy intake (kilocalories) by the estimated actual food costs for groups close to the poverty threshold. Then they add non-food consumption expenditures based on the expenditure patterns of groups closest to the poverty threshold. Adjusting the 1995 lines for prices (as explained previously) to get estimates for 2002, we can see that the poverty lines estimated in this thesis are located between lowest and highest estimates for rural as well as urban areas.

Ravallion and Chen (2007) derive poverty lines from a similar approach in collaboration with the NBS. Based on the consumption patterns of the poorest 15th and 25th percentile nationally, regional food bundles are defined and evaluated at regional prices. These are scaled to reach 2100 calories per person each day, with 75% of the calories coming from food grains. They use the same method as Khan and Riskin concerning non-food, while separating between urban and rural provinces as well. The urban/ rural poverty lines are simply the means of these regional lines, see table 10 below (Chen and Ravallion 2007). Their poverty lines correspond to the PWT line, which is located between the

[1] The provided 1993 PPP exchange rates are 1.03 Yuan/USD and 2.13 Yuan/USD (Minoiu and Reddy 2006). According to the calulations in chapter 3.3.1, the PWT PPP 1993 equals 1.48 and the calculated ICP PPP 1993 is equal to 2.74.

lower rural line and the urban for both years. China's National Bureau of Statistics (NSB) monitors only rural poverty (NBS 2004). The official poverty threshold for China in 1995 was 540 Yuan and 627 Yuan in 2002 (Khan and Riskin 2001; NBS 2004). Ravallion and Chen (2008) notes that this is one of the lowest lines in the developing world, which is confirmed by the included poverty lines – the official line for China is by far the lowest.

		1995			2002		
Table 11:Poverty lines		**National**	**Rural**	**Urban**	**National**	**Rural**	**Urban**
PWT-line		845	-	-	923	-	-
ICP-line		1571	-	-	1716	-	-
Official Poverty Line (NBS)		-	540	-	-	627	-
Chen & Ravallion (2007)		-	778	1099	-	850	1200
Khan & Riskin (2001)	lower line	-	810	1159	-	885	1266
	higher line	-	1604	2291	-	1752	2502
Reddy & Minoiu (forthcoming)	lower line	588	-	-	642	-	-
	higher line	1219	-	-	1332	-	-

B Additional equivalence scales

In chapter 3.3.2 we presented the OECD equivalence scale. In addition to the OECD scale, we have the following three scales (OECD 2009):

The "OECD-modified scale" used by the Statistical Office of the European Union (Eurostat 2005) is given as follows:

(16) $$EU^{ES} = 1 + (adults - 1) * 0.5 + children * 0.3$$

The “Square root scale” is given by:

(17) $$SQRT^{ES} = \frac{1}{\sqrt{adults + children}}$$

When deciding on which scale to apply to the income measure, this general formula for equivalence scale adjusted income from Saunders (2007) is useful, where *y* is income and α indicates the extent of economies of scale:

(18) $$y^{ES} = \frac{y}{(children + adults)^{\alpha}}$$

In this setting, α is assumed to be a number between 0 and 1, with close to one indicating a low degree of economies of scale. Per capita measures implicitly assume α to be equal to one. Saunders (2007) report that 0.55 is frequently used for industrialized countries, which is close to the square root scale. In the case of a household consisting of two adults and one child, the OECD-scale equals 2.2, the EU-scale 1.8 and the square root scale 1.7. The National Bureau of Statistics in China have estimated α to be equal to 0.89 for urban households in China (ibid.), yielding 2.7. As the OECD-scale provided the closest estimate to this, we chose this line.

C Migrants

Since reforms were initiated in 1978 and restrictions on population movements were reduced, China has possibly experienced the largest flow of migration from rural to urban areas seen in history (Zhang and Song 2003). The urban population in 1978 was 18 percent, while almost 45 percent in 2007 (NSB 2008). This number however, does not include migrants that do not have a permanent residence in cities. The city of Shenzhen for instance, had a population of 1.3 million people with local hukou in 2002. A census based on the same geographic boundaries revealed a resident population of 7 million (Chan 2009). This is often referred to as *the floating population,* and they are usually not included in official statistics for the urban population (Zhang and Song 2003). Recent estimates put the total number of migrant worker to 150 million people (Lu 2009).

These are not included in the 1995 household survey, since the surveys only incorporate people with an urban hukou (see chapter 2) residence permit in the urban sample. Interestingly, the 2002 survey includes a separate questionnaire addressing migrants specifically, a total number of 2000 households. The migrant section is not included in the main analysis, in order to ensure comparability over the years. But excluding migrants is likely to lead to underestimation of urban poverty. The hukou system does not allow most migrants to get a local residence permit in the cities they work in (Zhang and Song 2003). Migrant workers are mainly concentrated in low-skilled jobs with low wages. A local hukou is necessary in order to have access to urban welfare provisions such as housing, access to schools and health services (ibid.). As we have no such data for 1995, migrants can not be included in the main analysis, which is unfortunate.

D Can we identify the Balassa-Samuelson effect?

The Balassa-Samuelson hypothesis states that the price level is positively linked to income (see box for an explanation of this below). In this line of reasoning, we would expect that the estimated price coefficients from the regression analysis will be negative at large, since Beijing is the capital, one of the main trade centres in China and part of the coastal region in the east where reforms and trade liberalization were initiated first. A positive dummy variable would indicate that the price level is higher for the relevant province relative to urban Beijing, since the dummy variable of urban Beijing is by definition equal to zero[1]. This is the case for Guangdong, which gets a positive coefficient estimate for both 1995 and 2002. From table 5 we can see that of all the included provinces, only Guangdong has an estimated price level exceeding that of urban Beijing. Guangdong was one of the five elected areas to initiate reforms and liberalization in 1980, and has grown to become one of the main industrial centres of China (News Guangdong 2003). Thus, this is not an unreasonable result. Some estimates are close to zero – indicating a similar price level to that of urban Beijing.

The Balassa-Samuelson hypothesis

Research on international price levels has revealed that a country's price level is positively linked to real income per capita, when measured in a single currency. Bela Balassa and Paul Samuelson put forth a theory linking the price of nontradable goods to the determination of national price levels, giving rise to what are known now as the Balassa-Samuelson hypothesis. Nontradables are goods which can not be traded internationally with a profit due to large transportation costs, such as services.

This hypothesis assumes that there is not a big difference in the productivity of labour forces in rich and poor countries in the nontradables sector. Labour forces of poor countries are assumed to be less productive than those of richer countries in the traded goods sector, however. As prices of traded goods are assumed to be more or less equal in all countries, wages must be lower for workers in poor countries in the tradable's sector, assuming that wages are linked to productivity. As wages are lower for workers in a poor country, the relative cost of labour is cheaper, leading to lower production costs in the nontradeables sector. This is reflected in the lower price of nontradable goods in poor countries relative to rich countries.

Studies have applied the Balassa-Samuelson hypothesis on Chinese provinces order to explain provincial inflation rates, treating each province as a country (Guillamont and Hua 2002). Another common application is for exchange rate determination (Chang and Shao 2004; Dunaway and Xianming Li 2005; Frankel 2006; Gente 2006).

Source: (Krugman and Obstfeld 2003)

[1] As can be seen from the equation $P_p = e^{-\frac{(d_p)}{b}}$ presented in chapter 3.1.

The Balassa-Samuelson hypothesis can be tested by simply running a regression with the SPI as the dependent variable and price adjusted income as the explanatory variable. Average real income per capita for each province is calculated, separating between urban and rural areas. Comparing the average price level and the estimated SPI for 1995, we are able to identify the Balassa-Samuelson effect for Chinese households, when restricting the analysis to households consisting of two adults and one child, as shown in chapter 4.2.1.

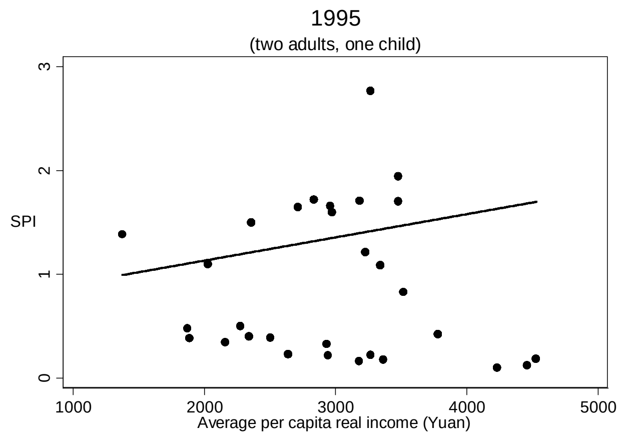

Figure 1: SPI and real incomes, restricting household composition (1995)

When investigating the relationship between the spatial price levels and real income for households consisting of two adults and one child for 2002, we find a small and barely significant negative effect:

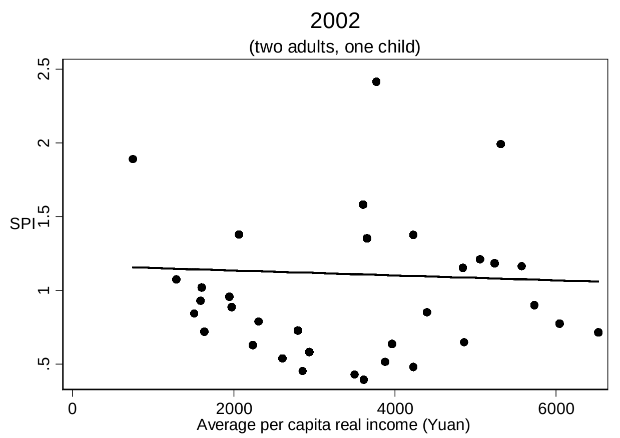

Figure 2: SPI and real incomes, restricting household composition (2002)

When expanding the number of observations by including all households in 2002, we find a positive and significant effect between the regional price level and real income:

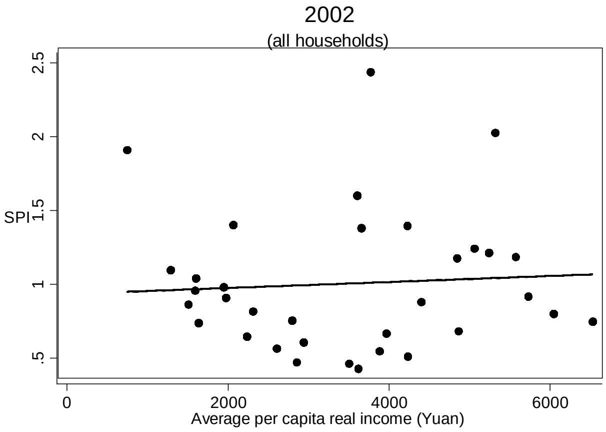

Figure 3: SPI and real incomes, all households (2002)

Thus, there is support of the Balassa-Samuelson hypothesis in 1995, but the results are mixed when it comes to 2002.

E All households: controlling for children and adults

	1995	2002
Log(Income)	-0.13	-0.18
	(53.25)**	(95.44)**
Constant	1.63	2.15
	(67.40)**	(103.17)**
Adults	0.00	0.02
	(0.90)	(15.53)**
Children	-0.01	0.02
	(2.71)**	(13.50)**
Age head of household	0.00	0.00
	(7.49)**	(9.33)**
Sex head of household	0.00	0.00
	(1.43)	(1.02)
Elders	0.02	0.01
	(9.58)**	(6.10)**
Observations	14595	15707
R-squared	0.26	0.48

Table 12: Regression results robust regression, all households

The table reports two sets of estimates, one for each year. Robust t-statistics are in parentheses. The robustness check include all households, and the number of children and adults as control variables. * significant at 5%; ** significant at 1%

Urban dummy variables	1995	2002
Shanxi	-0.07	-0.16
	(9.24)**	(24.35)**
Liaoning	-0.01	-0.06
	(1.62)	(9.54)**
Jiangsu	0.00	-0.04
	(0.26)	(6.46)**
Anhui	0.01	-0.05
	(1.09)	(7.69)**
Henan	-0.04	-0.14
	(5.59)**	(21.86)**
Hubei	-0.01	-0.07
	(1.18)	(11.28)**
Guangdong	0.06	0.04
	(7.57)**	(6.73)**
Chongqing	-	-0.06
		(8.06)**
Sichuan	-0.01	-0.06
	(0.81)	(8.52)**
Yunnan	-0.01	-0.04
	(1.39)	(5.56)**
Gansu	-0.01	-0.11
	(1.26)	(15.66)**

Table 13 : Regression results robust regression - all households, urban dummy variables

The table reports two sets of dummy variable estimates, one for each year. Chongqing is included only in 2002. The table reports two sets of estimates, one for each year. Robust t-statistics are in parentheses. The robustness check include all households, and the number of children and adults as control variables. Robust t-statistics are in parentheses. * significant at 5%; ** significant at 1%

Rural dummy variables	**1995**	**2002**
Beijing	-0.09	-0.21
	(6.55)**	(18.04)**
Hebei	-0.26	-0.25
	(25.58)**	(31.14)**
Shanxi	-0.27	-0.25
	(22.11)**	(30.29)**
Liaoning	-0.22	-0.20
	(18.46)**	(24.06)**
Jilin	-0.17	-0.26
	(15.49)**	(35.96)**
Jiangsu	-0.21	-0.17
	(21.34)**	(20.99)**
Zhejiang	-0.11	-0.14
	(11.25)**	(18.38)**
Anhui	-0.24	-0.18
	(23.25)**	(22.21)**
Jiangxi	-0.22	-0.16
	(21.03)**	(19.85)**
Shandong	-0.21	-0.20
	(22.58)**	(26.61)**
Henan	-0.34	-0.25
	(35.51)**	(30.21)**
Hubei	-0.30	-0.19
	(28.71)**	(22.56)**
Hunan	-0.23	-0.17
	(24.18)**	(21.64)**
Guangdong	-0.08	-0.07
	(7.88)**	(9.62)**
Guangxi	-	-0.19
		(22.40)**
Chongqing	-	-0.15
		(14.66)**
Sichuan	-0.30	-0.13
	(30.42)**	(15.77)**
Guizhou	-0.34	-0.18
	(25.39)**	(20.76)**
Yunnan	-0.20	-0.07
	(15.52)**	(6.37)**
Shaanxi	-0.36	-0.29
	(29.94)**	(32.14)**
Gansu	-0.36	-0.21
	(25.84)**	(21.09)**
Xinjiang	-	-0.15
		(17.26)**

Table 14: Regression results robust regression –all households, rural dummy variables

The table reports two sets of dummy variable estimates, one for each year. Guangxi, Chongqing, Xinjiang is included only in 2002. The robustness check include all households, and the number of children and adults as control variables. Robust t-statistics are in parentheses. * significant at 5%; ** significant at 1%

Urban	SPI 95	SPI 02
Anhui	1.97	1.41
Beijing	1.84	1.88
Chongqing	-	1.35
Gansu	1.70	1.01
Guangdong	2.84	2.38
Henan	1.33	0.86
Hubei	1.73	1.27
Jiangsu	1.87	1.48
Liaoning	1.69	1.35
Shanxi	1.09	0.77
Sichuan	1.77	1.38
Yunnan	1.70	1.54
Rural		
Anhui	0.27	0.70
Beijing	0.91	0.60
Chongqing	-	0.84
Gansu	0.11	0.59
Guangdong	0.98	1.26
Guangxi	-	0.68
Guizhou	0.12	0.71
Hebei	0.24	0.47
Henan	0.13	0.48
Hubei	0.17	0.68
Hunan	0.30	0.76
Jiangsu	0.35	0.75
Jiangxi	0.32	0.81
Jilin	0.49	0.44
Liaoning	0.33	0.61
Shaanxi	0.11	0.39
Shandong	0.36	0.62
Shanxi	0.22	0.47
Sichuan	0.17	0.95
Xinjiang	-	0.84
Yunnan	0.39	1.30
Zhejiang	0.77	0.85
Mean (national)	0.88	0.95
Mean (urban)	1.77	1.39
Mean (rural)	0.35	0.72
Coefficient of variation (national)	87.77	47.28
Coefficient of variation (urban)	24.35	31.25
Coefficient of variation (rural)	73.43	32.73

Table 15 : Spatial Price Index rural/ urban provinces robust regression, all households

The table provides two sets of spatial price level estimates, one for 1995 and one for 2002. The price level estimates have furthermore been split in an urban and a rural section. The province price level estimates are presented alphabetically. A summary of the average price level and variation for the rural and urban SPIs is provided below.

F All households, more control variables

	1995	2002
Log(Income)	-0.12	-0.18
	(50.32)**	(92.15)**
Constant	1.62	2.18
	(65.04)**	(100.07)**
Age head of household	0.00	0.00
	(2.15)*	(1.46)
Age spouse	0.00	0.00
	(0.00)	(1.07)
Children	-0.01	0.02
	(3.01)**	(13.68)**
Adults	0.00	0.02
	(0.50)	(14.31)**
Elders	0.02	0.01
	(9.07)**	(5.62)**
SexHH	-0.01	0.00
	(1.84)	(0.24)
Share of females in the household	-0.02	-0.01
	(2.45)*	(2.19)*
Education (medio)	0.00	-0.01
	(0.74)	(3.23)**
Education (high)	-0.05	-0.02
	(9.96)**	(4.66)**
Observations	14194	15280
R-squared	0.27	0.48

Table16: Regression results robust regression - all households, including more control variables

The table reports two sets of estimates, one for each year. Robust t-statistics are in parentheses. * significant at 5%; ** significant at 1%

Urban dummy variables	1995	2002
Shanxi	-0.07	-0.16
	(9.23)**	(23.53)**
Liaoning	-0.01	-0.06
	(2.01)*	(9.44)**
Jiangsu	0.00	-0.04
	-0.22	(6.36)**
Anhui	0.00	-0.05
	-0.38	(7.54)**
Henan	-0.05	-0.14
	(5.98)**	(21.03)**
Hubei	-0.01	-0.07
	-1.85	(11.24)**
Guangdong	0.05	0.04
	(6.84)**	(5.97)**
Chongqing	-	-0.06
		(7.70)**
Sichuan	-0.01	-0.06
	-1.25	(8.66)**
Yunnan	-0.01	-0.04
	-1.77	(5.27)**
Gansu	-0.01	-0.11
	-1.12	(14.86)**

Table 17 : Regression results robust regression - all households, more control variables, urban dummy variables

The table reports two sets of dummy variable estimates, one for each year. Chongqing is included only in 2002. Robust t-statistics are in parentheses. * significant at 5%; ** significant at 1%

Rural dummy variables	1995	2002
Beijing	-0.10	-0.22
	(7.39)**	(18.42)**
Hebei	-0.27	-0.26
	(25.96)**	(31.21)**
Shanxi	-0.28	-0.26
	(22.58)**	(30.60)**
Liaoning	-0.22	-0.21
	(18.87)**	(24.43)**
Jilin	-0.18	-0.27
	(16.19)**	(35.97)**
Jiangsu	-0.22	-0.18
	(21.87)**	(21.37)**
Zhejiang	-0.12	-0.15
	(12.24)**	(18.79)**
Anhui	-0.25	-0.19
	(23.62)**	(22.69)**
Jiangxi	-0.23	-0.16
	(21.48)**	(20.35)**
Shandong	-0.21	-0.21
	(23.17)**	(26.83)**
Henan	-0.34	-0.26
	(35.79)**	(30.36)**
Hubei	-0.31	-0.19
	(29.22)**	(22.90)**
Hunan	-0.24	-0.17
	(24.49)**	(22.10)**
Guangdong	-0.09	-0.08
	(8.84)**	(10.19)**
Guangxi	-	-0.19
		(22.88)**
Chongqing	-	-0.16
		(15.33)**
Sichuan	-0.31	-0.13
	(30.42)**	(16.48)**
Guizhou	-0.35	-0.19
	(25.57)**	(21.32)**
Yunnan	-0.20	-0.08
	(15.81)**	(7.03)**
Shaanxi	-0.37	-0.29
	(30.34)**	(32.41)**
Gansu	-	-0.22
		(21.48)**
Xinjiang	-0.36	-0.16
	(25.97)**	(17.96)**

Table 18: Regression results robust regression - all households and more control variables, rural dummy variables

The table reports two sets of dummy variable estimates, one for each year. Guangxi, Chongqing, Xinjiang is included only in 2002. Robust t-statistics are in parentheses. * significant at 5%; ** significant at 1%.

Urban	SPI 95	SPI 02
Anhui	1.97	1.42
Beijing	1.92	1.91
Chongqing		1.38
Gansu	1.78	1.05
Guangdong	2.89	2.37
Henan	1.33	0.87
Hubei	1.73	1.27
Jiangsu	1.90	1.50
Liaoning	1.71	1.36
Shanxi	1.11	0.79
Sichuan	1.79	1.39
Yunnan	1.72	1.57
Rural		
Anhui	0.25	0.39
Beijing	0.84	0.44
Chongqing	-	0.46
Gansu	0.10	0.47
Guangdong	0.91	0.47
Guangxi	-	0.58
Guizhou	0.11	0.59
Hebei	0.22	0.60
Henan	0.12	0.62
Hubei	0.15	0.67
Hunan	0.27	0.67
Jiangsu	0.32	0.68
Jiangxi	0.30	0.69
Jilin	0.45	0.74
Liaoning	0.31	0.75
Shaanxi	0.10	0.79
Shandong	0.34	0.82
Shanxi	0.20	0.82
Sichuan	0.16	0.84
Xinjiang	-	0.92
Yunnan	0.37	1.24
Zhejiang	0.72	1.27
Mean (national)	0.87	0.95
Mean (urban)	1.80	1.41
Mean (rural)	0.33	0.71
Coefficient of variation (national)	91.04	32.61
Coefficient of variation (urban)	24.39	30.49
Coefficient of variation (rural)	74.08	32.41

Table 19 : Spatial Price Index rural/ urban provinces all households and more control variables
The table provides two sets of spatial price level estimates, one for 1995 and one for 2002. The price level estimates have furthermore been split in an urban and a rural section. The province price level estimates are presented alphabetically. A summary of the average price level and variation for the rural and urban SPIs is provided below.