The Inequality-Growth Nexus in the Short- and Long- Run: Empirical Evidence from China

Guanghua Wan

UNU-WIDER, Finland

Ming Lu and Zhao Chen

Fudan University, China

Abstract

This paper argues that the conventional approach of data averaging is problematic for exploring the growth-inequality nexus. It introduces the polynomial inverse lag (PIL) framework so that the impacts of inequality on investment, education and ultimately on growth can be measured at precisely-defined time lags. Combining PIL with simultaneous systems of equations, we analyze the growth-inequality relationship in post-reform China, finding that this relationship is nonlinear and is negative irrespective of time horizons.

The Inequality-Growth Nexus in the Short- and Long- Run:

Empirical Evidence from China

1. Introduction

The literature on the relationship between inequality and growth is large and still growing. Yet, theoretical and empirical evidences are mixed (Banerjee and Duflo 2003, and references therein). Typically, cross-section regressions yield a negative relationship (Benabou 1996b) while the contrary is found using panel data models with fixed effect (Forbes 2000, Li and Zou 1998). In addition, Barro (2000) indicates that the relationship is non-significant when rich and poor countries are pooled together.

As asserted by Forbes (2000), the changes of sign in the inequality-growth relationship can be explained by the difference in the time horizon considered. She concludes that in the long-run the relationship is negative while it is positive in the short- or medium- run. This assertion is supported by Banerjee and Duflo (2003). However, neither of these studies considers short-, medium- and long- run relationships in one unified framework. In fact, no previous attempts have been made to incorporate very short-run effects into growth regressions[†].

Can this assertion or conclusion be used to reconcile the mixed empirical results? Answering this question is not only helpful in settling the intensive debate among academics. It is also vital for policy-makers. If the assertion is valid, possibilities exist for inter-temporal trade-offs of growth by manipulating income distributions. Otherwise, a low (high) inequality must be targeted in order to achieve growth if the short- medium- or long- run relationship is negative (positive).

The conventional approach to discovering the long-run vs short-run relationship is by averaging growth over different time horizons. For example, Forbes (2000) uses data

[†] By “very short-run”, we mean instant impacts without any delay or with delay by one time period.

averaged over a 5-year interval in a growth regression and claims that this is a medium- or short- run relationship. Subsequently, she also reports results using 10-year averages, which show an insignificant relationship. Based on these, Forbes asserts that the short-run positive relationship may not contradict the long-run negative relationship. She seems to imply that if a longer time horizon, say 20-year, is considered for averaging data, the relationship may become negative. Meanwhile, Barro (2000) relies on averages over a 10-year interval to estimate long-run relationships.

This practice of averaging is questionable on a number of grounds. First, no consensus exists regarding what time horizon defines the short-, medium- or long- run concepts. For example, a five-year interval can be considered as short-run by some and medium-run by others. Further, if conflicting results are obtained with a 20-year and a 25-year averaging, can one attribute these to medium- and long- run differences? What happens if 5- and 6-year averaging gives rise to different results? It is important to note that if the true relationship does involve a change in sign at all, there must be a point where such a change occurs (say from period t to period $t+1$). In this case, one can state that the relationship is positive (or negative) over time horizon t periods and becomes negative (or positive) over the time horizon $t+1$ periods. An appropriate approach should enable identification of this turning point or possibly multiple turning points. In this regard, the conventional averaging procedure is problematic, if not inapplicable at all.

Second, averaging data is usually justified on the ground that it takes away business cycle effects on growth[‡]. However, business cycles differ in length for the same economy over

[‡] Another argument against using annual data is that they are subject to shocks and may cloud the underlying true relationship. This argument seems untenable given the inclusion of the disturbance term in any econometric equations, which could accommodate shocks and other errors. In passing, it is noted that Barro (2000) opt to use averaged data but for different reasons, namely unavailability of high frequency data for some variables and the inability of current theories in establishing very short-run associations between growth and its determinants. Nevertheless, the inability should not prevent one from modeling empirical short-run relationships. After all, the medium- or long- run relationships are built on the short-run counterpart. The former do not exist without the latter.

time and for different economies. They start and end at different time points for different economies as well. Simply applying one time interval in averaging data, for one country over time or for different countries, may not help eliminate the cycle effects. In other words, taking averages is useful only when business cycles are properly identified. In this case, the cycles must be completely synchronized among different economies under consideration and they must be of precisely the same length over time. These are unlikely to be true even if difficulties in business cycle identification can be left aside.

Third, short-, medium- and long- run relationships between inequality and growth are different aspects of a same underlying economic or growth process, which corresponds to a particular data generating process (DGP). A DGP is exactly what an econometric model intends to capture or describe. When estimating different regressions, forced by arbitrarily chosen different time-intervals for averaging data, one might model different DGPs rather than different aspects of the same DGP. From this perspective, the changes in sign may not reflect difference between the long- and short- run. It may be caused by the use of different averages and by other differences inherent in different regression models.

Fourth, as pointed out by Attanasio *et al.* (2000), annual data provide information that is lost when averaging. This practice is particularly puzzling as paucity of data is often cited as a major hurdle in estimating growth regressions (Durlauf 2001). It can be easily ascertained that with 5 or 10 year-averaging, 80 or 90 per cent of sample observations are lost. Finally, it is illogical to make short-, medium- and long- run mutually exclusive as far as model specification is concerned. To explain, these different “runs” are embedded in a common DGP or common regression equation. In any case, it is desirable to develop a framework which allows for identification of the growth-inequality relationship over all possible time horizons. One can then discuss findings with precise definition of time intervals. Under this circumstance, results from different studies can be compared even if data used are of different frequencies. For example, one does not have to stick to 5-year averages in order to compare her/his results with Forbes (2000).

The main purpose of this paper is to introduce such a modeling framework which enables identification of the short-, medium- and long- run effects in one model. A second purpose is to extend Barro (2000) and Lundberg and Squire (2003) by adding important equations and by combining the simultaneous model with the newly introduced framework. In particular, education is endogenized in this paper, possibly for the first time in the inequality-growth literature. This extension is justified both theoretically and empirically. Finally, we use annual data from within China to explore the inequality-growth nexus.

2. Theories on the Inequality-Growth Nexus and the Modeling Framework

Several mechanisms are theorized to yield negative effects of inequality on growth. First, under imperfect capital market, higher inequality means more individuals facing credit constraints. Consequently, they cannot carry out productive investments in physical or human capital (Galor and Zeira 1993, Fishman and Simhon 2002). These can take place in the short-run or long-run. Second, a worsening inequality generates a rise in the fertility rate among, and less investment in human capital of, the poor (De la Croix and Doepke 2004). This is mostly likely to happen in the long run. Third, a more unequal income distribution causes weaker domestic demand which may slow down the economy, as occurring in China since late 1990s. This demand-related impact is expected to prevail mostly in the short-run. Fourth, growing inequality increases redistributive tax pressures, which deters investment incentives (Alesina and Rodrik 1994, Persson and Tabellini 1994, Benabou 1996b). Finally, a worsening inequality may lead to a more unstable socio-political environment for economic activities (Benhabib and Rustichini 1996). The last two mechanisms require certain time duration for the effects to materialize.

On the other hand, Galor and Tsiddon (1997a, 1997b) develop two theories predicting a positive inequality-growth relationship. Also, Benabou (1996a) shows that when human capitals of heterogeneous individuals are strongly complementary within localities, more inequality is inductive to growth, at least in the short run. In addition, a high or rising

inequality prompts the middle-class to vote for changes in taxation rate. Both higher (Saint-Paul and Verdier 1993) and lower taxation rates (Li and Zou 1998) could promote economic growth. Finally, conventional wisdom states that high inequality implies more savings or more investment (Galor and Omer 2002). All these positive effects can materialize in the short- or long- run.

Clearly, these theories indicate that the overall impact of inequality on growth cannot be set a priori (Aghion *et al.* 1999). More pertinent to this paper, the short- and long- run effects may well differ in magnitudes as well as in signs. As noted above, the very short-run effect is so far ignored in the empirical literature, despite its existence and importance. Even medium- and long-run effects are not modeled appropriately. It is important to point out that the existing theories implicitly or explicitly assume that inequality affects growth through its impacts on physical and human capital formation. This point is taken up later in this paper when our empirical model is specified.

To enable identification of the inequality effects over different time horizons, distributed-lag models can be used. Among the alternative distributed-lag structures, the polynomial inverse lag (PIL) of Mitchell and Speaker (1986) is preferred as it possesses two attractive features: its flexibility in uncovering the true lag structure and its easiness in estimation. The second feature is especially important as we will combine PIL with simultaneous equations.

Let Y denote growth and X denote inequality, the PIL model can be written as:

$$Y_t = b + \sum_{i=0}^{\infty} w_i X_{t-i} + e_t \quad (1)$$

where

$$w_i = \sum_{j=2}^n \frac{a_j}{(i+1)^j}, i = 0, \dots, \infty. \quad (2)$$

In the above model, w_i are the distributed lag weights, indicating the impacts of X on Y over time interval i . The notation a_j represents parameters to be estimated and n is the degree of polynomial. Substituting (2) into (1) and rearranging yield:

$$Y_t = b + \sum_{i=0}^{\infty} w_i X_{t-i} + e_t = b + \sum_{j=2}^n \sum_{i=0}^{m-1} \frac{a_j}{(i+1)^j} X_{t-i} + \underbrace{\sum_{j=2}^n \sum_{i=m}^{\infty} \frac{a_j}{(i+1)^j} X_{t-i}} + \Lambda + e_t. \quad (3)$$

The underlined term on the right hand side of (3) becomes negligible for t greater than eight thus can be omitted, as suggested by Mitchell and Specker (1986). Based on (3), one can obtain the effects of X on Y over any time interval, such as five or eight years.

The instant impact is given by $w_0 = \sum_{j=2}^n a_j$, the lagged impacts are given by w_i , ($i = 1, 2, \dots, \infty$), and the cumulative impacts are given by $\sum_i w_i$, depending on how short- and long-run are defined. In particular, we can use the infinite sum to indicate the very long-run impact.

An expended version of (3) with the underlined term omitted is:

$$\begin{aligned} Y_t = & b + a_2 \left[X_t + \frac{1}{4} X_{t-1} + \frac{1}{9} X_{t-2} + \Lambda + \frac{1}{m^2} X_{t-m+1} \right] + \\ & a_3 \left[X_t + \frac{1}{2^3} X_{t-1} + \frac{1}{3^3} X_{t-2} + \Lambda + \frac{1}{m^3} X_{t-m+1} \right] + \\ & \dots + a_n \left[X_t + \frac{1}{2^n} X_{t-1} + \frac{1}{3^n} X_{t-2} + \Lambda + \frac{1}{m^n} X_{t-m+1} \right]. \end{aligned} \quad (4)$$

Expressions in the square brackets are PIL terms associated with different degrees of polynomial n .

One can set $m = 9$, add variables other than X s in (4), and use the resultant regression to analyze the inequality-growth relationship. However, the issues of heterogeneity, measurement errors and endogeneity have received considerable attention in the literature. These must be addressed (Durlauf *et al.* 2004, Atkinson and Brandolini 2001) before

empirical estimation. In particular, Banerjee and Duflo (2003) argue why cross-country data are deficient due to differences in cultural structure, technology level and financial institutions. While not claiming absence of heterogeneity, this problem is less severe in this paper because data from within China will be used. More importantly, China remains a socialist country with strong institutional, cultural, political and even economic controls across regions. Despite so, some dummy variables will be incorporated into our empirical model to further address the heterogeneity issue.

Regarding measurement errors, this is largely related to the inequality variable; not or less applicable to other variables (Barro 2000). To be more precise, inequality data used in most cross-country regressions are calculated under different concepts of income (GDP, wage, disposable income or expenditure), different income recipients (individual, household or family) and different sampling procedures (proportional sampling, stratified sampling) or even different coverage of population (national, sub-national, regional or small-scale survey). In this paper, the regional urban-rural income ratio will be used to measure inequality. Both rural and urban income data are based on household surveys conducted by the National Statistical Bureau (NSB) of China. Therefore, we do not consider measurement errors as a major problem, at least insofar as variable definitions, population coverage and sampling techniques are concerned. Using the urban-rural income ratio as inequality indicator is justified on the ground that the urban-rural income gap constitutes over 70 per cent of overall regional inequality (Kanbur and Zhang 2005). And, no regional inequality data are available to us. Wei and Wu (1999) adopt the same practice. Bouguignon and Morrison (1998) find that the urban-rural labor productivity ratio is highly correlated with overall inequality.

The endogeneity problem is resolved by specifying and estimating simultaneous systems of equations, not by replying on lagged variables and the GMM estimation technique in a single equation. Recall the brief review of various growth theories in Section 2 of this paper, the impact of inequality on growth is mainly channeled through its effects on physical and human capital formations[§]. Thus, it is necessary to include investment and

[§] Demand-related impact of inequality must eventually work through capital and labour inputs.

education equations in the system. Consequently, we end up with a four-equation system after adding the usual growth and inequality equations. In contrast, Barro (2000) and Lundberg and Squire (2003) did not endogenize the human capital variable in their model. It is noted that estimating the inequality equation permits testing of the controversial Kuznets hypothesis.

Using INE_{PIL} to denote the inequality terms associated with PIL (INE_{PIL} = RHS of (4) excluding b), $incm$ to denote income level lagged by one year, the systems of equations are specified as (detailed definitions of all variables are provided in the Appendix):

$$incmgr = f_1(popgr, invt, edu, gov, cpi, trade, urbangr, private, incm, incmsq, central, west)$$

$$invt = f_2(INE_{PIL}, gov, cpi, trade, urban, private, incm, incmsq, central, west)$$

$$edu = f_3(INE_{PIL}, peduexp, urban, incm, incmsq, central, west)$$

$$inequality = f_4(incmgr, trade, agrexp, urban, private, incm, incmsq, central, west)$$

The first equation in the system explains per capita income growth ($incmgr$) which is determined by population growth as a proxy of labor input, investment expressed as proportion of GDP ($invt$), and human capital defined as average years of schooling (edu). These are standard growth determinants. Following Barro (2000) and Clarke (1995), we add government expenditure as a ratio of GDP (gov) and inflation (cpi) to this equation. The former represents government interference in economic activities and the latter may capture macro-economic conditions or business cycle effects. Also controlled are openness ($trade$), urbanization ($urbangr$) and privatization ($private$) variables. The convergence literature appeals for the inclusion of the initial income level ($incm_0$). Location dummy variables for central and western provinces are used in this and all other equations to contain heterogeneity.

In specifying the investment function, the most relevant variable, besides inequality (*INE*), is lagged per capita income (*incm*) as a proxy of savings plus its square (*incmsq*). As with the growth equation, government interference (*gov*) and macroeconomic conditions (*cpi*) are important independent variables. Little is necessary to justify the inclusions of openness, urbanization and privatization the investment model.

Although various growth theories indicate that inequality matters for human capital formation, no earlier attempts were made to specify the education equation. It is understandable, however, to consider lagged income ($INCM_{PIL}$). Also, education is likely to be affected by government spending on education, culture and health (*peduexp*). Needless to say, more urbanized regions enjoy better education thus the urbanization variable is relevant.

Leaving income growth and location dummy variables aside, five other variables are included in the inequality model. The Kuznets hypothesis dictates that the income variable and its square ought to be considered. Privatization is included as it is commonly perceived to be a cause of inequality in China. On the other hand, openness and urbanization are included as Wan et al. (2005) and Lu and Chen (2004) respectively find that they contribute to regional inequality. Given that the inequality variable is defined as the urban-rural income ratio, government support to agriculture (*agexp*) is expected to help narrow the urban-rural income gap.

3. Empirical Evidence from China

China represents a very interesting case for studying the inequality-growth relationship. Except the urban-rural disparity, pre-reform China was basically an egalitarian society. The low inequality was identified as a strain on economic growth. This is why Deng Xiaoping, at the onset of economic reforms, famously stated that: let some get rich first. The reform period has seen remarkable growth accompanied by fast rising inequality, not only in terms of rural-urban income gap. This growth is preceded by fairly low initial

inequality in the pre-reform period. From this perspective, the inequality-growth relationship seems to be negative. However, the Chinese experience depicts a positive correlation when pre- and post- reform periods are examined separately.

There is more. In early 2004, the Primer of China announced a growth target of 7% which is lower than any of the growth rates in China since economic reform began in 1978 (excluding the unusual period from 1989 to 1990). Such a move is unprecedented and represents a major policy shift to address, at least partly, the inequality problem in China. The high and rising inequality is perceived to hurt the national economy from the perspective of slacking domestic demand and political instability. Directing resources to the rural sector and non-coastal regions is expected to slow down growth in the short run, but may help achieve sustainable growth in the long run. Clearly, policy-makers in China, past and present, see both (short-run) positive and (long-run) negative effects of inequality on growth.

These observations appeal for a proper analytical study. Towards this end, data at the regional or provincial level for 1987-2001 are used to estimate the systems of equations outlined in Section 2. Though desirable, earlier data are too incomplete to be useful. Excluding Taiwan, Hong Kong and Macao, China has 31 provinces or regions, including four autonomous municipal cities. Hainan province was created in 1988 and is merged with Guangdong. Congqing is the youngest region in China. Fortunately, data for Congqing are available. Most data for Xizang (Tibet) are missing. Therefore, our sample consists of 29 regions. All observations in value terms are deflated by rural and urban CPIs respectively. For details on data sources and data construction, see the Appendix.

The systems of equations are estimated with three stage least squares after setting $m = 9$. To determine the degree of polynomial n , the general-to-specific approach is followed. This approach is also recommended by Mitchell and Speaker (1986). We started with $n = 6$, where high collinearity leads to automatic drops of some PIL terms by STATA. When n is reduced to five, the PIL term in the investment equation is insignificant. Once this

term is removed, all PIL terms are significant at the 10% level. The estimation results are reported in Table 1.

[Table 1 here]

The estimated models are of good quality with most parameters significantly different from zero. Notwithstanding that little can be said a priori about the signs of many estimates, the positive and significant impacts of physical and human capital investments, trade, urbanization and privatization on growth are consistent with economic theory. Government expenditure is found to be detrimental to growth (when investment is held constant) but helpful in increasing investment. These are acceptable since this variable is included as proxy for government intervention, particularly in bank lending. See Clarke (1995) and Partridge (1997). As far as the education equation is concerned, higher income is found to cause more human capital formation and urbanization is positively related to regional education level; both findings corroborate well with normal expectations.

One interesting finding relates to the income terms in the growth equation. It shows that growth does not depend on income levels in China, at any conventional levels of statistical significance. This is different from Barro (2000) who shows that the growth-inequality relationship is conditional on the level of development; it is positive across developed economies and negative in the developing world. It is noted, however, that inequality does not enter the growth equation directly in our model. Another income-related finding is that the Kuznets hypothesis is rejected. On the contrary, a U-pattern is supported by Chinese data, which is in line with Wan (2004).

Some comments on the inequality equation are in order. As indicated by the coefficients of the location dummy variables, the urban-rural divide is more severe in western than in central regions, which in turn is more severe than in coastal regions. This is understandable as urban China is more equal across locations while development in rural China is heavily reliant on geographic conditions. When everything else is the same, the

rural west usually lags behind with the east leading. Also, Table 1 indicates that in addition to the variables of government support to agriculture and income growth, privatization helps to reduce the rural-urban income gap. This is justified because TVEs in China, a major component of the privatization index, represent an important driving force in narrowing down the rural-urban gap, although they may contribute to growing inequality among rural regions (Wan and Zhu 2005). Consistent with Wan *et al.* (2005), trade is an inequality-increasing variable.

Now, attention is turned to the focal issue of this paper: how does inequality affect growth? Since the impact is channeled through investment and education, we first examine the relationship between inequality and these two factors. Referring to equation (1), the marginal effects of inequality are given by w_i . These are shown in Figure 1(a). In particular, the instant impact is given by w_0 , which is negative in the investment equation but positive in the education model. The impact of inequality on investment turns to be positive after one year and remains so for a number of years. It reverts to be negative after four years and reaches the negative peak in year six, before eventually converging to zero. On the other hand, inequality seems more beneficial to human capital formation over all time horizons except in years three and four. The positive effect reaches a peak in year seven and then converges to zero.

[Figure 1 here]

It is also useful to sum w_i to obtain the cumulative effects over different time horizons. These are plotted in Figure 1(b), which demonstrates that inequality is detrimental to investment no matter what time interval is considered. On the contrary, inequality always promotes human capital formation. Given the alternative theories predicting opposite effects, it seems unnecessary to try to justify these two findings. However, the first finding must be related to the lack of investment skills and various deficiencies in the Chinese financial system. The second finding must be related to the discrimination of secondary and higher education against rural residents. When the rural-urban gap decreases implying higher rural income, lack of education facilities in the countryside

prevents more youth from gaining education. When the gap expands implying higher urban income, more families in cities can afford to access readily available education. These match well the prediction of Perotti (1992), who concludes that rising inequality enables the rich to obtain education first when tuition is high relative to income, as in China.

Once the impacts of inequality on education and investment are identified, it is straightforward to simulate the inequality-growth relationship by allowing inequality to increase at a certain margin or percentage. Figure 2 shows the instant or lagged as well as cumulative impacts on growth when China's urban-rural income ratio is raised by 0.1 unit. The instant and lagged effects fast decrease to zero after an initially negative and then positive influences in the first four years. The cumulative line demonstrates a negative relationship between inequality and growth and this relationship holds no matter what time horizon is considered. Most interestingly, the relationship is found to be nonlinear, a key point underlying the theory proposed by Banerjee and Duflo (2003).

[Figure 2 here]

4. Summary

In this paper, we introduce the polynomial inverse lag (PIL) model in order to accommodate, within one unified framework, potentially differing impacts of inequality on growth over different time horizons. Applying simultaneous equations incorporating the PIL to data from one country, namely China, our results are expected to suffer less from the problems of heterogeneity, endogeneity and measurement errors, commonly encountered in cross-country growth regressions.

Despite the seemingly positive association between growth and inequality in post-reform China, our empirical results unequivocally point to the negative effects of inequality on growths in the short-, medium-, and long- run. The negative effects stem from the strong and negative influence of inequality on physical investment, which consistently

overweight the mostly positive impacts of inequality on human capital. The inequality-growth relationship is found to be nonlinear, so are the inequality-investment and inequality-education relationships.

As with any other study, this paper can be improved along many dimensions such as data quality, model refinements and better estimation techniques. One particular avenue for future research lies in the development of bootstrapping or other tools in order to attach statistical significance to the identified effects of inequality on growth. Another issue which is yet to be dealt is to conduct robustness test of our research findings. This could be difficult given the open-ended nature of growth theories (Brock and Durlauf 2001).

References

- Aghion, Philippe, Caroli, Eve and Garcia-Penalosa, Cecilia, 1999, "Inequality and Economic Growth: the Perspective of the New Growth Theories," *Journal of Economic Literature*, 37, 4, December, 1615-1660.
- Alesina, Alberto, and Dani Rodrik, 1994, "Distributive Politics and Economic Growth," *Quarterly Journal of Economics*, 109, 2, 465-490.
- Atkinson, A.B., and A. Brandolini, 2001, "Promise and Pitfalls in the Use of 'Secondary' Data-Sets: Income Inequality in OECD Countries as a Case Study," *Journal of Economic Literature*, 39, 3, 771-799.
- Attanasio, O.P., L. Picci and A. Scorcu, 2000, "Saving, Growth and Investment: a Macroeconomic Analysis Using a Panel of Countries", *Review of Economics and Statistics*, 82(2): 182-211.
- Banerjee, A. V. and E. Duflo, 2003, "Inequality and Growth: What Can the Data Say?", *Journal of Economic Growth*, 8: 267-299.
- Barro, Robert J., 2000, "Inequality and Growth in a Panel of Countries," *Journal of Economic Growth*, 5, 1, March, 87-120.
- Benabou, R., 1996a, "Heterogeneity, Stratification, and Growth: Macroeconomic Implications of Community Structure and School Finance", *American Economic Review* 86(3): 584-609.
- Benabou, R., 1996b, "Inequality and Growth," in Bernanke, B.S. and Rotemberg, J.J. (eds.) *NBER macroeconomics manual*. Cambridge, MA: MIT Press.
- Benhabib, J. and A. Rustichini, 1996, "Social Conflict and Growth," *Journal of Economic Growth*, 1, 1, 129-146.
- Brock, William A and Durlauf, Steven N, 2004, "Growth Empirics and Reality", *The World Bank Economic Review* 15(2): 229-272.
- Bouguignon, F. and C. Morrisson, 1998, "Inequality and Development: the Role of Dualism," *Journal of Development Economics*, 57, 233-257.
- Clarke, G. R. G., 1995, "More Evidence on Income Distribution and Growth," *Journal of Development Economics*, 47, 2, 403-427.

- Durlauf, Steven N, Paul A. Johnson and Jonathan R. W. Temple, 2004, Growth Econometrics, The University of Wisconsin working paper.
- Durlauf, Steven N, 2001, "Manifesto for a Growth Econometrics", *Journal of Econometrics* 100: 65-69.
- De La Croix, David and Matthias Doepke, 2004, "Inequality and Growth: Why Differential Fertility Matters," *American Economic Review*, 93, 4, 1091-1113.
- Fishman, A. and A. Simhon, 2002, "The division of Labor, Inequality and Growth," *Journal of Economic Growth*, 7, 117-136.
- Forbes, Kristin J., 2000, "A Reassessment of the Relationship between Inequality and Growth," *American Economic Review*, 90, 4, 869-887.
- Galor, Oded and Joseph Zeira, 1993, "Income Distribution and Macroeconomics," *Review of Economic Studies*, 60, 35-52.
- Galor, Oded and Omer Moav, 2002, "From Physical to Human Capital Accumulation: Inequality in the Process of Development," Brown University Working Paper No. 99-27.
- Galor, Oded and Tsiddon, Daniel, 1997a, "The Distribution of Human Capital and Economic Growth", *Journal of Economic Growth*, 2(1): 93-124.
- Galor, Oded and Tsiddon, Daniel, 1997b, "Technological Progress, Mobility, and Economic Growth", *American Economic Review*, 87(3): 363-82.
- Kanbur, R. and X. Zhang (2005). 'Fifty Years of Regional Inequality in China: A Journey Through Central Planning, Reform and Openness', *Review of Development Economics*.
- Li, Hongyi, and Heng-fu Zou, 1998, "Income Inequality is not Harmful for Growth: Theory and Evidence," *Review of Development Economics*, 2, 3, 318-334.
- Lundberg, Mattias and Lyn Squire, 2003, "The Simultaneous Evolution of Growth and Inequality," *The Economic Journal*, 113 (April), 326-344.
- Mitchell, D. W. and P. J. Speaker, 1986, "A Simple, Flexible Distributed Lag Technique: The Polynomial Inverse Lag," *Journal of Econometrics*, 31, 329-340.
- Patridge, Mark D., 1997, "Is Inequality Harmful for Growth? Comment," *American Economic Review*, 87, 1019-1032.

- Perotti, Roberto, 1992, "Income Distribution, Politics and Growth," *American Economic Review*, Vol. 82(2), 311-317.
- Persson, Torsten, and Guido Tabellini, 1994, "Is Inequality Harmful for Growth? Theory and Evidence," *American Economic Review*, 84, 600-621.
- Saint-Paul, Gilles and Verdier, Thierry, 1993, "Education, Democracy, and Growth", *Journal of Development Economics*, 42(2): 399-407.
- Wan, Guanghua and Yuchun Zhu, 2005, "Chinese TVEs as an Informal Sector and Employment Generation", WIDER conference paper.
- Wei, Shangjin and Yi Wu, 2001, "Globalization and Inequality: Evidence from within China," NBER working paper, 8611.
- Wan(万广华), 2004, 《转型经济中的不平等和经济发展：非线性模型是否必须?》, 《世界经济文汇》, 第4期。"Inequality and Economic Development in Transition Economies: Are Nonlinear Models Needed?", *World Economic Papers*. 2004(4): 1-13.
- Wan *et al.* (万广华、陆铭、陈钊), 2005, 《全球化与地区间收入差距：来自中国的证据》, 《中国社会科学》, 第3期, 17-26。"Globalization and Regional Inequality in China", *Social Sciences in China*. 2005(3): 17-26.

Data Appendix

- (1) Unless indicated otherwise, data for the period 1987-1998 are all from *Comprehensive Statistical Data and Materials for 50 Years of New China* (NBS, 1999). Data for years 1999-2001, unless indicated otherwise, are from *China Statistical Yearbook, 2000, 2001 and 2002* (NBS, various years).
- (2) *popgr* = population growth rate. Except for Hebei, Heilongjiang and Gansu, 1999-2001 data of agricultural and non-agricultural population are from provincial statistical yearbooks. Population data of Hebei, Heilongjiang and Gansu in 2000 are from *China Statistical Yearbook, 2001*. For these three regions, the 1999 population data are the averages of the neighboring two years, and the 2001 data are forecast based on data in 2000 and the growth rate during 1999-2000.
- (3) *incm* = per capita income lagged by one year. Regional income is the weighted average of urban and rural per capita incomes, with non-agricultural and agricultural population shares as weights. Both urban and rural incomes are deflated by regional urban and rural CPIs. For Shanghai, Beijing and Tianjin, urban and rural CPIs are the same.
- (4) *Incmsq* = *incm* squared.
- (5) *urgap* = urban-rural income gap. It is defined as urban-rural per capita income ratio.
- (6) *incmgr* = income growth rate. It is calculated based on *Incmsq*.
- (7) *invmtgdp* = Investment/GDP ratio. It is computed as total fixed capital investment over GDP.
- (8) *edu* = Education. *China Population Yearbooks* report regional population by education attainment as from 1987. Unfortunately, such data were not published for 1989, 1991 and 1992, and data for 1987 and 1988 are incomplete as illiterate population are not reported. Also, unlike data for other years, the 1994 data did not consider population below the age of 15. To estimate data for these years, we compute average years of schooling using data for the other years and then fit the model:

$$\ln(edu) = f(\cdot) + \mu,$$

where edu is per capita years of schooling, $f(\cdot)$ is simply a linear function of time trend and regional dummies, μ the error term. This model is estimated by GLS technique, allowing for heteroskedasticity in the panel data. The R^2 of the estimated equation is 0.966. Denote the predicted value by $\hat{\cdot}$, we have:

$$\hat{edu} = \exp[\ln(\hat{edu})] \exp(0.5 \hat{\sigma}^2),$$

where $\ln(\hat{edu})$ denotes the predicted values of $\ln(edu)$ and $\hat{\sigma}^2$ is the estimated variance of μ . Data for 1987-89, 1991, 1992 and 1994 are estimated by the above model.

- (9) *gov* = governmental consumption ratio, exclusive of expenditure on culture, education, science & health care. Unlike in the existing literature we cannot exclude education and defense expenditures as these sub statistics are not available at the regional/provincial level.
- (10) *Trade* is computed as the trade/GDP ratio. Trade data are converted into RMB.
- (11) *cpi* are used to proxy inflation. CPIs of Qinghai are from provincial statistical yearbook.
- (12) *agexp* = proportion of provincial fiscal expenditure on agriculture.
- (13) *peduexp* = per capita government expenditure on culture, education science and health care.
- (14) *Private* = Privatization, computed as the proportion of workers and staff in non-state-owned entities.
- (15) *urban* = Urbanization, defined as the proportion of non-agricultural population in the total.
- (16) *urbangr* = growth rate of *urban*.
- (17) *center, west*: Location dummies for central and western China, respectively.

Consistent with most of the literature, central provinces refer to *Shanxi, Inner Mongolia, Jilin, Heilongjiang, Anhui, Jiangxi, Henan, Hubei, Hunan* and *Guangxi*, and western provinces include *Sichuan, Guizhou, Yunnan, Shaanxi, Gansu, Qinghai, Ningxia*, and *Xinjiang*.

Figure 1. Impact of inequality on investment (dashed line) and education (solid line).

Figure 2. Inequality-growth nexus: instant and lagged effects (dashed line), and cumulative effects (solid line)

Table 1. Estimation Results

Right-hand side variables	Equation									
	Growth		Investment			Education		Inequality		
	Coefficient	T-ratio	Coefficient	T-ratio		Coefficient	T-ratio	Coefficient	T-ratio	
PIL (n=2)			-0.285	-1.707	*	0.137	1.930	*		
PIL (n=3)			1.563	2.312	**	-1.071	-1.818	*		
PIL (n=4)			-1.351	-2.598	***	2.287	1.716	*		
PIL (n=5)						-1.351	-1.652	*		
peduexp						0.0006	0.300			
agrex								-0.041	3.727	***
incmgr								-0.043	3.071	***
popgr	-0.068	-1.015								
inv	1.58	7.215								
edu	3.561	3.028								
gov	-1.529	-6.141	1.16	7.733	***					
cpi	-0.442	-4.604	0.262	3.011	***					
urbangr	0.375	5.282								
urban			-0.073	-1.352		0.042	8.400	***	-0.013	4.333
trade	0.064	2.667	-0.025	-1.136					0.007	7.000
private	0.805	6.765	-0.416	-5.012	***				-0.02	3.333
incm	-53.468	-1.308	6.629	0.161		7.399	1.735	*	-12.322	4.444
incmsq	0.698	0.234	1.058	0.352		-0.476	-2.817		0.874	4.348
central	7.956	3.716	-5.48	-3.914	***	0.282	2.541	**	0.272	2.989
west	-5.557	-2.621	4.833	2.271	**	-0.403	-2.385	**	0.76	6.129
constant	292.929	2.003	-76.008	-0.524		-22.97	-1.581		46.571	4.782

***, **, * = significant at the 1%, 5% and 10% levels of significance.