REGIONAL INCOME DIVERGENCE IN CHINA:
A NON-STATIONARY PANEL APPROACH

by

JEREMY WERTZER, AUTHOR

Professor Peter Pedroni, Advisor

A thesis submitted in partial fulfillment
of the requirements for the
Degree of Bachelor of Arts with Honors
in Economics

WILLIAMS COLLEGE

Williamstown, Massachusetts

5/19/2006

Abstract

Using newly developed econometric techniques, Pedroni and Yao (2006) show that China has been experiencing regional income divergence since the Reform Period, and that traditional explanations explaining divergence are insufficient. This paper considers the role of labor mobility in the context of endogenous growth with positive externalities to human capital. Using provincial level data on income, migration and population, non-stationary panel techniques are applied to test for regional convergence conditional on migration rates.

Acknowledgements

First and foremost, I want to give my deepest thanks to Professor Peter Pedroni for the enormous amount of time and energy he devoted to helping me with this project. He spent many, many hours making sure I learned the econometric techniques in time to complete my thesis. I also want to thank Professors Alan de Brauw, Jon Bakija, and Douglas Gollin for their helpful comments and suggestions throughout the process. And finally, thanks to my family and friends for their unconditional support.

Table of Contents

Introduction

Recent empirical studies suggest that China has experienced regional income divergence since reforms began in the late 1970s. In most instances, the per capita incomes of regions within a country have been shown to converge (for example, the U.S. states, and Japanese prefectures)[1]. Therefore, China's regional divergence, and particularly its exacerbation in the presence of more liberal, market-oriented reforms is perplexing. Various hypotheses have been extended to account for this phenomenon, including geographical factors and preferential government treatment towards certain regions. However, using recently developed empirical techniques, Pedroni and Yao (2006) conclude that neither of these explanations is sufficient and that other possible causes should be examined.

Razin and Yuen (1997) suggest that, in the context of positive externalities to human capital, income level convergence depends on labor mobility. Moreover, labor mobility in China could take on added importance due to the constraints on capital that have existed even throughout the reform period. In this paper, I expand on the analysis of Pedroni and Yao in two primary ways. First, I use data that has been updated through 2004 as opposed to 1997, which allows us to account for the more recent developments that have occurred in China, particularly continued liberalization of factor markets, increased volume of international trade due to accession into the World Trade Organization, and persistent expansion of the private sector. Second, and more importantly, I seek to control for labor mobility in testing for regional convergence. This avenue of research is suggested by Pedroni and Yao,

[1] Barro and Sala-I-Martin (1995), pp. 387-409

as their own analysis determines that traditional explanations of convergence are insufficient.

This research is important in a number of respects. In a narrow sense, it contributes to our understanding of China's economic development over the last 30 years, as well as its future prospects. China is quickly becoming one of the most formidable economic and political entities on the global landscape. The increased regional income inequality poses a danger to China's continued economic (and by extension, political) stability.[2] Moreover, a better understanding of the impact of labor mobility on income growth is important in the context of economic unions that are considering following the EU's lead in relaxing cross-border labor mobility restrictions.[3]

Finally in the broadest sense, this research could provide a modest contribution to our understanding of economic growth. As will be discussed in Section II, modifications to theoretical growth models have partly responded to empirical findings about convergence. For example, Lucas (1988), in proposing his endogenous growth model, claims to have been motivated in part by a desire to account for the lack of observed cross-country convergence predicted by the Solow model. Barro and Sala-I-Martin (1992) state that further research should be done regarding "the effects involving mobility of capital and labor across economies" on convergence.[4] If it were shown that labor mobility plays a unique role in income

[2] Pedroni and Yao (2006), p. 24
[3] Pedroni and Yao (2006), p. 1
[4] Barro and Sala-I-Martin (1992), p. 247

convergence, this might shed some light on our theoretical understanding of growth (for example, with respect to the role of human capital).

The remainder of this thesis is divided into four main sections. The next section is a review of some of the existing literature relevant to this topic. This review will begin with a derivation of two of the basic theories of economic growth: one based on the Solow model with technological progress, and the other based on the Lucas model with positive externalities to human capital. I have included this review because it gives a theoretical context to what is primarily an empirical project, thus justifying the econometric analysis that will follow.

The relevance of the theoretical models will become apparent as the literature review progresses and I explore existing research on convergence. In this discussion, I will begin with an explanation of traditional empirical techniques used to test for convergence, and will then discuss some of the studies that have tested for convergence across and within countries. I will also examine convergence in the context of the theoretical growth models. The exogenous Solow model tends to predict convergence in a closed economy, and even more so in the presence of capital mobility. For these reasons, it might not be an adequate framework through which to explain developments in China. Alternatively, the endogenous Lucas model does not necessarily imply level convergence and therefore offers a more compelling explanation for regional income divergence. Razin and Yuen (1997) argue that labor mobility plays an important role in the latter context.

I will conclude the literature review by evaluating convergence specifically as it relates to China. As we will see, there has been some disagreement on whether

China has or has not experienced regional income divergence since the reform period. However, the most recent empirical techniques suggest that it has. I will also review the literature that exists regarding migration in China throughout the reform period.

In section two, I examine in some detail the non-stationary econometric techniques that my analysis employs. This section will include a brief overview of time series econometrics, highlighting the concept of stationarity. I will then provide a justification for using panel techniques instead of traditional cross sectional tests to test for convergence. Finally, I explain the different tests for panel unit roots and how they can be employed to test for income convergence.

Section Three explains the data that I will use in my empirical analysis. This data includes provincial GDP per capita in the period 1978 to 2004, as well as various estimates for the degree of interprovincial migration. Unfortunately, there is no panel available that explicitly documents labor migration or mobility. However, using census data from 1990 and 2000, we can get a sense of which provinces have relatively open labor markets.

In Section Four, I present my empirical findings. I begin with a test for convergence conditioning only on fixed effects. I then employ the data on migration and test for convergence conditional on labor mobility by testing clusters of provinces separately. In Section Five, I summarize my conclusions and suggest avenues for further research.

Section I: Literature Review

Solow Growth Model

In 1956, Robert Solow published his influential paper 'A Contribution to the Theory of Economic Growth,' in which he explains economic growth as arising from a combination of factor accumulation and unexplained technical progress. In addition to Solow (1956), the following derivation is based on Jones' (1998) textbook presentation of the Solow model.

The Solow Model assumes an economy with a Cobb-Douglas production function given by

$$Y = F(K,L) = K^{\alpha}L^{1-\alpha} \qquad (1.1)$$

and a capital accumulation equation given by

$$\dot{K} = sY - dK, \qquad (1.2)$$

where $\dot{K}$ is the change in capital, K is the economy's stock of capital, L is the labor force, sY is the amount of gross investment, and dK is the amount of depreciation per period. Equation (1.1) can be rewritten in per capita terms as

$$y = k^{\alpha} \qquad (1.3)$$

where y=Y/L and k=K/L. And Equation (1.2) can be rewritten as

$$\dot{k} = sy - (n + d)k, \qquad (1.4)$$

where n is the rate of growth in the labor force (Appendix Figure 1.1 illustrates these equations graphically).

Over time, this economy will approach a steady state where the amount of capital per capita remains constant, such that k = k*. At points to the left of the steady state, sy > (n+d)k, which implies that $\dot{k} > 0$ and k is increasing. At points to the right of k*, sy < (n+d)k, which implies that $\dot{k} < 0$ and k is decreasing. In either case, the economy's level of capital per worker moves toward the steady-state level k*, where sy = (n+d)k, and therefore the capital-labor ratio is stable.[5]

We can solve for k* by setting equation (1.4) equal to zero, and substituting k^{α} for y, which yields

$$k^* = \left(\frac{s}{n+d}\right)^{1/(1-\alpha)}$$

Substituting this value into equation (1.3) gives us the steady-state value of per capita output

$$y^* = \left(\frac{s}{n+d}\right)^{\alpha/(1-\alpha)} \qquad (1.5)$$

It is worth taking a moment to examine some of the implications for cross-country comparisons of income levels and growth rates implied by this model. Based on equation (1.5), if we assume that depreciation rates are the same across countries, then the model predicts that per capita income levels should differ among countries only based on differences in population growth and the savings rate (since these variables determine the level of capital per worker for each economy's steady state). Another implication of the model as it stands is that growth *rates* in per capita income in the long-run should be zero. Growth rates in income can vary, insofar as countries are outside of their steady-states, but once a country reaches k*, $\dot{k} = 0$, and therefore $\dot{y} = 0$. In sum, thus far, we can explain differences in income levels among closed

[5] Jones (1998), pp. 20-26; Solow (1956), pp. 66-70

economies through differences in capital per worker (as implied by different savings rates and population growth rates), but we cannot explain differences in per capita growth rates across countries or non-zero long-run growth rates.[6]

In light of the implausibility of zero long-run growth, Solow extended the model to include exogenous technological change, such that the production function is now expressed as

$$Y = F(K, AL) = K^{\alpha}(AL)^{1-\alpha} \qquad (1.6)$$

where A is the level of technology, which we can think of as adding to the productivity of labor.[7] We assume that technology is growing at a constant rate such that $\frac{\dot{A}}{A} = g$. Equation (1.6) expressed in per capita terms is

$$y = k^{\alpha}A^{1-\alpha}.$$

Taking logs and differentiating, we get

$$\frac{\dot{y}}{y} = \alpha\frac{\dot{k}}{k} + (1-\alpha)\frac{\dot{A}}{A}. \qquad (1.7)$$

Equation (1.2) still describes capital accumulation.

The counterpart to the steady state of our earlier model is known as 'the balanced growth path.' When an economy reaches its balanced growth path, capital and output grow at a constant rate. Dividing equation (1.2) by K, we see that $\frac{\dot{K}}{K}$ is constant if and only if $\frac{Y}{K}$ is constant. For this to be true, $\frac{y}{k}$ must be constant which

[6] Jones (1998), p. 28

[7] Solow himself used a 'Hicks-neutral' technology variable, such that Y=AF(K,L). The "labor augmenting" form used above (for simplicity purposes) is borrowed from Jones (1998).

means that y and k must grow at the same rate. Substituting this fact into equation (1.7), we find that $\frac{\dot{y}}{y} = \frac{\dot{k}}{k} = \frac{\dot{A}}{A}$ which, by definition, is equal to g.[8]

We can solve for the steady state capital and income levels for the Solow model with technology by expressing variables as their ratio to A. We define the 'capital-technology ratio' $k_e \equiv K/AL$ and rewrite the production function

$$y_e = k_e^{\alpha}, \qquad (1.8)$$

where y_e is the 'output-technology ratio.' We can rewrite the capital accumulation equation as

$$\dot{k}_e = sy_e - (n+g+d)k_e. \quad (1.9)$$

The remaining analysis is analogous to the previous model except that now we are dealing with capital-technology ratios. At points where $sy_e > (n+g+d)k_e$, $\dot{k}_e > 0$ and the capital-technology ratio will be rising. At points where $sy_e < (n+g+d)k_e$, $\dot{k}_e < 0$ and the capital-technology ratio will be falling. When $sy_e = (n+g+d)k_e$, the economy has reached its steady state value k_e^* and the capital-technology ratio is constant over time. This result is depicted graphically in Figure 1.2 of the Appendix. Note that the graph is analogous to Figure 1 with the important distinction of variable-interpretation.[9]

Setting equation (1.9) to zero and solving for k_e^*, we get

$$k_e^* = \left(\frac{s}{n+g+d}\right)^{1/(1-\alpha)}. \qquad (1.10)$$

---

[8] Jones (1998), pp. 30-34
[9] Jones (1998), pp. 34-35

Substituting equation (1.10) into equation (1.8), we arrive at

$$y_e* = \left(\frac{s}{n+g+d}\right)^{\alpha/(1-\alpha)}. \quad (1.11)$$

This equation suggests that the output-technology ratio does not grow in the long-run. However, we are more interested in per capita income, which is the output per *worker*. Equation (1.11) implies that

$$y*(t) = A(t)\left(\frac{s}{n+g+d}\right)^{\alpha/(1-\alpha)} \quad (1.12)$$[10]

Adding technology to the neoclassical growth model has several implications. First, as equation (1.12) illustrates, there is now the possibility of long-run growth in per capita income, determined by the growth rate of technology, g. Second, in addition to savings rates and population growth rates, countries can differ in their level of output per worker because of differences in their level of technology. The principal weakness of this exogenous growth model, however, is that it fails to explain the 'engine of growth;' it does not explain *why* changes in technology occur, it simply takes g as exogenously determined. We will examine what this model implies about convergence in more detail later, but let us pause for a moment and note that the neoclassical model implies that countries will converge to their unique steady states regardless of initial endowments. Each country's steady state is determined by its structural characteristics (such as technologies, preferences, population growth, etc.). Insofar as countries differ in these structural parameters, they will not converge to the same steady state income levels. However, *controlling*

[10] Jones (1998), pp. 34-36

for these variables, countries should converge regardless of initial factor endowments, a phenomenon known as 'conditional convergence.'[11]

Endogenous Growth Model: Externalities to Human Capital

Robert Lucas, writing in his 1988 seminal paper, 'On the Mechanics of Economic Development,' proposes that the neoclassical model implies convergence levels that are unobservable in the real world. Lucas notes the ambiguity of the term 'technology' as used in the neoclassical context. Technology as defined as the 'stock of useful knowledge' is unlikely to be growing much faster in one country than in another because, in Lucas' words, "Human knowledge is just human, not Japanese or Chinese or Korean." Instead, Lucas chooses to focus on the "knowledge of a particular people, or perhaps particular subcultures of people," which he calls 'human capital.' Technology, in contrast, he defines as "something common to all countries, something 'pure' or 'disembodied.'"[12]

As we alluded to above, insofar as technology, so defined, is the same across countries, the Solow model predicts a strong tendency towards equality in income and growth rates. Lucas notes that this is true both without factor mobility and even more so when factor mobility is permitted. With mobility of labor or capital, either or both factors will move to where they earn the highest returns, thereby equalizing factor prices and the capital-labor ratio. Indeed this effect will equalize per capita income levels even in the context of differences in population growth and savings rates. Lucas makes the important point that "in the model as stated, it makes no difference

[11] Galor (1996), p. 1057
[12] Lucas (1988), p. 15

whether labor moves to join capital or the other way around."[13] However, in the last century, in the presence of restrictions on immigration, capital has not moved in such a way as to equalize factor prices.

Thus we have three main weaknesses of the neoclassical model: first, it does not explain long-run growth but rather takes it as exogenously determined; second, it fails to account for the lack of convergence across countries; and finally, it fails to explain why capital moves from poor to rich countries. To correct for the first deficiency, Lucas added human capital to the model and explained its accumulation, thereby endogenizing the engine of growth instead of taking it as exogenous. Lucas also addressed the second two deficiencies by assuming a positive production externality to human capital. We will first derive a simplified version of the Lucas model as presented by Doepke (2003), and then we will add the externality to human capital and explore its relevance to convergence theory.

We assume that the production function now takes the form

$$Y_t = K_t^{\alpha}(uH_t)^{1-\alpha}. \qquad (1.13)$$

Here H is human capital and u is the fraction of time spent in production (not in developing human capital). Capital develops as it did in the Solow model such that

$$K_{t+1} = (1-d)K_t + sY_t.$$

We can also tell a story about the development of human capital such that

$$H_{t+1} = B(1-u)H_t, \qquad (1.14)$$

[13] Lucas (1988), p.16

where (1-u) is the amount of time spent accumulating human capital and B is the productivity of the education sector. We can use equation (1.14) to determine growth in human capital as follows:

$$\frac{H_{t+1}}{H_t} - 1 = B(1-u) - 1 = \gamma, \qquad (1.15)$$

such that γ is the constant growth rate in H_t. To solve for the balanced growth path, we assume $k_t = \frac{K_t}{H_t}$. Dividing equation (1.15) by H_t, we get

$$\frac{K_{t+1}}{H_t} = (1-d)\frac{K_t}{H_t} + s\left(\frac{K_t}{H_t}\right)^{\alpha} u^{1-\alpha}. \qquad (1.16)$$

Noting that $H_{t+1} = (1+\gamma)H_t$ we can rewrite equation (1.16) as

$$\frac{K_{t+1}}{H_{t+1}/(1+\gamma)} = (1-d)\frac{K_t}{H_t} + s\left(\frac{K_t}{H_t}\right)^{\alpha} u^{1-\alpha}, \text{ or}$$

$$(1+\gamma)k_{t+1} = (1-d)k_t + sk_t^{\alpha}u^{1-\alpha}.$$

Plugging in our steady state level k=k* we get

$$(1+\gamma)k^* = (1-d)k^* + sk^{*\alpha} u^{1-\alpha}.$$

Finally, solving for k*, we arrive at

$$k^* = \left(\frac{s}{\gamma+d}\right)^{\frac{1}{1-\alpha}} u.[14] \qquad (1.17)$$

To solve for output on the balanced growth path, we can rewrite the production function in terms of k. Multiplying equation (1.13) by $\frac{H_t^{\alpha}}{H_t^{\alpha}}$ we get

$$Y_t = \left(\frac{K_t}{H_t}\right)^{\alpha} u^{1-\alpha} H_t.$$

[14] Doepke (2003), pp. 13-14

Substituting in k_t, we arrive at

$$Y_t = k_t^{\alpha} u^{1-\alpha} H_t. \qquad (1.18)$$

Thus, when k is constant, output grows at the same rate as human capital. Moreover, we see that human capital is functionally different from physical capital. Two countries with identical parameters except for differing amounts of initial physical capital will, in the long-run, converge to the same level of physical capital and, consequently, output. Contrastingly, two countries with identical parameters except for different initial levels of human capital will not converge in income. To see why, recall from equation (1.15) that $\gamma = B(1-u)-1$ and recall from equation (1.18) that, in the steady state, output grows at the rate of growth in human capital (γ). Thus, two economies with identical parameters will exhibit the same constant growth rate in human capital, equal to gamma, and so the country with the initially higher level of human capital will continue to have a higher level of human capital and consequently, higher income.[15]

We see that this model improves on the neoclassical model by endogenizing the growth rate. Lucas also added to the neoclassical model by incorporating a positive externality to human capital in production of final output. To show this effect, we can rewrite the production function as

$$Y_t = AK_t^{\alpha}(uH_t)^{1-\alpha}\bar{h}_t^{\varepsilon} \qquad (1.19)$$

where $\bar{h}_t$ is the average level of human capital in the economy and, as before, A is a production coefficient representing technology.[16] The external effect of human capital is not inconsistent with economic theory or real life observations. As David

[15] Doepke (2003), p. 15
[16] Razin and Yuen (1997a), p. 229

Weil notes, a study in Ethiopia found that "more than half the benefit of an individual's going to school for another year accrued to people other than the person attending school."[17] Moreover, the externality explains why both labor and capital would want to move in the same direction, which is often the case in the real world. And as we will see later, this spillover effect will have important implications for convergence in the context of labor mobility.

Convergence Literature

It is worth taking a moment to explore the most common techniques used in measuring convergence. There have traditionally been two measurements: beta convergence and sigma convergence. Beta convergence can be examined by regressing the growth rate against a constant and the log of initial income level:

$$\log\left(y_{it} / y_{i,t-1}\right) = a - (1 - e^{-\beta}) \cdot \log(y_{i,t-1}) + u_{it}, \quad (1.20)$$

where t denotes the year, i denotes the country or region, and u_{it} is a random error term. A positive value for beta implies that countries or regions with lower initial levels of income grow faster than those with higher levels and therefore converge over time. We look for sigma convergence by examining the cross-economy variance of income at time t (σ_t^2) such that

$$\sigma_t^2 = e^{-2/\beta} \cdot \sigma_{t-1}^2 + \sigma_{ut}^2. \quad (1.21)$$

If the coefficient of variance declines over time, there is said to be sigma convergence. Note that beta convergence is a necessary, but not sufficient condition for sigma convergence.[18] This is an important distinction, as sigma convergence in essence measures absolute dispersion in income, while beta convergence can be

[17] Weir and Knight (2000); as cited in Weil (2005), p. 177
[18] Barro and Sala-I-Martin (1995), pp. 383-385

measured while accounting for other variables and can therefore detect conditional convergence.[19]

Various studies have been conducted to determine the degree of convergence both between countries and within them. Baumol (1986) was one of the most important early papers to empirically test for cross-country convergence. Looking at data on per capita income and productivity for 16 countries from 1870-1979, Baumol found strong evidence of convergence. Regressing per capita growth rate over that period against the initial GDP per work-hour indicated a strong inverse correlation between the two variables. Moreover, a great degree of variation in growth rates was explained by initial income levels:

$$\textit{Growth Rate}\ (1870\text{-}1979) = 5.25 - 0.75\ln\ (\textit{GDP}\ \text{per}\ \textit{WorkHr},\ 1870),\ \text{R-sq} = 0.88$$

In light of this result, Baumol hypothesized that a productivity spillover occurs between industrialized countries. Specifically, if one country develops an innovation in production, similar industries in other countries will be under tremendous pressure to obtain access to that innovation. However, when using data on a wider range of countries, there is no longer evidence of convergence, and in fact, Baumol concludes that the poorest countries have been growing most slowly. Baumol suggests that this may be a result of differences in product mix and education impeding the spillover of technology.[20]

A major criticism of Baumol's convergence finding was that he selected countries based on the availability of accurate data, which suggests an *ex post* selection of successful economies since those countries with easily available

---

[19] Raiser (1998), p. 3
[20] Baumol (1986), pp. 1075-1080

historical data are today's rich countries. As William Easterly points out, it's the rich countries that can afford economic historians who reconstruct time series income data. And, since these countries are all rich today regardless of where they began in 1870, the model will be predisposed to predict convergence.[21]

Another major article in the convergence literature is Barro and Sala-I-Martin (1992). In this paper, the authors study two phenomena: convergence among the 50 US states and convergence among a large set of countries within the framework of the neoclassical growth model. The empirical results suggest absolute convergence in income and product at a speed of about 2% per year for the states.[22] The authors express surprise at the equal rates of convergence for income and product, when, in the context of free factor mobility (which we expect to be characteristic of the US states) there should be faster convergence in product than income. The intuition for this is that capital stocks should converge more rapidly since some capital will flow from the low marginal product of capital (MPK) states to the high MPK states. And since the income derived from that capital would flow back to the low MPK states, income would converge more slowly than product. The authors hypothesize that this might be an indication of imperfect capital mobility if capital is broadly defined as both human and physical capital. Such a theory is consistent with the fact that foreigners cannot own domestic human capital.[23]

Barro and Sala-I-Martin also look for convergence across Japanese prefectures over the period 1930-1990. They estimate a beta coefficient of .0279,

---

[21] Baumol (1986), p. 1075; Easterly (2002), p. 64

[22] Barro and Sala-I-Martin (1992), p. 245

[23] Barro and Sala-I-Martin (1992), pp. 239-241; For a further discussion, see Barro et al. (1995), pp. 109-114

without controlling for structural variables (such as savings rate or population growth), a result that strongly indicates absolute beta convergence. They also test for convergence in each five year sub-period, and find a negative correlation between initial income and growth rate in every period when controlling for structural variables. The authors also test for sigma convergence across the prefectures. They conclude that, after increasing during World War II due to massive military spending, the coefficient of variance declined steadily until the 1980s and has remained stable since.[24]

The results for convergence across countries differ markedly from those for US states and Japanese prefectures. Not only is there little correlation between growth rate and initial income level, but the beta coefficient is actually negative, suggesting a small tendency for rich countries to grow faster than poor ones.[25] Only when the authors control for other variables does the data exhibit a convergence rate similar to that of the states. This implies that there is *conditional* convergence among countries, which recall means that each country converges to its unique steady state income level. In order to control for steady states, the authors include variables for school enrollment, the average ratio of government consumption expenditure to GDP, political stability, and market distortions. The implication of the study is that countries have different steady-state values. Interestingly, the authors note, “the absence of substantial labor mobility across countries reinforces the possibility of

[24] Barro and Sala-I-Martin (1995), pp. 393-398
[25] Barro and Sala-I-Martin (1992), p. 241

substantial divergences in these steady-state values." In contrast, the US states do not differ significantly in the steady-state values.[26]

Growth Theories and Convergence

Thus far we have examined two basic models for economic growth with differing implications for cross-country variance in income levels. As we have already seen, the neoclassical exogenous growth model predicts a tendency towards long-run convergence in both income levels and growth rates across countries, particularly when capital is allowed to flow freely across borders. Barro et al. (1995) states that in a neoclassical model with perfect capital mobility "the predicted rates of convergence are infinite," which essentially means that convergence should be complete and instantaneous.[27]

Let's consider, in the context of the neoclassical model, two economies, A and B, with identical parameters but differing initial endowments of capital. Because they have identical parameters (technology, population growth, depreciation rate, and savings rate), they converge to the same steady state capital-labor ratio, k*. As the above Barro et al. quotation suggests, in the presence of factor mobility, convergence will occur even faster. If economy A is below its steady state per capita income level and economy B is above, the MPK in economy A will be higher and capital will flow from B to A (or labor will flow from A to B), lowering income in B and raising it in A until $k_A = k_B = k^*$.[28] Again, this progression will occur even in autarky, but less

[26] Barro and Sala-I-Martin (1992), p. 243
[27] Barro et al. (1995), p. 109
[28] Galor (1996), p. 1062

quickly. Thus, in conditions of either autarky or factor mobility, economies will experience conditional convergence.

Endogenous growth models with externalities to human capital, though they also predict convergence in growth rates, tend to predict divergence in income levels. Tamura (1991) builds an endogenous growth model with spillovers to human capital that predicts income convergence. We will forgo a detailed derivation of his model, simply noting his conclusion that, if this spillover transfers across two economies, their income levels will converge in the long-run.[29] This idea is extended by Assaf Razin and Chi-Wa Yuen in two 1997 papers "Factor Mobility and Income Growth: Two Convergence Hypotheses" and "Income Convergence within an Economic Union: The Role of Factor Mobility and Coordination." Razin and Yuen hypothesize that level convergence in an endogenous growth model framework is possible but has more stringent requirements than in the exogenous neoclassical model. Specifically, in order to see level convergence in an endogenous growth model with positive externalities to human capital, there must be labor mobility (which Tamura's model explicitly does not require). Recall that this requirement was also absent in the exogenous model.[30]

Razin and Yuen's analysis builds on the Lucas model that we examined earlier with the production function of equation (1.19): $Y_t = AK_t^{\alpha}(uH_t)^{1-\alpha}\bar{h}_t^{\varepsilon}$. Let's consider the implications of this model for convergence. Again imagine two economies, A and B, with identical structural parameters but differing initial endowments of human capital, such that $\bar{h}_A < \bar{h}_B$. Because of the spillover effect, the

[29] Tamura (1991), p. 524
[30] Razin and Yuen (1997), p. 171

marginal product of capital and the marginal product of labor will be higher in economy B when capital-labor ratios are equal. Thus, capital will move from A to B, equalizing interest rates and growth rates, but not income levels. With equal rates of return to capital, there will be a higher capital-labor ratio in country B, as well as a higher $\bar{h}$, which implies a higher wage in economy B and therefore a lack of income level convergence.[31]

Now let's observe what happens if we allow for labor to move across economies. In this case, due to the spillover effect, the marginal product of labor is higher in B and so wages there are higher. Responding to the higher wage, workers in country A will migrate to country B. This has the dual effect of raising the wage in economy A and lowering the wage in economy B. The wage increases in economy A because there is a decreased supply of labor and, *ceteris paribus*, a higher capital-labor ratio. The wage in country B declines because the migrants increase the supply of labor and decrease the average level of human capital. Moreover, the higher wage in country A will act as an incentive to invest more in human capital (since the rate of return to human capital is higher) and likewise, the lower wage in economy B will be an incentive to decrease the rate of human capital accumulation. This process will occur until wages and human capital levels are equalized. Thus, in the context of externalities to human capital, labor mobility is essential for income level convergence.[32] Without labor mobility, the model implies 'club convergence,' which we will define as when "per capita incomes of countries that are identical in their structural characteristics converge to one another in the long-run provided that their

[31] Razin and Yuen (1997a), p. 232
[32] Razin and Yuen (1997a), pp. 233-234

initial conditions are similar as well."[33] In the presence of labor mobility, we are back to our familiar conditional convergence, since initial endowments of physical and human capital can differ.

It is necessary to note, however, an alternative theory, which proposes that labor mobility can actually cause *divergence*. This theory supposes that the positive externalities to high levels of human capital mostly extend to individuals with large amounts of human capital. Insofar as this is the case, we might expect that an economy with low levels of average human capital will lose migrants with from the highest part of their human capital distribution, rather than the middle. More specifically, suppose we have two economies—A and B, where A has a high average level of human capital and B has a low average level of human capital. In this case, the individuals who gain most from migration are those in country B who have higher than average human capital accumulation, since the positive externalities in economy B benefit them. Therefore, the average level of human capital in country B might actually go down due to migration. This phenomenon is known as 'brain drain.' On the other hand, the opportunity to possibly realize these benefits to human capital accumulation through migration might induce many more people to increase their human capital who then end up not migration. In this case, migration (or rather, the possibility of migration) can still promote convergence, even when the positive human capital externalities extend only to those who themselves have high levels of

[33] Galor (1996), p. 1056; See also Johnson and Takeyama (2003) for a discussion of the distinctions between difference types of convergence

human capital. [34] It is worth keeping these alternative theories in mind as we consider migration's effects on China's regional incomes.

China's Regional Income Divergence

The existence of regional income divergence in China has been widely debated. Chen and Fleisher (1996) conclude that China has actually experienced modest regional convergence in both levels and growth rates since the reforms of the late 1970s. They base their regressions on a Solow model augmented to include human capital, similar to that proposed by Mankiw et al. (1992). The authors conclude that the liberalization of economic policies has increased the tendency towards convergence but that this convergence is conditional on several variables, most importantly coastal location.[35] Insofar as the lack of convergence between coastal and non-coastal provinces concerns structural differences, and not differences in initial factor endowments, we can interpret this as conditional convergence. However, it would be useful to include the true variables (such as foreign direct investment) that are hypothesized to be causing divergence rather than merely including a dummy variable for whether the province is coastal or non-coastal.

Raiser (1998) reaches similar conclusions, arguing that developments in regional incomes in China fit the theoretical growth model prediction that increased liberalization—and therefore greater free-market movement of capital—should exert pressure towards income convergence. Raiser finds that sigma convergence does occur over the period 1978-1992, but that the decline in variation slows in the mid-1980s. He also examines beta convergence in different periods and finds that it is

[34] Beine et al. (2001), p. 282
[35] Chen and Fleisher (1996), pp. 148-153

higher in the late 70s and early 80s than in the late 80s and early 90s. A hypothesis test assuming equal coefficients over these two periods fails at the 95% confidence level.[36] Raiser attributes the decline in the rate of convergence to differences in growth between coastal and non-coastal regions and to preferential inter-regional government transfers, which cause capital to flow to relatively rich regions rather than to poor ones.[37]

Other scholars have reached differing conclusions to those of Chen and Fleisher (1996) and Raiser (1998). Using cross-sectional techniques, Yao and Zhang (2001) reject absolute and conditional provincial income convergence. However, when controlling for distance from "growth centers," which is uniformly applied to provinces within three regions (east, central, and west), the authors find what can be best interpreted as club convergence, as divergence between regions is accounted for by distance from growth centers and not merely structural differences. Using panel data, the authors also construct a unit-root test for each of the three regions separately.[38] When including a time trend, they conclude that club convergence applies to the eastern and western regions.[39] That is, convergence applies between provinces within these clubs. However, it should be noted that the methodology employed is seriously flawed, particularly in the assumption of homogenous dynamics and parameters of interest. The methodology employed in Pedroni and Yao (2006) and in this paper will relax those tenuous assumptions.

---

[36] Raiser (1998), pp. 4-5
[37] Raiser (1998), p. 13
[38] For explanation of unit-root, see Section III
[39] Yao and Zhang (2001), pp. 478-479

Démurger et al. (2002) test for convergence conditioning on geographic location and an index of preferential government policies. They find 'weak' convergence (statistically insignificant), and therefore conclude that geography and preferential policies are important causes of regional divergence. Young (2000) argues that the (partial) reform process created incentives for local governments to construct barriers to trade thereby fragmenting regional markets, which has led to decreased specialization and diverging factor intensities. This process, he predicts, results in divergence of income across regions.[40]

Analysis in Pedroni and Yao (2006) affirms Young's finding of regional divergence in China since 1978. The authors improve significantly on the existing literature by using newly developed empirical techniques that account for the time series properties of the data. An explanation of these non-stationary panel techniques and their advantages will be presented in Section III. Avoiding the specific econometric methodology for now, Pedroni and Yao find that provincial income levels are diverging and that this divergence persists even when segregating into the hypothesized 'convergence clubs' determined by geographic location or degree of preferential treatment. That is, the differences in patterns of growth *cannot* be accounted for by these usual explanations. The authors suggest that labor mobility may play an important role and recommend further research in this area.[41]

Labor Mobility in China and Classification of Migrants

The Communist Party instituted the *hukou* system of household registration in the early 1950s, recording households' locale and classifying each as either urban or

[40] Young (2000), p. 1092
[41] Pedroni and Yao (2006), p. 23

rural. Individuals received social benefits that were tied to the location of their *hukou* and became ineligible for benefits if they moved. The system was tightened over the ensuing years, particularly after the Great Leap Forward (1958-1960), and was remarkably effective in restricting migration. Since China was a planned economy, and consequently lacked markets, the government was able to link the *hukou* system to the allocation of housing, jobs, food and other necessities.[42] According to Wu and Treiman (2004), “this tight administrative control…virtually eliminated unauthorized rural-to-urban migration in the pre-reform era.”[43] Zhao (2004) concurs, arguing that the government’s absolute control over the economy “made it almost impossible for people without local *hukou* to live in urban areas.”[44]

In order for a residency change to be deemed official, it was necessary to obtain a government sanctioned transfer of one’s *hukou*. During the pre-reform period, these transfers were rarely authorized unless it was motivated in support of state-initiated programs. As Chan (2001) notes, “an approval for self-initiated relocation to a city from the countryside was only a dream for ordinary peasants.”[45] Thus, prior to the beginning of reforms in the late 1970s, it was both extremely difficult to obtain a change in one’s *hukou* and nearly impossible to migrate without obtaining one.

When the Reform Period (*gaige kaifang)* began in 1978, migration restrictions started to erode, and a new ‘regime’ in Chinese migration based more on market

[42] Wu and Trieman (2004), p. 365; Zhao (2004), p. 287
[43] Wu and Treiman (2004), p. 364
[44] Zhao (2004), p. 287
[45] Chan (2001), p. 128-129; See also Chan and Yang (1992), p. 4

forces than central planning began (Chan and Yang, 1992).[46] Although the increase in migration from rural to urban areas was slow at first, it accelerated in the late 1980s. Several policy factors led to an increase in rural-urban migration. First, the introduction of the Household Responsibility System and the concurrent abolition of the commune system meant that peasants were no longer tied to their land and increased the availability of food in the urban free market. As food markets grew in the late 1980s, a private sector in urban areas was born, which stimulated labor demand. These developments provided an opportunity for rural migrants to survive in urban areas and increased demand for migrant businesses and cheap migrant workers.[47] In addition to a *de facto* deterioration of the *hukou's* effectiveness, the system itself underwent institutional changes after the economic reforms, making it increasingly possible, though still difficult, for rural residents to officially change their permanent registration status.[48] As the private sector flourished in the 1990s and the *danwei*-based rationing system was abandoned, migration continued to increase (Liang and Ma, 2004). In 1988, the central government officially eased restrictions on rural-urban migration, allowing farmers to migrate if they could provide their own food and were financially self-sufficient.[49] Restrictions on *hukou* transference have been even further liberalized since 2000. [50]

A by-product of the more market-oriented migration regime in the post-Reform Era is its greater degree of complexity. Therefore, analyzing migration in

---

[46] Chan and Yang (1992), p. 4

[47] Wu and Treiman (2004), p. 365; Zhao (2004), p. 288

[48] Ping and Pieke (2003), p. 16

[49] Yang and Zhou (1999), p. 11

[50] For a detailed summary of recent policy changes see Ping and Pieke (2003), pp. 16-17

China can quickly become a bewildering task. It is imperative that we explicitly define the various categories of migrants to avoid any confusion of terms. As we have seen, under the command economy, unofficial migration was almost impossible. So the definition of migrant during that time is fairly straightforward: one who attains a change in *hukou* status. In contrast, since the market reforms, categorizing migrants is much trickier. Most fundamentally, we must distinguish between those who obtain an official change in residence and those who are *de facto* migrants, that is, they are registered in one locale, but inhabit another. The former are often referred to as 'permanent migrants' while the latter are referred to as 'floating' or 'temporary' migrants. However, this terminology can be misleading; migrants without official change in registration often stay in their destination for many years, and should thus be considered 'permanent migrants' in the literal sense.[51] Also, note that floating migration was severely limited during the pre-reform period because local *hukou* registration was necessary for basic survival. However, floating migrants now constitute an increasingly large portion of total migration. For example, according to the 2000 census, almost 75% of total intercounty migrants who migrated between 1995 and 2000 were floating migrants.[52]

While many floating migrants should be considered permanent, others do remain in their destination for only a short period. This reality introduces a further complexity into the categorization of migration, which is determining the minimum duration of stay required in order to be considered a migrant. Adding to the confusion is the fact that this threshold has changed over time. For example, in the

---

[51] Liang and Ma (2004), p. 469
[52] Liang and Ma (2004), p. 472

1990 census, persons were considered migrants only if they had resided in the destination community for over a year. On the other hand, in the 2000 census, the minimum duration of stay was only six months.[53] Stays of less than six months are more appropriately thought of as '"circulation" rather than migration.[54]

Although most migration can be characterized as rural to urban migration, some migrants leave their home province (interprovincial migration), whereas others stay in their home province, perhaps moving to a county town or the provincial capital (intraprovincial migration).[55] Analysis of the 1990 census puts interprovincial migration at around 32.42% of total migration in China, while estimates based on the 2000 census put that figure at 26.4%.[56] For the purposes of this paper, we are primarily interested in inter-provincial migration because we are testing for income convergence across provinces. That being said, insofar as locales within a province are more or less receptive to intraprovincial migration, we would expect them to be similarly receptive to inter-provincial migration. Thus, information about intraprovincial migration may still be of some use in inferring information about migration across provinces.

Migrant Characteristics

The demographic characteristics of China's migrants have been thoroughly researched. As alluded to above, there is broad consensus in the literature that the vast majority of migration is from rural to urban areas. Moreover, rural-urban migration continues to grow spectacularly, increasing from two million migrants in

[53] Johnson (2003), p. 25
[54] Chan Yang (1992), p. 2
[55] Liang and Ma (2004), p. 478
[56] Zhao (2004), p. 289

the mid 1980s to as many as 70 million in the mid-1990s, and then to as many as 94 million in 2002.[57] To put this into the context of overall migration, between 70% and 80% of rural migration is to urban destinations.[58] Consequently, much of our analysis of migration generally will focus on rural-urban migration specifically.

There is some ambiguity in the literature as to the degree of gender differentials in the migrant population. Furthermore, the gender composition of the migrant labor force may be changing over time. According to Zhao (2004), men account for a majority of migrants. He speculates that the reason for this gender imbalance is due to the traditional expectations of women belonging in the home, as well as the higher demand in urban areas for male migrants in industries such as construction.[59] Similarly Ping and Pieke (2003) estimate that only around one third of migrants are women.[60] However, Zhang et. al. (2004), conclude that women are becoming increasingly represented in the migrant population. Using a logit estimator based on rural household level surveys, the authors conclude that between 1980 and 1990, being male made migration by 11.09 times more likely, whereas, between 1990 and 2000, that figured dropped to just 3.13 (though still significant at the 5% level). Moreover, among younger migrants, the gender differential is virtually non-existent.[61]

Data on the interaction of education and migration is also inconclusive.[62] Chan (2001) concludes that migrants are generally more educated than the overall

---

[57] Ping and Pieke (2003), p. 6
[58] Zhao (2004), p. 290
[59] Zhao (2004), p. 290; See also Hare (1999), p. 9(correct)
[60] Ping and Pieke (2003), p. 8
[61] Zhang et al. (2004), pp. 239, 242
[62] Zhao (2004), p. 292

population. Based on data from the 1990 census, he reports that 2.2% of non-*hukou* migrants attained a college education, compared to just 1.6% of the national population.[63] Similarly, based on surveys conducted in 1995, Rozelle et. al. (1999), find that migrants are characteristically “relatively well-educated and becoming increasingly so.”[64] Zhang et al. (2004) find that the probability one is in the migrant labor force increases by 10% for every additional year of education.[65]

In contrast, Hare (1999), using data collected in 1995 in Henan province, finds that “years of formal education do not appear to have an important effect on the probability of migrating.”[66] Zhao (1999a) looks at household level surveys conducted in Sichuan province in 1994 and 1995 and finds a statistically significant *negative* effect of household’s years of schooling on likelihood of migration.[67] Similarly Zhao (1999b) distinguishes between migration and rural non-farm work, and finds that the effect of an additional year of high school on the likelihood of migration is not statistically significant, but the effect on choosing rural non-farm labor is significant and non-negligible. This finding suggests that highly educated rural workers prefer local non-farm work to migration.[68]

The education of migrants is important given our theoretical assumptions about migration and convergence. The model proposed in Razin and Yuen (1997a) assumes that human capital accumulated by migrants is equal to the average level of human capital accumulation in the source economy as a whole. On the other hand, if

[63] Chan (2001), p. 134
[64] Rozelle et al. (1999), p. 378
[65] Zhang et al. (2004), pp. 239-240
[66] Hare (1999), p. 9
[67] Zhao (1999a), p. 284
[68] Zhao (1999b), pp. 775-770

the most educated of the source population migrate the predictions of convergence may fall apart, as the average level of human capital, and perhaps the average wage, will fall in the source economy, while the average level of human capital and the average wage in the destination economy will not necessarily drop. This scenario would be more consistent with the 'brain drain' model we examined earlier, and could actually facilitate divergence. While the above analysis suggests that the exact effect of educational attainment on migration is ambiguous, it is likely that migrants are at least slightly more educated that the average rural worker. However, Zhao's (1999b) analysis suggests that, in general, the most educated rural workers tend to choose local non-farm work over migration, which would mean that 'brain drain' is not a serious danger. That descriptive statistics show migration rising with education levels could be the result of some other correlated variable, such as youth.[69]

One other education effect needs to be considered. Razin and Yuen (1997a) suggest that wage increases in the source economy consitute higher returns to labor, thus facilitating greater 'investment' in labor in the form of education. This mechanism should cause the level of human capital to rise in the source economy. However, as de Brauw and Giles (2005) demonstrate, we must consider that greater access to migration and higher local wages raise the opportunity cost of attending school, and might thus act as disincentive for education. Using household surveys conducted in fifty-two villages of four provinces during 2004, the authors conclude that an increase within a village in opportunity for migration decreases the probability of its residents attending high school. They hypothesize that this result may be

[69] Rozelle et al. (1999), p. 389

partially due to the fact that migrant workers are often restricted from employment in skilled-industries, thereby reducing the returns to schooling for rural workers.[70] This finding partially undermines our model's predictions about migration and convergence.

On the other hand, recent research on the effects of migration on living standards in the source community provides support for our hypothesis that migration promotes convergence. Taylor et al. (2003) find that migration has "an unambiguously positive" effect on per capita household income in the source economies. Specifically, they estimate that migration causes household income to increase between 16% and 43% depending on the specific characteristics of the household. Du et al. (2005) estimate that having a migrant increases household per capita income in poor counties by between 8.5% and 13%.[71] In addition, Chan (2001) notes that migration facilitates the transfer of technology from modern urban centers to previously isolated rural areas, and is "an effective and cheap way to siphon off surplus rural labor and ease pressure on local land and resources."[72] These assessments of the effects of migration on source economies are largely consistent with our theoretical hypothesis of convergence conditional on labor mobility.

Determinants of Mobility

Economic motivations are the most important driving force behind migration. We assume that potential migrants are rational actors and will migrate if the expected benefits of relocating exceed the expected costs. In the case of rural to urban

[70] de Brauw and Giles (2005), pp. 25-26
[71] Du et al. (2005), p. 706
[72] Chan (2001), p. 146

migration in China, the primary benefit is almost always the increased wage that can be earned in urban areas.[73] The costs of migration, however, are more complex and can vary considerably by source locale. First, there are the obvious monetary expenses, such as the cost of transport, urban housing, and temporary government permits. One survey conducted in 1995 found these costs to be 721.7 yuan for the average migrant.[74] We might expect that areas with relatively easy access to roads and buses would have greater incidences of migration, however, Rozelle et al. (1999) find that ease of transportation is not a statistically significant determinant of migration.[75]

As we saw earlier, prior to the Reform Period, institutional barriers were extremely important factors in inhibiting migration. The *hukou* system of household registration prevented rural workers from migrating by denying them access to basic necessities. Though the *hukou* system has become seriously loosened, certain institutional costs remain. Officially registered urban residents still receive many benefits as a result of their *hukou* status, including free education for their children, and greater access to housing and health care.[76] As a result, some scholars view the *hukou* system as continuing to act a severe barrier to migration.[77] However, in a survey conducted by Du et al. (2005), only 1.66% of males and .38% of women cited a lack of public services due to not being registered residents as the most important

[73] Ping and Pieke (2003), p. 12; See also Zhu (2002), pp. 227-228
[74] Zhao (1999), p. 777
[75] Rozelle et al. (1999), p. 388
[76] Yang and Zhou (1999), p. 12; Zhou (1999), p. 778
[77] For example, Whalley and Zhang (2004)

factor affecting their migration decision.[78] This suggests that we should be wary of placing too much emphasis on the *hukou* system as the primary impediment to migration.

There are also institutional factors on the rural side that discourage migration. Under the household responsibility system, families in rural areas have land-use rights but not ownership rights per se. That is, if they decide to leave, they cannot sell the land, but must instead return it to the local government.[79] This loss of future earnings constitutes a considerable cost to migration. Increasingly though, some rural areas are developing land rental markets. Rozelle et al. (1999) find that the ability to rent is positive and statistically significant determinant of the decision to migrate.[80] More developed markets can also facilitate migration in other ways. It is possible that the poorest farmers, who would gain the most benefit from migration, cannot do so because they lack the initial funds necessary to relocate. In support of this hypothesis, Rozelle et al. (1999) find that farmers are significantly more likely to migrate if they come from a village where informal credit markets exist.[81]

Perhaps the most important costs to migration are the least tangible ones, such as those deriving from uncertainty. Much research in this area has focused on the impact of existing migrant networks in supplying information and providing support to potential migrants. These networks exist in communities where there has already been a history of migration to a particular destination area. The importance of these connections has been well documented. A random sample conducted in 1995 of

---

[78] Du et al. (2005), p. 708
[79] Yang and Zhou (1999), p. 12
[80] Rozelle et al. (1999), p. 389
[81] Rozelle et al. (1999), p. 389

migrant workers in Jiangsu and Shanghi province found that over 75% had been assisted by relatives and friends during their first migratory trip.[82] That same year, a sample of migrants in Shangdong found that 70% had prearranged jobs before migrating.[83] Rozelle et al. find that a village's migration network, as approximated by lagged migration, is "a highly significant" determinant of migration.[84] Finally, in the survey conducted by Du et al. (2005), over 18% of male respondents said that "lack of information and social networks" was the most important factor affecting their migration decision.

In sum, the expected benefit of rural to urban migration is almost always the superior wage available in urban areas. The costs to migration can vary considerably depending on the source and destination locales. In addition to the explicit transportation costs, there are opportunity and informational costs associated with migration. These can be mitigated depending on the pervasiveness of markets and migration networks.

---

82 Zhao (2003), p, 501
83 Zhao (2003), p. 501
84 Rozelle et al. (1999), p. 388

Section II: Econometric Techniques

Autoregressive and Moving Average Models

As a first step in understanding this paper's econometric techniques, it is worth examining a brief overview of time series econometrics. A variable is modeled as an autoregressive (AR) process if its contemporaneous value is a function of its previous values. A simple example of such a model is a first order AR process, represented as

$$y_t = \alpha + \phi y_{t-1} + \varepsilon_t \qquad (2.1)$$

We see that the realization of y in period t is explained by a constant, its value lagged one period, and a random term, which is assumed to be independent and identically distributed (i.i.d.). Intuitively, we are considering each value of y to be a single realization from a distribution of possible values. The epsilon terms reflects the fact that the particular value of y that is observed in period t is partly random. AR models can become more complex by including additional lags. For example, including a second lag would constitute a second order AR process, denoted AR(2). More generally, an AR(N) process is represented as

$$y_t = \alpha + \phi_1 y_{t-1} + ... + \phi_N y_{t-N} + \varepsilon_t \qquad (2.2)$$

The series y_t can also be represented in moving average (MA) form, as a function of previous disturbances. For example, a first order MA process would be written

$$y_t = \alpha + \theta_1 \varepsilon_{t-1} + \varepsilon_t \qquad (2.3)$$

Analogously to the AR model, we can extend an MA process to an arbitrary number of lags, Q, such that

$$y_t = \alpha + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q} + \varepsilon_t \qquad (2.4)$$

$$= \sum_{j=0}^{Q} \theta_j \varepsilon_{t-j} \text{ where } \theta_0 = 1$$ [85]

Importantly, these two forms are generally invertible. For example, an autoregressive process for y_t can be rewritten as a function of previous shocks, since y_{t-1} itself depends on the shock in period t-1. Specifically, dropping the constant for simplicity, we can express y_{t-1}

$$y_{t-1} = \phi y_{t-2} + \varepsilon_{t-1} \qquad (2.5)$$

Substituting in equation (2.2), we can rewrite equation (2.1) as

$$y_t = \phi(\phi y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t$$

$$= \phi^2 y_{t-2} + \phi\varepsilon_{t-1} + \varepsilon_t \qquad (2.6)$$

Note that this exercise can be continued and, as the limit approaches infinity, we have

$$y_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j} \qquad (2.7)$$

In progressing from equations (2.5) to (2.7), we have illustrated that AR processes can be represented as infinite order MA processes, in terms of the AR coefficient (phi). Similarly, in most cases, we can also convert the MA to AR form.[86]

Unit-roots and Stationarity

A concept of particular importance for our analysis is that of stationarity. Informally, a variable is stationary if the impact of a random shock to the variable

---

[85] Pedroni (2006), Topic 2, pp. 7-8
[86] Pedroni (2006), Topic 2, p. 17

eventually dissipates over time. More formally, let us consider again the AR(1) process depicted in equation (2.1).

For a stochastic process to be considered stationary, the following conditions must hold

$$E(y_t) = E(y_{t-s}) = \mu$$

$$E[(y_t - \mu)]^2 = E[(y_{t-s} - \mu)^2] = \sigma^2$$

$$E[(y_t - \mu)(y_{t-s} - \mu)] = E[(y_{t-j} - \mu)(y_{t-j-s} - \mu)] = \gamma_s$$ [87]

The first condition requires that the mean of y be constant for all time periods. The second requires that the variance be constant at all time periods as well. Finally, the third condition implies that the covariance between time periods of equal distance should be constant, regardless of the absolute time period being considered. In other words, time matters only relatively, not absolutely.[88]

It is easy to see that whether or not equation (2.1) is stationary depends on the value of phi. Iterating backwards similarly as we did in equations (2.5) through (2.7), only this time allowing $\alpha \neq 0$, we see that the solution to the difference equation is

$$y_t = \alpha \sum_{j=0}^{t-1} \phi_1^j + \phi_1^t y_0 + \sum_{0}^{t-1} \phi_1^j \varepsilon_{t-j} \qquad (2.8)$$

The expected value of y_t is

$$E(y_t) = \alpha \sum_{j=0}^{t-1} \phi_1^j + \phi_1^t y_0 \qquad (2.9)$$

and the expected value of y_{t+s} is

$$E(y_{t+s}) = \alpha \sum_{j=0}^{t+s-1} \phi_1^j + \phi_1^{t+s} y_0 \qquad (2.10)$$

---

[87] Enders (2004), p. 53

[88] Pedroni, (2006), Topic 2, p. 12

Taking the limit as t approaches infinity, it is clear from these expressions that the time series is stationary only if $|\phi_1| < 1$. In this case, the mean converges to $\alpha/(1-\phi_1)$, which is time independent. Likewise, the variance of y_t, equal to $\sigma^2/(1-(\phi_1)^2)$, is also time independent. Finally, the autocovariances converge to $\sigma^2(\phi_1)^s/\left[1-(\phi_1)^2\right]$, which is, again, time-independent. Thus, we have shown that stationarity requires $|\phi_1| < 1$. In contrast, when $\phi_1 = 1$, the mean, variance, and autocovariances are all time dependent, and y_t will not revert back to a constant mean. Intuitively, this makes sense, as phi equaling one implies that the effects of a shock to y are permanent. Moreover, if phi is greater then one, the effects of a shock become greater over time, which means that the dynamic process of y will explode. The case in which phi equals one is a vital one, and is know as a 'unit-root' process.

Suppose we want to test whether a series contains a unit root and is thus non-stationary. The most common method of testing for a unit root is the augmented Dickey-Fuller (ADF) test. Recall that we can estimate an autoregressive process using simple OLS. In testing for a unit root, the estimated parameter of interest is the first autoregressive coefficient. For example, returning to the AR(1) process in equation (2.1):

$$y_t = \alpha + \phi y_{t-1} + \varepsilon_t$$

We see that testing the hypothesis of stationarity would require the following null and alternative hypothesis:

$$H_0 : \phi = 1$$
$$H_A : \phi < 1$$

If we transform the equation slightly, subtracting y_{t-1} from both sides, we arrive at:

$$\Delta y_t = (\phi - 1) y_{t-1} + \varepsilon_t \qquad (2.11)$$

We can rename the phi minus one term as rho, and in doing so it is clear that the new null hypothesis is that rho equals zero (in which case a unit-root is present), which conveniently, is just the t-statistic on y_{t-1}.

Note however, that if y is not well approximated by a first order autoregressive process, then the epsilon terms will be serially correlated, distorting our test results. In order to eliminate this potential serial correlation, we must include lags of the difference of y.[89] There are a few methodologies for determining the appropriate number of lags to include. There is a tradeoff, of course, between including too few lags, and therefore having serial correlation, and including too many, reducing the power of the test by wasting degrees of freedom. As it turns out, unit root tests are especially sensitive to the former, and therefore, the most appropriate method is the most conservative, known as 'the step down approach.' This approach consists of beginning with an arbitrarily large number of lags and testing for the significance of the last lag. If the last lag is insignificant, then we run the regression again, this time with one fewer lags and we repeat this process until the last lag is significant.[90] In summary, the augmented Dickey-Fuller test estimates the following regression:

$$\Delta y_t = \rho y_{t-1} + \sum_{j=1}^{p} \phi_j \Delta y_{t-j} + \varepsilon_t \qquad (2.12)$$

And then testing the following hypothesis:

[89] Enders (2004), pp. 181-191
[90] Pedroni (2006), Topic 3, p. 35

$$H_0 : \rho = 0$$
$$H_A : \rho < 0$$

Where the null hypothesis is the existence of a unit root. One final note is necessary, which is the issue of critical values. The critical values will actually depend on whether a constant and/or time trend is included. At the 95% confidence level, the t-statistic is –3.45, -2.89, and –1.95 for models with both a time trend and constant, just a constant, or neither, respectively.[91]

Integrated and Cointegrated Processes

We have shown that, when y_t contains a unit-root, it is non-stationary. Often a non-stationary series can be made stationary by 'differencing' it. For example, while y_t is not stationary, Δy_t could be. To give a simple example, consider again the unit-root process $y_t = y_{t-1} + \varepsilon_t$. Taking the first difference would yield $y_t - y_{t-1} = \varepsilon_t$, which by assumption is i.i.d., and therefore necessarily a stationary process.[92] If the difference of a unit-root is stationary, we say the series is integrated of order one, notated $y_t \sim I(1)$.

Cointegration is the analogous principle applied to the multivariate case. That is, if a linear combination of two variables, each of which contains a unit-root individually, is stationary, we say that the two variables are cointegrated. Intuitively, we can think of cointegrated variables as having a stationary *relationship* in the long-run, even though the variables individually contain a unit-root.[93]

Non-stationary Panel Techniques

---

[91] Enders (2004), p. 183
[92] Ashenfelter et al. (2003), p. 251
[93] Pedroni (2006), Topic 6, p. 12

Panel data has both a cross-sectional and time series component. We call the cross-sectional units of the panel 'members' and generally denote this the 'ith' dimension. By now we are familiar with the time series dimension which we denote 't.' Finally, we can have data on multiple variables for each member over time; we denote this dimension 'm.' Broadly speaking, there are three types of cointegrating relationships in working with panel data. The first is known as 'panel cointegration,' in which two variables are cointegrated *within* the panel members. In this case, we are not interested in how variables are related across members. Why then include the cross-sectional dimension and not just conduct the test for a single member? The reason we use a panel in this case is that it greatly increase the statistical power of our tests. If we had *enough* time series observations for a single member, then a simple unit-root test for that member would be sufficient. However, we rarely have sufficient time series observations to conduct analysis in this manner. But by using panel data, we are able to derive additional information from the ith dimension observations. Importantly, and in contrast to traditional microeconomic panel approaches, non-stationary panel techniques do not require homogeneity in parameters across members, as long as some commonality exists.[94]

The second type of cointegrating relationship is known as 'cross member cointegration,' which is the type that we will be most interested in for the purposes of this paper. This type of analysis examines whether there is a cointegrating relationship among the same variable *across* members. In the case of income convergence, we would be interested in whether income across regions converges to a

[94] Pedroni (2006), Part 2, p. 31

long-run steady state value. Conversely, if there is no such relationship, that is, if the appropriate linear relationship still contains a unit root, this would be evidence of income divergence. We will discuss this methodology in greater detail below.

The final type of cointegration is a combination of the previous two. Specifically, the relationship between a particular variable across members can be stationary *conditional* on one or more other variables. Note however, that this is different than a fixed effects model insofar as we are not assuming that income converges conditioning on a constant. Rather, the additional variable(s) is itself non-stationary.[95]

Advantages to Non-Stationary Panel Approaches to Testing for Convergence

The shortcomings of traditional tests for convergence, such as those discussed in the literature review, are well documented. Recall that these methods regress the average growth rates over a certain time frame on the initial income levels for a cross-section of countries. For illustrative purposes, let's consider a similar (though more simplified) equation to equation (1.20):

$$g_i = \alpha + \beta y_{i0} + \gamma' x_i + \nu_i \qquad (2.13)$$

where i is the index of countries, g_i is the average growth rate between periods 0 and T, x_i is a vector of variables that control for cross-country heterogeneity and ν_n is an error term. Evans (1997) demonstrates that the estimates of β will be inconsistent except under the most stringent circumstances. Specifically, Evans demonstrates that it must be true that deviations between y_{it} and $\bar{y}_t$ are first order autoregressions with a common autoregressive parameter σ such that $0 < \sigma < 1$. If this condition fails,

[95] Pedroni (2006), Part II, p. 13

then it must be the case that y_{i0} and ν_i are correlated. The second condition is that the vector of control variables must control for all cross-country heterogeneity. Any omitted control variable will be correlated with both y_{n0} and ν_n.[96] It should be apparent that the likelihood of these conditions being satisfied in practice is extremely low.

It is also helpful to examine more informally why we should intuitively expect the cross-sectional beta-convergence test to be a poor tool for analyzing convergence. In essence, the cross-sectional approach attempts to use a snapshot in time to infer something that in reality is dynamic. The estimate of beta assumes that the relationship between initial income values and growth rates holds *at any point in time*. In reality, we might expect that incomes will deviate from their growth path, for example due to fluctuations in the business cycle. Recognizing the time series characteristics of the data allows us to relax the assumption that economies stay close to their growth paths or that the growth paths are linear. Moreover, it is apparent that the arbitrariness of the selection of initial income levels can potentially distort the estimates in the beta-convergence test. In any particular year, an economy could, for various reasons, have an unusually high or unusually low level of income.

Testing for Unit Roots in Panel Data

The most common panel unit root tests apply the ADF test to panel data. However, there are multiple ways in which to accomplish this task. In working with panels, we want to pool the information extracted from the time series dimension and the cross section dimension. We will consider two of the major tests for panel unit

[96] Evans (1997), pp. 219-220

roots, each of which uses a different methodology in the pooling of data. The first test, proposed by Andrew Levin, Chien-Fu Lin, and Chia-Shang James Chu (hereafter, the LLC test), is classified as a pooled 'within-dimension' test. This test allows for heterogeneity in the dynamics across members, but importantly, assumes homogeneity in the parameter of interest, that is, rho.

The LLC test proceeds as follows. First an ADF test is conducted on each member individually. The residuals are used to compute an estimated residual variance for each of the members, which we will later use as a weighting mechanism. Next, two new series are computed for each member using the following estimations:

$$\Delta y_{it} = \tilde{\alpha}_i + \sum_{k=1}^{K_i} \phi_{ik} \Delta y_{it-k} + \varepsilon_{1,it} \qquad (2.14)$$

$$y_{it} = \tilde{\alpha}_i + \sum_{k=1}^{K_i} \phi_{ik} \Delta y_{ik} + \varepsilon_{2,it} \qquad (2.15)$$

We compute the new series as:

$$\Delta \hat{y}_{it}^{*} = \hat{\varepsilon}_{1,it}$$
$$\hat{y}_{it}^{*} = \hat{\varepsilon}_{2,it}$$

Let us take a moment to consider what this step is doing, as it is an important one. Each of these regressions is extracting out the heterogeneous dynamics of the different members from the residuals. Notice, that these dynamics are allowed to vary, that is, we are not restricting the model such that each member includes the same number of lags. We then compute two additional series $\Delta \hat{y}_{it}^{**}$ and $\hat{y}_{it}^{**}$, which are simply $\Delta \hat{y}_{it}^{**}$ and $\hat{y}_{it}^{**}$ weighted by the variances of the residuals for each member.

Using these new series, we estimate the panel regression

$$\Delta \hat{y}_{it}^{**} = \rho \hat{y}_{it}^{**} + \mu_{it}^{**} \qquad (2.16)$$

and compute a t-statistic (t_{ρ_i}) for rho. However, since we do not know the distribution of this statistic, we compute the transformation:

$$Z_{NT}^{LLC} = \nu^{-1/2}(t_{\rho_{NT}} - \mu\sqrt{N}) \qquad (2.17)$$

$$\text{Where } \mu = \frac{E[num(t_{\rho_i})]}{E[den(t_{\rho_i})]} \text{ and } \nu = \frac{Var([num(t_{\rho_i})]}{(E[num(t_{\rho_i})])^2}$$

This test statistic follows a standard normal distribution, so we simply use the one-tailed critical values, and large negative values indicate a rejection of the null of a unit root. It is important to note, and should be clear from the above analysis, that the LLC test assumes that the parameter of interest (rho) is the same for all members of the panel.[97]

It is often the case that this assumption of homogeneity in rho is unreasonable. Thus, it is generally preferable to use the so-called Im, Pesaran and Shin (hereafter IPS) test, categorized as a group mean 'between-dimension' test. Informally, this test pools information by combining the t-statistics from the individual unit root tests of each of the members to form an average. In many ways, the IPS test is much simpler computationally than the LLC test. Specifically, we begin by performing individual ADF tests on each of the members, allowing dynamics to be heterogeneous. We then take the t-statistics estimated for rho and combine them as follows:

$$\bar{t}_{\rho} = N^{-1}\sum_{i=1}^{N} t_{\rho_i} \qquad (2.18)$$

The final test statistic, Z_{NT}^{IPS}, is constructed as:

---

[97] Pedroni (2006), pp. 57-63

$$Z_{NT}^{IPS} = \sqrt{N/\nu}(\bar{t}_{\rho} - \mu)$$ [98] (2.19)

where $\mu = E[t_{p_i}]$ and $\nu = Var[t_{\rho_i}]$

Z_{NT}^{IPS} follows a standard normal distribution, with the appropriate one-sided critical values. It is particularly important to note that the IPS test does *not* assume heterogeneity in the parameter of interest. That is, it could be the case that each member follows a stationary process but with a differing value of rho, say -.5 for one and -.3 for another. This feature of the test because it often the case that we will want to allow for heterogeneity. For example, in testing for convergence, we might want to test for convergence while allowing for different regions to be converging at different rates.

Testing for Income Convergence in Panel Data

With the above analysis, we are almost ready to execute the income convergence test. Recall that, in testing for convergence, we are in practice testing whether there is a cointegrating relationship between income levels in different regions. In other words, we are testing whether $(y_{it} - y_{jt}) \sim I(0)$ for all pairs of i,j. It can be shown that such a test is administered by simply including a common time effect variable, which is convenient because it does not alter the limiting distributions of the unit root test, and moreover, this way we do not have to conduct many pair wise tests. To see how, note that if convergence holds for each pair i,j, then the sum of these pairs must also be stationary:

$$N^{-1}\sum_{j=1}^{N}(y_{it} - y_{jt}) \sim I(0)$$

This expression is equivalent to

[98] Pedroni (2006), pp. 65-66

$$N^{-1}\sum_{j=1}^{N}(y_{it}) + N^{-1}\sum_{j=1}^{N}(y_{jt}) = y_{it} - \bar{y}_t.$$ [99]

So, when $(y_{it} - y_{jt})$ is stationary for all i,j, it must be the case that $(y_{it} - \bar{y})$ is also stationary for all i. And conversely, when $(y_{it} - \bar{y})$ is stationary for all i, it must be the case that all members converge pair-wise. Stated another way, $(y_{it} - \bar{y})$ will be I(1) if $(y_{it} - y_{jt}) \sim I(1)\ \forall$ i, j, and $(y_{it} - \bar{y})$ will be I(0) if, and only if, $(y_{it} - y_{jt}) \sim I(0)\ \forall$ i, j.[100] This result is very useful in our analysis because it means we can test for convergence by simply conducting a panel unit root test on $(y_{it} - \bar{y})$. Note that we will always include a member specific intercept term, which can be interpreted as controlling for fixed effects. Therefore, a result of convergence should be interpreted as conditional, and not absolute, convergence.

It is vital that we be clear about our null and alternative hypotheses when conducting this test. The null hypothesis is that subtracting out y-bar renders a unit-root process for all members. Algebraically, we represent this as follows:

$$\Delta(y_{ti} - \bar{y}_t) = \rho_i(y_{it-1} - \bar{y}_{t-1}) + \sum_{j=1}^{p}\phi_j\Delta(y_{it-j} - \bar{y}_{t-j}) + \varepsilon_t$$

$$H_0 : \rho_i = 0\ \forall i$$

Note that this is not the same as saying that no members are cointegrated pair wise. In fact, the tricky part is realizing, even if we allow for the possibility that some members are converging and some are not, the design of the test statistic is such that our alternative hypothesis must be that *all* (and not merely some) members are cointegrated with y-bar. The reason is that if just one member is diverging from the

[99] Pedroni (2006), Part II, p. 89
[100] Pedroni (2005), p. 9; Evans (1998), p. 297

rest, then the true cointegrating relationship between the other members is not y-bar, but rather some other series. Y-bar is, in a sense, contaminated by the divergent member, as that member is included in the calculation of y-bar. Therefore, our alternative hypothesis, when specified relative to the relationship between y_{it} and $\bar{y}$, becomes:

$$H_A : \rho_i < 0 \ \forall i$$

Note that in economic terms, our null hypothesis is merely that at least one member is diverging (since this is all that is necessary to render $(y_{it} - \bar{y})$ non-stationary for all i). Whereas, our alternative hypothesis is that all regions are converging.

However, although the test can be expected to reject the null in favor of the alternative when a single pair diverges asymptotically, whether it does so in practice depends on the small sample properties of the test. Therefore, it might be the case that when only a small number of members are diverging, we could incorrectly reject the null more often than we should. That is, the contamination effect of including a diverging member in the calculation of y-bar might be so small that, with limited data, our test incorrectly concludes, at conventional significance levels, that some members are cointegrated with y-bar.

To test how sensitive our test is to different numbers of diverging members, we must employ Monte Carlo simulations. For example, as mentioned above, asymptotically, we should only reject the null hypothesis five times out of one hundred at the 5% significance level when even just a single member is diverging from the rest. To test what happens in practice, I constructed 1000 realizations of a simulated panel of 25 members and 30 time periods. I computed how often we reject the null of divergence with a critical value of –1.64, which represents the nominal 5% significance level of the test, for all cases ranging from just one member diverging to all 25 members diverging.[101] These results are presented in Figure 2.1. I also computed the corresponding critical values at the 5% percent significance level for different cases. These critical values are reported in Figure 2.2. Note that for simplicity, I limited this investigation to the use of the IPS test for the null of no cointegration.

Figure 2.1

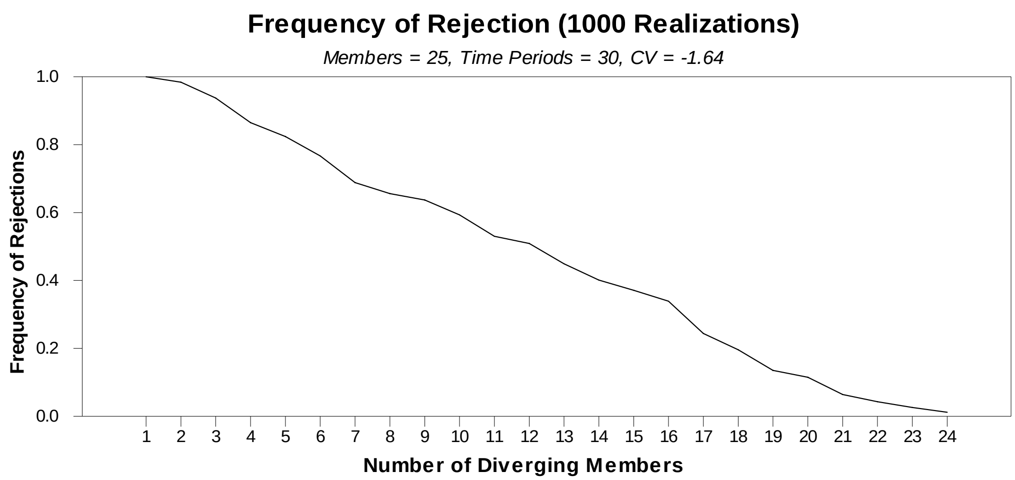

Figure 2.2

Members Diverging	Critical Values at 5% significant level		Members Diverging	Critical Values at 5% significant level
1	-8.87914		13	-4.96527
2	-8.58374		14	-4.27177
3	-8.19135		15	-3.81893
4	-7.58768		16	-3.76738
5	-7.13898		17	-3.28200
6	-7.02734		18	-2.91533
7	-6.60004		19	-2.45215
8	-6.59789		20	-2.12893
9	-5.93545		21	-1.77385

[101] The RATS code for these simulations is included in the appendix

10	-5.83806		22	-1.57098
11	-5.32256		23	-1.36525
12	-5.05790		24	-1.06241

Clearly, with a conventional critical value, we reject the null that $(y_{it} - \bar{y})$ far too often. For example, when just one member is diverging, y-bar is no longer the true cointegrating series for the other members, and therefore, we should conclude that $(y_{it} - \bar{y})$ is non-stationary for all i. In practice, we almost always reject the null. As shown in Table 2.1, the true critical value at the 5% significance level is –8.87914. At first it would seem that a simple solution would be to use this value as or critical value. That way, when we reject the null, we can truly conclude with 95% confidence that all members are converging. However, a Monte Carlo simulation in the case where all members are converging indicates that a test with this critical value has only about 10% power. That is, using –8.87914 as our CV, which would be valid as a test of the null that at least one country is diverging, we only reject the null about 10% of the time in favor of the alternative that they are all converging. This renders the test ineffective against such a minor difference between the null and alternative hypotheses. Rather, a more appropriate strategy would be to amend our null and alternative hypothesis. For example, at our conventional critical value of –1.64, we can safely reject the null hypothesis that all members are diverging pair wise, and conclude that at least some members are converging. If our test statistic is, say, -1.8, we can conclude with great confidence that at least four members are converging.

It might seem odd at first that once 22 or more members are diverging, our critical value is actually *below* the asymptotic critical value. However, upon closer inspection, this finding is not surprising. Note that we are using the asymptotic adjustment values to construct our IPS test statistic. These are appropriate for very large sample sizes, but with more limited data, the true distribution of t-bar could be slightly off. Therefore, that it would differ from asymptotic critical values is to be expected.

Section III: Data

There are several data sets relevant to this project. Our variable income variable is based on real per capita provincial GDP data from 1978 to 1995 (in 1995 yuan) constructed by Hsueh and Li (1999). These figures are updated through 2004 using data from the various editions of the Chinese Statistical Yearbook. We take nominal provincial GDP and divide by reported population to get nominal per capita figures. We then use data on provincial CPI to adjust the nominal data into 1995 yuan so that it is consistent with the Hsech and Li set.

Satisfactory data on Chinese migration is very much more difficult to ascertain. Data from the 2000 census explicitly records provincial migrant populations, however, we need to be careful not to cluster based solely on migrant populations, as this would only capture provinces with significant *in*-migration. Since these provinces are likely to be the wealthiest, we would expect that they should be converging *a priori*. Rather, what we are interested in is whether rich and poor countries are converging with each other in the long-run due to flows of labor. Therefore, we want include those provinces with significant in- *and* those with significant out-migration. We also would like to focus primarily on inter- rather than intraprovincial migration. While intra-provincial migration may provide some information about general conditions of labor mobility within a province, our true interest is in interprovincial migration since we are testing for income convergence across provinces.

The data source that best satisfies our requirements is that constructed by Johnson (2002). Using population data from the 1990 censuses and natural growth rates for the decade, he constructs estimates of the ‘expected’ population for each province.[102] By subtracting this estimate from the observed population recorded in the 2000 census, he constructs an estimate of net in- or out-migration for each province over the period 1990-2000.[103] Dividing this number by total population yield migration estimates as a percentage of population. A negative number suggests that province experienced net out-migration, while a positive number suggests net in-migration. This data is presented in Appendix Figure 3.1. As Johnson himself notes, this data is hardly ideal, and indeed results in much underreporting of out-migration. That being said, it is the best estimate available in the literature for our purposes. In any event, we are more interested in recognizing which provinces are generally open to mobility, as opposed to the exact figures themselves.

[102] Note: Natural Growth Rate = Birth Rate – Death Rate

[103] I tried a similar technique to construct net migration per province for each year. However, the resulting estimates were impossibly small. The most likely reason is that the populations recorded yearly in the China Statistical Yearbooks are based on place of registration and not actually residence. Insofar as this is the case, our estimate would only record migration associated with a permanent change in registration, which would miss the vast amounts of floating migration.

Section IV: Empirical Analysis

Panel Unit Root Test

As a first step in our empirical analysis, we must first establish that the natural log of provincial income is itself non-stationary. Otherwise, testing for cointegration will be meaningless, as the linear combination of two stationary variables is necessarily stationary. Consequently, we begin with a panel unit root test, excluding any common time effects, and, as always, including fixed effects. Figure 4.1 contains the results for the panel unit root tests.

Figure 4.1

Panel Unit-Root Test on Log GDP *No Common Time Effects*	
Test	**Statistic**
Levin-Lin ADF	7.40215
IPS ADF	10.09841

Recall that in order to reject the null we must have test statistics less than -1.28 and -1.64 at the 10% and 5% confidence levels, respectively. Since the test results are both highly positive, we clearly accept the null of a unit root, and consequently, nonstationarity.

Since we have failed to reject the hypothesis that a unit root is present for each member, we can now include a variable for common time effects in order to test for convergence. Recall that this technique essentially tests whether a cointegrating relationship exists between provincial income for each pair of provinces. Figure 4.2 contains the results for the cross-member cointegration test.

Figure 4.2

Cross-member Cointegration Test	
Test	**Statistic**
Levin-Lin ADF	2.58046
IPS ADF	0.17265
Includes all 28 provinces in dataset from 1978-2004; max lags=5	

Since we observe positive test statistics for both the within- and between-dimension tests, we fail to reject the null that all provinces are diverging pair wise. Interestingly, our IPS statistic is noticeably lower than Pedroni and Yao's corresponding statistic of 3.45, based on data from 1978-1997. This difference suggests that by including recent years, provinces have been converging to a greater extent that before. Insofar as factor mobility has increased over the course of the 1990s and 2000s, this finding is consistent with our hypothesis that increased mobility in factors of production would facilitate convergence. Again though, our positive IPS statistic means that we still accept that a significant amount of divergence is occurring in the full set of Chinese provinces.

It is worth taking a moment to illustrate, more concretely, what exactly took place in computing this output. For example, for the IPS statistic, we calculated individual rho values for each province including common time effects. Specifically, we estimated

$$\Delta(\tilde{y}_{it}) = \alpha_i + \rho(\tilde{y}_{it-1}) + \sum_{j=1}^{p} \phi_j \Delta(\tilde{y}_{it-j}) + \varepsilon_t$$

where $\tilde{y}_{it} = y_{it} - \bar{y}_t$ and recall that ρ can differ for each member. Presented below in Figure 4.3 are the individual t statistics for each estimate of ρ:

Figure 4.3

Province	Estimated Rho	t-value on rho-hat
Anhui	-1.200	-4.63
Beijing	-.039	-.357
Fujian	-.053	-1.68
Gansu	-.372	-3.07
Guangdong	-.048	-1.06
Guangxi	-.137	-1.6076
Guizhou	.0012	.0211
Hebei	.036	-0.8479
Heilongjiang	-.080	-1.7106
Henan	.267	-1.3285
Hubei	-.621	-3.905
Hunan	-.108	-0.985
Inner Mongolia	-.272	-2.3323
Jiangsu	-.034	-0.8779
Jiangxi	-.199	-1.4945
Jilin	-.568	-3.131
Liaoning	-.242	-2.2816
Ningxia	-.034	-0.673
Qinghai	-.033	-1.2503
Shaanxi	-.290	-2.5663
Shandong	-.0161	-0.4954
Shanghai	-.189	-2.0222
Shanxi	-.095	-1.2547
Sichuan	-.027	-0.2688
Tianjin	-.093	-1.5924
Xinjiang	-.034	-0.4179
Yunnan	.035	0.3371
Zhejiang	-.024	-0.853

Recall that the IPS test takes these values and computes an average for the entire panel. The final test statistic is then constructed by subtracting the mean of the asymptotic distribution of the ADF t-test and multiplying by the square root of N divided by the theoretical variance of t. At first glance, it may seem surprising that many of the individual test statistics are negative, yet our group-mean statistic for the panel is positive. Recall though, that these individual t-statistics follow a non-standard distribution, and therefore, the p-values of these statistics are not those we associate with say, a standard t-distribution. Moreover, the ADF t-bar distribution is not centered at zero, which is why we must use the adjustment values to construct our final statistic.

We now turn to the possible role of labor mobility in facilitating convergence. As our theoretical analysis suggested, in case of positive externalities to human capital and the absence of labor mobility (irrespective of the mobility of capital) income levels across economies can converge. Labor mobility, it is hypothesized, is the necessary ingredient for convergence. In the case of China, capital mobility is restricted as well, in which case labor mobility could play a role even without positive externalities to human capital. This paper seeks to test, empirically, whether relative degree of labor mobility is facilitating convergence among certain provinces in China. As we saw in chapter two, there are numerous structural and informational barriers to migration, many of which vary by locale. If we are able to separate provinces into those that are characterized by relatively free flow of labor from those where labor continues to be severely restricted, we can test whether the group of provinces with relatively high labor mobility is converging. This technique is known as cluster analysis, as we are 'clustering' provinces based on degree of labor mobility.

We begin by including only those provinces with out-migration greater than 1% of the population or in-migration greater than 2%. This leaves us with 20 provinces; Hunan, Henan, Inner Mongolia, Quinghai, Shaanxi, Lianoning, Hubei and, Hebei are excluded. We again construct a bootstrap to see how sensitive our test is to different combinations of non-cointegrated members. That is, in economic terms, we want to see how likely we are to reject the null of divergence for different numbers of converging members. These results are presented in Figures 4.4 and 4.5.

Figure 4.4

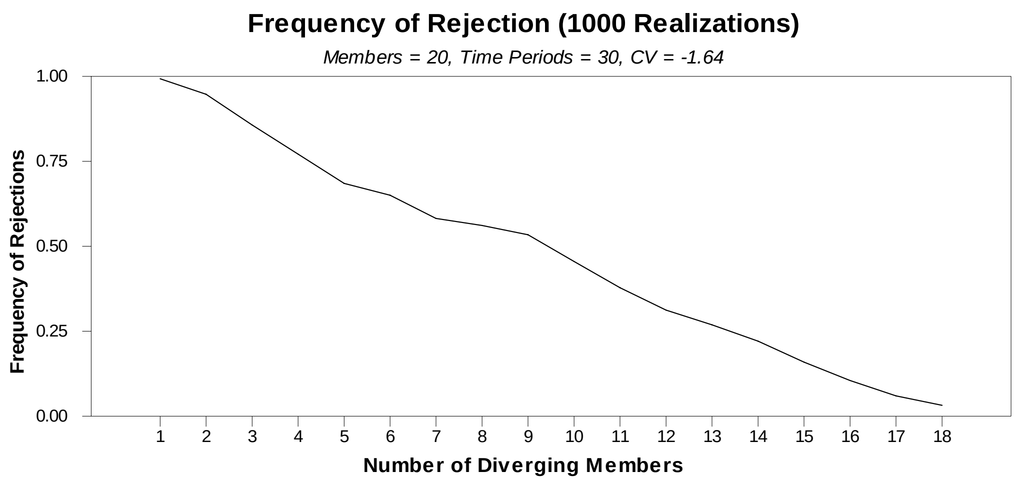

Figure 4.5

Members Diverging	Critical Values at 5% significant level		Members Diverging	Critical Values at 5% significant level
1	-7.842		10	-4.634
2	-7.477		11	-4.039
3	-7.193		12	-3.812
4	-6.617		13	-3.166
5	-6.088		14	-2.730
6	-6.025		15	-2.467
7	-5.803		16	-2.120
8	-5.783		17	-1.716
9	-5.029		18	-1.376

Again, the null in this case at the conventional critical value is that virtually every province is diverging. Thus, a rejection of the null would indicate that a least a handful is converging. Our test results based on the actual data are presented in Figures 4.6 and below.

Figure 4.6

Cross-member Cointegration Test *Cluster A*	
Test	**Statistic**
Levin-Lin ADF	2.29
IPS ADF	1.09
	Includes 20 provinces from 1978-2004; max lags=5

Again for both tests, we fail to reject the null at even the –1.64 critical value. In fact, our IPS statistic increased from that based on all 28 provinces. This means we cannot reject the null hypothesis that all members are diverging.

As a next step, we will try narrowing our cluster even more, this time only including those with greater than 4% out-migration or 3% in-migration. This excludes Jiangxi, Gansu, Yunnan, Jilian, and Shanxi, in addition to those excluded from Cluster A. The results from this test are displayed in Figure 4.7. We then construct Cluster C by excluding all but the top 5 in-migration and top five out-migration provinces. These results are presented in Figure 4.8.

Figure 4.7

Cross-member Cointegration Test *Cluster B*	
Test	**Statistic**
Levin-Lin ADF	2.49
IPS ADF	2.40
	Includes 15 provinces from 1978-2004; max lags=5

Figure 4.8

Cross-member Cointegration Test *Cluster C*	
Test	**Statistic**
Levin-Lin ADF	0.8122
IPS ADF	-1.15525
	Includes 10 provinces from 1978-2004; max lags=5

Cluster C is close to a rejection of the null, but not sufficient. Figure 4.9 lists the critical values based on the Monte Carlo simulations. As we can see, even rejection of all members diverging pair wise, requires a test statistic less than -1.55. Moreover, even a rejection of the null would be suspect considering that the small sample properties of the test could render our distribution completely inaccurate with much fewer than 15 members. Consequently, we are again forced to accept the null that all members are diverging.

Figure 4.9

Members Diverging	Critical Values at 5% significant level
1	-5.775
2	-5.285
3	-4.844
4	-4.326
5	-3.684
6	-3.088
7	-2.569
8	-1.797
9	-1.554

Section V: Conclusions and Avenues for Future Research

In this paper we have considered the possibility that regional divergence in China might be explained by restrictions on labor mobility. This exercise is justified based on a theory, proposed by Razin and Yuan (1997), that labor mobility is necessary for income level convergence in the presence of positive externalities to human capital. As we saw in Section I, internal movement has been stringently restricted throughout much of China's history. Over the last decade or so, these restrictions have increasingly eroded in parts of the country, although significant barriers to movement still exist.

This paper has sought to exploit these developments in testing empirically whether Chinese regional divergence can be explained by a lack of labor mobility. Specifically, we grouped provinces according to the relative pervasiveness of migration and tested whether those with relatively free labor mobility constitute a cluster of converging economies.

To this end, we have utilized newly developed, non-stationary panel techniques, which are superior to traditional, cross-sectional tests for convergence. These panel tests exploit the time series characteristics of the data and recognize that economies frequently deviate from their long-run growth paths. More specifically, we tested for convergence using panel unit-root tests with extracted common time effects. As we showed, the interpretation of the null and alternative hypotheses for these tests can be quite complex. The way the test is constructed, asymptotically, we should accept the null that $(y_{it} - \bar{y})$ follows a unit-root for *all* i, if just a single member is diverging from the rest. Note however, that such an interpretation is conceptually different from accepting that all members are diverging pair wise. To test the extent of this difference in practice, that is, with small sample sizes, we found that in fact the ‘contamination’ effect on y-bar due to individual members diverging, is minimal. Rather, using the asymptotic 5% critical value, we are, in effect, approximately testing the null of divergence for all members against the alternative of convergence for some.

Our empirical analysis concluded that, as a whole, Chinese provinces have continued to diverge into the new millennium. Moreover, based on our tests, we cannot conclude confidently that those provinces with relatively high degrees of labor mobility are not diverging as well. There are numerous possible conclusions that could follow from this result. The first is that labor mobility truly fails to facilitate convergence. As we saw in Section I, some scholars have proposed that migration could actually exacerbate divergence, as occurs in the case of so-called ‘brain drain.’ Moreover, there is some support of this possibility based on evidence that Chinese migrants are, on average, more educated than the general rural population. In addition, as de Brauw and Giles (2005) show, migration could discourage education in source communities by raising its opportunity cost.

However, it would be premature and irresponsible to necessarily conclude, based on our analysis, that freer migration does not lead to regional convergence. First, our clusters are based on a highly imperfect approximation of migration levels. Regrettably data on migration in China leaves much to be desired. Our analysis has made due with the best possible approximation, but hopefully in the future there will be much more accurate and consistent data available. There is also the problem of measurement error in our construction of per capita provincial income. The population figures in the denominator quite possibly reflect population according to local household registration, which would exclude migrants. Insofar as this occurs, it would bias the results against finding convergence.

Second, as was clear from Section I, restrictions on labor mobility still persist, to some degree, in *all* provinces of China. That is, while degrees of mobility differ by region, no two provinces can claim to have anything close to a "free" flow of labor. Even within provinces, rural to urban migration involves significant costs that should result in persistent wage gaps. It would be useful to conduct this paper's analysis again after several years, as factor markets continue to be liberalized and the effects of previous reforms can better take effect.

Finally, it is possible that our clusters are, in reality, exhibiting some convergence, but that our tests failed to pick up on this. Our use of asymptotic correction terms is of questionable accuracy. This concern becomes even greater as we reduced the number of members in the panel as part of our cluster analysis. For example, in Cluster C, we included just 10 members, which made our use of asymptotic values particularly tenuous.

That being said, we have arrived at some notable conclusions. First of all, after updating provincial GDP data for the last seven years, we have reaffirmed the conclusion in Pedroni and Yao (2006) that China continues to experience regional divergence in the post-Reform period. Moreover, while we have not necessarily drawn any broader theoretical conclusions, we can be fairly confident that, in the specific case of China, relative degrees of labor mobility have not (yet) contributed significantly to convergence. Finally, our Monte Carlo simulations of the test for cross-member convergence have provided additional insights into the behavior of these tests in the presence of small sample sizes that has thus far been absent from the literature.

References

Ashenfelter, Orley, Phillip Levine, and David Zimmerman (2003), *Statistics and Econometrics: Methods and Applications,* USA: Wiley & Sons.

Barro, Robert J. and Xaver Sala-I-Martin (1992), 'Convergence,' *The Journal of Political Economy*, 100 (2), 223-251.

Barro, Robert J. and Xaver Sala-I-Martin (1995), *Economic Growth*, New York: McGraw-Hill.

Barro, Robert J. et al. (1995), 'Capital Mobility in Neoclassical Models of Growth,' *The American Economic Review,* 85 (1), 103-115.

Baumol, William J. (1986), 'Productivity Growth, Convergence, and Welfare: What the Long-Run Data Show,' *The American Economic Review,* 76 (5), 1072 1085.

Chan, Kam Wing and Yunyan Yang (1992), "Inter-provincial Migration in China in the Post-1949 Era," *Working Paper.*

Chan, Kam Wing (2001), "Recent Migration In China: Patterns, Trends, and Policies," *Asian Perspective* 25(4), 127-155.

Chen, Jian and Belton M. Fleisher (1996), 'Regional Income Inequality and Economic Growth in China,' *Journal of Comparative Economics,* 22, 141 164.

China National Bureau of Statistics, *China Statistical Yearbook,* various issues, Beijing: China National Statistics Press.

de Brauw Alan and John Giles. (2005), 'Migrant Opportunity and the Educational Attainment of Youth in Rural China,' *Working Paper.*

Demurger, Sylvie et al. (2002), "The Relative Contributions of Location and Preferential Policies in China's Regional Development," *China Economic Review,* 13, 444-465.

Doepke, Matthias (2003), Unpublished Notes, [http://www.econ.ucla.edu/doepke/teaching/c32/sec3.pfd].

Du, Yang, et al. (2005), "Migration and Rural Poverty in China," *Journal of Comparative Economics,* 33, 688-709.

Easterly, William (2002), *The Elusive Quest for Growth,* Cambridge, MA: The MIT Press.

Enders, Walter (2004) *Applied Econometric Time Series: 2^{nd} Edition*, USA: Wiley & Sons.

Evans, Paul (1997) "How Fast do Economies Converge?" *Review of Economics and Statistics*, 79, 219-225.

Evans, Paul (1998) "Using Panel Data to Evaluate Growth Theories," *International Economic Review*, 39(2), 295-306.

Galor, Oded (1996), 'Convergence? Inferences from Theoretical Models,' *The Economic Journal*, 106 (437), 1056-1069.

Hare, Denise (1999), "Push Versus Pull Factors in Migration Outflows and Returns," *Journal of Development Studies*, 35(3), 45.

Johnson, D. Gale (2003), "Provincial Migration in China in the 1990s," *China Economic Review*, 14, 22-31.

Johnson, Paul and Lisa Takeyama (2003), "Convergence Among the U.S. States: Absolute, Conditional, or Club?" *Working Paper*.

Jones, Charles I. (1998), *Introduction to Economic Growth*, New York: W.W. Norton & Company.

Liang, Zai and Zhongdong Ma (2004), "China's Floating Population: New Evidence from the 2000 Census," *Population and Development Review*, 30(3), 467-488.

Lucas, Robert E. Jr. (1988), 'On the Mechanics of Economic Development,' *Journal of Monetary Economics*, 22, 3-42.

Mankiw, N. Gregory, et al. (1992), 'A Contribution to the Empirics of Economic Growth,' *The Quarterly Journal of Economics*, 107 (2), 407-437.

Pedroni, Peter and James Yudong Yao (2006), 'Regional Income Divergence in China,' *Journal of Asian Economics*, 17, 294-315.

Pedroni, Peter (2004), 'Social Capital, Barriers to Production and Capital Shares: Implications for the Importance of Parameter Heterogeneity from a Nonstationary Panel Approach,' [http://www.williams.edu/Economics/wp/pedronisocialcapital.pdf].

Pedroni, Peter (2006), 'Empirical Methods for Macroeconomics: Unpublished Lecture Notes."

Ping, Huang, and Frank Pieke (2003), "China Migration Country Study," *Department* *for International Development.*

Raiser, Martin (1998), ‘Subsidising Inequality: Economic Reforms, Fiscal Transfers, and Convergence across China,’ *Journal of Development Studies*, 34 (3), 1-26.

Razin, Assaf and Chi-Wa Yuen (1997a), ‘Income Convergence within an Economic Union: the Role of Factor Mobility and Coordination,’ *Journal of Public Economics*, 66, 225-245.

Razin, Assaf and Chi-Wa Yuen (1997), ‘Factor Mobility and Income Growth: Two Convergence Hypotheses,’ *Review of Development Economics*, 1 (2), 171-190.

Rozelle, Scott et al. (1999), “Leaving China’s Farms: Survey Results of New Paths and Remaining Hurdles to Rural Migration,” *The China Quarterly* 158, 367-393.

Solow, Robert M. (1956), ‘A Contribution to the Theory of Economic Growth,’ *The Quarterly Journal of Economics*, 70 (1), 65-94.

Tamura, Robert, (1991), ‘Income Convergence in an Endogenous Growth Model,’ *Journal of Political Economy*, 99, 522-540.

Taylor, J. Edward, et al. (2003), “Migration and Incomes in Source Communities: A New Economics of Migration Perspective from China,” *Economic Development and Cultural Change*, 52(1), pp. 75-101.

Weil, David N. (2005), *Economic Growth*, New York: Addison Wesley.

Weir, Shandra and John Knight (2000), “Education Externalities in Rural Ethiopia: Evidence from Average and Stochastic Frontier Production Functions,” *Centre for the Study of African Economies,* Oxford University.

Whally, John and Shunming Zhang (2004), “Inequality Change In China and (Hukou) Labor Mobility Restrictions,” *Working Paper*.

Wu, Xiaogang and Donald J. Treiman (2004), “The Household Registration System and Social Stratification in China: 1955-1996”, *Demography*, 41 (2), 363-384.

Yang, Dennis Tao and Hao Zhou (1999), “Rural-urban Disparity and Sectoral Labour Allocation,” *Journal of Development Studies*, 35(3), 105-133.

Yao, Shujie and Zongyi Zhang (2001), “On Regional Inequality and Diverging Clubs: A Case Study of Contemporary China,” Journal of Comparative Economics 29, 466-484.

Young, Alwyn (2000), ‘The Razor’s Edge: Distortions and Incremental Reform in the People’s Republic of China,’ *Quarterly Journal of Economics*, 115 (4), 1091 1135.

Zhang et al. (2004), “China’s Rural Labor Market Development and its Gender Implications,” China Economic Review 15, 230-247.

Zhao, Yaohui (1999a), “Leaving the Countryside: Rural-to-Urban Migration Decisions in China,” *The American Economic Review*, 89(2), 281-286.

Zhao, Yaohui (1999b), “Labor Migration and Earnings Differences: The Case of Rural China,” *Economic Development and Cultural Change*, 47(4), 767-782.

Zhao, Yaohui (2003), “The Role of Migrant Networks In Labor Migration: The Case of China,” *Contemporary Economic Policy*, 21(4), 500-511.

Zhao, Zhong (2004), “Migration, Labor Market Flexibility, and Wage Determination in China: A Review,” *The Developing Economies*, 43(2), 285-312.

Zhu, Nong (2002), “The Impacts of Income Gaps on Migration Decisions in China,” *China Economic Review*, 13, 213-230.

Appendix

Figure 1.1

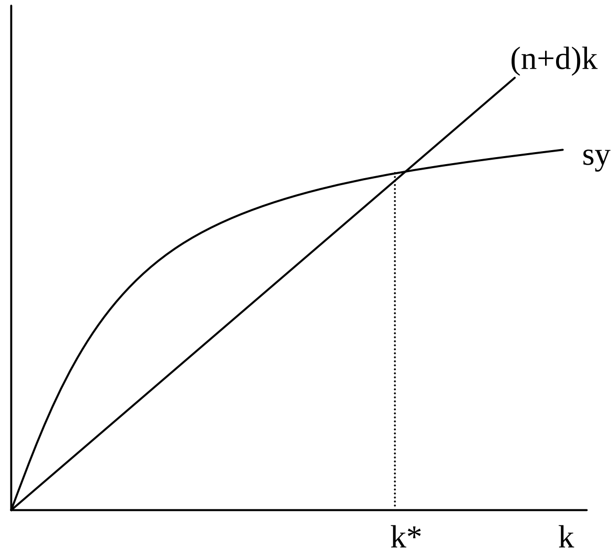

Figure 1.2

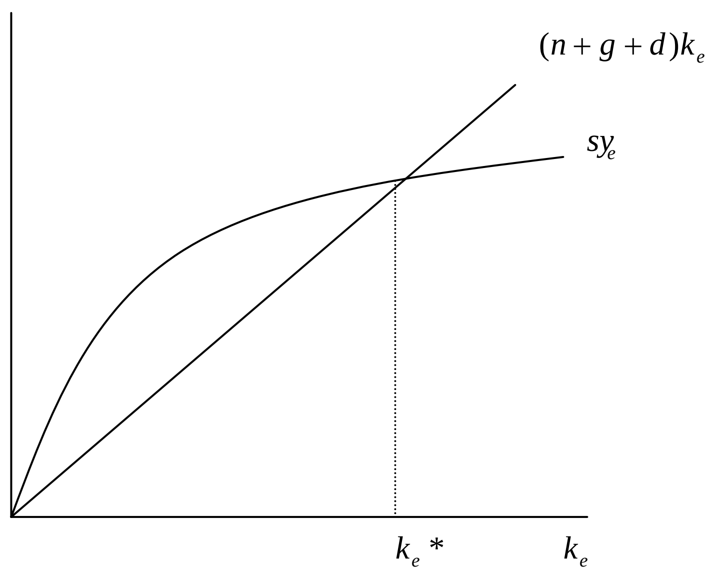

Figure 4.1

Province	Estimated Migration (percent)
Beijin	24.4
Tianjin	9.3
Hebei	1.9
Shanxi	3.1
Inner Mongolia	-0.7
Liaoning	0.2
Jilin	3.1
Heilongjian	-3.9
Shanghai	26.2
Jiangsu	4.5
Zhejiang	6.2
Anhui	-4.1
Fujian	5.3
Jiangxi	-3
Shandong	4.6
Henan	-0.7
Hubei	1.4
Hunan	-0.8
Guangdong	23.9
Guangxi	-5.9
Hainan	3.8
Chongqing	-3.8
Sichuan	-4.6
Guizhou	-7
Yunnan	2
Tibet	-1.8
Shaanxi	-0.7
Gansu	-1.2
Qinghai	-0.7
Ningxia	-4.1
Xinjiang	13.2

RATS Instructions

```
compute nsecs = 10 ; *enter number of members
compute tperiods = 30
decl vect[series] Y(nsecs)
decl vect[series] ytilde(nsecs)
decl vect[series] dytilde(nsecs)
decl vect[real] tstats(nsecs)
decl vect[series] Ytemp(nsecs)
decl vect[series] zstatseries(9) ; *MAKE SURE DIMENSION EQUAL TO
LARGEST U
decl vect[real] CV(9) ; *MAKE SURE DIMENSION EQUAL TO LARGEST U

do u=1,9 ; *U is number of DIVERGING MEMBERS
```

```
compute reject = 0
compute maxco = nsecs - u ;*enter number of converging members
compute mlag = 5 ;*enter starting lags
compute div = nsecs - maxco
allocate 130
seed 2002

do l=1,1000

*SIMULATE SERIES
set xtemp 1 130 = 0.0
set xtemp 2 130 = xtemp{1}+%ran(1)
set x 1 tperiods = xtemp(t+100)

do i=1,maxco
 set Y(i) 1 tperiods = 0.0
 set Y(i) 1 tperiods = x + %ran(1)
end do i

do i=maxco+1,nsecs
 set Ytemp(i) 1 130 = 0.0
 set Ytemp(i) 2 130 = Ytemp(i){1} + %ran(1)
 set Y(i) 1 tperiods = Ytemp(i)(t+100)
end do i

set ybar 1 tperiods = 0.0
do i=1,tperiods
 do k=1,nsecs
 set ybar i i = ybar + Y(k)
 end do k
end do i
set ybar = ybar/nsecs

do i=1,nsecs
set ytilde(i) = y(i) - ybar
end do i

do i=1,nsecs
 dif ytilde(i) / dytilde(i)
end do i

do j=1,nsecs
 do llags=mlag,1,-1
  linreg(noprint) dytilde(j)
  # dytilde(j){1} dytilde(j){1 to llags} constant
  compute mtratio = %beta(llags+1)/sqrt(%seesq*%xx(llags+1,llags+1))
```

```
  if abs(mtratio)>= 1.64
   { ; compute maxlag=llags ; break ; }
  end do llags

  if llags == 1 .and. abs(mtratio)<1.64
  { ; compute maxlag = 0 ; }

 if maxlag == 0
 { ; linreg(noprint) dytilde(j)
 # ytilde(j){1} constant ; }
 else
 { ; linreg(noprint) dytilde(j)
 # ytilde(j){1} dytilde(j){1 to maxlag} constant ; }

 compute tstats(j)= %tstats(1) ;*This is the vector of all the rho tstats
end do j

compute tbar = %avg(tstats)
compute zstat = (sqrt(nsecs/.71))*(tbar+1.54)

if zstat<=-1.64 ; { ; compute reject = reject+1 ; }

set zstatseries(u) l l = zstat

end do l

statistics(fractiles,noprint) zstatseries(u)

set rejectseries u u = reject
compute CV(u) = %fract05

end do u

set rejectseries2 = rejectseries / 1000

graph(header='Frequency of Rejection (1000 Realizations)',$
  subheader='Members = 20,Time Periods = 30, CV = -1.64',$
  hlabel='Number of Diverging Members',$
  vlabel='Frequency of Rejections') 1
# rejectseries2
display CV
```