UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

THREE ESSAYS ON GROWTH AND DEVELOPMENT

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

NORMAN AARON MAYNARD
Norman, Oklahoma
2011

THREE ESSAYS ON GROWTH AND DEVELOPMENT

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF ECONOMICS

BY

Dr. Kevin Grier, Chair

Dr. Scott Linn

Dr. Robin Grier

Dr. Benjamin Keen

Dr. Daniel Hicks

DEDICATION

to

My parents,

Frank and

Mary Maynard; and

My wife

Jessie Maynard

For

All their love and support on a long journey

Acknowledgements

I would like to thank the members of my committee, Yifei Ding, Brian Piper, Beatriz Maldonado-Bird, and seminar participants at the University of Oklahoma. Any remaining errors are my own.

Contents

List of Tables

List of Figures

Chapter 1

Long-Run Growth Differences and the Neoclassical Growth Model

1.1 Introduction

The neoclassical growth model (hereafter NGM) has had some success in explaining differences in income levels across countries.[1] When combined with the criticism of endogenous growth models by Jones (1995a,b), this success has confirmed the NGM as the cornerstone of long-run macroeconomics. The NGM is far less satisfactory when considering long-run growth rates. One of the key implications of the standard NGM is that there is a single, world-wide long-run growth rate of output due to the nature of technological knowledge. This view of technology growth as a pure public good is well articulated by Mankiw et al. (1992, MRW hereafter): "We assume that g [is] constant across countries. g reflects primarily the advancement of knowledge, which is not country-specific." Yet Lee et al. (1997) provide evidence that per capita

[1] I use the term NGM to mean predominantly the general exogenous savings model originally proposed by Solow (1956), along with the added simplifying assumptions suggested by Solow (1957) and generally made in the subsequent literature.

output growth rates do in fact differ across countries. The challenge to the NGM is compounded when considering, as pointed out by Grier and Grier (2007), that the world income distribution is diverging despite clear non-divergence (and in most cases significant convergence) in the standard determinants of income levels.

So far, despite the criticisms, many economists still favor the NGM for the simplicity of its analysis and the clear policy implications that are the hallmark of the model. In order to address some of the NGM's shortcomings, economists often modify or augment the model in order to account for the empirical anomalies mentioned above. One of the more common alternatives is to identify factors not usually included in the theoretical model that are important for explaining growth. Examples include human capital as in MRW, institutional differences as in Hall and Jones (1999), and geography as in Sachs (2001).

The approach of this paper is unique in that rather than augmenting the model or estimating atheoretical growth regressions, I relax some of the common simplifying assumptions of the NGM that are not economically motivated. The most common simplifying assumptions are

- The production function is Cobb-Douglas
- Factor markets are perfectly competitive (that is, factor elasticities in the production function are also factor income shares)
- There are constant returns to scale in capital and labor
- Factor shares are identical across all countries, allowing pooling of the data in cross-country analysis
- Technology is Harrod neutral, or labor-augmenting

While this paper retains the assumptions of Cobb-Douglas production and perfectly competitive factor markets, it relaxes the latter three assumptions. These simple changes give the result that the growth rate of GDP per capita is no longer equal to the growth rate of labor-augmenting technology, as is implied by the standard NGM. Output growth with this more general production function also depends on the output elasticities of labor and natural resources, which under competitive markets are equal to their respective factor income shares.[2]

In particular, if technology growth is exogenous and common across all countries but countries differ in their output elasticities for various factors, they will still have differing long-run growth rates in output. Such differences could potentially explain the continued divergence in output levels despite convergence in output determinants as documented in Grier and Grier (2007).

This paper also empirically examines how well factor elasticity differences across countries can explain differences in output growth rates, and whether the predicted differences in growth rates are sufficient to explain the divergence pattern in incomes.[3] I utilize the structure of the theoretical model (namely, that pooling the data in a cross-section is appropriate if technology growth rates are treated as parameters), along with data on factor income shares for 51 countries (using the best measure of capital), to determine whether the variance in long-run growth rates projected by the model can account for the observed variance in long-run growth rates across countries.[4]

Although several measures of factor income shares are presented for completeness, the preferred results utilize factor share data for total capital and reproducible

[2]This result is relatively intuitive, since labor augmenting technology is no longer improving all non-capital resources.

[3]Zuleta (2008), using an endogenous growth model, also makes a connection between factor income shares and long-run growth, but the spirit and mechanics of his model differ substantially from mine.

[4]Although factor shares do vary between countries, I assume that they are constant over time within countries (a result consistent with the claims of Gollin, 2002). This insures that analysis along the balanced growth path is valid.

capital from Caselli and Feyrer (2007). This data extends the work of Gollin (2002), who argues that this labor share data is less likely to suffer significant measurement error than more conventional measures based on corporate employee compensation. The Caselli and Feyrer data also provides the best fit for the model and the most reasonable parameter estimates.

I find that, after relaxing the simplifying assumptions regarding factor income shares, technology, and returns-to-scale, the model is able to account for 45% to 75% of the variation in long run output per capita growth rates between countries (depending on the specification). In addition, the preferred results imply a labor-augmenting technology growth rate of 1.6% to 2.1% per year, and equally plausible estimates of technology growth augmenting capital and land. I perform several robustness checks, finding that in all cases the parameter estimates as a group are statistically significant at any level desired.

This result suggests that these seemingly innocuous assumptions can signficantly affect the results of cross-country comparisons, and that empirical work in this area needs to take their affects into account. It also indicates that the NGM can account for a substantial portion of cross-country growth differences while retaining the assumption of Cobb-Douglas production, despite a significant literature to the contrary (discussed below). Although this analysis does not address major swings in within-country growth rates, this version of the NGM is able to account for a sizable portion of the variation in cross-country growth.

The rest of the paper is organized as follows. The next section discusses some of the literature on the NGM and presents the previous work most directly related to my approach. Section 3 presents the basic theory behind the standard Solow model and the version which relaxes the simplifying assumptions. Section 4 describes the empirical version of the model and discusses the data used. Section 5 presents the results and offers some discussion. Section 6 concludes.

1.2 Previous Literature

Following the initial work of Solow (1956, 1957), one of the most influential recent works on the NGM is MRW, which showed that the NGM (when augmented by including human capital along with physical capital) does a good job of accounting for differences in cross-country income level differences. This approach was extended to a panel framework by Islam (1995) and Caselli et al. (1996). The implications of the model are supported by Sala-i Martin (1996) and Barro and Sala-i Martin (2002). In contrast to MRW, Lee et al. (1997) provide evidence that countries belong to multiple growth regimes, as opposed to the single growth regime suggested by the NGM.

The case against the NGM has received more attention in recent years.[5] Easterly and Levine (2001) argue that many predictions of the NGM, such as a constant cross-country steady state growth rate which is independent of policy variables, are unquestionably inconsistent with empirical patterns of growth; a similar argument is made by Hausmann et al. (2005). This claim certainly seems in line with the evidence from Pritchett (2000), Jerzmanowski (2006), and Jones and Olken (2008).

Bernanke and Gurkaynak (2002) build on MRW by conducting a more targeted test of the restrictions imposed by the Solow model, rejecting the implication that the savings rate is independent of output. Grier and Grier (2007) show that income continues to diverge over time, despite strong evidence that the determinants of virtually every version of the NGM are either converging or at least holding constant over time. All of these criticism suggest that one or more of the assumptions driving the neoclassical results are in error.

[5] I do not offer an exhaustive list of the criticisms of the NGM here. What follows is only a small sampling from an extensive literature.

A significant recent literature has focused on the Cobb-Douglas production function as the possible source of this counterfactual. Duffy and Papageorgiou (2000), using panel data, reject the Cobb-Douglas specification in favor of a heterogeneous CES, in which the elasticity of substitution between physical and human capital is less than one for developing countries but greater than one for the wealthiest countries. Masanjala and Papageorgiou (2004) and Miyagiwa and Papageorgiou (2007) provide subsequent support for these conclusions. Klump et al. (2007) reach different conclusions on the value of the elasticity of substitution but still reject the unitary elasticity of substitution implied by Cobb-Douglas.[6] In contrast to these results, I retain the assumption of Cobb-Douglas aggregate production and identify how much of the variation in growth rates may be accounted for by relaxing assumptions related to other facets of production.

In particular, I use a more complex version of the NGM's Cobb-Douglas production function. I then combine the theoretical structure with the empirical data on cross-country factor share differences from Caselli and Feyrer (2007) to produce empirical results.

1.3 Theoretical Background

1.3.1 The Standard Neoclassical Growth Model

In a typical representation of the Solow model, output is created by a concave production function with capital and effective labor as inputs, $Y(t) = F(K(t), A(t)L(t))$. Assuming the production function is Cobb-Douglas, the production process is as fol-

[6] A more complete survey of this literature can be found in Chirinko (2008).

lows:

$$Y(t) = K(t)^{\alpha}(A(t)L(t))^{\beta} \tag{1.1}$$
$$\alpha, \beta > 0, \quad and \quad \alpha+\beta = 1,$$

where Y denotes output, K denotes capital stock, A denotes labor-augmenting technological progress, and L denotes labor. By also assuming that markets are perfectly competitive (as is typically done in this literature), α and β represent capital and labor shares of national income, respectively.

Since this production process is constant returns-to-scale (CRS), a typical approach is to divide both sides by effective labor and analyze the model in intensive form. However, an equally valid approach is to examine the dynamics of the model directly and identify the balanced growth path. Since this approach will simplify the generalization I use in the following section, I also follow it below.

Assuming the standard laws of motion and common parameterizations,[7] finding a balanced growth path means the growth rate of capital must be constant, which requires that Y/K be constant.[8] This implies that on the balanced growth path the growth rates of output and capital must be equal, or $g_Y(t) = g_K(t)$. Applying this implication to production function (having taken logs of both sides and differentiated with respect to time) yields

$$g_Y(t) = \alpha g_Y(t) + \beta(n+g), \tag{1.2}$$

[7]That is, $\dot{K}(t) = sY(t) - \delta K(t)$, $\dot{L}(t) = nL(t)$, and $\dot{A}(t) = gA(t)$, where dots represent time derivatives of the respective variables; $s > 0$ and $\delta > 0$ are the savings and depreciation rates, respectively; $n > 0$ is the long-run growth rate of the population; and $g > 0$ is the long-run growth rate of technology.

[8]Since $g_K(t) = \dot{K}(t)/K(t) = sY(t)/K(t) - \delta$.

and given $\beta = 1 - \alpha$, this implies

$$\begin{aligned} g_Y &= n + g \\ \Rightarrow g_{Y/L} &= g. \end{aligned} \tag{1.3}$$

Thus, in the standard NGM the long-run growth rate of output per worker is simply the long-run growth rate of labor-augmenting technology. If knowledge is a purely public good, as is typically suggested by the neoclassical framework, then all countries should converge to the same long-run growth rate, and only ongoing divergence in the determinants of steady state income should lead to widening income differences. Essentially, the standard NGM can account for zero variation in long-run growth rates. Unless we are willing to say that all of the variation in observed growth rates is due to transition dynamics (despite convergence in the determinants of income levels), some modification to the standard NGM is needed.

1.3.2 The General Technology Case

The above is only true if technology is Harrod-neutral and output is CRS in capital and labor. However, the Solow model can operate with a more general production process.[9] Consider the Cobb-Douglas production function

$$Y(t) = A(t)\, K(t)^{\alpha}\, L(t)^{\beta}\, T(t)^{\gamma} \tag{1.4}$$

$$\alpha, \beta, \gamma \geq 0, \quad and \quad \alpha + \beta + \gamma = 1,$$

where T denotes productive land (or any other non-reproducible input to production). Assume that the Hicks-neutral technological progress $A(t)$ can be decomposed into

[9] The following is based on Nordhaus (1992) and Romer (2006).

components augmenting each input to production according to the following equation:

$$A(t) = A_K(t)^{\alpha} A_L(t)^{\beta} A_T(t)^{\gamma}. \tag{1.5}$$

If $A_K(t) = A_T(t) = 1$ and γ is zero (implying α and β sum to one), $A(t)$ is the standard labor-augmenting technology / CRS in capital and labor case. Substituting (1.5) into (1.4) and decomposing this into growth rates yields

$$\begin{aligned} g_Y(t) &= g_A(t) + \alpha g_K(t) + \beta g_L(t) + \gamma g_T(t) \\ &= [\alpha g_{AK}(t) + \beta g_{AL}(t) + \gamma g_{AT}(t)] \\ &\quad + \alpha g_K(t) + \beta g_L(t) + \gamma g_T(t), \end{aligned} \tag{1.6}$$

where the growth rates of each input augmenting technology may be unique.

In the absence of ongoing changes to national borders, no country can expand its productive land. To reflect this fact, I assume $g_T = 0$. As before, in the steady state capital and total output must grow at the same rate, which implies the following steady state growth in total output

$$g_Y(t) = \frac{\alpha}{1-\alpha} g_{AK}(t) + \frac{\beta}{1-\alpha} g_{AL}(t) + \frac{\gamma}{1-\alpha} g_{AT}(t) + \frac{\beta}{1-\alpha} n, \tag{1.7}$$

which implies that the long-run growth rate of output per worker is

$$g_{Y/L} = \frac{\alpha}{1-\alpha} g_{AK} + \frac{\beta}{1-\alpha} g_{AL} + \frac{\gamma}{1-\alpha} g_{AT} - \frac{\gamma}{1-\alpha} n. \tag{1.8}$$

This result differs from the standard result in two ways.

First, the long-run growth rate of output per worker depends on factor income shares and population growth, not just technology growth. This result is due to the introduction of land to the production function, which eliminates the CRS in K and L assumption. With only capital and labor in the production process, the long-run growth equation would be

$$g_{Y/L} = \frac{\alpha}{1-\alpha} g_{AK} + g_{AL}. \tag{1.9}$$

In this case, in addition to the growth due to labor-augmenting knowledge, some additional amount of growth would result due to better technology augmenting capital goods. One plausible interpretation of this capital augmenting technology is an improvement in the average quality of capital goods, although this is by no means the only possible meaning of the growth term.

Second, the long-run growth rate of output per worker depends on multiple technology growth rates, not just labor-augmenting technology. This is due to the general multiplicative technology used. Without this generalization, the long-run growth equation would be

$$g_{Y/L} = \frac{\beta}{1-\alpha} g_{AL} - \frac{\gamma}{1-\alpha} n. \tag{1.10}$$

Aside from the drag caused by population growth, notice that the growth rate of output per worker in this case is not directly proportional to the growth rate in technology. The reason for this is straightforward: since the model only has labor augmenting technology, land's contribution to output does not develop as knowledge improves.

1.4 Empirical Approach and Data

Most empirical models of the NGM have some similar elements to the approach described above, but they differ in several important respects. First, they are typically estimated on levels rather than growth rates. Second, growth rates in population and technology (as well as savings and depreciation rates) are treated as data while the estimated parameters are functions of the output shares, thus implying that all countries have the same αs and βs (and γs, if land were to be an included factor).[10] This second aspect is particularly problematic, since previous research in a panel setting has indicated that pooling the output elasticities across countries is not supported by the data (see Grier and Tullock, 1989; Durlauf and Johnson, 1995).

If the output elasticities (which are also factor income shares when markets are competitive) are taken to be identical for all countries, then the NGM still predicts all countries must have the same long-run growth rate, and the observed divergence in per capita incomes can only be explained by ongoing divergence in the determinants of the balanced growth path, such as the savings rate, population growth rate, and depreciation.[11] This must be the case since the NGM assumes that technology is a Samuelsonian public good, and thus any increase in knowledge must also be identical for all countries.

While maintaining the assumption of knowledge as a public good, I relax the assumption that all countries must have the same factor income shares. I reverse the usual paradigm by using data on factor shares, which differ across countries, to estimate factor-specific technological growth rates, which should be pooled across countries according to the NGM.

If the output elasticities vary across countries, as they most certainly do, equation

[10] Owen et al. (2009) estimate a growth rate equation based on a neoclassical structure and estimate the number of growth regimes (implicitly determining the number of distinct αs and βs) using a finite mixture model. They primarily focus on the evidence that there are multiple components to the mixture rather than on implications for the NGM.

[11] Although differences in population growth rates in the above model would lead to differences in ouput growth, the term $\gamma/(1-\alpha)$ scales down these small differences even more, making it exceedingly unlikely that growth rates would differ enough to see the profound divergence observed in the data.

1.8 above becomes

$$\begin{aligned} g_{Y/L,i} &= g_{AK}\frac{\alpha_i}{1-\alpha_i}+g_{AL}\frac{\beta_i}{1-\alpha_i}+g_{AT}\frac{\gamma_i}{1-\alpha_i}-n_i\frac{\gamma_i}{1-\alpha_i}+\varepsilon_i \\ \Rightarrow \tilde{g}_i &= g_{AK}\frac{\alpha_i}{1-\alpha_i}+g_{AL}\frac{\beta_i}{1-\alpha_i}+g_{AT}\frac{\gamma_i}{1-\alpha_i}+\varepsilon_i, \end{aligned} \tag{1.11}$$

where subscripts indicate countries, $\tilde{g}_i = g_{Y/L,i} + \frac{\gamma_i}{1-\alpha_i}n_i$, $\varepsilon_i \sim N(0, \sigma_\varepsilon)$ is a mean-zero i.i.d. country specific error term, and all other parameters are defined as in Section 2. The above cross-sectional regression can be estimated using OLS.

If the estimated technology growth rate coefficients are jointly insignificant, then the model fails even a basic test of applicability to real world data. Since even previous versions of the NGM have passed this test in cross-sectional data, such a result may lead us to question the appropriateness of the data or the transformation suggested by the analysis above rather than the NGM itself. Given that the parameters are jointly significant, how well the variation in predicted growth rates matches the observed variation (as indicated by the R^2 value) can be seen as indicating how much of the divergence in income levels can be accounted for by dropping simplifying assumptions, rather than depending on more fundamental alterations to the model.

1.4.1 Data

I use several measures of factor income shares. The first are based on the "naïve labor shares" (NLS) from Gollin (2002). Gollin's NLS data construct labor shares based directly on corporate employee compensation data, which he argues will overestimate cross-country differences by failing to properly account for the income of small business employees, especially in developing countries. Nevertheless, these measurements are standard in the literature and are thus included as a baseline for comparison. The dates for the NLS factor income measurements vary among the countries from

1982 to 1992. The use of factor shares in these various periods reflects the common assumption that factor shares do not vary significantly over time within countries.

In order to determine land shares of income, I utilize the result from Valentinyi and Herrendorf (2008) that in the US land makes up approximately 1/3 of capital share in the agricultural sector, but is only about 15% of capital share for nonagriculture sectors. I assume that this relationship holds for all countries, reflecting the fact that agricultural technology should be similar across countries under the Solow model. Although this assumption is restrictive, it allows me to estimate the empirical model using the NLS.[12] The data on agricultural sector share of GDP in 1993 also come from Gollin.

I use a version of Gollin's preferred data from Bernanke and Gurkaynak (2002), who provide an expanded list of countries. Caselli and Feyrer (2007) further improve this dataset by distinguishing the factor share of *reproducible* capital (RC) from total capital (TC) for 53 countries.[13] I thus have three sets of factor income data: The first uses NLS data, and calculates capital and land share using data on the agricultural sector; the second uses TC for the capital share, which also calculates land share using data on the agricultural sector, and defines the residual as labor share; the third uses RC as the land share, TC minus RC as the capital share, and the residual as labor share. This final measure is the preferred one, as it does not risk the measurement error in labor shares from the NLS data or in land shares from the assumption regarding agricultural sector shares. The NLS and TC results are included for comparison.

Data on output per capita and population come from the Penn World Tables of Heston et al. (2009). [14] For population and output variables, I determine the long-run

[12] The preferred results use the reproducible capital data of Caselli and Feyrer (2007), which does not require this assumption.

[13] Due to the availability of agricultural sector share, the TC data only include 37 of these countries. The RC data includes 51, dropping two due to output data availability.

[14] The output variables are chain-weighted real per capita GDP (RGDPCH) and real GDP per worker

growth rates using two methods: in the first (AVG) I take the difference between the natural logs of the first and last period and divide by the number of periods; in the second (REG) I calculate a simple linear regression on time over the sample periods for each variable within each country.

I use 18-year average growth rates from 1990-2007 for all regressions reported.[15] These years are selected so that most factor income shares are calculated prior to the growth data, avoiding possible arguments that the growth is driving the factor income shares rather than the other way around; although this is not a concern if factor income shares are roughly constant over time, it still serves as a helpful hedge against a potential data problem. As a robustness check against the possibility that 18 years is insufficient to reflect steady state growth, I also run the regressions beginning 20 years before the last factor income observation, from 1973-2007, and find similar results. Recent evidence by McQuinn and Whelan (2007) suggests that covergence speeds are substantially higher than earlier estimates indicate, and thus the 18 year sample is still preferred.

Descriptive statistics for all variables are available in Table 1.1.[16] It is clear from this table that the empirical growth rates of output per capita vary substantially, with the NLS sample varying from -6% up to +5% growth. Although the TC and RC samples have a smaller range—from around -2.5% to +4% depending on the measure—they still show substantial variation. Most of the countries in the sample fall within 2.5 percentage points on either side of the sample mean. Given the economic significance of a 5% long run growth rate difference among countries, the size of the gap underscores the importance of accounting for the variation.

Two brief comments on the other descriptive statistics are worth making at this

(RGDPWOK).

[15]Due to data availability, Bahrain's growth rates are from 1990-2004.

[16]A full list of countries and which regressions include them is available in Table 1.4.

Table 1.1: Descriptive Statistics

NLS (N=60)	MEAN	ST DEV	MAX	MIN
$\alpha/1-\alpha$	1.038	0.540	2.635	0.2211
$\beta/1-\alpha$	0.758	0.162	0.940	0.184
$\gamma/1-\alpha$	0.242	0.162	0.816	0.060
n (AVG)	0.014	0.010	0.049	-0.011
$g_{Y/L}$ (RGDPCH AVG)	0.015	0.017	0.051	-0.051
$g_{Y/L}$ (RGDPWOK AVG)	0.011	0.017	0.046	-0.055
n (REG)	0.015	0.011	0.053	-0.011
$g_{Y/L}$ (RGDPCH REG)	0.016	0.018	0.060	-0.054
$g_{Y/L}$ (RGDPWOK REG)	0.011	0.018	0.047	-0.057
TC (N=37)				
$\alpha/1-\alpha$	0.426	0.173	0.857	0.207
$\beta/1-\alpha$	0.913	0.036	0.956	0.827
$\gamma/1-\alpha$	0.087	0.036	0.173	0.044
n (AVG)	0.013	0.010	0.049	0.001
$g_{Y/L}$ (RGDPCH AVG)	0.018	0.015	0.049	-0.023
$g_{Y/L}$ (RGDPWOK AVG)	0.012	0.016	0.035	-0.023
n (REG)	0.014	0.010	0.053	0.002
$g_{Y/L}$ (RGDPCH REG)	0.018	0.016	0.060	-0.027
$g_{Y/L}$ (RGDPWOK REG)	0.012	0.016	0.043	-0.028
RC (N=51)				
$\alpha/1-\alpha$	0.238	0.113	0.613	0.031
$\beta/1-\alpha$	0.807	0.102	0.929	0.489
$\gamma/1-\alpha$	0.193	0.102	0.511	0.071
n (AVG)	0.013	0.010	0.049	-0.001
$g_{Y/L}$ (RGDPCH AVG)	0.019	0.014	0.056	-0.023
$g_{Y/L}$ (RGDPWOK AVG)	0.013	0.015	0.038	-0.023
n (REG)	0.014	0.010	0.053	-0.002
$g_{Y/L}$ (RGDPCH REG)	0.019	0.015	0.065	-0.027
$g_{Y/L}$ (RGDPWOK REG)	0.013	0.015	0.044	-0.028

The output variables are chain-weighted real per capita GDP (RGDPCH) and real GDP per worker (RGDPWOK).

point. First, the output growth rate data are very similar whether using the AVG method or the REG method (roughly 1.0% to 1.9% average output growth, with 1.3% to 1.5% average population growth), although the latter leads to slightly higher standard deviations. This suggests that the results from either should be similar, although both are included in all samples for completeness. Second, note that by construction the variables $\beta/1-\alpha$ and $\gamma/1-\alpha$ always sum to 1, whether summing the means or adding the max of one variable to the min of the other. Thus these variables do not represent independent information; rather, they indicate how the $1-\alpha$ share of income is split between labor and land / non-reproducible capital.

1.5 Results and Discussion

The estimation results of equation (1.11), $\tilde{g}_i = g_{AK}\frac{\alpha_i}{1-\alpha_i} + g_{AL}\frac{\beta_i}{1-\alpha_i} + g_{AT}\frac{\gamma_i}{1-\alpha_i} + \varepsilon_i$, are presented in Table 1.2. The four columns indicate the different measures of output, using PPP adjusted constant-price GDP per capita and PPP adjusted constant-price GDP per worker. The table is further broken down into three sections representing the different measures of factor income shares: the standard Naïve Labor Shares and agriculture share of GDP decomposition (NLS); the Total Capital shares calculated by Bernanke and Gurkaynak (2002), also decomposed using agriculture share of GDP (TC); and the Reproducible Capital breakdown from the Caselli and Feyrer (2007) data (RC).

In all cases, the growth rate of labor-augmenting technological progress is positive and statistically significant at the 1% level, ranging between 1.6% and 2.8%. Also in all cases, the parameters are jointly significant at any level ($Prob > F = 0.000$), and the model accounts for roughly 45% to 75% of the variation in the data (depending on specification).

Table 1.2: OLS Regression Results, 18-Year Growth Rates

NLS	RGDPCH (AVG)	RGDPWOK (AVG)	RGDPCH (REG)	RGDPWOK (REG)
g_{AK}	0.008	-0.008	0.003	-0.013
	(0.010)	(0.011)	(0.011)	(0.011)
g_{AL}	0.018***	0.018***	0.020***	0.020***
	(0.004)	(0.004)	(0.005)	(0.005)
g_{AT}	-0.120	0.040	0.006	0.054
	(0.038)	(0.041)	(0.041)	(0.043)
N	60	60	60	60
R^2	0.592	0.454	0.584	0.453

TC	RGDPCH (AVG)	RGDPWOK (AVG)	RGDPCH (REG)	RGDPWOK (REG)
g_{AK}	0.123**	0.117**	0.121*	0.119**
	(0.054)	(0.051)	(0.060)	(0.056)
g_{AL}	0.025***	0.025***	0.028***	0.027***
	(0.006)	(0.005)	(0.006)	(0.006)
g_{AT}	-0.648**	-0.693***	-0.665**	-0.720***
	(0.246)	(0.227)	(0.270)	(0.247)
N	37	37	37	37
R^2	0.730	0.629	0.720	0.617

RC	RGDPCH (AVG)	RGDPWOK (AVG)	RGDPCH (REG)	RGDPWOK (REG)
g_{AK}	0.031*	0.028	0.027	0.029*
	(0.017)	(0.018)	(0.017)	(0.017)
g_{AL}	0.017***	0.016***	0.021***	0.018***
	(0.006)	(0.006)	(0.006)	(0.006)
g_{AT}	0.004	-0.022*	-0.003	-0.030**
	(0.014)	(0.012)	(0.012)	(0.012)
N	51	51	51	51
R^2	0.754	0.652	0.744	0.664

The output variables are chain-weighted real per capita GDP (RGDPCH) and real GDP per worker (RGDPWOK).
White (1980) heteroscedasticity-consistent standard errors are in parentheses
*** indicates statistical significance at the 1% level; ** at the 5% level; * at the 10% level

For the NLS data, the growth rate of labor-augmenting technology is the only parameter that is statistically significant. The other parameters vary widely and change sign depending on output specification. However, if these data introduce noise into the measurement of labor shares, as Gollin and others suggest, this could lead to problems in parameter estimation. There is evidence of this in the NLS data descriptive statistics, as the mean value of $\alpha/1-\alpha$ implies that the average reproducible capital share is greater than 0.5; in contrast, the maximum value implied by the TC data is about 0.46. This is in line with various critiques suggesting that the naïve labor data frequently underestimates true labor shares. For this reason, the NLS estimates should only be seen as a first pass at examining these data.

The TC data produce much stronger results, with every parameter individually significant at the 10% level or higher, and the lowest R^2 (0.617) above the highest for the NLS data (0.592).[17] However, these estimates seem to suffer from problems of their own. First, the sample size is somewhat small (37 countries), which leaves open the possibility that the goodness of fit is an artifact of this particular sample. Second, the growth rate of capital-augmenting technology is implausibly high (11.7% to 12.3% annual growth), while the growth rate of land-augmenting technology is implausibly large and negative (-64.8% to -72% annual growth). Since on average the TC data on $\alpha/1-\alpha$ are roughly five times the magnitude of the data on $\gamma/1-\alpha$, and the parameter estimates reverse this magnitude (and are always of opposite sign), these estimates are likely the result of improper splitting of capital into reproducible and non-reproducible capital.

I next turn to the RC data, which should substantially reduce measurement error problems over the previous two and is thus the preferred measure of factor shares. I find that the growth rate of labor-augmenting technology is again significant at the 1%

[17] Of course, directly comparing the R^2 for different samples is not appropriate; the point is made here for descriptive purposes rather than inference.

level and of plausible magnitude in all cases. Capital-augmenting technology ranges from 2.7% (but statistically insignificant) to 3.1% and significant at the 10% level. Land-augmenting technology ranges from +0.4% but insignificant to -3% and significant at the 5% level. For the various measures of output, the R^2 ranges from 0.652 to 0.754. Thus the best data suggests that as much as 75% of observed differences in growth rates can be accounted for by the Solow model once unnecessary simplifying assumptions are dropped.[18]

A bit of caution is needed when interpreting the coefficients, especially g_{AT}. What precisely does a -3% growth rate on land-augmenting technology really mean? Certainly it indicates that countries depending more heavily on non-reproducible capital will have lower growth rates, but it does not go very far in explaining why. Is it because countries that depend heavily on agriculture have fewer opportunities for investment? Or is it because natural resources are not merely experiencing zero growth, but are actually being used up over time? This suggests some promising areas for future research, as dealing with these details is beyond the scope of this paper.[19]

At this point a final comment on coefficient significance is needed. In the preferred specification, although all three coefficients are significant at the 10% level or higher for some measure of output, the only coefficient that is robust across measures is labor-augmenting technology, g_{AL}. This may suggest that allowing factor income shares to differ across countries and allowing the production function to have decreasing returns to scale in capital and labor are driving the results, and that relaxing the assumption of Harrod-neutral technology is not as important. Retaining the assumption of Harrod-

[18] A possible concern here is that pooling developed countries with less developed countries may not be appropriate. At the 5% significance level, a Chow test for splitting the sample into OECD and non-OECD countries fails to reject the appropriateness of pooling in all four regressions using the RC data.

[19] An additional point on interpretation is that the regressions only include those variables which the theoretical model suggests are important. These results would not be meaningful outside the context of testing the model presented, since a model suggesting additional variables would likely imply an omitted variable bias in the results presented here.

neutrality would imply that the correct model is the one in equation (1.10) rather than the more general equation (1.8) used above. Some caution is needed here as well, though. It is also plausible that labor-augmenting knowledge diffuses quickly across national borders, while technology augmenting other factors of production does not diffuse as readily. As an example, improved management techniques may be readily applied almost anywhere, whereas insect-resistant crops may only provide benefits in the specific land and climate where it was developed.[20] In this case, I would expect the g_{AL} coefficient to be statistically significant in the pooled sample, while g_{AK} and g_{AT} may not be.

The estimation model assumes that the 18-year growth rates represent the balanced growth paths of the countries in the dataset. However, if convergence to the steady state is a slow process, as MRW and others have suggested, then this may not be a long enough period to capture output growth rates in the long run.[21] As a robustness check against this, I construct 35-year growth rates (1973-2007) in order to capture growth rates over a longer run. The results appear in Table 1.3. For the NLS and TC data, the signs, significance, and general magnitude of all estimates are similar to the 18-year data. In terms of goodness of fit, the NLS data improve for all specifications, while the TC data stay roughly the same of suffers somewhat. The most notable change is for the RC data: the magnitude of g_{AL} is cut in half, causing the estimates to become insignificant, while g_{AK} rises to the 4.1% to 5.1% range and is significant at the 5% level for all specifications. The R^2 increases slightly for the AVG output measures, but falls for the REG measures. While this may suggest that technology growth has shifted from capital-augmenting to labor-augmenting over the last several decades, such an interpretation would require a very strong belief that output shares do not

[20] It may be possible to explicitly test this explanation using a technology lag similar to Comin et al. (2008), but this is left to future research.

[21] I am grateful to an anonymous referee for pointing this out.

Table 1.3: OLS Regression Results, 35-Year Growth Rates

NLS	RGDPCH (AVG)	RGDPWOK (AVG)	RGDPCH (REG)	RGDPWOK (REG)
g_{AK}	0.007	-0.005	0.003	-0.011
	(0.010)	(0.011)	(0.011)	(0.012)
g_{AL}	0.021***	0.019***	0.019***	0.019***
	(0.004)	(0.004)	(0.004)	(0.005)
g_{AT}	-0.009	0.031	-0.002	-0.044
	(0.034)	(0.038)	(0.039)	(0.043)
N	60	60	60	60
R^2	0.715	0.608	0.618	0.476

TC	RGDPCH (AVG)	RGDPWOK (AVG)	RGDPCH (REG)	RGDPWOK (REG)
g_{AK}	0.129***	0.114**	0.121*	0.128**
	(0.044)	(0.045)	(0.060)	(0.050)
g_{AL}	0.021***	0.016***	0.028***	0.015**
	(0.005)	(0.006)	(0.006)	(0.007)
g_{AT}	-0.618***	-0.568***	-0.665**	-0.649***
	(0.197)	(0.183)	(0.270)	(0.212)
N	37	37	37	37
R^2	0.745	0.5941	0.643	0.437

RC	RGDPCH (AVG)	RGDPWOK (AVG)	RGDPCH (REG)	RGDPWOK (REG)
g_{AK}	0.046**	0.041**	0.051**	0.047**
	(0.020)	(0.021)	(0.020)	(0.022)
g_{AL}	0.012*	0.008	0.010	0.007
	(0.006)	(0.006)	(0.007)	(0.008)
g_{AT}	0.009	-0.003	0.000	-0.014
	(0.011)	(0.012)	(0.012)	(0.014)
N	51	51	51	51
R^2	0.790	0.670	0.718	0.537

The output variables are chain-weighted real per capita GDP (RGDPCH) and real GDP per worker (RGDPWOK).
White (1980) heteroscedasticity-consistent standard errors are in parentheses
*** indicates statistical significance at the 1% level; ** at the 5% level; * at the 10% level

change even over long time horizons. Since the sample includes many developing countries that we know have experienced significant shifts in their economies since 1973, I prefer not to make this claim over a 35-year horizon and thus prefer the 18-year estimates.[22]

1.6 Conclusion

In this paper I demonstrate that the NGM is able to account for a sizeable part of cross-country growth rate differences. This is made possible by relaxing the assumptions of common cross-country factor income shares, Harrod-neutral technological progress, and CRS in labor and capital—assumptions implemented more for mathematical convenience than economic realism. This suggests that researchers conducting cross-country growth comparisons should be particularly careful about using these assumptions. I am also able to retain the standard assumptions of a Cobb-Douglas aggregate production function and perfectly competitive factor markets, and show that much of the discrepancy between the NGM and observed output growth can still be reconciled.

I estimate a model in which individual countries' output growth rates depend on factor income shares and population growth rates. Utilizing several measures of country-specific factor income shares, I find that relaxing simplifying assumptions can account for between 45% and 75% of observed differences in output growth rates. Although it is important to use caution when interpreting the coefficients, the implied growth rate of labor-augmenting knowledge falls into a very plausible range around

[22]One might also suspect that the R^2 values for the preferred model are high because adding the term $\frac{\gamma}{1-\alpha}n$ to the empirical growth rates soaks up much of the variation. Although this would be an interesting result in itself, it turns out not to be the case. I estimated the regression without correcting for this term, in which case the coefficient on $\frac{\gamma_i}{1-\alpha_i}$ represents $g_{AT} - \bar{n}$ where $\bar{n}$ represents the world-wide average population growth rate. The R^2 falls somewhat, as should be expected, but the model still accounts for 35% to 73% of the variation in all cases.

1.5% to 2% per year.

The model still leaves some questions unanswered. Because of the cross-sectional nature of the analysis, it cannot address the major swings in within-country growth rate trends as documented by Pritchett (2000) and Jones and Olken (2008). Further, depending on the specification the model still leaves between a quarter and half of the variation in long-run growth rates unexplained.

Nevertheless, I show that the weight of evidence brought against the NGM is not as burdensome as it at first appears. The conclusion that the NGM implies all countries should have the same rate of output growth along the balanced growth path depends critically on simplifying assumptions. When these assumptions are relaxed, the NGM falls much more in line with empirical reality.

Table 1.4: Country List

COUNTRY	NLS	TC	RC	COUNTRY	NLS	TC	RC
Algeria	X	X	X	Mauritius	X	X	X
Australia	X	X	X	Mexico	X	X	X
Austria	X	X	X	Morocco			X
Belgium			X	Namibia	X		
Benin	X			Nepal	X		
Bolivia			X	Netherlands	X	X	X
Botswana	X	X	X	New Zealand			X
Bulgaria	X			Niger	X		
Burundi	X	X	X	Nigeria	X		
Cameroon	X			Norway	X	X	X
Canada			X	Panama	X	X	X
Colombia	X	X	X	Papua New Guinea	X		
Congo	X	X	X	Paraguay	X	X	X
Costa Rica	X	X	X	Peru	X	X	X
Cote d'Ivoire	X	X	X	Philippines	X	X	X
Denmark	X	X	X	Portugal			X
Ecuador	X	X	X	Romania	X		
Egypt			X	Rwanda	X		
El Salvador			X	Sierra Leone	X		
Fiji	X			Singapore			X
Finland	X	X	X	South Africa	X	X	X
France	X	X	X	Spain			X
Germany, FRG	X			Sri Lanka	X	X	X
Ghana	X			Sweden	X	X	X
Greece	X	X	X	Switzerland			X
Honduras	X			Thailand	X		
Hong Kong				Trinidad and Tobago	X	X	X
Hungary	X			Tunisia			X
Ireland	X	X	X	Turkey	X		
Israel			X	Ukraine	X		
Italy	X	X	X	United Kingdom	X	X	X
Jamaica	X	X	X	United Rep. of Tanzania	X		
Japan	X	X	X	United States	X	X	X
Jordan			X	Uruguay	X	X	X
Kenya	X			Venezuela	X	X	X
Korea, Rep. of	X	X	X	Viet Nam	X		
Malaysia		X	X	Zambia	X	X	X
Mali	X			Zimbabwe	X		

Chapter 2

Beyond Twin Peaks: Development and Polarization in the World Income Distribution

2.1 Introduction

The influential work of Quah (1993, 1996, 1997) encouraged us to think about "twin peaks" and dynamics in the cross-country distribution of per-capita income. His work popularized the view that this distribution was moving from being uni-modal to bi-modal, with a hollowing out of the middle. Subsequent research has worked to refine and extend Quah's results. In this paper we study the evolution of the distribution of per capita income over 135 countries and 6 decades using variable dimension mixture models. The model assumes that the observed income distribution is composed of an unknown number of individual component densities.

We use a version of the reversible jump Markov Chain Monte Carlo (MCMC) model presented in Richardson and Green (1997) and further developed in Cappé et al. (2003). This approach uses Bayesian methods to probabilistically determine (a) the number of distinct component densities in the overall distribution, (b) the means and variances of each component density, (c) the weight of each component density in the overall mixture, and (d) an allocation of countries to individual component densities. To our knowledge, this is the first application of reversible jump MCMC to the cross-country income distribution, and our novel method produces novel and provocative results.

Our findings imply that as early as 1950, there were multiple components in the cross country distribution of per-capita incomes, and rather than a hollowing out of the middle, more recent decades feature a large middle class of countries. It also seems that the development process fundamentally changed in the 1970s.

For the 1950s and 1960s, the distribution of per-capita income is most likely composed of two separate component densities.[1] We call this period "development", because a large number of countries move from likely being in the poorer component to likely being in the richer component, even as the mean of the richer component is increasing.

During the 1970s, the distribution changes into one that is most likely composed of three individual component densities, with a very large gap opening between the means of the poorest and richest components. This gap increases notably in the 1980s. We label this period "polarization", as the richest group of countries gets smaller and further away from the low and middle income groups.

In the 1990s and 2000s, the allocation of countries to component groups is relatively stable as is the percentage gap between the means of the components. We refer to this period as "hysteresis" because low inter-group mobility means that the poor and middle income countries are seemingly stuck in their relatively disadvantaged

[1]More precisely, the posterior density of the number of components in the mixture has a strong mode at 2.

positions.

Our work speaks to several distinct areas of growth and development. First, we document a different version of unconditional divergence than has been studied in the literature (see Pritchett, 1997, and Grier and Grier, 2007, for examples). In 1950 the gap between the mean income level of the poorest group of countries and the richest group in the distribution is around 5, while by 2008 it is around 22. Our results are also relevant to the poverty trap literature (see Azariadis and Drazen, 1990, and Bloom et al., 2003). We find that the mean income level of the bottom group of countries barely moves between 1950 and 2008 and that 32 countries are sorted into the bottom group with a probability of .66 or greater for all 7 decades that we study. Our results also have implications for club convergence (see Baumol, 1986; Galor, 1996). We consistently find strong evidence of recognizable groups or clubs of countries. However in the 1950s and 1960s we see considerable inter-group mobility. Later in our sample, the number of distinct groups rises but the amount of inter-group mobility falls, suggesting that club convergence may be occurring in our data since 1980.

The rest of the paper proceeds as follows. In Section 2.2 we review some of the important empirical work on the evolution of the world income distribution. Section 2.3 contains a description of mixture models and a brief description of reversible jump MCMC, along with the explicit setup that we use in our analysis. Section 2.4 contains our main results along with discussions of the performance of our algorithm and a sensitivity analysis on some of our priors. Section 2.5 concludes.

2.2 Previous Literature

There are two branches in the empirical literature on the evolution of the world income distribution. The first branch follows Quah in looking for multi-modality via kernel methods. The second branch models the distribution using finite mixture methods.

2.2.1 Multi-Modality papers

Quah's pioneering work used stylized graphs, kernel density plots, and Tukey box plots to forcefully argue that the world income distribution was becoming multi-modal and that the "middle class" of countries was hollowing out. Quah, however, did not provide formal statistical tests for multi-modality. This was first done by Bianchi (1997), whose non-parametric tests confirmed an emerging multimodality over time in the world income distribution.

Henderson et al. (2008) provide an excellent review of this literature along with their own work, which applies a variety of multi-modality tests to a range of different macro variables. Henderson et al. show that the finding of multiple modes in the income distribution is sensitive to the type of test employed, especially when using unweighted data.

While this literature has unquestionably changed how we think about growth and development, it is fair to point out that rejecting uni-modality does not give us very specific information about exactly how many modes or clubs exist in the income distribution.

2.2.2 Finite mixture models

The first paper to apply a finite mixture model to the world income distribution is Paap and van Dijk (1998). They estimate a series of two component mixture models using a Weibull density and a truncated normal density as the underlying components for a sample of 120 countries from 1960 to 1989. While Paap and van Dijk do experiment with what types of densities to use as their two components, they do not consider whether the correct number of components is two, or whether the number of underlying components changes over time.

Pittau et al. (2010) estimate a series of finite mixture models for a sample of 102 countries from 1960 to 2000. They make the important point that the number of underlying components in a mixture model does not have to be the same as the number of visible modes in a Kernel estimate and they argue that a mixture modeling approach may find "hidden" groups that would not be seen from what they call "bump-hunting" exercises. Pittau et al. estimate mixture models with 1,2,3, and 4 components and then use likelihood ratio tests to try and determine the preferred number of mixture components. They exclusively use normal components in their mixture models. Pittau et al. claim that the number of components in the mixture is 3 throughout their sample and find very little movement of countries from one component to the other.

We agree with Paap and van Dijk that, since an income distribution is truncated at zero and skewed to the right, the underlying components should not necessarily be normal. We also agree with Pittau et al. that it is important to estimate the number of components from the data rather than imposing a certain number *ex ante*. Our approach is to directly endogenize the number of components in the mixture by using a reversible jump MCMC method for finite mixtures as pioneered by RG and extended by Cappé et al. (2003). This method is fully Bayesian and produces estimates of the full posterior densities of the model parameters.

The next section briefly explains mixture modeling and then describes how we use reversible jump MCMC to endogenize the number of underlying components in the mixture.

2.3 Empirical Approach

In this section, we present the basic intuition behind our empirical approach, including a discussion of how a reversible jump MCMC algorithm can be used to identify the number of components in the mixture model. We then present the model more formally and specify the prior distribution.

2.3.1 Finite, fixed dimension mixture models

The object of our analysis is a non-standard distribution, and there are theoretical reasons to believe that it is composed of several distinct but unobserved groups (such as the convergence clubs of Baumol, 1986). This sort of problem lends itself naturally to finite mixture modeling, in which a complex distribution is treated as a combination of several more conventional distributions, and the particular component any given observation comes from is unknown.[2]

Typically, the chief objectives of analyzing any mixture model are to estimate (1) the parameters of each component density, (2) the weight of each component in the overall mixture, and (3) the probabilities that any particular observation is drawn from a given component in a given year. The structure of these models lends itself particularly well to using data augmentation and Gibbs sampling to solve them.[3]

Our approach extends the usual Bayesian framework by also estimating (4) the probability of different numbers of component densities in the overall distribution in the model.

[2]It is possible to use mixture modeling to approximate any target distribution with an arbitrary degree of precision by using a large enough number of components. However, we focus on small mixtures because our objective is to sort observations into specific components which are interpretable.

[3]An introduction to this mixture analysis is presented by Lavine and West (1992), Koop (2003), and most general texts on Bayesian analysis. For a more complete discussion of mixture analysis in a Bayesian context, see Robert and Casella (2004) and Früwirth-Schnatter (2006).

2.3.2 Reversible Jump MCMC and Model Selection

Green (1995) developed a generalization of the standard MCMC methods that is able to endogenize the model selection process. Green's reversible jump MCMC technique enables the estimation of the posterior probability of each model among a finite set of alternatives, enabling the statistician to select the best model(s) for inference (usually those with the highest probabilities). This technique has been used to identify the number of components in a mixture by RG and we describe it below.

The problem of determining the number of clubs is essentially a problem of model selection. That is, the statistical question is whether to use a mixture model with one normal component, or three, or any other number. Once the appropriate number of distributions is chosen using the reversible jump MCMC, the rest of the inference can proceed in the standard Bayesian fashion using a Gibbs sampler.

Since the reversible jump MCMC techniques simulate over the entire posterior model space, once the chain has converged the draws from each model represent draws from the true posterior. Thus, the posterior probability of that mixture contains J components may be estimated by

$$P(\mathscr{M} = J \,|\, y) = \frac{1}{T} \sum_{i=1}^{T} I_{\{i=J\}}, \tag{2.1}$$

where T is the total number of draws from the posterior and $I_{i=J}$ is an indicator function that takes the value 1 whenever the model being drawn from contains J components.

2.3.3 Specification of the Model

We model the distribution of the per capita income levels across countries as a finite mixture of lognormal components. We use lognormal components instead of normal components because, given that we are studying an income distribution, we expect it to be left truncated at zero and generally have positive skewness. Thus data from a single lognormal density could easily show up as being made up of multiple normal components, as that would be the only way the normal mixture model could deal with the asymmetry. It is easy to implement a mixture of lognormals model by simply using mixtures of normals to capture the distribution of the log of per-capita incomes.

More formally, we assume that the log-levels of income are drawn from the mixture distribution

$$y_i \sim \sum_{j=1}^{J} \eta_j \mathscr{N}\left(\mu_j,\ \sigma_j^2 \mid \mathscr{M} = J\right), \tag{2.2}$$

which indicates that, conditional on the number of components in the mixture $\mathscr{M}$, any particular country y_i is drawn (with probability η_j, also called the weight of component j) from a Normal distribution with mean μ_j and variance σ_j^2. The weights must each be non-negative and must sum to one.

The techniques we use treat each observation (the log of per capita GDP for a country in a given year) as having been drawn from a single distribution. Since the "club membership" of each observation is unobserved, we use a latent allocation variable z_i for each observation. This allows us to divide the estimation process between estimating the parameters of each component and, given these parameters, estimating the allocation of each observation y_i such that

$$y_i \mid z_i = j,\ \mathscr{M} = J \sim \mathscr{N}\left(\mu_j,\ \sigma_j^2\right). \tag{2.3}$$

Because the allocation variable is randomly drawn from the sampler, it is possible in any particular pass through the sampler (which we will call a 'sweep') that one or more components will have no observations allocated to it, creating an "empty component." The presence of empty components, which will be more frequent when particular components have a low probability of allocation η_j, is essential to the reversible jump algorithm.

For updating parameters given a particular model, we use the Gibbs sampler for within model steps using data augmentation. This step takes the number of components as given, and draws from the posterior of each parameter conditional on previous values of all the others. For jumps between models, we employ the birth and death moves based on RG. This approach begins with J components, some of which might not have any observations allocated to them. An increase in the number of components is accomplished by adding a new empty component, called a 'birth' move, or by removing an existing empty component, called a 'death' move. Whether a proposed move is a birth ($J+1$ components) or a death ($J-1$ components) is chosen randomly with equal probability, and the proposed move is accepted or rejected probabilistically in a manner similar to the Metropolis-Hastings algorithm. A proposed birth move is accepted with probability $\pi = \min\{1, A\}$ (or $\pi = \min\left\{1, A^{-1}\right\}$ for death moves), while the previous draw from the sampler (from the current model) is repeated otherwise.[4]

Unlike Metropolis-Hastings samplers, though, it is not straightforward to say that this can create a Markov chain that must converge to draws from the true posterior distribution across models with different parameters, and especially with different numbers of parameters as we have here. The key advance produced by Green (1995) was to show that we can draw a vector of random variables as part of the jumping procedure, and use those additional variables to match the dimensionality of the smaller

[4]Richardson and Green also utilize split-combine moves in order to increase the efficiency of the sampler and allow it to decrease the number of components even when there are no empty components. However, Cappé et al. (2003) show that the reversible jump algorithm will still converge to the true posterior across models with only the birth-death moves. Since including the split-combine moves greatly complicates the technical aspects of the algorithm, and the efficiency gains are negligible in our application, we have chosen to omit them.

model space to that of the larger. This dimension matching procedure insures that the move is reversible, and thus maintains the detailed balance condition of an MCMC sampler. The reversible jump MCMC thus modifies the conventional Metropolis-Hastings formulation of A as follows:

$$A = \frac{p(\theta' \mid y)}{p(\theta \mid y)} \times \frac{r_m(\theta')}{r_m(\theta)\, q(u)} \times \left| \frac{\partial \theta'}{\partial (\theta,\, u)} \right|, \tag{2.4}$$

where the first term is the ratio of posteriors; the second term is the ratio of the probability of jumping up to the higher model (with parameters θ') to the probability of jumping from the higher dimension space down to the lower space (with parameters θ and random draws u); and the final term is the determinant of the Jacobian of the proposal function that matches the dimensions of the two spaces. In our case, since the allocation of data points do not change, the first term is just the ratio of priors, while the second is decreasing in the number of empty components in the previous draw from the current model. In this way both the priors and the data directly affect the probability of a jump, and the reversible jump algorithm allows us to probabilistically sample across different models in essentially the same manner as Metropolis-Hastings.

On a less technical note, we follow the earlier literature by focusing on the unconditional distribution rather than conditioning on various sources of income such as human and physical capital levels. The estimates we present thus do not represent any particular theoretical growth model, but rather constitute a thorough description of the world income distribution.

2.3.4 Prior Specification and Hyperparameter Selection

Utilizing the reversible jump MCMC technique requires specifying proper prior distributions for all parameters. Our priors follow RG; they assign a uniform prior probability to every possible number of components up to the maximum number considered (ten), and tune the priors on component parameters (which are identical for all components regardless of the number of components) to the data.[5]

The priors for the component means are Normal with means ξ and variances κ, where ξ is set to the midpoint of the observed data and κ is the square of the observed range of the data (so that one standard deviation is the full range of the observations R). The component variances are distributed inverse Gamma with shape parameter α and scale parameter β.[6] Although we set $\alpha = 2$, for β we wish to avoid strong influence of the hyperparameter. This is achieved by imposing a hierarchical prior such that β is itself a random variable which follows a Gamma distribution. The shape parameter g is set to 0.2, while the scale parameter h is set to $100 \times g / \left(\alpha \times R^2\right)$. These hyperparameters are chosen to have a very low influence on the posterior results.

The prior for the weights η_j is a Dirichlet distribution with an identical parameter for every component, $\delta = 1$. Increases in the parameter δ increase the likely number of observations sorted into each component, bounding this number away from zero. Since the birth-death step that the estimation relies on can only decrease the model size if there are empty components, RG suggest using the value here, which readily allows the existence of empty components while still permitting full exploration of higher parameter spaces.

The prior structure we use is not a 'natural conjugate' prior discussed in the

[5]For a detailed discussion of the necessity of proper priors, as well as the common forms these priors take, see Green (1995).

[6]Following RG, we use the parameterization of the Gamma distribution such that α/β is the mean and α/β^2 is the variance.

Bayesian literature, but it is very similar and does provide many of the benefits of conjugacy. The conjugate structure is not necessary for our approach; as with all MCMC methods, any prior desired may be used (so long as they are proper). However, conjugacy is convenient and not particularly restrictive as long as the hyperparameters are chosen to keep the influence of prior information low.

2.4 Results

A discussion of the performance of the sampler using a well known dataset and simulated data is available in the appendix. This analysis shows that the sampler is capable of replicating previous work and is able to recover all components from a three mixture distribution even when the third component is not observable in a kernel density plot. Given the good performance of our sampler on these two problems, we now move to examining our main question of interest, namely what is the posterior density of the number of components in the empirical distribution of per capita income across countries and how has this density evolved over time?

2.4.1 Data and details

The per capita income data we use come from Maddison (2010).[7] We use this dataset instead of the Penn World Tables because of its larger collection of countries over a longer time period. Having a larger number of countries is important in order to be able to label what we study as the "world" income distribution. The Maddison data allows us to have countries like North Korea, Cuba, and many eastern European countries in our data as of 1950.

[7] This data adjusts for purchasing power parity using the Geary-Khamis (GK) method and uses constant dollars with 1990 as the benchmark year.

We have a sample of the same 135 countries in the years 1950, 1960, 1970, 1980, 1990, 2000, and 2008. For each cross section we run our sampler for 300,000 draws, discard the first 100,000 as burn-in, and use the last 200,000 for analysis. We do this for each of the seven cross sections listed and then examine patterns and changes in the number of components or the allocations over the time period. Our priors are set following RG, as we described in Section 2.3.4 above.

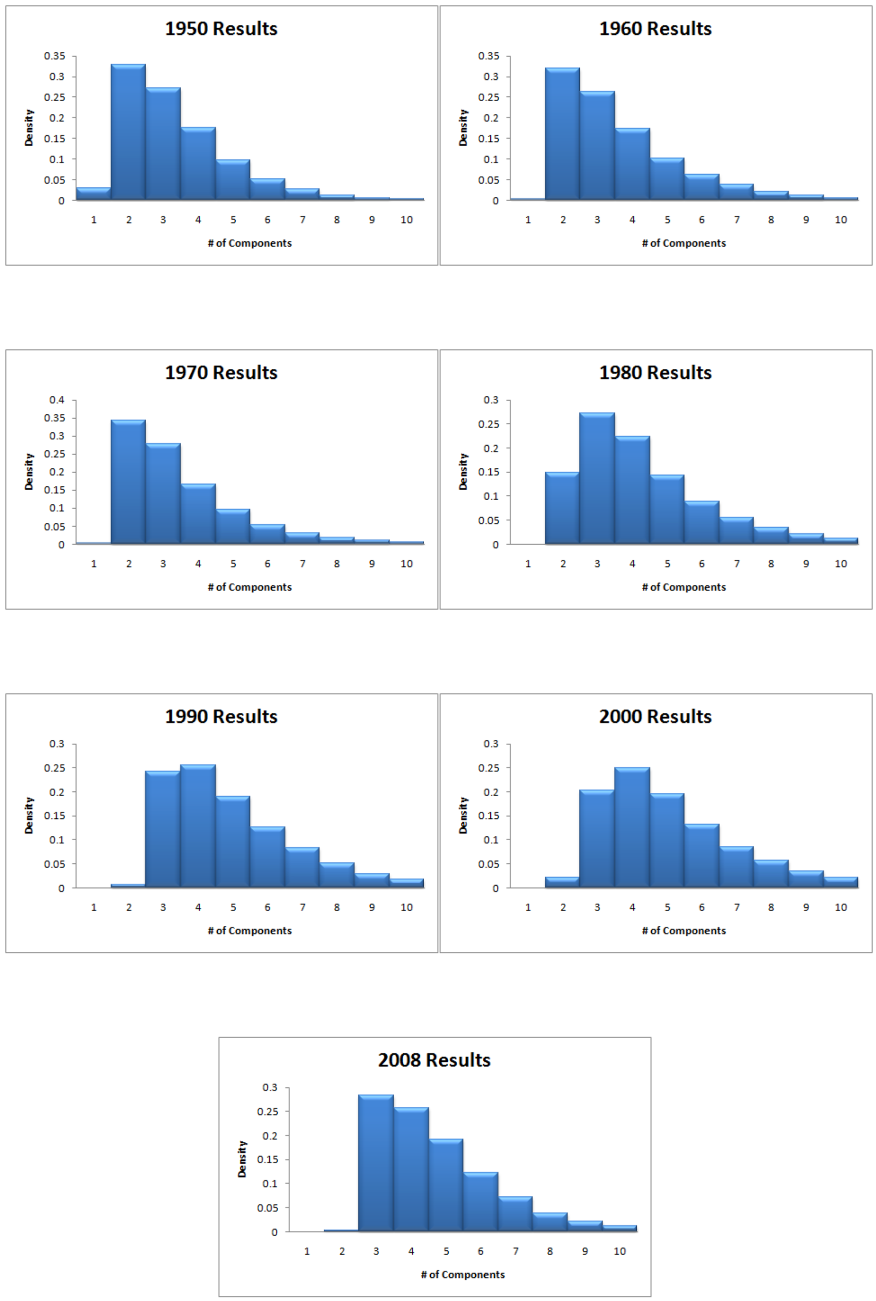
2.4.2 The posterior densities of the number of components

We have argued that we can capture the number of groups or clubs in the world income distribution by the number of individual lognormal components in the best mixture model for that distribution. We construct the marginal posterior density for the number of components in the mixture model using the fraction of the time our sampler spends in each of the 10 possible models ($k = 1, \ldots, 10$). Figure 2.1 presents the posterior densities of the number of lognormal components in the data as calculated by our sampler for each of the cross sections we use.

Several features merit attention here. First, even in 1950 and 1960, there is almost no probability attached to the $k = 1$ point. There were "twin peaks", in the sense we study here, from the beginning of our sample. Second, for the first three cross sections, the mode of the posterior is at $k = 2$, and the next most likely answer is $k = 3$. Third, in 1980, the mode clearly switches to $k = 3$, with $k = 4$ being the second most likely answer. Fourth, in the last 3 cross sections the mode is either at $k = 3$ or $k = 4$, though both command almost the same level of posterior probability. Fifth, point four notwithstanding, over 40% of the posterior probability in 2008 falls in the range $k > 4$.

Thus, we can say that, since 1950, there have always been multiple components to this distribution of per-capita incomes across countries and that the number of compo-

Figure 2.1: Histograms of Posterior Probability of Number of Components



nents comprising the distribution has risen over time. Since 1990 there is essentially no probability that the distribution is composed of 2 components, and while the modal answer is 3 or 4 components, there is a fair amount of probability on even higher dimensional models.[8]

2.4.3 The evolution of the world income distribution from 1950 - 1970

As noted, in the first three cross sections, the posterior density of k has a strong mode at two, meaning the overall distribution is composed of a weighted average of two component lognormal densities. In this section we describe the evolution of the means of the two component densities and the weights on each one in the overall distribution. This information is presented in Table 2.1. In 1950, the low income component has a mean of $1,194 while the high income component has a mean of $5,613. The overall distribution is a weighted average with a weight of .66 on the low distribution and .34 on the high.

By 1970, we can see that the mean of the low distribution is almost unchanged, at $1,206, while the mean of the high distribution has risen to $6,778. More significantly though, the weights for the overall distribution have change to .45 on the low and .55 on the high. In practical terms, while the spread between the low and high components increased slightly, much more of the overall distribution is coming from the high component.

We thus think of this period as being one of development, with countries "jumping" from the poor to the rich group while the rich group still gets richer. Specifically, 45 countries move from having over a 50% chance of being from the poor component

[8] In what follows below, we will concentrate our analysis from 1980-2008 on the three component model, rather than the four-component model. We explain the reasons for this decision in section 5.5 below. None of the substantive conclusions in our analysis are affect by the choice of k=3 vs. k=4.

Table 2.1: Parameter Means of Lognormal Component Densities

PANEL A:	mean		standard deviation		weight	
Year	low	high	low	high	low	high
1950	1,194	5,613	880	6,663	0.661	0.339
1960	1,087	5,017	621	5,464	0.478	0.522
1970	1,207	6,778	651	6,689	0.452	0.548

PANEL B:	mean			standard deviation			weight		
Year	low	mid	high	low	mid	high	low	mid	high
1980	936	4,069	12,587	366	2,934	6,673	0.323	0.422	0.255
1990	953	4,533	16,170	357	3,031	4,633	0.343	0.483	0.174
2000	1,110	5,638	19,650	641	4,308	6,142	0.374	0.438	0.188
2008	1,113	6,217	22,670	571	4,881	6,093	0.312	0.479	0.208

Means and standard deviations are in 1990 dollars adjusted for purchasing-power parity.

in 1950 to having more than a 50% chance of being in the rich component in 1970. Of these, 22 move from having more than a 60% chance of being in the poor group to more than a 60% chance of being in the rich group. We see these moves mostly in Southern and Eastern Europe, South Asia, Latin America and the Middle East. Figure 2.2 (panel A) illustrates this upward mobility by plotting the probability of being in the top component in 1950 on the horizontal axis and the corresponding probability in 1970 on the vertical axis. Countries above the 45 degree line have increased their probability of being in the rich group over these two decades.

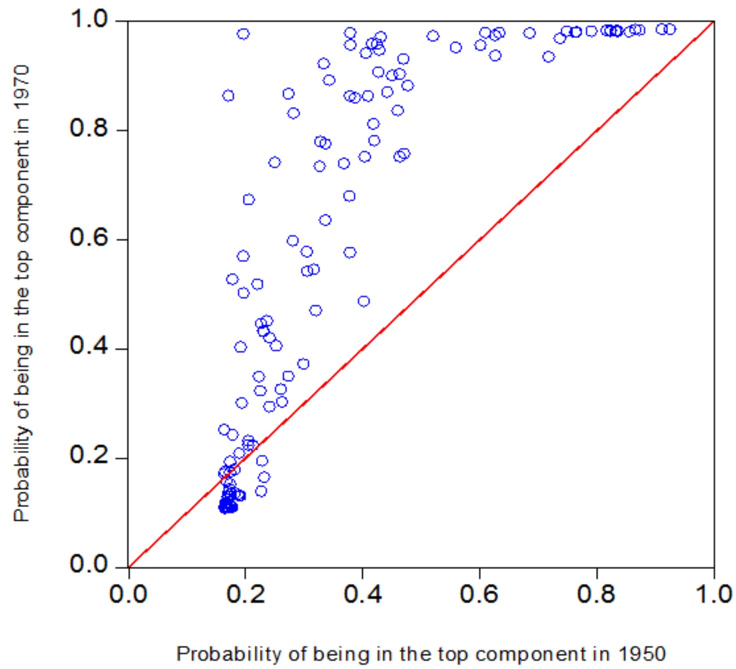
Some specific examples include Greece (62% chance of being in the low group in 1950, 95% chance of being in the high group in 1970), Romania (75% chance of being in the low group in 1950, 74% chance of being in the high group in 1970), Brazil (66% chance of being in the low group in 1950, 77% chance of being in the high group in 1970), Panama (62% chance of being in the low group in 1950, 86% chance of being in the high group in 1970), Japan (62% chance of being in the low group in 1950, 97% chance of being in the high group in 1970), Taiwan (79% chance of being in the low group in 1950, 67% chance of being in the high group in 1970), South Korea (80% chance of being in the low group in 1950, 57% chance of being in the high group in 1970), and Turkey (67% chance of being in the low group in 1950, 78% chance of being in the high group in 1970).

2.4.4 The evolution of the world income distribution from 1980 – 2008

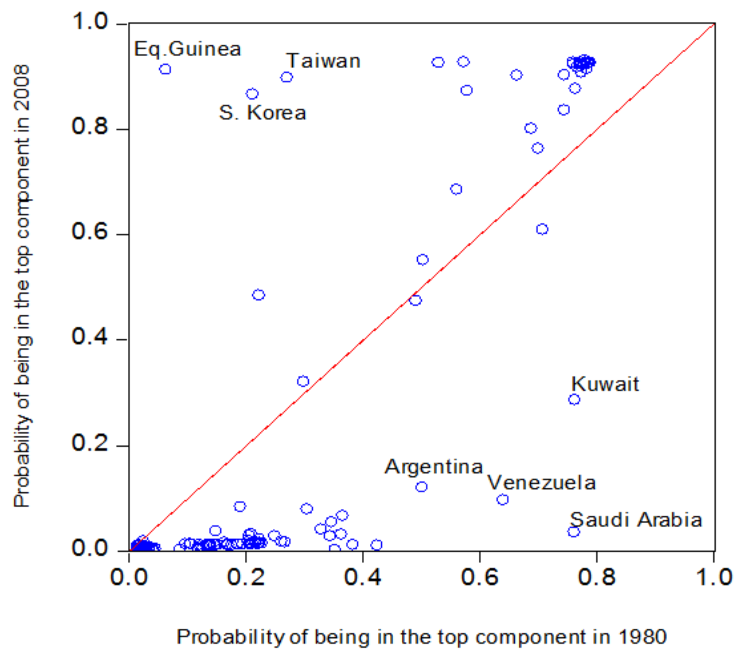
As noted above, our reversible jump sampler indicates that the number of separate lognormal components in the world income distribution increased from 2 to 3 between 1970 and 1980. The empirical distribution in 1980 is a weighted average of these three

Figure 2.2: Development vs. Polarization

PANEL A:



PANEL B:



components with a weight of .32 on the low component, .42 on the middle component and .26 on the high component. The means of the three components in 1980 are $936, $4,068, and $12,587 respectively. The substantive change from 1970 to 1980 is the pulling away of the upper group with an average income level over 12 times greater than that of the low group.

In 1990, the weights on the three groups comprising the empirical distribution change notably. The weight on the rich group falls from .26 to .175, while the weight on the poor group rises slightly and the weight on the middle group rises from .42 to .482. Greece, Portugal, Argentina, Venezuela, Kuwait, and Qatar are examples of countries that fall from being most likely in the top component in 1980 to being most likely in the middle component in 1990 (though Greece and Qatar return to being most likely in the top component by 2008).

The ratio of the mean of the poor component to the mean of the rich component falls from .075 to .06 while the ratio of the middle component to the rich component falls from .32 to .28. Thus the 1970s and 1980s are decades of polarization, as a shrinking group of rich countries move farther away from larger groups of middle income and poor countries.

In 2000 and 2008, the results do not change nearly as dramatically. The weight on the rich component rises very slightly, the percentage gap in means between rich and poor rises very slightly, and the percentage gap in means between the middle class and rich remains quite stable. In 2008 the mean of the low component is $1,113, the middle mean is $6,217 and the high component's mean is $22,670.

We categorize these last two decades as ones of stability or hysteresis in the income distribution. The poor and middle income countries seem increasingly locked in to their relatively disadvantaged positions. Figure 2.2 (panel B) illustrates the relative lack of mobility in the world income distribution since 1980. The probability of being in the rich group in 1980 is on the horizontal axis and the same probability in 2008 on the vertical. Compared to the 1950 – 1970 period, many more countries here are below the 45 degree line indicating that they are less likely to be in the richest group in 2008 than they were in 1980. Only Equatorial Guinea, South Korea and Taiwan move into the top group in this 28 year period, while Argentina, Venezuela, Kuwait, and Saudi Arabia fall out.

Of course there is also some mobility in and out of the middle group during this period. 9 countries (China, India, Burma, Pakistan, Cambodia, Vietnam, Cape Verde Islands, Lesotho, and Mozambique) move from the bottom to the middle group, while 5 countries (Nicaragua, North Korea, Iraq, Ivory Coast, Dijbouti, and Sao Tome & Principe) fall from the middle to the bottom.

Overall, there are 22 moves between groups in this 29 year period and half of them are downward moves, compared to 45 moves between groups, all of which were upward, in the 21 years from 1950 – 1970.

2.4.5 Discussion

There is some good news in these results: (1) While the lowest component's mean in 1950 was $1,198 and is $1,130 in 2008, the weight on the lowest component in the empirical distribution has fallen from .66 in 1950 to .32 in 2008. (2) Over time there has developed a "middle class" component, which has a weight of around 45% in the empirical distribution of country incomes per capita. (3) The mean of the highest component in the empirical distribution has risen from $5,588 in 1950 to $22,634 in 2008.

The bad news of course is that (1) 32% of the weight in the 2008 world income distribution is still coming from a component whose mean has not budged in 59 years. (2) The middle class component mean is not catching up at all to the mean of the

Table 2.2: The Perpetually Rich and the Chronically Poor

PANEL A:	Prob. of being in top component always greater than .66			
Australia	Belgium	Canada	Finland	France
Netherlands	New Zealand	Norway	Sweden	Switerland
UK	US			

PANEL B:	Prob. of being in bottom component always greater than .66			
Bangladesh	Benin	Burkina Faso	Burundi	Cameroon
Central Afr. Rep.	Chad	Comoros	Ethiopia	Gambia
Guinea	Guinea Bissau	Madagascar	Malawi	Haiti
Kenya	Sierra Leone	Sudan	Mali	Mauritania
Nepal	Niger	Nigeria	Rwanda	Tanzania
Togo	Uganda	Zaire (Congo)	Zimbabwe	

highest component over the 1980 – 2008 period.

There are a small group of countries that are highly likely to be in the top component throughout the entire sample period. Specifically, panel A of Table 2.2 lists the 13 countries whose probability of being in the top group is always greater than .66. More significantly though, there are a much larger group of countries whose probability of being in the bottom component of the distribution is always greater than .66. Panel B of Table 2.2 lists those 32 countries. While the majority of them are Sub-Saharan African countries, not all of them are and not all of the Sub-Saharan African countries are in this chronically poor group. It is perhaps not an exaggeration to say that these 32 chronically poor countries are caught in a poverty trap.

Of course the interpretation of the results in the last three cross sections should be tempered by the fact that there is almost as much posterior probability on $k = 4$ as there is on the $k = 3$ case that we discuss above. We prefer to emphasize the three-component model over the four-component one because the added component does not provide any additional substantive insights. When we increase k to 4, the middle component in the three-component model is split into the second and third groups, with the summed probability of these groups roughly equal to the probability of being in the middle group when $k = 3$. While this is not a problem in itself, the four-component model does not do a very good job of sorting countries into the middle two components. In particular, no country is put into component 2 with a probability of greater than .45.

Given that the move from $k = 3$ to $k = 4$ almost exclusively consists of fairly imprecisely dividing the middle class into two groups, we have concentrated our efforts in this paper on describing the composition and evolution of the three group model. Overall we see the stagnation of the lowest component, the incredible increase in the mean of the upper component and, since 1980, the existence of a robust middle component.

2.4.6 On the behavior of the sampler

The previous results depend on appropriate mixing behavior in the MCMC sampler. Figure 2.3 provides some information on the mixing of our sampler for each of the 7 cross-sections by graphing the last 50,000 draws for k (the number of components in the mixture) from each of the 7 chains of 300,000 that we ran. As can be seen the sampler is moving through the full space, except for $k = 1$ in the early years and $k = 1$ or 2 in the later years.

Figure 2.4 presents some information on how stable our estimates of the number of components are over the chains. It is easy to see how the probabilities move a fair amount during the burn-in period (the first 100,000 sweeps) and settle down to be stable during the last 200,000 sweeps that we use for analysis. As an experiment, we ran one of our cross-sections for 500,000 sweeps (an additional 200,000) and the results we found in that case were basically the same as in our baseline case.

It is also worth noting that our sampler does a good job of assigning countries to

Figure 2.3: Mixing of Draws across Number of Components

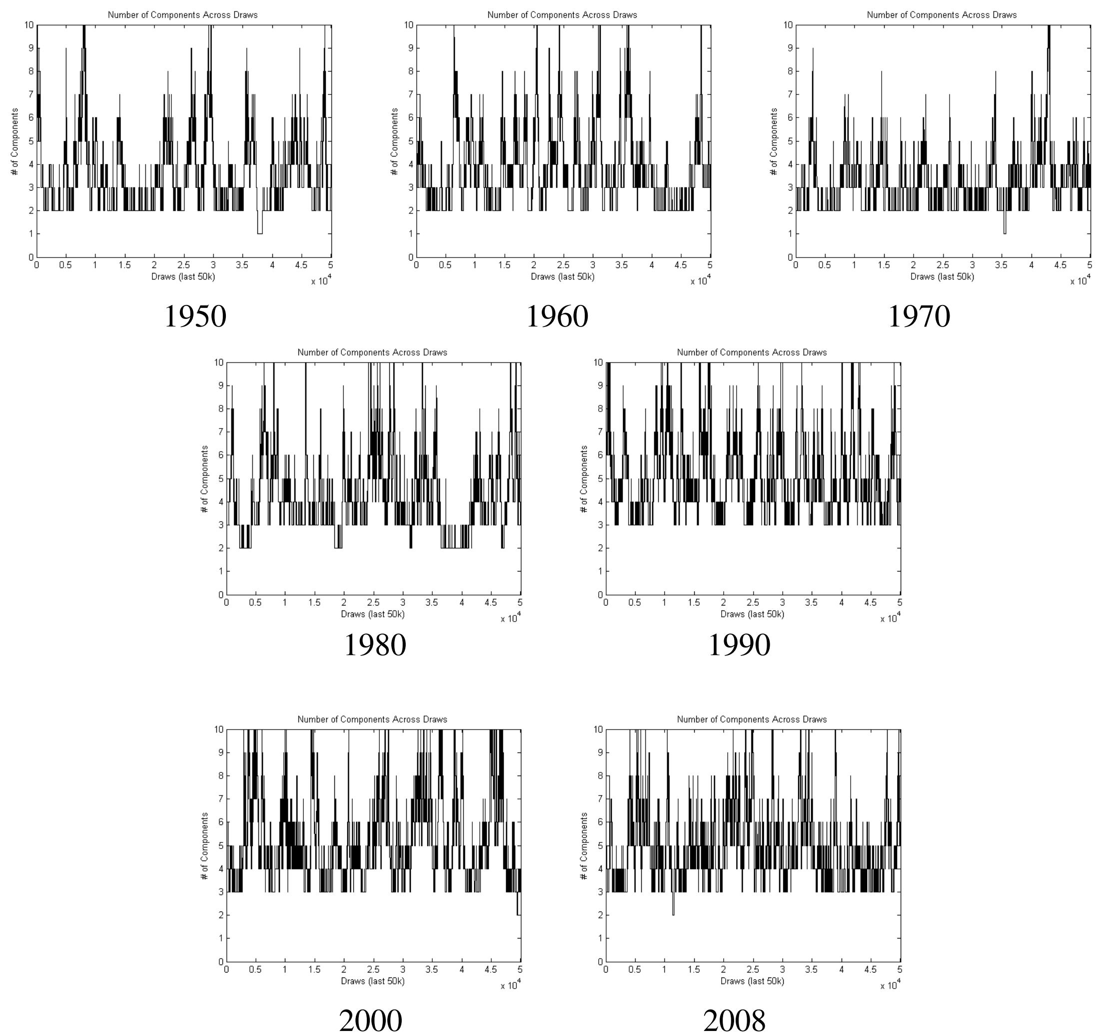
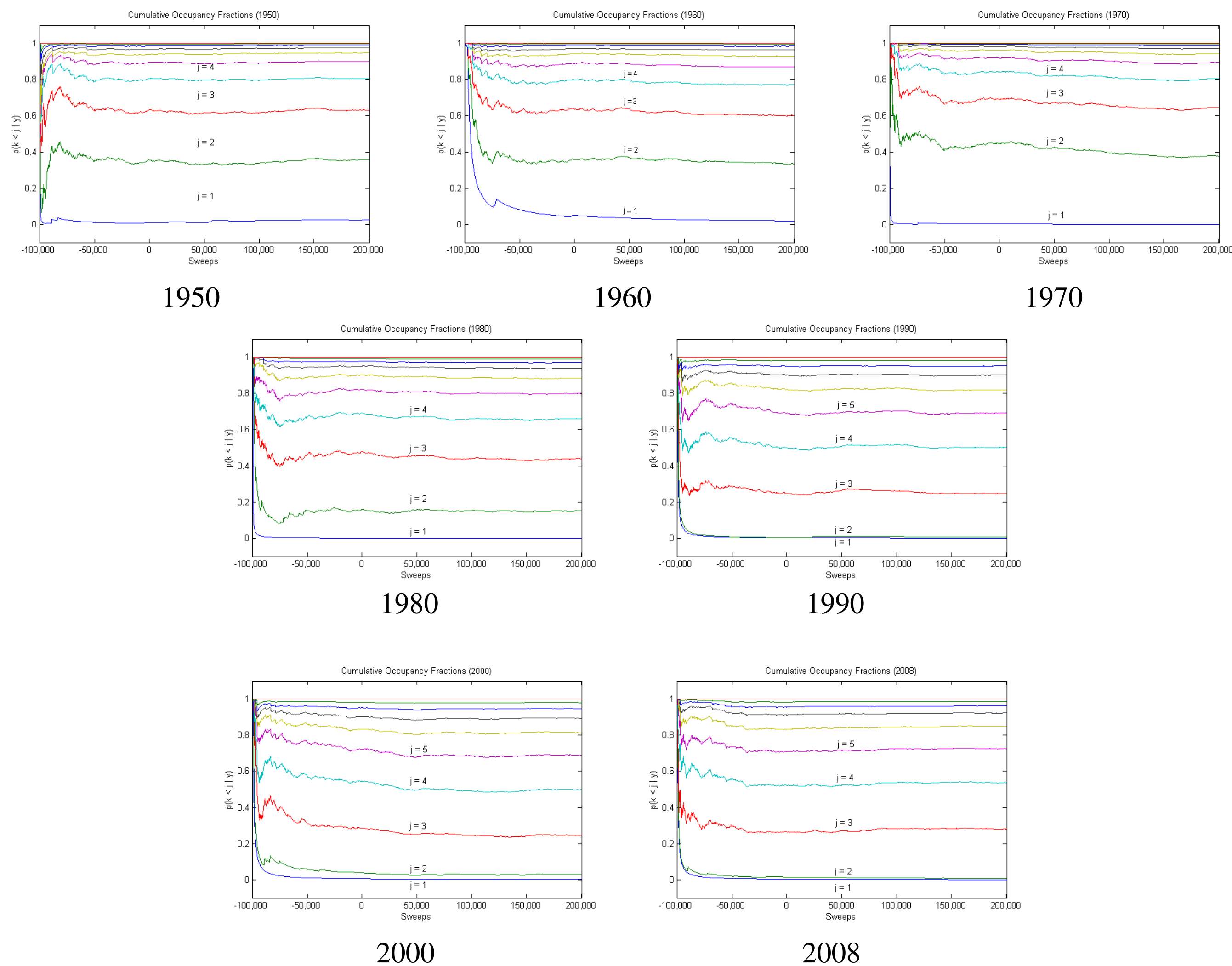


Figure 2.4: Cumulative Occupancy of Number of Components across Sweeps



a particular component density. In 1970 (the last cross section with a posterior mode at $k = 2$), 113 of our 135 countries can be assigned to one of the components with probability of .66 or above. In 2008, using $k = 3$, 124 of our 135 countries can be assigned to one of the three components with probability of .66 or above. That is 84% in 1970 and 92% in 2008.

In sum, we believe our reversible jump MCMC sampler is fully exploring the parameter space for the number of components in the mixture and that the draws we use for analysis represent draws from converged chains. The draws for the means and variances of the components are straightforward Gibbs steps and we do not present information on their evolution over draws of the chain here.

2.4.7 Prior sensitivity analysis

In this section we study the sensitivity of the posterior density of k, the number of components, to two types of changes in our priors. We undertake this analysis for the 1960 and 1990 cross sections of our data. Our uniform prior for the number of mixture components places no direct restrictions on the results (except for our restriction that $k \leq 10$), so the main candidates for influential priors are the prior on the variance of the variances of the components and the prior on the variance of the mean of the components. Intuitively, if the variance of the components is restricted to be small, then it will take more components to describe a given dataset other factors held constant. Similarly, if the means of the components are restricted to be close together, i.e. are drawn from a distribution with a small variance, then it will take more components to describe a given dataset.

As described above, the prior on the variance of the variances of the components is hierarchical in order to minimize its effect on the posterior density of the number of components. We draw the scale parameter for the gamma density prior from another gamma density. In our sensitivity analysis, we first double and then halve the variance of scale parameter in the hierarchical gamma density and check what effect that has on our posterior for the number of components.

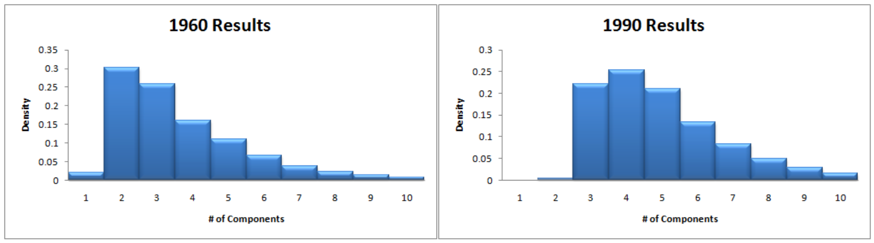
Figure 2.5 (panel A) shows the posterior for the number of components when the scale parameter in the hierarchical prior is doubled for 1960 and 1990. Panel B reports the same results for the case where the scale parameter is halved. In all cases, the mode of the posterior remains the same as that found in our baseline case and overall, the histograms are quite similar. This supports the idea that making this prior hierarchical does limit its effect on the main results.

The variance parameter for the normal prior distribution of the means of the components is, for the reasons described above, another candidate to be an influential prior. In our sensitivity analysis we first double and then half the standard deviation of the density we use to draw the means of the components.

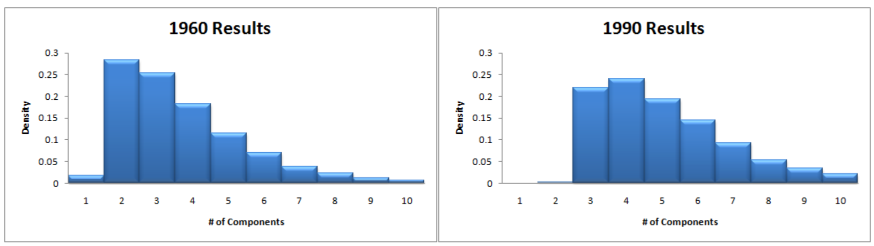
Figure 2.5 panel C shows that doubling the standard deviation concentrates the posterior for the number of components at the lower values, though the mode does not change from our baseline case. Panel D shows that halving the standard deviation shifts the posterior markedly to the higher values with no clear mode in the distribution.

We think that our baseline choice of prior for this parameter, where the standard deviation for the prior distribution of the component means is set equal to the range of the data, is a "Goldilocks" prior, where the means aren't forced either too close together or too far apart. We can say though that, if anything, our results are conservative in that they more likely understate rather than overstate the number of components in the overall data.

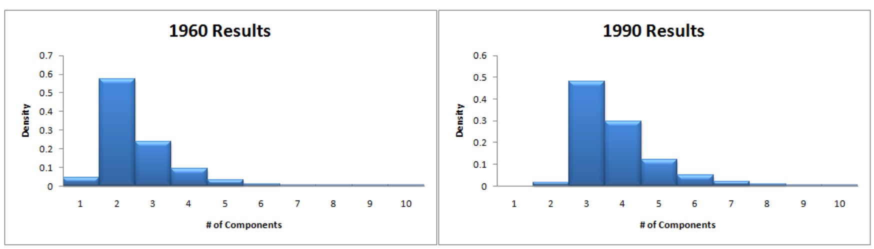
Figure 2.5: Prior Sensitivity Analysis



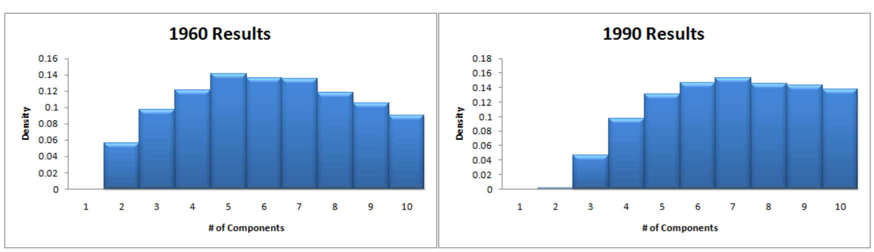
PANEL A: $2 \times h$



PANEL B: $h/2$



PANEL C: $4 \times \kappa$



PANEL D: $\kappa/4$

2.5 Conclusion

In this paper we use Bayesian variable dimension mixture models to characterize the evolution of the distribution of per capita incomes across 135 countries from 1950 – 2008. Our work is the next step in a natural progression from Quah's seminal contributions to the formal multi-modality tests of Bianchi (1997) and Henderson et al. (2008) to the mixture models of Paap and van Dijk (1998) and Pittau et al. (2010).

In our paper, for the first time in the literature, the number of individual components or groups in the overall distribution, the means and variances of each group, and the allocation of countries into groups are all being estimated in a single comprehensive model.

Our new approach leads to new insights. We show that the 1950s and 1960s were decades of development with two groups in the distribution and a significant amount of movement from the poor to the rich group. This configuration changed in the 1970s to a three group setting and the 70s and 80s showed polarization as the richest countries moved farther away from the middle and bottom income groups. In the 1990s and 2000s, the percentage gaps between groups stabilize, but the bottom two groups are stuck in their relatively disadvantaged positions.

Our work raises interesting questions for future research. What events in the 1970s influenced the formation of a third group and the start of the polarization of the income distribution? What happened over the last 18 years to significantly slow down that polarization? It also illustrates a very useful method for studying the number and structure of hidden component groups in overall empirical distributions.

Chapter 3

Miracles, Disasters, and Clubs in Steady-State Growth

3.1 Introduction

Theoretical models of economic growth often make claims about the underlying growth trend in per capita income levels. A major problem with investigating these underlying (or steady-state) growth clubs is that steady-state growth is unobservable. We would like to be able to identify, for example, whether some countries have essentially zero underlying growth, with only transitory deviations from this trend. We might also like to confirm that countries which we consider close to the technological frontier all grow at something close to the same underlying rate, with movements in relative income levels possible but not major, lasting differences in growth rates. To answer these kinds of theoretically motivated questions using empirical observations, economists must find meaningful ways to approximate steady-state growth.

I utilize two estimation procedures not commonly seen in the growth and development literature to address the problem. Following previous work by Hausmann et al. (2005) and Jones and Olken (2008) to identify underlying growth trends, I first estimate the steady-state growth rates of income per capita using a dynamic mixture model approach from the business cycle literature. My approach expands on theirs by explicitly trying to capture the gradual convergence to the steady state common in most growth models. I then characterize the distribution of steady-state growth rates using a variable-dimension mixture of normal components previously used to capture the evolution of the world income distribution by Grier and Maynard (2010).

Using country-level data on annual per capita GDP growth rates from 1952 to 2008, I find a wide variety of growth experiences in the estimated steady-state growth rates. Broad historical events are captured by the estimates, including wide-spread growth deceleration in the mid-1970s, severe reversals of growth around 1980, and a large number of accelerations or growth 'miracles' amid the smaller shifts and constant steady-steady growth rates found in some countries.

Despite this large variation, over 85% of the steady states are well described by a single normal component. The remaining steady states occupy the tails of another normal component, suggesting the latter may represent a 'noise' component or may be compensating for a true distribution with heavier tails than a single normal distribution can account for. I find no evidence of distinct groups of stagnant or frontier growth rates in the unconditional steady states. This places the results at odds with models of growth clubs—and in particular poverty traps—which generally imply sharp country groupings, often associated with key variables being on opposite sides of some threshold. At the same time the large standard deviation of the primary component (around 2%) is at odds with neoclassical and similar models, which suggests that in the steady state growth rates should be very tightly distributed around the growth rate at the technological frontier. My results instead suggest that more attention should be paid to theoretical models in which variables or groups of variables with heavy tailed distributions affect determine steady-state growth rates across countries.

The rest of the paper is organized as follows: Section 3.2 outlines the related literature and highlights the contribution of this paper. Section 3.3 explains the first stage, within-country mixture model I use to estimate steady-state growth rates and discusses the first-stage results. Section 3.4 shows how these results can be used in a variable dimension mixture of normal components to compare long-run growth rates across countries. The section then describes the cross-country results and discusses their implications for the multiple steady state model. Section 3.5 concludes.

3.2 Previous Literature

Early research on multiple balanced growth paths by Murphy et al. (1989), Azariadis and Drazen (1990), Galor (1996), and others have emphasized the possibility of multiple balanced growth paths and/or poverty traps as an explanation for the varied growth experiences of different countries. More recent work by Klump and De La Grandville (2000) and Howitt and Mayer-Foulkes (2005) have continued this emphasis on differences in long-run growth paths across countries, and in particular in connecting them with the aggregate production function.

These models frequently imply the existence of groups of countries which grow at the same rate in the steady state. In the neoclassical growth model and semi-endogenous growth models built on the work of Jones (1995a), all countries should have the same steady-state growth rate. Models with poverty traps suggest we should see a group of countries with steady-state growth of zero. Empirical work related to these models have generally investigated differences in income levels rather than in growth rates. Research approximating the steady-state growth rates proposed by theoretical models is rare. Beginning with Durlauf and Johnson (1995), a branch of the literature has focused on using sophisticated techniques to identify groups of countries that seem to follow a similar growth processes, but even this research program has generally avoided statements about steady-state growth.[1]

Pritchett (2000) and Jerzmanowski (2006) have shown that multiple long-run growth paths are not just a cross-country phenomenon; many countries seem to switch among distinct growth regimes over time. To accomodate these insights, I use dynamic mixture modeling tools more frequently applied in business cycle analysis to estimate the steady-state growth rates of output per capita. These tools allow the steady states to experience structural shifts.[2] The within-country estimates of steady-state growth are similar to the work of Hausmann et al. (2005) and Jones and Olken (2008), in which the data determines when shifts in the underlying growth trend occur, rather than assuming a constant steady-state growth rate or breaking the data into segments of predetermined length. However, rather than treating annual growth rates as i.i.d. random variables around some mean, the dynamic mixture model allows annual growth rates to adjust slowly toward their steady state. The estimated underlying growth rates are thus more closely linked to theoretical steady-state growth rates than previous measures have been. Using these estimated steady-state growth rates as data, I then use Bayesian mixture models to identify any distinct groupings of steady states, such as poverty traps or growth along the technological frontier.

Paap et al. (2005) attempt to identify the number of distinct growth clubs using data on GDP per capita, finding evidence for three growth classes. They use a likelihood-based latent class model which imposes the assumption that all countries within a cluster have identical long-run growth rates which must be constant for the entire sample. In contrast, my approach allows steady-state growth rates to vary over

[1] Recent contributions to this literature include Bloom et al. (2003) and Canova (2004).

[2] The sources of these shifts from one steady-state path to another is an interesting empirical question that is beyond the scope of this paper.

time within countries, and the growth rates themselves are grouped into latent classes by the mixture of normal components.

This paper is similar to Alfo et al. (2008) and Owen et al. (2009) in that all three analyze growth rates using mixture models. Unlike this paper, Alfo et al. and Owen et al. focus on estimating parameters for a standard, neoclassical growth equation which are allowed to differ across components within the mixture. They use information criteria to select the number of components and use maximum likelihood methods to estimate the parameters of the growth regression clubs. Both papers use five-year average growth rate data. While this approach may be sufficient on a descriptive level, theoretical growth models make claims about underlying steady-state growth, which is almost certainly not reflected in simple averages taken over arbitrary periods. In order to develop stylized facts about economic growth more suitable for guiding modeling decisions, I introduce a first-stage dynamic mixture model, providing a more systematic way to approximate steady-state growth rates within countries.

3.3 Dynamic Mixture Models and Steady-State Growth

In the first stage, I utilize a dynamic mixture model within individual countries to estimate the steady-state growth rate and any shifts that might take place in that variable.[3] The next section briefly discusses dynamic mixture modeling before describing the results of the first stage. A brief review of Bayesian statistical methods and Markov Chain Monte Carlo (MCMC) simulations is available in the appendix.

[3]This technique, rather than a regime switching model, is used in order to keep from restricting the model to a preconceived number of regimes.

3.3.1 Approximating steady-state growth Rates

Since theoretical growth models make claims about the steady-state growth of per capita income, which is inherently unobservable, I estimate steady-state growth rates from the annual data. A more common solution is to take an average growth rate over several years. Averaging over all available data, however, implies that there must be a single steady-state growth rate at all times. But just as we could conceivably see different groups of growth rates across countries, we could also see variation within a country in the underlying growth rate over time. For this reason, it is also common to have many observations for each country by averaging over a set number of years, typically five or ten years.

While this approach is not unreasonable from a purely descriptive perspective, it faces a couple of weaknesses if the goal is to identify the underlying growth rate which might be described by a theoretical model. First, steady-state growth over the sample period may shift, but both the timing and number of shifts are unknown. Second, if a country is slow to converge to its balanced growth path, we would expect growth which is far from steady-state growth to slowly adjust between the previous period's growth and the steady state.

In order to capture these features, I estimate the following dynamic mixture model:

$$\begin{aligned} \tilde{y}_t &= \rho g_t + (1-\rho)\tilde{y}_{t-1} + \varepsilon_t \\ g_t &= g_{t-1} + S_t \eta_t, \end{aligned} \tag{3.1}$$

where $\tilde{y}_t$ represents per capita GDP growth in year t, g_t represents the underlying steady-state growth rate in year t, and ε_t and η_t are normally distributed random variables.[4] Dynamic mixture models are similar to the time varying parameter models

[4] Fagiolo et al. (2008) suggest that ε_t should be a fat-tailed distribution, rather than normal. I have

seen in applied macroeconometrics, but instead of allowing the parameter of interest to shift every period, dynamic mixture models assume that the parameter only shifts when the (unobserved) indicator variable S_t equals one. S_t is meant to capture structural shifts in the steady-state growth rate g_t, which we assume are relatively infrequent. The coefficient ρ, which I assume is between zero and one, captures how quickly the growth rate returns to its steady state after a deviation in the previous period.

The dynamic mixture model imposes very little structure on the observed data. Unlike regime switching or multiple change-point models, it does not impose a minimum duration on a given value of g_t or a minimum change in g_t to be classified as a shift.[5] The dynamic mixture model also differs from a regime switching model by treating all moves in g_t as structural breaks rather than moves from one well-defined regime to another. That is, if g_t shifts up by 0.03 but later shifts down by the same value, this model treats this as a third value of g_t, not a return to the previous regime.

I use Bayesian estimation methods for dynamic mixture models pioneered by Gerlach et al. (2000) and further developed by Giordani et al. (2007).[6] Their methodology allows for taking draws from the posterior of S_t without conditioning on g_t, substantially reducing the complexity of the estimation process. Conditional on the location and number of breaks, the steady-state growth rates are estimated using the algorithm of Carter and Kohn (1994). The other parameters to be estimated (ρ, σ_ε^2, and σ_η^2) can be drawn from their conditional distributions, which are conventional (The conditional posterior of ρ is distributed normally, truncated between 0 and 1, while both variance

estimated the results allowing for ε_t to switch between a low- and high-variance distribution, which is similar to allowing more weight in the tails than a normal distribution can accomodate. This approach increased computation time and complexity without having a significant effect on the estimated steady-state growth rates or the timing of structural shifts in g_t.

[5] Using the Bayesian methods described below, I use the priors to impose a penalty on steady states lasting two years or less.

[6] These techniques have been used for analysis of higher frequency data by Primiceri (2005), Giordani and Kohn (2008), and Koop et al. (2009).

Table 3.1: Summary Statistics for Maddison Data and First Stage Results

SERIES	# OBS	MEAN	STD	MAX	MIN	BEGIN	END
Maddison (Annual Growth)	7,638	1.8%	5.4%	51.0%	-95.4%	1952	2008
Maddison (Steady States)	243	1.5%	4.7%	19.5%	-27.1%		

terms are conditionally drawn from independent inverse gamma distributions). The process used is thus a Gibbs sampler, and is iterated through until a sufficiently long chain (less some burn-in period to allow for convergence of the Markov chain) is produced. I estimate the above model for each country in the sample (134 countries) using annual growth rates from 1952 through 2008.[7]

3.3.2 Prior and Comments on the Dynamic Mixture Model

Using per capita GDP from Maddison (2010), I estimate the dynamic mixture model for each country in my sample (summary statistics are available in Table 3.1). I have a total of 57 annual observations per country.[8] I run the sampler for 1200 draws and drop 200 for burn in, leaving 1000 draws for inference. Using the diagnostics of Raftery and Lewis (1992) on several countries suggests that this exceeds the requirements for convergence and inference.

At this stage it is worth briefly considering some of the distinctive features of Bayesian analysis. In Bayesian estimation, the object is not selecting parameters to maximize a likelihood function, but rather to identify the entire distribution of the parameters conditional on the data observed. This distribution is called the *posterior*

[7]The Maddison income per capita data is available for these countries from 1950–2008. I lose one observation by constructing growth rates as the difference between the natural log of income in a given year and the natural log of income in the previous year. I lose another observation constructing the lagged dependent variable $\tilde{y}_{t-1}$.

[8]Maddison measures income in Geary-Khamis constant dollars using 1990 as the base year.

distribution. The posterior is typically described as being proportional to the likelihood function times the he *prior distribution*, or

$$p(\theta \,|\, y) \propto p(y \,|\, \theta) \times p(\theta), \tag{3.2}$$

where $p(\theta)$ is the prior distribution, which incorporates any previous knowledge the researcher may have about the parameters in question. In practical terms, the prior may be based on a naive, preliminary analysis of the data, and in general priors may be designed to have little influence on the posterior, rather than being informative, by insuring that the variances of parameter priors are relatively diffuse.

One common concern with Bayesian anaylsis is the choice of density for the priors. One of the most convenient choices is conjugate priors, which means that the prior density is chosen such that the posterior, which derives its form from the prior times the likelihood function, will be of the same functional form as the prior. This implies that as more data is added, the functional form of the posterior will not change. Conjugate priors can sometimes allow for closed form solutions to the posterior, and were thus essential for Bayesian analysis prior to the advent of numerical simulation methods. Conjugacy is no longer necessary in order to estimate posteriors, but it is still convenient and is not particularly restrictive as long as the variances chosen are diffuse. I use independent conjugate priors which are somewhat tuned to the data, but are not particularly informative.

The prior for g_t is normal with mean set equal to the mean growth rate in the data and standard deviation equal to half of the observed range of the data. The prior for the autoregressive parameter ρ is standard normal truncated to the interval $[0, 1]$. The variances have their priors independently distributed inverse gamma with identical scale parameters equal to twice the observed variance, and shape parameter set so that the mean of the prior is half the observed variance in the data. The results are quite insensitive to changes in any of these parameters.

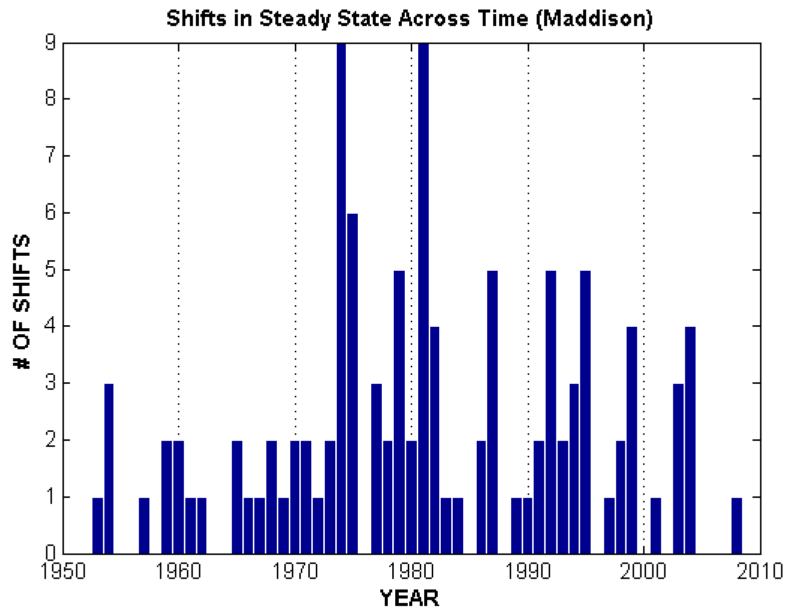
The probability that in any period there is a structural shift could be sensitive to the value of a strict prior probability on the frequency of shifts. To some degree this is desireable, since I assume that shifts in the steady-state growth rate are somewhat rare (and we wish to limit the effect of large but brief shifts in annual growth); on the hand, priors which are too influential could overwhelm the information provided by the data. In order to balance these issues, I use a hierarchical prior on S_t with a relatively small variance. The hierarchical prior follows a beta distribution with parameters four and seventy six, which can be interpreted as if the researcher had eighty observations prior to the sample, four of which are known to include a shift.[9]

The average probability of shifts across years varies widely from country to country, making inference trickier than simply selecting a probability above which a year is counted as a shift. Consistent with trying to identify rare shifts, I count a year as containing a shift when the probability of a shift for a given year is three standard deviations from the mean for the entire series. For most observations the actual probability is more than four standard deviations from the mean, and closer to six is not uncommon. Varying this cutoff slightly on either side does not have much effect on the final results.[10]

[9]These parameters restrict the variance of the hierarchical prior, essentially penalizing shifts which are too frequent. However, the results are nearly unchanged if the parameters are one and nineteen, which retains the same average value but allows the prior on S_t to vary closer to 0 and 1.

[10]On a few occassions, the data transition gradually enough that the estimator has difficulty selecting between a shift in one year or a shift in an adjacent year. This can lead to two adjacent years having a lower frequency of shifts than would otherwise be the case, although in the sampler there is never a shift in both periods in the same sweep. In these cases, I use a cutoff of 2.5 standard deviations, and select whichever year has a higher frequency of shifts as the true shift. The results are not particularly sensitive to either of these choices.

Figure 3.1: Shifts in Steady-State Growth Rates (Maddison Data)



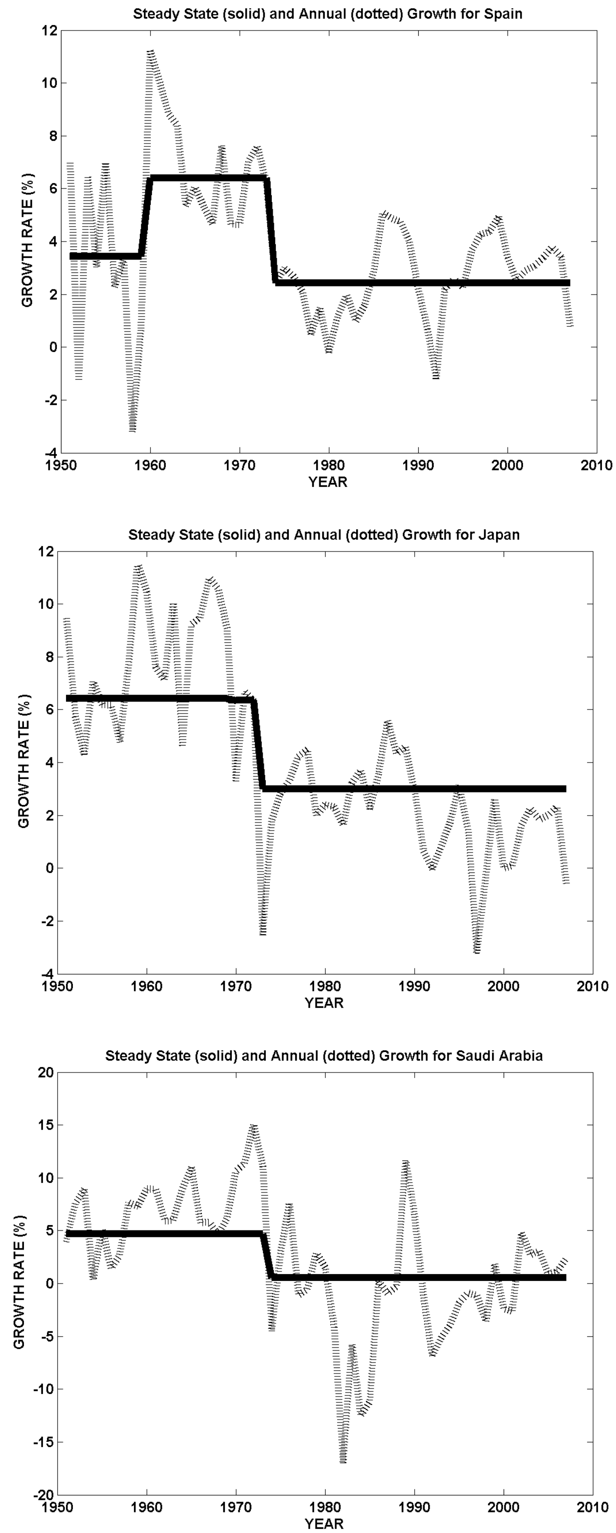
109 total shifts in Maddison dataset.

3.3.3 Results of First Stage Estimation

The distribution of steady-state growth shifts over time is shown in Figure 3.1. Two prominent periods stand out as times of unusually instability in steady-state growth: the mid 1970s, and 1979-1982. In the 1970s we see seven Western European economies (see Figure 3.2 for examples and Table 3.2 for a full list) slowing down, with growth rates ending up roughly within a one percentage point range for all of them. We see a similar pattern in three non-European countries still associated with the economic frontier—Israel, Puerto Rico, and Japan. In four additional countries—Comoros Islands, Jamaica, North Korea, and Saudi Arabia—we see a shift from moderate or high growth to low or even negative steady-state growth. Only one country sees a strong and sustained shift to higher growth during this period: Botswana moves from 2.4% to 6.4% steady-state growth in 1975.

A second wave of growth decelerations arrived between 1979 and 1982, coincid-

Figure 3.2: Growth Decelerations: The Mid-1970s



Dotted lines represent annual growth rates in per capita GDP (Maddison). Solid lines are estimated steady-state growth rates. Note that vertical axes differ for each country.

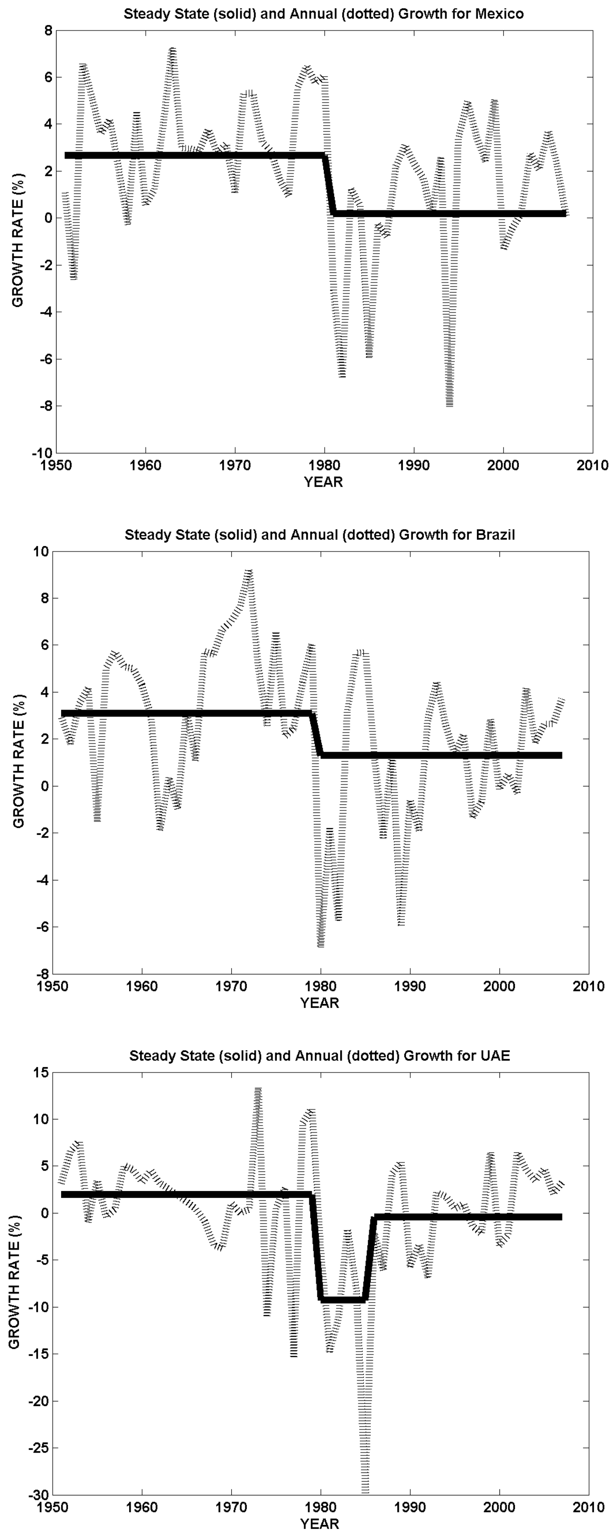
Table 3.2: Growth Decelerations: The Mid-1970s

COUNTRY	YEAR	STEADY STATES PREV.	STEADY STATES NEW	COUNTRY	YEAR	STEADY STATES PREV.	STEADY STATES NEW
Italy	1970	4.9%	2.7%	Israel	1973	5.3%	1.9%
Austria	1974	4.1%	2.9%	Puerto Rico	1974	5.0%	2.2%
France	1974	3.4%	1.8%	Japan	1974	6.3%	3.0%
Portugal	1974	4.2%	1.9%				
Belgium	1975	4.1%	2.2%	Comoros Is.	1972	2.3%	-3.6%
Netherlands	1975	3.3%	1.9%	Jamaica	1974	5.3%	-0.9%
Spain	1975	6.4%	2.4%	N. Korea	1974	5.9%	0.7%
				Saudi Arabia	1975	4.7%	0.6%

ing with major international events such as rising interest rates in the developed world, the Latin American debt crisis, and the recession of the early 1980s. As expected, a number of Central and South American countries experience decreases in steady-state growth during this this period (see Figure 3.3 for examples and Table 3.3 for a full list). A number of African and Middle Eastern countries are affected, as well. Unlike the deceleration of the early to mid 1970s, however, this wave is perhaps better described as a reversal of growth. Of all the countries affected, only Brazil's steady-state growth exceeded 0.5%, and most of the countries involved switched from periods of sustained positive growth to sustained negative growth. Nicaragua and Gabon experienced short, extreme periods of negative growth (in the Nicaraguan case associated with political revolution) in the late 1970s, Yemen experienced high growth throughout the 1970s, while the United Arab Emirates entered a severe downturn that lasted most of the 1980s. In all four cases, these shorter periods of dramatic change give way after 1982 to new steady-state growth rates lower than they had been in the 1950s and 1960s.

In contrast to the bad news of the mid 1970s and around 1980, there are a num-

Figure 3.3: Growth Reversals: 1979-1982



Dotted lines represent annual growth rates in per capita GDP (Maddison). Solid lines are estimated steady-state growth rates. Note that vertical axes differ for each country.

Table 3.3: Growth Reversals: 1979-1982

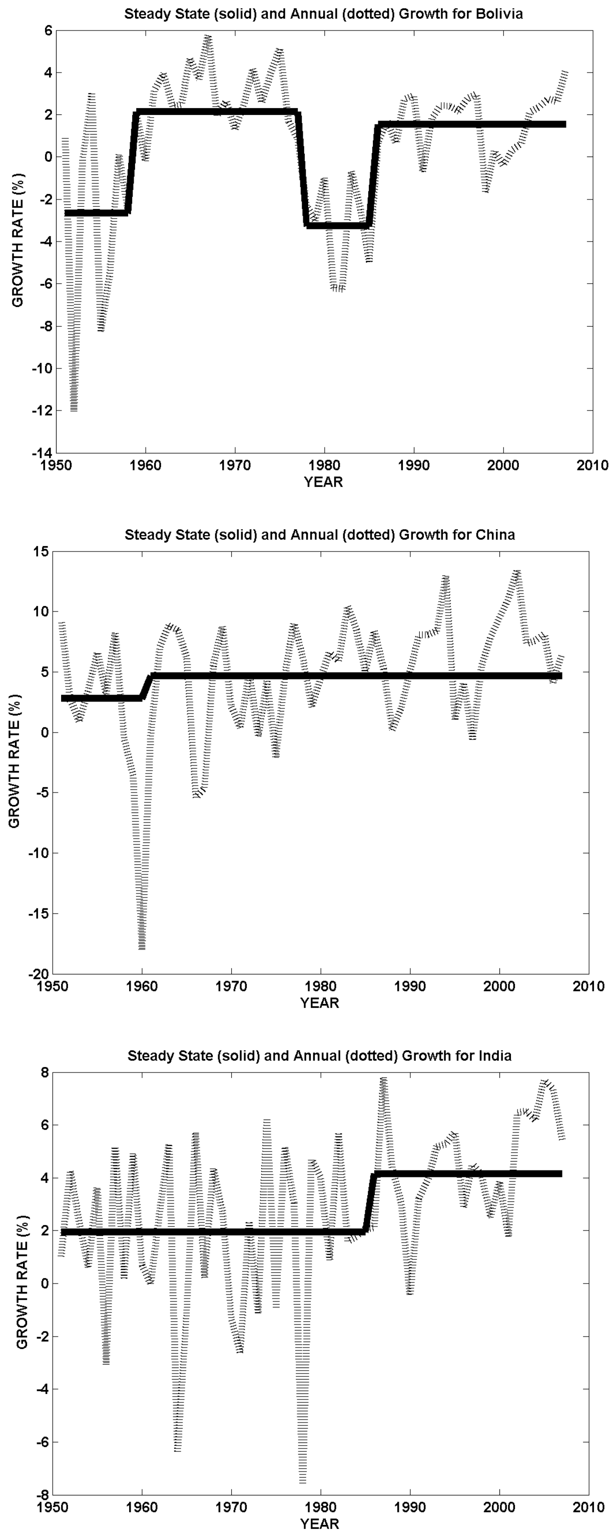
COUNTRY	YEAR	STEADY STATES PREV.	STEADY STATES NEW	COUNTRY	YEAR	STEADY STATES PREV.	STEADY STATES NEW
Bolivia	1979	2.1%	-3.3%*	Gabon	1979	3.5%†	-2.2%
El Salvador	1979	2.2%	-4.9%*	Togo	1980	1.9%	-2.4%
Honduras	1979	1.0%	0.1%	S. Tomé & P.	1981	1.9%	-1.4%
Nicaragua	1980	2.5%†	-2.6%	Côte d'Ivoire	1982	2.1%	-3.6%
Brazil	1981	3.1%	1.3%	Congo	1982	1.6%	-0.7%
Haiti	1981	-0.2%	-1.7%				
Paraguay	1981	1.4%	-0.2%	Yemen	1979	6.6%	0.5%
Mexico	1982	2.7%	0.2%	UAE	1981	2.0%	-9.3%

* El Salvador returned to positive steady-state growth in 1983 (and Bolivia in 1987), but at a rate of 1.1% (1.5% in Bolivia).
† Nicaragua actually has growth drop for two years in 1978 to a rate of -24.9%, while Gabon experienced a similar drop in 1977 of -27.1%. They are included on this list because the shift from a gradual upward trend prior to 1977/78 to steady decline after 1980 is still noteworthy.

ber of positive stories to be found as well. Many of these stories fall roughly into two broad categories: accelerations, movements from low growth (less than 0.5%) to moderate positive growth; and growth miracles or take offs, movements from moderate to high growth (4.0% or higher). A list of countries in these two categories is presented in Table 3.4, with examples in Figure 3.4. In the table we see that Kuwait began to grow moderately following the 1991 Gulf War after decades of decline. Bolivia actually experiences two distinct periods of acceleration: in 1960, towards the end of the decade long rule of the Revolutionary Nationalist Movement, Boliva shifted from -2.7% growth to positive 2.1%; after the political turmoil and economic decline begining in 1979, the Bolivian economy shifted back into positive 1.5% steady-state growth in 1987.

As expected, the growth miracles category includes such countries as China, Singapore, and India; the model dates their periods of economic take off to 1962, 1967,

Figure 3.4: Economic Take-Offs



Dotted lines represent annual growth rates in per capita GDP (Maddison). Solid lines are estimated steady-state growth rates. Note that vertical axes differ for each country.

Table 3.4: Economic Take-Offs

COUNTRY	YEAR	STEADY STATES PREV.	STEADY STATES NEW	COUNTRY	YEAR	STEADY STATES PREV.	STEADY STATES NEW
Belgium	1959	2.0%	4.0%	Poland	1992	2.8%	4.1%
Spain	1961	3.4%	6.4%	Albania	1993	2.6%	5.2%
China	1962	2.8%	4.7%	Angola	1994	0.3%	5.4%
Singapore	1967	1.9%	7.1%	Bulgaria	1998	2.7%	5.1%
Yemen	1970	1.1%	6.6%	Uruguay	2003	0.6%	5.7%
Botswana	1975	2.4%	6.4%	Argentina	2004	1.2%	5.2%
India	1987	1.9%	4.2%	Cambodia	2004	2.2%	5.3%
Burma	1992	1.8%	6.2%	Ethiopia	2004	1.2%	5.6%
Bolivia	1960	-2.7%	2.1%	Kuwait	1991	-2.2%	3.7%
Sri Lanka	1967	0.4%	2.1%	Cuba	1994	-0.1%	3.9%
Bolivia	1987	-3.3%	1.5%	Afghanistan	1995	0.4%	3.0%
				Tanzania	1999	0.2%	3.5%

and 1987, respectively.[11] Several countries—such as Belgium, Spain, and Yemen—experience sustained periods of miracle-like steady-state growth before returning to modest levels of growth later on. We see two Sub-Saharan African countries fit into this category as well: Botswana shifts from a modest 2.4% growth rate to 6.4% steady-state growth in 1975, and Angola shifts from near-zero steady-state growth to 5.4% in 1994. Finally, we see several countries—Argentina, Uruguay, Cambodia, and Eritrea & Ethiopia—experiencing what seems to be a take off in the mid-2000s, although interpretations of the shift must be tempered by the fact that the dataset ends in 2008.

Finally, it is worth noting patterns in the *lack* of shifts. Out of 134 countries,

[11] It may seem odd to date the take-off of China in 1962 rather than the more commonly accepted date of 1978. However, after the Great Leap Forward of 1958-1961, China did undertake several key reform measures. Those measures were undermined or surpressed during the Cultural Revolution, but were reinstated along with many others in 1978. Thus we must either classify the strong positive growth c. 1962-1966 as an aberation to a pre-1978 steady state, or we must classify the Cultural Revolution as an aberation from a post-1962 steady state. The dynamic mixture model selects the latter.

52 hold to the same steady-state growth rate for the entire 57 year sample. These countries are drawn from all continents, including countries like Colombia, Costa Rica, the Dominican Republic, Lesotho, Morocco, Malaysia, Thailand, Tunisia, and Turkey. Despite the broad range of countries in this category, we do see two noticable country groupings that never shift their steady-state growth rates: the Nordic countries (Sweden, Finland, Norway, and Denmark), and the UK and its offshoots Austraila, Canada, and New Zealand.[12] Among the remaining 82 countries, there are 109 growth shifts, or an average of 1.3 shifts per country.

The approach and results of this first stage are similar in spirit to those presented by Hausmann et al. (2005) and Jones and Olken (2008). The emphasis in all three efforts is to classify growth within each country around underlying trend growth which may experience shifts. Unlike the previous literature, however, the dynamic mixture model I utilize allows annual growth rates to return to trend slowly, thus assigning to the trend properties we think of in the theoretical literature as applying to steady-state growth. This difference, along with the use of the Maddison dataset instead of the Penn World Table, leads to some differences in the estimated results. While Jones and Olken find the economic take off of Botswana occurs in 1966, and Hausmann et al. find it in 1969, I find that steady-state growth does not shift until 1975; I find that China's take off, at least in terms of steady states, takes place in 1962 (following the Great Leap Forward), rather than in 1978 (following the Cultural Revolution) as Jones and Olken and Hausmann et al. do; and while Jones and Olken find Côte d'Ivoire experiences a downward shift in 1979, I find that steady-state growth does not become negative until 1982.

The steady-state growth rates estimated here capture many of the broad patterns of cross-country economic growth described by Pritchett (2000). We see mountains

[12]The United States is not in this category, interestingly, because of an increase in steady-state growth from 1.3% to 2.3% in 1959.

like Côte d'Ivoire, plateaus like Mexico, hills like Costa Rica, accelerators like India, and plains like Chad. This holds even allowing for (1) an unknown number of shifts in the underlying growth rate and (2) gradually adjusting transitional dynamics after a shift in the steady state.

Once we have identified the steady states, though, we wish to know if we see multiple distinct clubs as theoretical models like Howitt and Mayer-Foulkes (2005) and Davis (2008) seem to suggest. Do steady states with near-zero growth form a clearly defined 'poverty trap' club, while a separate group describes growth of countries on the technological frontier? To answer this question, I represent the steady-state growth rates as a mixture of normally distributed components and estimate the number of components (clubs) in the mixture.

3.4 Cross-Country Comparison using Mixtures of Normals

The first stage dynamic mixture model identifies the steady-state growth rate for 134 countries over 57 years, resulting in 243 unique steady states.[13] In the second stage, I estimate a finite mixture of normal components, a technique which approximates a non-standard distribution by using a probabilistic combination of more conventional ones. This allows for the estimation of both the component parameters and the allocation of observations across components. In addition, I use the reversible jump Markov Chain Monte Carlo (MCMC) algorithm developed by Green (1995) and specifically

[13] Using only unique steady states implies that countries that have followed a constant steady state for the entire sample, such as the United States, would provide only one observation, whereas countries that have shifted many times, such as Bolivia, would contribute more. However, using all 7,638 country-years would result in a histogram with many high local modes at values that are repeated exactly for a large number of time periods; estimating a mixture of normals across all these observations would almost certainly result in an over-fitting model.

applied to mixture models by Richardson and Green (1997) to estimate the number of components.

In what follows I briefly discuss mixtures of normal components in general and the reversible jump MCMC algorithm used to estimate the correct number of components.[14] I then turn to the estimation results.

3.4.1 Mixtures of Normals and Reversible Jumps

Finite mixture models represent complex distributions as a weighted combination of several more conventional distributions. For a mixture of normal distributions, we can write the overall distribution in terms of observation x_i as

$$\tilde{y}_i \sim \sum_{j=1}^{J} w_j \mathcal{N}\left(\mu_j,\ \sigma_j^2\right), \tag{3.3}$$

where J is the total number of components in the mixture, μ_j and σ_j^2 are the mean and variance of component j, and weight parameter w_j is the probability that any given observation will be drawn from component j. The weights are all non-negative and must sum to one.

Although mixture models can be used strictly as a means of fitting an observed distribution, often the observations in the non-standard distribution are interpreted as being drawn from the individual components distinctly, but the identity of the true component is not available. It is convenient in this case to represent the observations using the conditional distribution

$$\tilde{y}_i \mid a_i = j \sim \mathcal{N}\left(\mu_j,\ \sigma_j^2\right), \tag{3.4}$$

[14] A more complete discussion of the reversible jump methodology is available in the appendix.

where a_i is a latent allocation variable produced using data augmentation. I use this latent variable approach in the estimation below.

Conditional on the appropriate number of components J, it is straightforward in Bayesian statistics to estimate the means, variances, and weights of the components, as well as the posterior probability that each observation is drawn from any given component, using Gibbs sampling.[15] Identifying the correct number of components is a more complicated matter. I utilize the reversible jump MCMC technique pioneered by Green (1995). This technique allows the sampler to probabilistically jump among a finite set of separate models with distinct parameter spaces. It generates random draws from a proposal generating density in order to match the parameter spaces of two models, then uses something very similar to a Metropolis-Hastings step to either move to the new model or to continue sampling from the current model. The approach used here is identical to that of Grier and Maynard (2010), and the interested reader is referred there for the details of the algorithm used.[16]

I run the model for 200,000 sweeps and remove the first 50,000 as burn in, leaving 150,000 sweeps for inference.[17] Each sweep consists of a Gibbs move updating all parameters conditional on the number of components and a reversible jump move which probabilistically increases or decreases the number of components by one. My priors in all cases are tuned to the data as in Richardson and Green (1997) and Grier and Maynard (2010). Several graphs illustrating the performance of the sampler are presented in Figure 3.5.

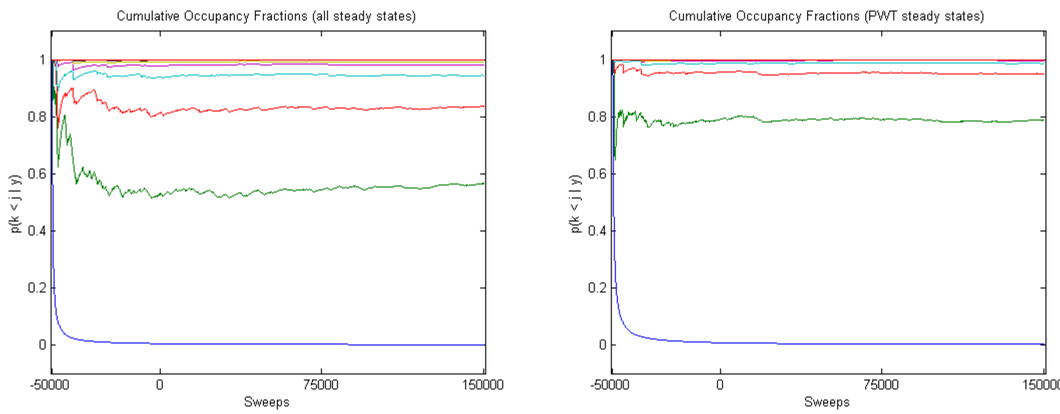
[15] In order to address the potential problem of 'label switching,' I utilize the Equivalent Class Reordering method of Papastamoulis and Iliopoulos (2010); in practice, however, almost no relabeling occurs, at least for the number of components given the highest probability by the reversible jump algorithm.

[16] For a more thorough discussion of reversible jump MCMC in the context of finite mixture models, see Richardson and Green (1997) and Früwirth-Schnatter (2006).

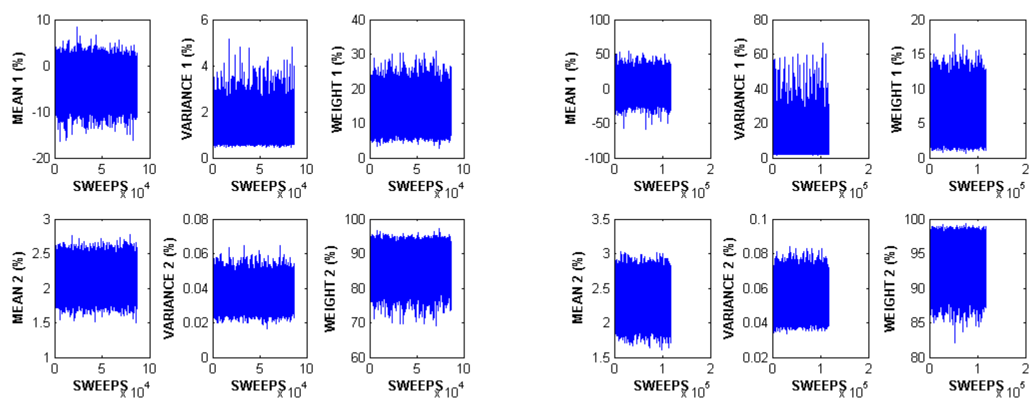
[17] I have run the model for 150,000 sweeps (leaving 100,000 for inference) and 300,000 sweeps (with 75,000 burn in, leaving 225,000 for inference) respectively, and found the same results.

Figure 3.5: Graphs of Simulator Performance

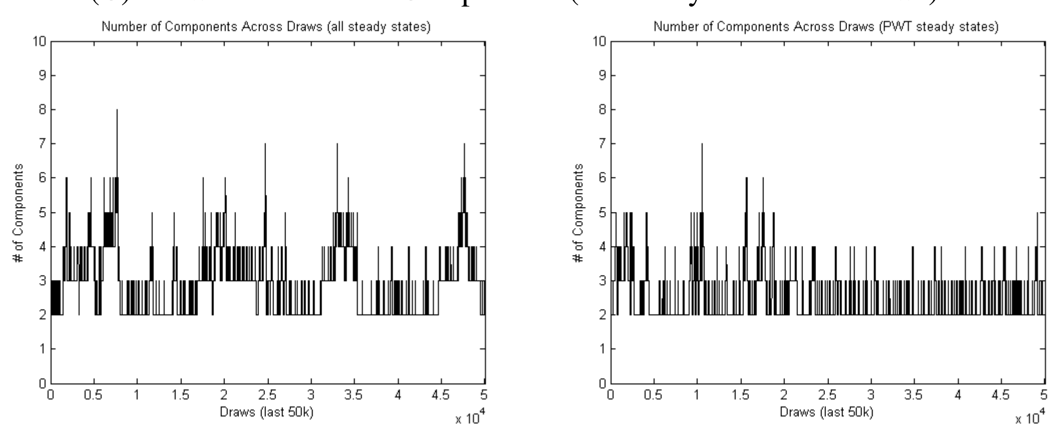
(A) Cumulative Occupancy Fractions



(B) Draws of Posterior Parameters



(C) Draw of Number of Components (Last Fifty Thousand Draws)



The left column shows results for the Maddison dataset, while the right column shows results for the Penn World Table 6.3.

Table 3.5: Number of Normal Components (Maddison)

DATA SERIES	OBS	p(1)	p(2)	p(3)	p(4)	p(5)	p(6)	p(7–10)
Steady State	243	0	**0.580**	0.265	0.103	0.035	0.012	0.006
Average growth	134	**0.649**	0.231	0.082	0.023	0.007	0.005	0.003
5-year growth	1,474	0	0	**0.683**	0.261	0.045	0.010	0.001
10-year growth	670	0	0.026	0.188	**0.383**	0.237	0.114	0.053

Dataset includes 134 countries, 1952-2008. For each data series, the highest probability model is indicated in bold.

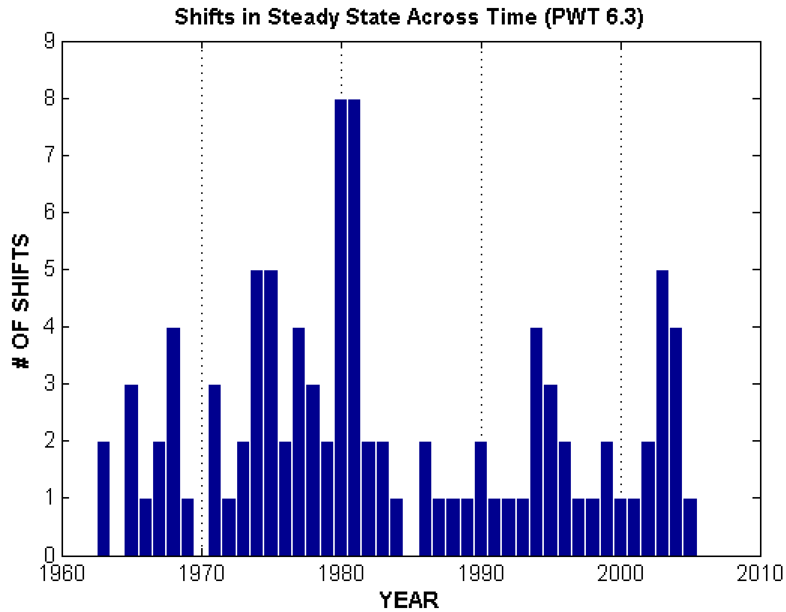
3.4.2 Mixture Model Results

The posterior probabilities of each possible number of components is presented in Table 3.5. Using the estimated steady-state growth rates from the first stage yields strong support for two normal components: this model receives more than half of the total probability, and more than double the probability of the next most likely model. After removing the initial draws, the reversible jump algorithm never drops down to a single normal component.

Table 3.5 also illustrates a significant drawback to approximating steady state growth by averaging over a set number of years—the number of components selected changes depending on the number of years used to construct average growth. If we use all 57 years to construct a single average for each country, we get strong support for a single normal component. But averaging over only five years provides equally strong support for three components, and an intermediate number of years (ten) yields an even higher number of components (four) as the most probable, and a substantial probability on even more.

This shortcoming is the result of failing to correctly identify the break dates in steady-state growth. For a given block of years, it's possible that half have an under-

Figure 3.6: Shifts in Steady-State Growth Rates (Penn World Table Data)



lying growth rate of -5%, while the other half have an underlying growth rate of +5%. Averaging across this break date will yield an estimate of 0% growth, failing to capture either period well. An accurate picture of how underlying growth is distributed, then, depends crucially on first estimating steady-state growth rates themselves.

This sensitivity is equally apparent when we consider a change of dataset. To demonstrate this, I ran the first stage estimation again using the Penn World Table 6.3 data of Heston et al. (2009). This dataset includes 110 countries from 1960 to 2007, thus excluding several countries and years contained in the Maddison dataset, but adding other countries.[18] The broad groupings for the steady state data—including an increased number of shifts in the mid-1970s, around 1980, and in the mid-2000s—remain, although shift dates and exact values for steady-state growth for particular countries may differ. A graph showing the distribution of steady-state growth shifts over time is shown in Figure 3.6, with summary statistics available in Table 3.6.

[18]For country-years in both datasets, many of the measurements are the same, although some will still differ due to data quality issues and alternative methods of calculating purchasing power parity.

Table 3.6: Summary Statistics for Penn World Table Data and First Stage Results

SERIES	# OBS	MEAN	STD	MAX	MIN	BEGIN	END
PWT 6.3 (Annual Growth)	5,060	2.0%	6.2%	78.0%	-56.1%	1962	2007
PWT 6.3 (Steady States)	208	2.4%	6.7%	62.9%	-41.2%		

Table 3.7: Number of Normal Components (PWT 6.3)

DATA SERIES	OBS	p(1)	p(2)	p(3)	p(4)	p(5)	p(6)	p(7–10)
Steady State	208	0	**0.787**	0.163	0.038	0.010	0.001	0.001
Average growth	110	**0.683**	0.206	0.070	0.023	0.009	0.005	0.003
5-year growth	990	0	0	**0.572**	0.286	0.102	0.031	0.009
10-year growth	440	0	**0.524**	0.311	0.111	0.036	0.013	0.006

Dataset includes 110 countries, 1962-2007. For each data series, the highest probability model is indicated in bold.

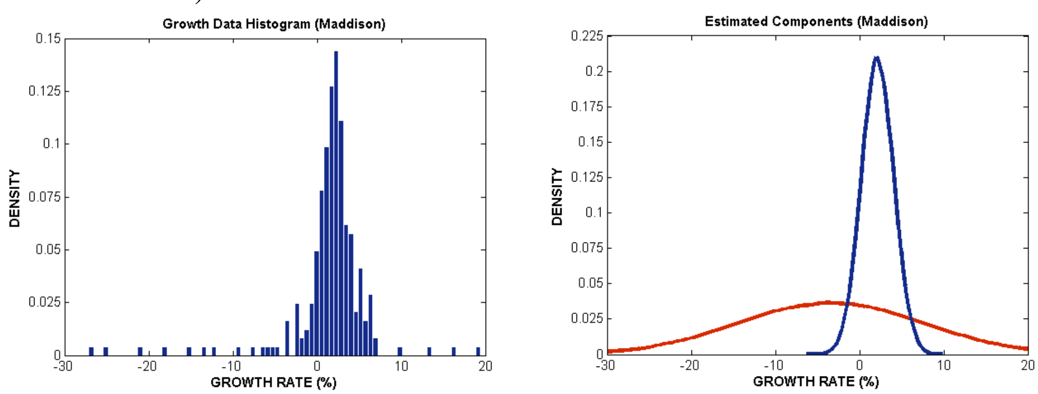
With this alternative data, we again estimate the mixture model for the estimated steady states, as well as for five, ten, and 57 year growth averages (the results are presented in Table 3.7). The model still selects two components for the distribution of steady states, and it again selects different numbers of components depending on the number of years used to calculate average growth. However, it is also worth pointing out that for the ten year averages, the estimated number of components is now two, as opposed to four using the Maddison dataset. Taking average growth over some number of years does not seem to describe underlying growth very precisely, since it is highly sensitive to the number of years chosen and the choice of datset. Estimating where shifts in trend growth take place seems to be particularly important in this context.

Table 3.8: Mixture Parameter Estimates

	First Component			Second Component		
Steady States	mean	variance	weight	mean	variance	weight
Maddison	2.2%	0.04%	86.9%	-3.3%	1.2%	13.1%
	(0.1%)	(0.01%)	(3.2%)	(2.2%)	(0.4%)	(3.2%)
PWT 6.3	2.3%	0.05%	94.5%	4.6%	7.3%	5.5%
	(0.2%)	(0.01%)	(1.8%)	(8.5%)	(3.8%)	(1.8%)

Point estimates are means of the posterior distributions (standard deviations of posteriors in parenthesis).

Figure 3.7: Steady-State Growth Histogram and Estimated Mixture Components (Maddison Data)

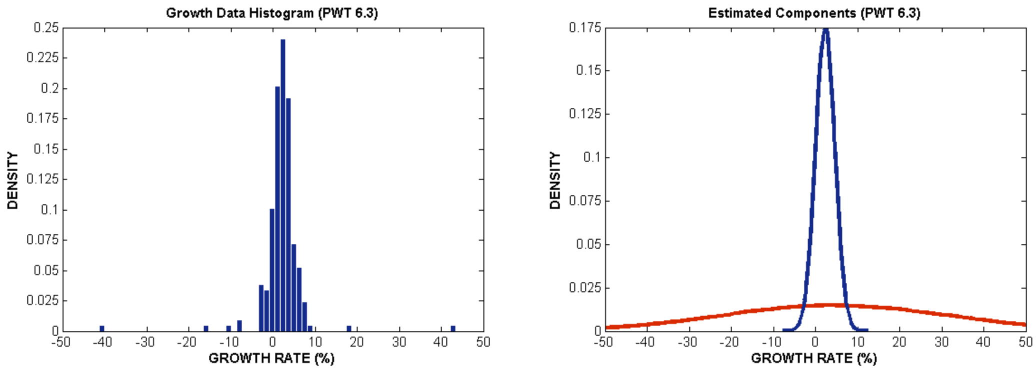


243 steady states in Maddison dataset.

3.4.3 Discussion

Having identified the number of normal components, I now turn to describing the components and country allocations themselves (parameter estimates for the mixture model are presented in Table 3.8). Histograms of the Maddison and Penn World Table datsets and the estimated components are can be seen in Figures 3.7 and 3.8. The steady states sort quite well, with every steady state in the Maddison dataset sorting into one group or another with at least 59% probability (79% for Penn World Table), and an average probability of 94% (98% for Penn World Table). For both

Figure 3.8: Steady-State Growth Histogram and Estimated Mixture Components (Penn World Table Data)



208 steady states in Maddison dataset.

the Maddison and Penn World Table datasets, most of the steady states (over 85% in both cases) seem to be drawn from a single normal component, and in both cases this component is centered in the 2–2.5% range with variances between 0.004 and 0.005, implying standard deviations around 2%. The posterior distribution of these parameters is relatively tight around the means in both cases. The mean and variance of the second component, in contrast, are imprecisely estimated, and differ substantially across datasets. This is not surprising given how few steady states are sorted into the second component: 21 of 243 steady states in the Maddison dataset, and only 6 of 110 in the Penn World Table dataset. Further, the observations are all drawn from the tails of the component, high and low, in both datasets. It is possible that this second component is thus serving as a 'noise' component, only absorbing outliers or possibly fatter tails to the distribution than a normal distribution can accomodate.[19]

In the Maddison data, the second component is clearly capturing only extreme growth rates, with every steady state either less than -3% or greater than 10%. As is

[19]Estimating a mixture using Student's t-distribution components rather than normals, I found the model overwhelming selected a single component with very fat tails (degrees of freedom less than two). However, using simulated data the estimation procedure used to take draws from Student's t-distributions required extremely large gaps between means in order to differentiate two components from a single component with excessively high degrees of freedom.

Table 3.9: Steady States Sorted into Noise Component (Maddison)

Country	Years Begin		Years End	Duration	Growth Rate
Kenya	1952	–	1953	2	-5.70%
South Korea	1952	–	1953	2	10.10%
Gabon	1977	–	1978	2	-27.10%
Nicaragua	1978	–	1979	2	-24.90%
Nigeria	1968	–	1970	3	16.20%
Cape Verde	1978	–	1980	3	13.40%
Hungary	1989	–	1991	3	-6.20%
Mongolia	1990	–	1992	3	-12.50%
North Korea	1994	–	1996	3	-15.30%
Sierra Leone	1995	–	1997	3	-21.30%
Iran	1977	–	1980	4	-13.70%
El Salvador	1979	–	1982	4	-4.90%
Qatar	1981	–	1986	6	-18.20%
UAE	1981	–	1986	6	-9.30%
Qatar	1974	–	1980	7	-3.40%
Bolivia	1979	–	1986	8	-3.30%
Cameroon	1987	–	1994	8	-7.80%
Zimbabwe	1999	–	2008	10	-5.40%
Equatorial Guinea	1995	–	2008	14	19.50%
Côte d'Ivoire	1982	–	2008	27	-3.60%
Comoro Islands	1972	–	2008	37	-3.60%

shown in Table 3.9, none of the countries represented spend the entire sample in the second component, and these extreme steady states tend to be quite short lived. The median duration of steady states in this component is four years, and of the 21 steady states in the component, only four—Comoro Islands beginning in 1972, Côte d'Ivoire beginning in 1982, Equatorial Guinea beginning in 1995, and Zimbabwe beginning in 1999—are sustained for ten or more years.

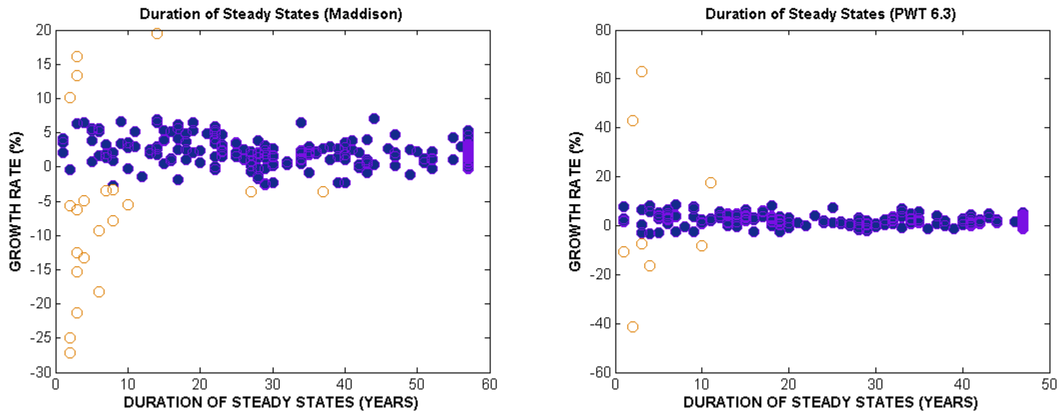
The result that more than 85% of steady state since 1952 appear to be drawn from a single normal distribution, with the remaining steady states well captured by a 'noise' or 'heavy-tails' component, is surprising. Theoretical research such as Murphy et al. (1989), Azariadis and Drazen (1990), and Azariadis (1996) has focused on the possibility of underdevelopment or poverty traps, with empirical applications such as Bloom et al. (2003) and Graham and Temple (2006) still occupying a prominent place in the development literature. Others such as Pritchett (2000), Paap et al. (2005), Jerzmanowski (2006), and Owen et al. (2009) have found a variety of growth patterns or groups across countries. The results presented here suggest that the range of growth patterns, patterns which are clearly visible when estimating steady-state growth rates within countries, can almost entirely be accounted for by a single normal distribution. Periods of economic take-off or decline, the miracles and disasters of economic growth, are drawn from a single diffuse distribution that may best be thought of as capturing noise or fat-tails in the growth process.[20] At least in unconditional steady-state growth rates, there does not appear to be a discrete distinction between countries with near-zero growth and positive growth.

Yet these results also run against the neoclassical growth literature and the idea that worldwide technological progress drives a common steady-state growth rate across countries. The variance of the main component in both datasets implies a standard deviation of 1.9–2.3%; this indicates a wide range in the steady-state growth rates themselves, not just annual rates around some common steady state. Further, although the steady states in the noise component tend to be more brief, the range in the main component does not fall for longer duration steady states (see Figure 3.9). There is a broad range of growth rates even in the most persistent steady states. This stands in contrast to the neoclassical model, which would imply that the standard deviation be small enough to only reflect measurement error in the sampler.

The results do not indicate discrete growth clubs, but neither do they suggest a distribution with very little variation around a worldwide steady-state growth rate. In-

[20]This suggests that heavy-tailedness could describe steady-states across countries in much the same way that Fagiolo et al. (2008) show it describes annual growth within countries.

Figure 3.9: Range of Growth Rates by Duration



Solid circles indicate steady states from the first component, while open circles indicate steady states from the noise component. The dispersion of steady states in the first component is relatively constant regardless of how long the country remains in the steady state.

stead, they suggest that there is substantial variability across countries in steady-state growth, and that this variability follows a single, potentially fat-tailed distribution. The key question this raises from a theoretical standpoint, then, is not how to account for growth clubs, but what places a country high or low within this distribution.

To illustrate the kind of model being described, consider the work of Davis (2008). This model includes a growth regime similar to an AK model, in which the growth rates of individual countries will be depend on institutional characteristics, and the distribution of growth rates within this group could be quite broad depending on the range of institutions represented.[21] Howitt and Mayer-Foulkes (2005) have similar implications for below-frontier growth, although their model implies that frontier growth should appear as a peak in the upper tail of the distribution while Davis' model

[21] AK models have fallen out of favor in economic growth, although Todo and Miyamoto (2002) have suggested that scale effects may be more relevant than previously thought. Recent work by Bond et al. (2010) suggests that the relationship between investment and long run growth suggested by AK models holds quite well for non-OECD countries, which is precisely in line with the implications of this model.

suggests that some countries with small but rapidly growing market size should experience steady-state growth well above that of the frontier. The empirical results presented here indicate that these models, which imply a broad distribution of steady-state growth rates, deserve more focus in the literature.

3.5 Conclusion

This paper uses a mixture of normal components to characterize the distribution of steady-state growth rates in per capita income across countries. In order to capture the possibility of changes in steady-state growth within a country over time, I first use a dynamic mixture model previously applied to business cycle analysis to estimate underlying steady-state growth rates. The first stage resembles the within-country estimation of Hausmann et al. (2005) and Jones and Olken (2008), but captures dynamic features generally associated with theoretical steady states. I find a wide range of growth experiences captured in the movements of the steady-state growth rates, including major cross-country decelerations and growth reversals in the mid-1970s and around 1980, as well as many examples of strong growth accelerations throughout the 1952-2008 sample.

The second stage provides insight into the literature on poverty traps as well as work on multiple growth paths such as Howitt and Mayer-Foulkes (2005). Despite the wide range of experiences captured by the estimated steady-state growth rates, I find that over 85% of steady-state growth rates are well described by a single normal distribution, with the remaining observations all drawn from the tails of another normal component, which is best thought of as a noise or heavy-tailedness component.

These results suggest that poverty-trap stagnation and technological frontier growth are not well-defined, distinct groups, at least without conditioning on variables which are able to account for the singular (but possibly fat-tailed) nature of the distribution of steady-state growth rates. What is driving the distribution of steady-state growth rates, and the mechanisms by which they affect the growth process, are interesting questions for future theoretical and empirical work.

Bibliography

Alfo, M., Trovato, G., Waldmann, R., 2008. Testing for country heterogeneity in growth models using a finite mixture approach. Journal of Applied Econometrics 23, 487–514.

Azariadis, C., December 1996. The economics of poverty traps part one: Complete markets. Journal of Economic Growth 1, 449–486.

Azariadis, C., Drazen, A., 1990. Threshold externalities in economic development. Quarterly Journal of Economics 105 (2), 501–526.

Barro, R., Sala-i Martin, X., 2002. Convergence. Journal of Political Economy 100, 223–251.

Baumol, W., 1986. Productivity growth, convergence, and welfare. American Economic Review 76, 1025–1054.

Bernanke, B., Gurkaynak, R., 2002. Is growth exogenous? taking Mankiw, Romer, and Weil seriously. In: Bernanke, B., Rogoff, K. (Eds.), NBER Macroeconomics Annual 2001. MIT Press, Cambridge, MA, pp. 11–57.

Bianchi, M., 1997. Testing for convergence: Evidence from non-parametric multimodality tests. Journal of Applied Econometrics 12 (4), 393–409.

Bloom, D., Canning, D., Sevilla, J., 2003. Geography and poverty traps. Journal of Economic Growth 8, 355–378.

Bond, S., Leblebicioğlu, A., Schianarelli, F., 2010. Capital accumulation and growth: A new look at the empirical evidence. Journal of Applied Econometrics n/a, n/a.

Canova, F., February 2004. Testing for convergence clubs in income per capita: A predictive density approach. International Economic Review 45 (1), 49–77.

Cappé, O., Robert, C., Rydén, T., 2003. Reversible jump, birth-and-death and more general continuous time Markov chain Monte Carlo samplers. Journal of the Royal Statistical Society 65 (3), 679–700.

Carter, C., Kohn, R., 1994. On Gibbs sampling for state-space models. Biometrika 81, 541–553.

Casella, G., George, E., August 1992. Explaining the gibbs sampler. The American Statistician 46 (3), 167–174.

Caselli, F., Esquivel, G., Lefort, F., Sep 1996. Reopening the convergence debate: A new look at cross-country growth empirics. Journal of Economic Growth 1, 363–389.

Caselli, F., Feyrer, J., July 2007. The marginal product of capital. Quarterly Journal of Economics 122 (2), 535–568, working Paper.

Chib, S., Greenberg, E., November 1995. Understanding the metropolis-hastings algorithm. The American Statistician 49 (4), 327–335.

Chirinko, R., 2008. σ: The long and short of it. Journal of Macroeconomics 30 (2), 671–686.

Comin, D., Hobijn, B., Rovito, E., 2008. Technology usage lags. Journal of Economic Growth 13 (4), 237–256.

Davis, L., 2008. Scale effects in growth: A role for institutions. Journal of Economic Behavior & Organization 66, 403–419.

Duffy, J., Papageorgiou, C., 2000. A cross-country empirical investigation of the aggregate production function specification. Journal of Economic Growth 5 (1), 87–120.

Durlauf, S., Johnson, P., 1995. Multiple regimes and cross-country growth behavior. Journal of Applied Econometrics 10 (4), 365–384.

Easterly, W., Levine, R., 2001. It's not factor accumulation: Stylized facts and growth models. World Bank Economic Review 15 (2), 177–219.

Fagiolo, G., Napoletano, M., Roventini, A., 2008. Are output growth-rate distributions fat-tailed? some evidence from oecd countries. Journal of Applied Econometrics 23, 639–669.

Früwirth-Schnatter, S., 2006. Finite Mixture and Markov Switching Models, 1st Edition. Springer Series in Statistics. Springer.

Galor, O., July 1996. Convergence? inference from theoretical models. Economic Journal 106, 1056–1069.

Gelman, A., Rubin, D., November 1992. Inference from iterative simulation using multiple sequences. Statistical Science 7 (4), 457–472.

Gerlach, R., Carter, C., Kohn, R., 2000. Efficient Bayesian inference for dynamic mixture models. Journal of the American Statistical Society 95 (451), 819–828.

Giordani, P., Kohn, R., 2008. Efficient Bayesian inference for multiple change-point and mixture innovation models. Journal of Business and Economic Statistics 26 (1), 66–77.

Giordani, P., Kohn, R., van Dijk, D., 2007. A unified approach to nonlinearity, structural change, and outliers. Journal of Econometrics 137, 112–133.

Gollin, D., Apr. 2002. Getting income shares right. The Journal of Political Economy 110 (2), 458–474.

Graham, B., Temple, J., 2006. Rich nations, poor nations: How much can multiple equilibria explain? Journal of Economic Growth 11, 5–41.

Green, P., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82 (4), 711–732.

Grier, K., Grier, R., 2007. Only income diverges: A neoclassical anomaly. Journal of Development Economics 84, 25–45.

Grier, K., Maynard, N., October 2010. From development to polarization: The evolution of the world income distribution using variable dimension mixture models, Working Paper.

Grier, K., Tullock, G., September 1989. An empirical analysis of cross-national economic growth, 1951-1980. Journal of Monetary Economics, 256–76.

Hall, R., Jones, C., Feb. 1999. Why do some countries produce so much more output per worker than others? The Quarterly Journal of Economics 114 (1), 83–116.

Hausmann, R., Pritchett, L., Rodrik, D., 2005. Growth accelerations. Journal of Economic Growth 10 (4), 303–329.

Henderson, D., Parmeter, C., Russell, R. R., October 2008. Modes, weighted modes, and calibrated modes: Evidenceof clustering using modality tests. Journal of Applied Econometrics 23, 607–638.

Heston, A., Summers, R., Aten, B., August 2009. Penn world table version 6.3. Tech. rep., Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania.

Howitt, P., Mayer-Foulkes, D., February 2005. R&D, implementation, and stagnation: A Schumpeterian theory of convergence clubs. Journal of Money, Credit and Banking 37 (1), 147–177.

Islam, N., Nov. 1995. Growth empirics: A panel data approach. The Quarterly Journal of Economics 110 (4), 1127–1170.

Jerzmanowski, M., 2006. Empirics of hills, plateaus, mountains and plains: A Markov-switching approach to growth. Journal of Development Economics 81, 357–385.

Jones, B., Olken, B., August 2008. The anatomy of start-stop growth. Review of Economics and Statistics 90 (3), 582–587.

Jones, C., August 1995a. R&D-based models of economic growth. Journal of Political Economy 103 (4), 759–784.

Jones, C., May 1995b. Time series tests of endogenous growth models. Quarterly Journal of Economics 110 (2), 495–525.

Klump, R., De La Grandville, O., March 2000. Economic growth and the elasticity of substitution: Two theorems and some suggestions. American Economic Review 90 (1), 282–291.

Klump, R., McAdam, P., Willman, A., 2007. Factor substitution and factor-augmenting technical progress in the United States: A normalized supply-side system approach. Review of Economics and Statistics 89 (1), 183–192.

Koop, G., 2003. Bayesian Econometrics. Wiley.

Koop, G., Leon-Gonzalez, R., Strachan, R., April 2009. On the evolution of the monetary policy transmission mechanism. Journal of Economic Dynamics and Control 33 (4), 997–1017.

Lavine, M., West, M., 1992. A Bayesian method for classification and discrimination. Canadian Journal of Statistics 20 (4), 451–461.

Lee, K., Pesaran, M. H., Smith, R., 1997. Growth and convergence in a multi-country empirical stochastic solow model. Journal of Applied Econometrics 12, 357–392.

Maddison, A., 2010. The World Economy Volume 2: Historical Statistics. First edition 2003. OECD Publishing.

Mankiw, G., Romer, D., Weil, D., May 1992. A contribution to the empirics of economic growth. The Quarterly Journal of Economics 107 (2), 407–437.

Masanjala, W., Papageorgiou, C., 2004. The Solow model with CES technology: Nonlinearities and parameter heterogeneity. Journal of Applied Econometrics 19, 171–201.

McQuinn, K., Whelan, K., 2007. Conditional convergence and the dynamics of the capital-output ratio. Journal of Economic Growth 12 (2), 159–184.

Miyagiwa, K., Papageorgiou, C., 2007. Endogenous aggregate elasticity of substitution. Journal of Economic Dynamics & Control 31, 2899–2919.

Murphy, K., Shleifer, A., Vishny, R., 1989. Industrialization and the big push. Journal of Political Economy 97 (5), 1003–1026.

Nordhaus, W., 1992. Lethal model 2: The limits to growth revisited. Brookings Papers on Economic Activity 2, 1–43.

Owen, A., Videras, J., Davis, L., 2009. Do all countries follow the same growth process? Journal of Economic Growth 14 (4), 265–286.

Paap, R., Franses, P., van Dijk, D., August 2005. Does Africa grow slower than Asia, Latin America and the Middle East? evidence from a new data-based classification method. Journal of Development Economics 77 (2), 553–570.

Paap, R., van Dijk, H., 1998. Distribution and mobility of wealth of nations. European Economic Review 42 (7), 1269–1293.

Papastamoulis, P., Iliopoulos, G., June 2010. An artificial allocations based solution to the label switching problem in bayesian analysis of mixtures of distributions. Journal of Computational and Graphical Statistics 19 (2), 313–331.

Pittau, M., Zelli, R., Johnson, P., 2010. Mixture models, convergence clubs, and polarization. Review of Income and Wealth 56 (1), 102–122.

Primiceri, G., 2005. Time varying structural vector autoregressions and monetary policy. Review of Economic Studies 72 (3), 821–852.

Pritchett, L., 1997. Divergence, big time. Journal of Economic Perspectives 11 (3), 3–17.

Pritchett, L., 2000. Understanding patterns of economic growth: Search for hills among plateaus, mountains, and plains. World Bank Economic Review 14 (2), 221–250.

Quah, D., 1993. Galton's fallacy and tests of the convergence hypothesis. Scandinavian Journal of Economics 95, 427–443.

Quah, D., July 1996. Twin peaks: Growth and convergence in models of distribution dynamics. Economic Journal 106 (437), 1045–1055.

Quah, D., 1997. Empirics for growth and distribution: Stratification, polarization, and convergence clubs. Journal of Economic Growth 2, 27–59.

Raftery, A., Lewis, S., November 1992. Comment: One long run with diagnostics: Implementation strategies for Markov chain Monte Carlo. Statistical Science 7 (4), 493–497.

Richardson, S., Green, P., 1997. On Bayesian analysis of mixtures with an unknown number of components. Journal of the Royal Statistical Society 59 (4), 731–792.

Richardson, S., Green, P., 1998. Corrigendum: On Bayesian analysis of mixtures with an unknown number of components. Journal of the Royal Statistical Society 60 (3), 661.

Robert, C., Casella, G., 2004. Monte Carlo Statistical Methods, 2nd Edition. Springer Texts in Statistics. Springer.

Romer, D., 2006. Advanced Macroeconomics, 3rd Edition. McGraw-Hill Irwin.

Sachs, J., 2001. Tropical underdevelopment, nBER Working Paper No. 8119.

Sala-i Martin, X., 1996. The classical approach to convergence analysis. Economic Journal 106, 1019–1036.

Solow, R., 1956. A contribution to the theory of economic growth. Quarterly Journal of Economics 70, 65–94.

Solow, R., August 1957. Technical change and the aggregate production function. Review of Economics and Statistics 39 (3), 312–320.

Todo, Y., Miyamoto, K., 2002. The revival of scale effects. Topics in Macroeconomics 2 (1).

Valentinyi, A., Herrendorf, B., 2008. Measuring factor income shares at the sectoral level. Review of Economic Dynamics 11, 820–835.

Waagepetersen, R., Sorensen, D., April 2001. A tutorial on reversible jump MCMC with a view toward applications in QTL-mapping. International Statistical Review 69 (1), 49–61.

White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48, 817–838.

Zuleta, H., 2008. Factor saving innovations and factor income shares. Review of Economic Dynamics 11, 836–851.

Appendices

Appendix A

Bayesian Analysis and Reversible Jump MCMC

A.1 Bayesian Analysis and MCMC

Classical estimation techniques revolve around identifying the set of parameters, which we will refer to as θ^*, which maximize the *likelihood function* of the data, $p(y\,|\,\theta)$. Typically a familiar functional form is assumed, and the likelihood maximizing parameters are found either analytically or through iterative numerical methods.

In Bayesian estimation, the object is not selecting parameters to maximize a function, but rather to identify the entire distribution of the parameters conditional on the data observed, or the *posterior distribution*. The posterior probability of the parameters is $p(\theta\,|\,y) = [p(y\,|\,\theta)\,p(\theta)]\,/\,p(y)$ by Bayes rule. The typical Bayesian formulation is

$$p(\theta\,|\,y) \propto p(y\,|\,\theta) \times p(\theta) \tag{A.1}$$

where $p(\theta)$ is the *prior distribution*, which incorporates any previous knowledge the researcher may have about the parameters in question. In practical terms, the prior may be tuned to the data being analyzed in a way that reflects what might be gleaned from naive analysis, and in general priors may be designed to have little influence on the posterior, rather than being informative.

Since these posterior distributions are often complex and cannot be expressed analytically except for the simplest cases, they are frequently simulated using Markov chain Monte Carlo (MCMC) simulation techniques. MCMC simulation methods come in two main varieties: Gibbs sampling takes random draws from the conditional posterior distribution of a single parameter (conditional on the current values of all other parameters), stores this draw as the new value of the parameter, then continues to draw from the conditional posteriors of each other parameter; Metropolis-Hastings sampling takes random draws from a more conventional distribution and, with probabilities based on the value of the posterior at this draw, probabilisticaly either accepts the draw and stores it as a draw from the posterior or rejects the draw and stores the previous draw.

In both cases the Monte Carlo simulation forms a Markov chain in that the value of the draws are serially dependent on the immediately preceding draw. As long as the probability of moving between two values in the Markov chain are the same regardless of the direction of movement, a property known as detailed balance, it has been shown that any MCMC method will converge to drawing from an equilibrium distribution which is exactly the posterior distribution of interest.[1]

[1]For a more thorough introduction to MCMC techniques for Bayesian statistics, see Casella and George (1992) and Chib and Greenberg (1995). Robert and Casella (2004) and Früwirth-Schnatter (2006) provide a more complete discussion of of Bayesian approaches to mixture analysis.

A.2 Reversible Jump MCMC

What follows is a brief technical summary of the steps in the reversible jump Markov chain Monte Carlo approach. Waagepetersen and Sorensen (2001) provide a tutorial of reversible jump MCMC, and Green (1995) offers a more thorough presentation.

1. Initially the parameter vector for the current model $\mathscr{M}_j$ should be updated according to some standard MCMC method (In my case the Gibbs sampler). The Gibbs sampler takes advantage of the fact that the conditional distributions of the parameters have simple closed form solutions, although the joint distribution can be quite complicated. A sweep takes some initial values of all parameters but one, then takes a random draw of this last parameter from its conditional distribution. This draw is then counted as the new value for that parameter. This is done for each of the parameters in turn until all have been drawn, in which case the new draws are treated as new initial values in another sweep. After a sufficient number of draws, this sampling procedure will converge to the true joint posterior, and subsequent draws from the conditionals will also be draws from the joint posterior. The exact conditional posteriors used in this paper are as follows:

   (a) $\eta_j \mid \cdots \sim \mathscr{D}(\delta + n_1, \ldots, \delta + n_k)$ where $n_j = \sum_{i=1}^{N} \mathscr{I}_{\{z_i = j\}}$;

   (b) $\mu_j \mid \cdots \sim \mathscr{N}\left\{\left(\sigma_j^{-2}\sum_{i=1}^{N}\mathscr{I}_{\{z_i=j\}} y_i + \kappa\xi\right) / \left(\sigma_j^{-2} n_j + \kappa\right),\ \left(\sigma_j^{-2} n_j + \kappa\right)^{-1}\right\}$;

   (c) $\sigma_j^{-2} \mid \cdots \sim \Gamma\left(\alpha + n_j/2,\ \beta + \frac{1}{2}\sum_{i=1}^{N}\mathscr{I}_{\{z_i=j\}}\left(y_i - \mu_j\right)^2\right)$;

   (d) $\beta \mid \cdots \sim \Gamma\left(g + k\alpha,\ h + \sum_{j=1}^{k}\sigma_j^{-2}\right)$; and

   (e) $p(z_i = j \mid \cdots) \propto \left(\eta_j/\sigma_j\right) exp\left\{-\left[y_i - \mu_j\right]^2 / 2\sigma_j^2\right\}$, which is normalized so that $\sum_{j=1}^{k} p(z_i = j \mid \cdots) = 1$;

2. For the between models moves, Richardson and Green (1997) utilize two steps: a split-combine move and a birth-death move. Cappé et al. (2003) show that using the birth-death step alone will also converge to the true posterior. As this step is substantially simpler to code, I procede without the split-combine step. A general formulation of the reversible jump algorithm is given below, followed by the specific birth-death application used.

   (a) Randomly choose (with predetermined probability p_{ij}) an alternative model $\mathscr{M}_j$ from the available options, and propose a move to the alternative parameter space of the randomly chosen model.

   (b) Determine the parameters to use for the proposal. Since the parameter spaces are of different sizes, if the size of the parameter space is greater in the proposed model, generate a random vector u_i from a proposal generating density $q_{ij}(u_i|\theta_i)$ such that $dim(u_i) = dim(\theta_j) - dim(\theta_i)$ and $\theta_j = \varphi_{ij}(\theta_i, u_i)$, where $\varphi_{ij}(\cdot)$ is an invertible mapping. If $dim(\theta_j) < dim(\theta_i)$, then for the proposed jump set $(\theta_j, u_j) = \varphi_{ji}^{-1}(\theta_i)$ (in such a case $u_j$is irrelevant, but is included here for completeness).

   (c) Make the jump from $\mathscr{M}_i$ to $\mathscr{M}_j$ with probability

$$\alpha = \min\left\{1, \frac{P(\mathscr{M}_j, \theta_j|y)p_{ji}}{P(\mathscr{M}_i, \theta_i|y)p_{ij}q_{ij}(u_1|\theta_i)}J\right\}$$

   where $P(\mathscr{M}_j, \theta_j|y)$ is the posterior probability of the model and parameters; in practice, since this is proportional to the likelihood multiplied by the prior, this is replaced with $P(y|\theta_j)P(\theta_j)$ as long as the models i and j are known up to the same multiplicative constant. J here is the determinant of the Jacobian of the transition function, or

$$J = \left| \frac{\partial \varphi_{ij}(\theta_i, u_i)}{\partial(\theta_i, u_i)} \right|,$$

and is necessary in order for the dimension matching of Green (1995) to occur.[2]

3. The birth-death step randomly chooses between two possible types of jump: increasing the number of components by one through the birth of an empty component (that is, a component with no observations allocated to it) and decreasing the number of components by one through the death of an empty component.

   (a) A birth move is chosen with probability b_k and a death move is chosen with probability $d_k = 1 - b_k$. Since a single component cannot have a death and once the maximum model size is reacher ther cannot be a birth, $b_1 = 1$ and $b_{kmax} = 0$, with d_1 and d_{kmax} defined appropriately. For all other jumps, I set $b_k = d_k = 0.5$.

   (b) For the birth step, a new component is generated with the parameters drawn from the prior distributions:

      i. $\mu_{j^*} \sim \mathcal{N}\left(\xi,\ \kappa^{-1}\right)$;

[2]Note that if we were to consider in our mapping from section 3 the possibility of mapping in both directions (which may be necessary in the case of non-nested models), we would additionally require random vector $u_j$from proposal generating density $q_{ji}(u_j|\theta_j)$ such that $dim(u_i) + dim(\theta_i) = dim(u_j) + dim(\theta_j)$. The non-unit expression in α would then need to be multiplied by the new proposal generating density. Combining this with the use of the likelihood and prior to determine the posterior yields the following intuitively paired interpretation of α:

$$\begin{aligned} \alpha &= \min\left\{1,\ \frac{P(y|\theta_j)}{P(y|\theta_i)} \times \frac{P(\theta_j)}{P(\theta_i)} \times \frac{p_{ji}}{p_{ij}} \frac{q_{ji}(u_2|\theta_j)}{q_{ij}(u_1|\theta_i)} \times \left| \frac{\partial \varphi_{ij}(\theta_i, u_i)}{\partial(\theta_i, u_i)} \right| \right\} \\ &= \min\{1,\ (likelihood\ ratio) \times (prior\ ratio) \\ &\quad \times (proposal\ ratio) \times (Jacobian)\} \end{aligned}$$

ii. $\sigma_{j^*}^{-2} \sim \Gamma(\alpha,\, \beta)$; and the weight drawn from

iii. $\eta_{j^*} \sim Beta(1,\, k)$where k is the number of components before the birth.

iv. Once the new component is created, the existing weights are mulitiplied by $1-\eta_{j^*}$ so that the weights of the new model sum to 1.

(c) For the death step, each empty component is given an equal probability of being chosen and removed (so the probability of a particular component being removed depends not only on the total number of components k but on how many of those components have $n_{j^*}=0$). Once component j^* has been removed, the existing weights are divided by $1-\eta_{j^*}$ so that the weights of the new model sum to 1.

4. When the proposed jump has been formulated, the algorithm probabilistically jumps to the new model with probability $\min(1,\, A)$ for a birth and $\min\left(1,\, A^{-1}\right)$ for a death, where A is defined below:

   (a) The prior ratio is $PR = p(\mathscr{M}_{k+1}) / [p(\mathscr{M}_k) \times B(k\delta,\, \delta)] \times (k+1) \times \eta_{j^*}^{\delta-1}\left(1-\eta_{j^*}\right)^{k\delta-k+N}$, where $B(\cdot,\, \cdot)$ is the Beta function and N is the total number of observations;

   (b) The proposal ratio is $PROPR = d_{k+1} / \left[(k_0+1) \times b_k \times beta_{1,k}\left(\eta_{j^*}\right)\right]$, where k_0 is the number of empty components before the birth and $beta_{1,k}(\cdot)$ is the density of $Beta(1,\, k)$;

   (c) The determinant of the Jacobian is $|J| = \left(1-\eta_{j^*}\right)^k$; and therefore

   (d) $A = PR \times PROPR \times |J|$ since the likelihood ratio is 1. Modification of these definitions for the death step is straightforward.

5. I iterate through this algorithm to generate a sufficiently long converged chain, remove some amount of burn-in at the beginning for which the chain has not yet converged, and conduct inference based on the posterior of each model and the parameters of the most likely model.[3]

A.3 Does our sampler work?

At the end of the day, what you get from MCMC is the output of the sampler. Typically, there are no goodness of fit metrics available. So it is important to establish that our sampler works well in situations where we already know something about the correct answer. To that end, we present a replication of a result in RG and also a small simulation experiment.

A.3.1 Replicating Richardson and Green's results using the Galaxy data

One of the three datasets RG use is the "Galaxy data" often used in other papers on mixture modeling. It comprises a velocity measure for 82 galaxies. Figure A.1 (left panel) presents a kernel density plot of the data. We use the same priors as RG, run our sampler for 300,000 draws and discard the first 100,000 as burn-ins. We differ from RG by limiting the maximum number of components to 10 (compared to their 30), and using only birth / death steps to change dimensions. In other respects our sampler is the same as RG, including labeling components for identification purposes by the order of the means.[4]

[3]See Raftery and Lewis, 1992 for a discussion of long chain Bayesian diagnostics, as opposed to the multiple sequence procedure advocated by Gelman and Rubin, 1992

[4]More specifically, we sort each draw in the order of the drawn means from lowest to highest. We do this in order to avoid the problem of "label switching," in which draws from distinct components cannot be readily distinguished. Although there are limitations to using the order of the means as the

Figure A.1: Kernel Plots of Galaxy and Simulation Data

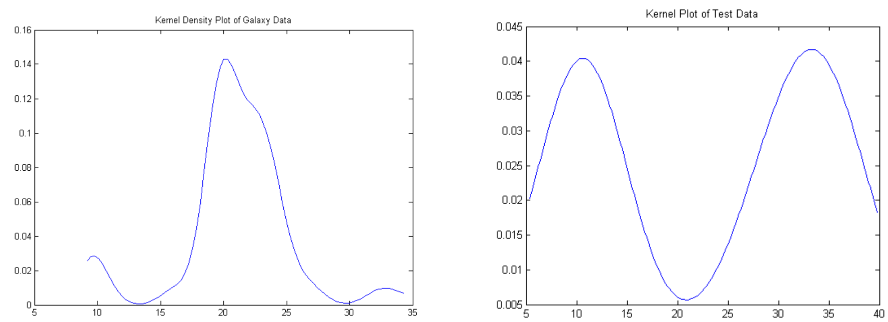


Table A.1: Posterior Probabilities of Test Data

PANEL A:	p(1)	p(2)	p(3)	p(4)	p(5)	p(6)	p(7)	p(8)	p(9)	p(10)
Galaxy (RG)	0	0	.061	.128	.182	**.199**	.16	.109	.071	.04
Replication	0	0	.071	.153	.194	**.198**	.164	.1096	.069	.041
PANEL B:										
Simulation	0	.301	**.379**	.184	.076	.036	.014	.0062	.0018	.0003

Percentage of jumps accepted: Galaxy RG–4%; Galaxy replication–16.6%; Simulation experiment–3.8%.

Table A.1 (panel A) compares the posterior distribution for the number of components, k, in the galaxy data presented in RG with the posterior we generate with our sampler.[5] As can be seen, our sampler puts a bit more posterior probability on a lower number of components than does RG's (which is likely because of the fact that we limit the maximum number of components to 10), but overall the two posterior densities are quite close. Both samplers agree that $k = 5$ or $k = 6$ are the most likely number of components and both put no probability on $k \leq 2$.

labeling criterion, in practice we do not encounter these difficulties for the models with the numbers of components we interpret.

[5]Richardson and Green (1998) point out that their original paper included an error in the acceptance probability equation. The changes to their results were quite small, and we have corrected for the error in our own code.

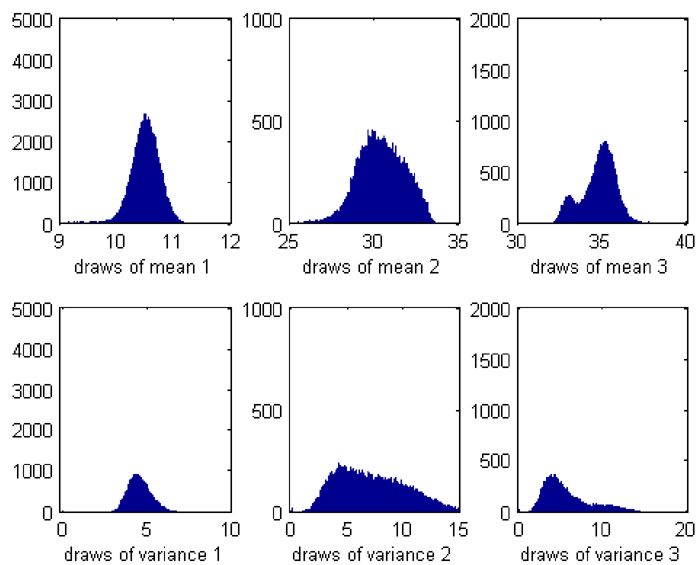
A.3.2 An experiment using simulated data

The above exercise shows that our sampler produces results similar to another well known, highly cited, sampler. Here we create a dataset that is a mixture of three underlying components and check to see if our sampler can recover the correct number of components, the parameters of the components, and the correct allocation of observations to components.

Specifically we create a 200 observation dataset that is a mixture of three normal components. 45% of the data are drawn from a normal with mean 10 and standard deviation 2, 22.5% from a normal with mean 30 and standard deviation 2.5, and 32.5% from a normal with mean 35 and standard deviation 2. Figure 1 (right panel) shows a kernel density plot of the data. Notice that even though we know there are three underlying components, the kernel density has only two modes.

Panel B of Table 1 shows the posterior distribution for the number of components in the data as generated by our sampler. We again form our priors as in RG, set k_{max} equal to 10, and take 300,000 draws discarding the first 100,000 as burn-in. Our sampler finds a strong mode at $k = 3$ (41% of the posterior probability is on this point), and almost no probability on $k > 4$. Figure A.2 presents the posterior distributions for the parameters of the three distributions estimated by the $k = 3$ output of the sampler, showing that these parameters are being estimated fairly well. Finally, we note that our algorithm correctly classifies 188 of the 200 observations in the sample, which is 94% success.

Figure A.2: Posterior Distributions of Simulated Data Parameters ($k = 3$)



The means of the posterior distributions for the means and variances are as follows: Component 1–mean 10.2, standard deviation 2.2; Component 2–mean 28.7, standard deviation 2.7; Component 3–mean 34.9, standard deviation 2.4.