GROWTH AND DISTRIBUTION

by

Danny Quah*
LSE Economics Department
April 2007

* I thank Lucien Foldes, Vinayak Nagaraj, and Chris Rogers for helpful advice.

GROWTH AND DISTRIBUTION

by

Danny Quah
LSE Economics Department

April 2007

ABSTRACT

[[Foreword or abstract goes here. Richard Stone Lectures in London. Various Econometric Society presentations. Who this book is for (graduate student background in mathematics and probability through to current research findings, and extensions and new directions for further work)]]

Keywords: convergence, diffusion, inequality, kernel, Markov, panel data, polarization, poverty, resolvent, stochastic kernel, transition

JEL Classification: C23, D30, D63, 010, O41, O57

Communications to: D. Quah, Economics Department, LSE, Houghton Street, London WC2A 2AE

Tel: +44/0 20 7955.7535, Email: dq@econ.lse.ac.uk
(URL) http://econ.lse.ac.uk/staff/dquah/

GROWTH AND DISTRIBUTION

by

Danny Quah*

LSE Economics Department

April 2007

1 Introduction

This manuscript describes some research on international patterns of economic growth and income distribution. It takes a unified approach to global growth and inequality by analyzing the dynamics of income distributions.

The chapters that follow do three things: First, they document the facts on economic growth, income distribution, and global inequality. Second, they describe the mathematical and statistical tools for modelling distribution dynamics. Third, they provide a number of analytical models designed to help understand the evolution of worldwide patterns of growth and inequality. No single all-encompassing picture emerges to be apposite everywhere but, nonetheless, several powerful conclusions hold.

Before laying out those messages, it is useful to set the stage by making explicit a simple and stylized version of the questions and potential conclusions that will be addressed in detail. Figures 1.1 and 1.2 provide a compact illustration.

In these Figures time evolves along the horizontal axis. The vertical axis maps not just a scalar—if so then the Figures would simply graph the dynamic profile of a timeseries variable—but instead in these Figures each time-t observation is the directly observed density of a cross-sectional distribution. Depending on the application these cross-sectional distributions can be defined over different economic variables—they can be the cross-section of per capita incomes across

* I thank Lucien Foldes, Vinayak Nagaraj, and Chris Rogers for helpful advice.

Figure 1.1 Emerging twin peaks in the cross-country income distribution: The labels describe *possible outcomes* and caricature several **named examples**.

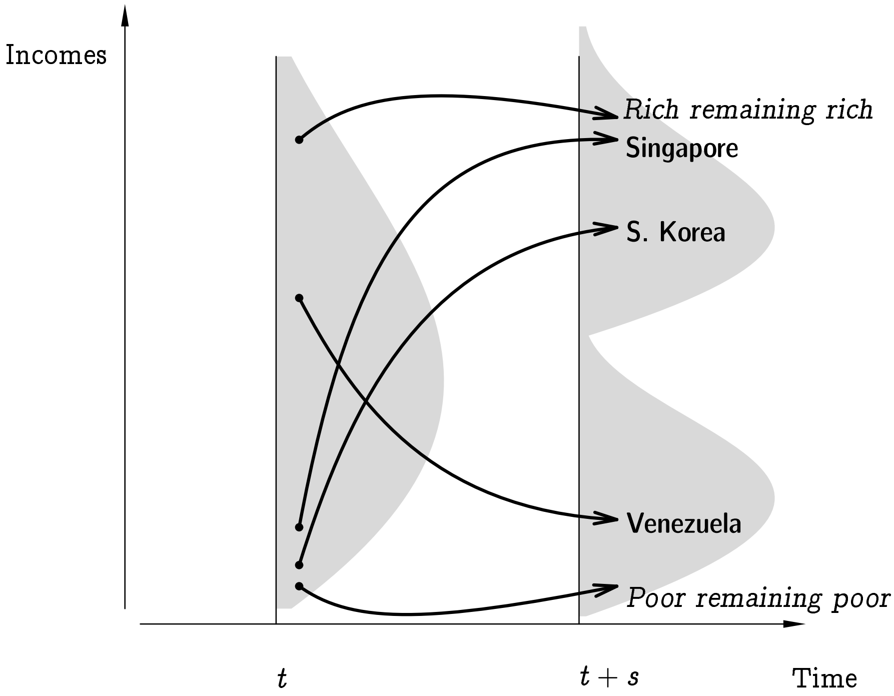
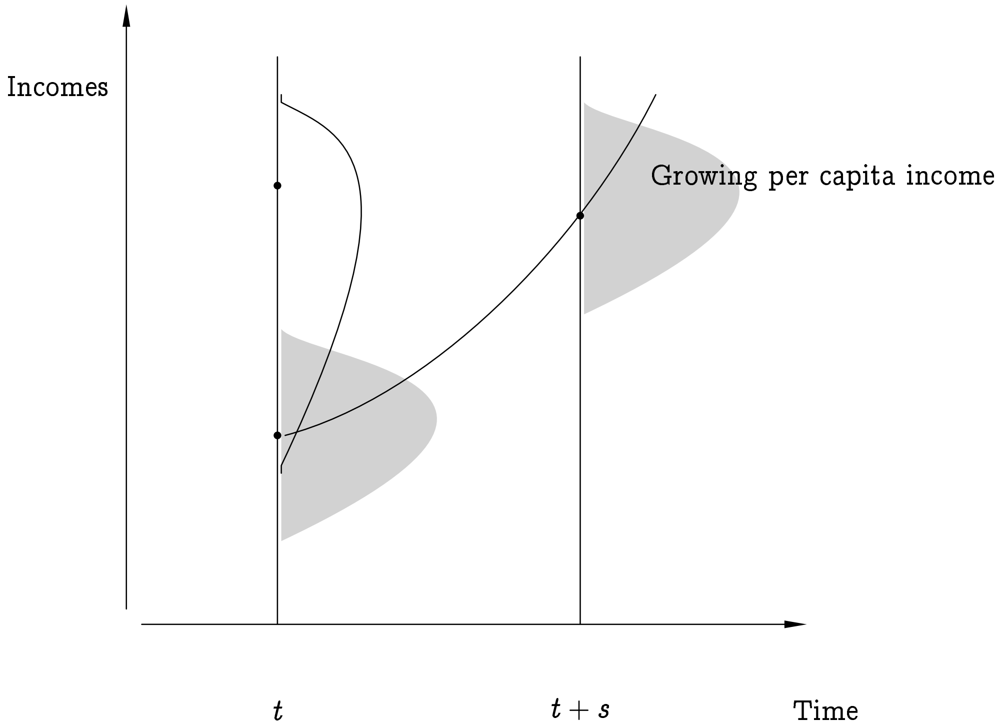


Figure 1.2 Individual country economic growth and distribution: The shaded areas describe the evolution in time of the income distribution within a given country, one whose average income grows as indicated in the labelled line. Each individual country in the cross-country distributions of Figure 1.1 has its own within-country income distributon dynamics, shown in this figure by a representative other time-t density around a different per-capita income observation. Individual country distributions in the Figure are usefully also scaled by population to better indicate the relative importance of different-sized countries. The curves plotted then are of course no longer densities.



countries; the cross-section of average productivity across economies; the cross-section of individual incomes across people; and so on.

Figure 1.1 depicts the distribution dynamics in cross-country income distributions. Here, each element in the distribution is an economy-wide per-capita income observation. In this Figure China, with over a billion inhabitants, and Singapore, with under 5 million, count equally in the cross-section per-capita income distribution. Figure 1.1 depicts a range of hypothetical behaviors. Some parts of the cross section originally rich can remain rich; those originally poor, remain poor: stratification occurs. Simultaneously the Figure illustrates, with no logical contradiction, intra-distribution transitions: some initially rich countries stagnate, decline to the bottom of the income distribution, and are later identified to be relatively poor; other, initially poor countries grow fast and are subsquently viewed as successful growth miracles. Thus, as illustrated in the Figure, both stratification and mobility simultaneously occur.

The shape of the distribution also evolves. In Figure 1.1 the earlier time-t distribution shows many middle-income countries clustering around the average and relatively few countries either very rich or very poor: the distribution is smoothly unimodal. By contrast the later time-$t+s$ distribution, as drawn, is sharply bimodal. Between times t and $t+s$ a pair of clusters have emerged, one grouped about an upper middle-income level and the other about a lower middle-income level. With timepaths continuous, this twinpeaks emergence can be interpreted as displaying "convergence clubs"-behavior, where countries sufficiently close to each other converge towards each other, whereas those sufficiently far apart, diverge.

Although correct the verbal description does not do full justice to the Figure: In the vanishing middle-income class that is shown, countries will end up diverging away from each other even though initially they had been relatively close. Thus, the Figure illustrates countries beginning alike each other and then continuing to converge towards each other, and yet others similarly close initially but then diverging away from each other over time.

In contrast to mapping the behaviour of country averages, Figure 1.2 takes individual per-capita income observations in the distributions previously graphed in Figure 1.1 and unpacks the income distribution around that average; at each instant in time each country's income distribution observation is scaled by its population. Doing this for every country in the the cross-section uncovers the income distribution across the more than 6 billion people on earth.[1]

Modelling *those* income distribution dynamics— done implicitly for the earlier and simpler Figure 1.1—would aid understanding the evolution of global poverty and inequality. Figure 1.2 can show how as an economy grows—i.e., as its macroeconomy develops and per capita income rises—different possibilities emerge: some people in society might be left behind; inequality across society might rise or fall; the very richest might pull ahead ever more, independent of summary measures of inequality; everyone within society might have their own incomes rise in tandem with the per-capita average; and so on.

A model of evolving distributions for the dynamics in Figures 1.1 and 1.2 provides three advantages over earlier extant approaches. First, working directly with the distribution itself leaves as late as possible in the analysis the decision on which inequality index to use. Indeed, for certain analyses a researcher can get away altogether without relying on any one index. Since every income inequality index—Gini, Theil, variance of the log, and so on—is an attempt to collapse the infinite-dimensioned information in a distribution func-

[1] To obtain the income distributions for a country in Figure 1.2 one does not need actual observations on individual incomes in that country. Nor does an income figure need to be imputed to each person in that county. Instead, summary statistics together with auxiliary assumptions—as in Bourguignon and Morrisson (2002), Quah (2003), Sala-i-Martin (2002a), or Chapter 6—will suffice to obtain useful estimates.

tion to just a summary scalar value, all inequality indexes necessarily discard information and thus have specific disadvantages, depending on the underlying distribution and the research question. But when using instead a law of motion for the entire distribution, a researcher doesn't have to rely on just a single index—any desired summary measures can instead be computed as desired, as the final step, not an initial one, in the analysis.

Second, a model for dynamics such as in Figures 1.1 and 1.2 permits bridging different literatures. Researchers in income inequality have traditionally worked on issues to which macroeconomists working in inequality and growth, say, have paid far less attention. For instance, income inequality research has developed axiomatic justification for different inequality indexes and considered the different benefits and disadvantages afforded by each such index; income inequality resarchers have worked on defining alternative notions of poverty, measuring (again indexes of) mobility, assessing the effect of imperfect data on measuring inequality, and so on (e.g., Cowell, 2000, 1995). On the other hand macroeconomists interested in growth and inequality have paid little attention to when different indexes might mislead or yet others might be more useful, but have looked mostly at their correlations with macroeconomic indicators such as growth and development.

The reason for this divergence is, of course, that the research questions differ across fields. But, nonetheless, since all these issues depend on the same underlying implicit model—distributions evolving through time–making that model explicit will allow greater cooperation and closer integration in research.

Third, many issues that arise ad hoc can be dealt with transparently in a unified way in a model of explicit distribution dynamics. As just one important example a perennial question is, How much world income inequality is due to income inequality within countries and how much due to macroeconomic income disparities between countries. How much is that changing through time?

Research attacking this problem with specific income inequality indexes then need to verify that intra-country inequality indexes and inter-country ones compose in appropriate ways (or, the opposite, that world income inequality measures decompose usefully into within-country ones). Sometimes, the researcher asks if the weighting scheme used to form a summary world inequality index should use country population weights—when China or India would then assume greater importance—or country income weights—when the US or Japan would then do so. If we construct the world income distribution explicitly from the underlying distributions, as done in the Figure, many of these concerns about the right way to do that accounting are simply irrelevant. There is one and only one way to form the dynamics of the world income distribution, and that is that used in Figure 1.2. (This does not diminish the importance of the conceptual problem of *decomposing* inequality measures, an issue to which we will return in Section 6.5 when we discuss the dynamics of measured world income inequality arising from inequality dynamics between countries and inequality dynamics within them.)

The same reasoning also clarifies what information is revealed when using cross-country per capita incomes rather than, say, cross-section distributions of individual incomes. It is not that the results obtained using only macroeconomic averages are biased—what are they biased for? Results obtained from cross-country average data in Figure 1.1 simply address questions other those treated when one uses within-country income distributions, Figure 1.2.

What probability and econometric models are appropriate for studying the distribution dynamics in Figures 1.1 and 1.2? A standard timeseries model might take individual countries or groups of individual countries and seek to characterize the dynamic properties of that vector. But this procedure—well-understood though it might be—neglects that it is an entire *distribution* whose dynamics we seek to understand. Thus, a standard timeseries approach does not tell us about, say, the emergence of multiple clusters; it is silent on what is happening to the interquartile range [the income distance between the 75th and 25th percentiles]; it does not inform on how different inequality and polarization measures are evolving through time.

On the other hand estimating a cross-section or panel data regression does characterize the behaviour of the conditional average in the cross-section. For addressing certain questions that might well be exactly what a researcher wishes to do. It would, however, be altogether uninformative for the dynamics of the entire distribution.

This book is concerned with modelling the dynamics of distributions in economic growth. The chapters to follow describe extant data on, provide mathematical tools to analyze, and draw conclusions about the kinds of questions raised in Figures 1.1 and 1.2. The natural mathematical framework for this work is that of differential equations taking values in a space of distributions. Just as a scalar differential equation might contain terms like $aX(t)$, where a is a constant real number and $X(t)$ is the value of the timepath at time t, the differential equations needed here will involve terms like $(\mathcal{T}^*)(\mathsf{F}_t)$ where $\mathcal{T}^*$ is an operator mapping distributions to distributions and F_t is the value at time t of the timepath of distributions. We use differential rather than difference equation models in the base framework because for the questions of interest here growth models are naturally set in continuous rather than discrete time (see, e.g., Merton, 1990).

The central empirical findings in this research are XXXX-fold:

(i) Over the last half-century, while economic growth has raised worldwide incomes by XX%, the cross-country income distribution has shown emerging twin peaks, i.e., both convergence and divergence have simultaneously occurred.

(ii) Over the last half-century, there have been both growth miracles and growth disasters. With some variation, growth disasters concentrate in the group of economies initially relatively poor; growth miracles, in that group initially already relatively better off.

(iii) Over the last half-century, while the cross-country income distribution has obviously evolved, some of its characteristics have not changed overly much. Its standard deviation has risen by XX%; the interquartile range by YY%; the 90-10 percentile spread by ZZ%. Contrast this with ...

(iv) Over the last half-century, the largest single factor driving inequality worldwide has been macroeconomic. For forming the worldwide distribution of income across people, the dynamics of country-wide average per-capita incomes have mattered much more than have the dynamics of inequality across people within countries.

This paper documents these empirical results and describes in the literature where similar such findings have been obtained. Those findings have added to interest in economic theories that emphasize certain kinds of discreteness and discontinuity, i.e., instances of economic nonconvexity, nonlinearity, and nonergodicity (e.g., Azariadis and Drazen, 1990; Durlauf, 1993, 1996; Galor and Zeira, 1993; Quah, 1996a). In turn, these theories have motivated yet further empirical analyses.

Such empirical approaches have also provided new perspectives on classical questions of inequality and growth (e.g., Quah, 2003; Sala-i-Martin, 2002a; Schultz, 1998); emphasized the importance of new features to consider in studying inequality and cross-section income distributions (e.g., Esteban and Ray, 1994; Wolfson, 1994); and motivated new models for analyzing spatial economic dynamics (e.g., Quah, 2002; Quah and Simpson, 2003). Although some of these issues fall outside traditional macroeconomic analyses of cross-country economic growth, they all relate to growth and distribution more generally, and so this paper will consider them where appropriate.

The remainder of this paper is organized as follows. To establish notation and to fix ideas, this rest of this section describes below neoclassical economic growth and convergence, set against a background of interest, not solely in the behavior of any one economy but, in the dynamics of a rich cross section of economies. This section is intended to be extremely terse and is included here principally so the reader identifies the notational convention used in the rest of the paper.

Section 2 documents some key lessons from the extant literature on the empirics of economic growth, inequality, and income distribution. Section 3 begins the analysis central to this paper. This section presents the basic mathematical structure underlying empirical analysis for distribution dynamics. It describes a number of the key findings for per capita incomes data across countries, and considers extensions where those results have been either confirmed or refuted. Section 4 then constructs several canonical theoretical models to explain key findings. Section 5 reviews and extends the empirical evidence on patterns of cross-country growth.

Section 6 considers extensions to the basic model of cross-country income distribution dynamics. By merging that model with the dynamics of within-country incomes, we obtain an integrative framework for, simultaneously, growth and inequality in particular, or growth and distribution more generally. This places growth and inequality on an equal footing and asks what the empirical evidence says on how they matter economically (subsection 6.2). Subsection 6.7 extends income distribution dynamics to spatial distribution dynamics. Finally, section 7 summarizes in one place the key lessons from this paper. The Technical Appendix, section 9, collects together additional mathematical discussion for some of the results in the paper.

While most of the technical results used here will be known to specialists, as far as I can tell the general reader or applied researcher will not have had access to them in a single convenient and easily accessible account. I hope that additional value derives from an approriate level of rigor and the distinctive technical approach here. For instance, every student of stochastic processes will have read accounts of Markov theory that begin "Consider a collection of random variables, either X_t, $t = 0, 1, 2, \ldots$ or $X(t)$, $t \in [0, \infty)$, defined on a probability space $(\Omega, \mathfrak{F}, \Pr)$". When working with income distributions, however, the researcher doesn't see just a realized sequence of random variables, as would happen say when a researcher studies asset prices, interest rates, GDP, or unemployment. Instead, the variable that the researcher naturally and directly deals with—empirically, analytically, and conceptually—is a distribution function. Most useful to the applied researcher then is an exposition of stochastic processes and dynamics that take as primitive the distributions themselves. This approach, explicit in all the technical sections that follow and implicit in the discussion otherwise, distinguishes the approach here from, say, purely mathematical accounts of Markov processes.[2] While obviously one development can be derived from the other, using a technical exposition adapted to one's specific research problem is both more pleasant and more efficient.

For instance, in many developments of stochastic process theory the infinitesimal generator (section 3.4 and Definition 3.15) is taken as the hypothesized collection of parameters in a stochastic differential equation. The properties of the resulting process are then derived from the infinitesimal generator. An applied researcher, however, sees first of all data and then, with a little work, the transition density or transition probability. It is much more natural and intuitive for that researcher to see the infinitesimal generator derived from a transition density than the other way round. Other examples abound.

> *So go ahead and do it; give those other examples. The Chapman-Kolmogorov property: Always hold for transition probability matrices and transition densities estimated from data? The applied researcher wants to un-*

[2] This difference, however, is easily over-emphasized. The most abstract and rigorous expositions of the mathematical theory make plain that one can approach the study of Markov processes from either perspective: see, e.g., Chung (1967, Section 2, Ch. 1) or Doob (1953, Ex. 3, p. 86).

derstand why. Analytical expositions, on the other hand, begin by hypothesizing it, and then go on to develop implications. Markov chain: Analytical expositions begin taking the set of discrete states as a priori given. On the other hand, the applied researcher wants to know what implications follow from particular discretizations, given a set of continuously-distributed data. Regularity conditions for estimating distributions or stochastic kernels: Practitioner wants to know a convenient set of generally applicable conditions. Analytical researcher is often interested in assumptions that can be shown theoretically to be yet weaker than previous ones used, but none of which is in practice verifiable. Mix estimation and analysis: Practitioner wants to take a given observed dataset, estimate underlying objects, and then work out implications. Excellent presentations of estimation end there—statistics and econometrics; excellent presentations of stochastic process theory don't discuss estimation. This is as it should, but the applied researcher needs to go look at and understand too many different conventions and notations; some of the developments which will never be useful to one group of researchers but of intense interest to another. Summarize: A specialist will find many things familiar in this manuscript, only sometimes the results and the development seem to be written backwards compared to what is typically found elsewhere. But the hope is, applied researchers will find this approach convenient for their work, and a relatively easy way to introduce themselves to hard, abstract ideas in stochastic process theory.

2 World growth and inequality

This section describes some key lessons from extant literature on economic growth and world income inequality. It provides our first look into the data on global income distributions and summarizes the extant literature.

Try to summarize extant results in an integrated coherent way, presenting the basic facts that are already known. Go crazy a little and give alternative visualizations. Bourguignon and Morrisson (2002) Chen and Ravallion (2004a) Deaton (2005) Firebaugh (2003) Jones (1997) Maddison (1989) Milanovic (2002) Schultz (1998) Pritchett (1997) Sala-i-Martin (2002a,b) World Bank (2000) World Bank (2002) UNU (2004)

---

3 Distribution dynamics and laws of motion

Why study the empirics of evolving patterns of economic growth—changing incomes and welfare—across countries as a model of the dynamics of distributions? Why not just calculate autoregressions or cross-correlations or model the dynamics of the panel through some regression?

The answer to this rests in whether models of distribution dynamics incite fresh economic thinking or raise new questions that more conventional approaches leave unidentified. Distributional changes, by definition, are not readily obvious when one studies only the variances or other moments of a cross-sectional distribution: Is the spread increasing between the top and bottom 10% of per capita incomes across countries but *decreasing* between the top and bottom 40%? The variance might at the same time rise or fall or be invariant, but is certainly uninformative of such stretching and compressing within the distribution.

Thinking further along those lines outlines the obvious empirical descriptions obtainable using a distribution-dynamics approach. Deforming the distribution through time can easily leave its first few moments invariant.

But what economic questions might attach to such distributional characterizations? As one important example, illustrations like that just given suggest particular kinds of nonlinearities or heterogeneities, where economies diverge from each other when they are sufficiently different but at income levels close to each other's (and at around the world average) they become progressively more similar. No dynamic correlations calculated without special foreknowledge would easily detect such dynamics, whereas the timepath of estimated distributions makes those features immediately conspicuous. A researcher can then design a more directed, parametrized econometric analysis to study further such heterogeneities.

Second, economists have long studied income distributions across people as an object of independent interest, with or without its possible connection to economic growth; see, e.g., the discussions surrounding the Kuznets curve in the Presidential Address by Atkinson (1997). Empirical work on economic growth that models cross-country per capita incomes as an evolving distribution can then potentially provide useful cross-fertilization with models of *personal* income distribution dynamics. For instance, Quah (1996a) studied cross-country income distribution dynamics using the structure that Galor and Zeira (1993) had previously developed for personal income distributions, reinterpreting the market for human capital loans as one for cross-country capital flows, and obtaining similar polarization and inequality implications. Conversely, the twin-peaked feature in the cross-country per capita income distribution (Quah, 1993a, 1997) has, in turn, been investigated for personal incomes (Zhu, 2003).

Yet a third benefit from a distribution-dynamics approach to economic growth is that it recasts discussion of inequality and growth in an integrative framework (Bourguignon, 2003; Quah, 2003; Sala-i-Martin, 2002a). This will be discussed at length in Section 6 but, simply put, the idea is that beyond seeking to determine if inequality across people is causal for aggregate economic growth or vice versa, a researcher can instead study whether inequality and macroeconomic growth co-evolve in particular ways.

Distribution dynamics considers not just the timepath of income distributions—each income distribution treated as a point-in-time snapshot—but also a law of motion or a mechanism for how the distribution at one timepoint evolves into that at a later time.

Let $x_j(t)$ denote per capita income in economy j at time t, and let $\mathsf{F}_{X,t}$ denote its cross-section income distribution at time t. To study the dynamics $\{\,\mathsf{F}_{X,t} : t \geq 0\,\}$, a researcher might begin by tracking, say, the mean, standard deviation, possibly higher-order moments, and other characteristics such as inter-quartile spread, of $\mathsf{F}_{X,t}$ as this cross-section distribution evolves through time. But this conceals the intra-distribution dynamics, the movement of economies from one part of the distribution to another. For such information, we turn to the stochastic kernel (section 3.2). If $\mathbb{F}$ denotes the collection of (cross-section) distributions and $\mathcal{T}^*_{t,s}$, for $t, s \geq 0$, are operators mapping the space $\mathbb{F}$ to itself, distribution dynamics considers

$$\mathsf{F}_{X,t+s} = \mathcal{T}^*_{t,s}\mathsf{F}_{X,t},$$

a law of motion for $\{\,\mathsf{F}_{X,t} : t \geq 0\,\}$. Sections 3.2–3.5 will treat this in greater detail but, for concreteness, an explicit example of this transition law, using the stochastic kernel $\mathcal{M}$, is

$$\mathsf{F}_{X,t+s}(x^\dagger) = \int_{-\infty}^{\infty} \mathcal{M}_{t,s}(\,x, (-\infty, x^\dagger]\,)\, d\mathsf{F}_{X,t}(x),$$

or, when $\mathsf{F}_{X,t}$ has density $\mathsf{f}_{X,t}$,

$$\mathsf{f}_{X,t+s}(x^\dagger) = \int_{-\infty}^{\infty} p_{t,s}(\,x, x^\dagger\,) \times \mathsf{f}_{X,t}(x)\, dx,$$

where $p_{t,s}$ is the transition density from time t to time $t+s$. The stochastic kernel $\mathcal{M}$ is a representation of the operator $\mathcal{T}^*$. Often the researcher needs to consider stochastically-perturbed versions of these transitions. How best to do so remains an open question, and researchers often leave only implicit how they consider the disturbances in such equations.

These equations are often used to describe the stochastic dynamics of an observed scalar random process $X(t)$ that in turn has the hypothesized but unobserved underlying distribution $\mathsf{F}_{X,t}$. Here, conversely, it is $\mathsf{F}_{X,t}$ that is measured directly—the cross-country income distribution—while the associated underlying $X(t)$ is only hypothesized. The two perspectives are, of course, mathematically equivalent (Doob, 1953, p. 255).

In distribution dynamics the limit as $s \to \infty$ of $\mathsf{f}_{X,t+s}$ in such laws of motion holds special significance. It is (the density of) what the cross-country income distribution tends towards, should the system continue along its historical path. Yet other characteristics—e.g., first-passage times across different parts of the distribution; evolving shapes in $\mathsf{f}_{X,t}$; the changing spreads—can also be used to get further insight on the evolution of these cross-country income distributions.

The remainder of this section describes statistical economic models for these laws of motion. Section 3.1 is concerned with the estimation problem, going from observed data through distributions to a stochastic kernel. Section 3.6 describes the analysis of long-run behavior and other characteristics arising from these distribution dynamics. Section 4 describes economic models that all have a catch-up or convergence feature. Whereas the economic ideas differ considerably across these models, all of them generate implications for distribution dynamics. Put technically, all these models restrict $\mathcal{M}$, and thus allow the latter's economic interpretation.

3.1 Statistics

The discussion here cannot pretend to be a rigorous exposition of all the underlying ideas; the associated statistical literature is vast.[3] But the presentation will attempt to convey enough key results that an applied empirical researcher can confidently use these tools. Because many published descriptions elsewhere of these results draw motivations from statistical or econometric theory, they typically seek the weakest possible assumptions. What applied researchers might find most interesting turns out to be only a special case of much deeper results. This can make what is useful to an applied researcher seem obscure or difficult to apply. Here, our concern instead is to develop the key results using relatively familiar assumptions.

Moreover, rather than cover a vast range of different ways to estimate densities, this article concentrates on just one class of such estimation techniques (the kernel estimator) but gives enough detail on it that the researcher can expertly apply the technique. No one class of techniques will necessarily uniformly dominate another class, and so in practice the researcher might well need to seek specialized expertise for a particular problem. However, thoroughly understanding one class of techniques means that the researcher always knows a way to get reasonable empirical results (the further research problem then might be to sharpen those findings).

Index economies by the integer $j \in \{1, 2, \ldots, J\}$. Suppose time t is continuous, with $t \in [0, \infty)$. Let $y_j(t)$ denote per capita income in economy j at time t and $N_j(t)$ be the corresponding population. World per capita income $\overline{y}(t)$ is

$$\overline{y}(t) \stackrel{\text{def}}{=} \left(\sum_{j=1}^{J} N_j(t) \right)^{-1} \times \sum_{j=1}^{J} y_j(t) N_j(t).$$

[3] A list of overviews alone would already include Devroye (1987); Devroye and Györfi (1985); Hardle and Linton (1994); Pagan and Ullah (1999); Prakasa Rao (1983); Silverman (1986).

Denote by $y_{\max}(t)$ per capita income in the leading economy at time t, i.e.,

$$y_{\max}(t) = \max_j \ y_j(t).$$

The identity of that leading economy can vary over time, as a once-lagging economy leapfrogs over yet others to get to the frontier. Such leapfrogging behavior appears naturally in some models of technological change (e.g., Aghion and Howitt, 1998).

For analyzing the cross-section distribution, it will often be useful to take as the primitive data $x_j(t)$, an appropriately normalized version of per capita incomes. For instance, we might normalize per capita income relative to the leading economy $y_j(t)$, i.e.,

$$x_j(t) \stackrel{\text{def}}{=} y_j(t) \times y_{\max}(t)^{-1} \in (0, 1].$$

Alternatively, we might take normalization relative to the world average, i.e.,

$$x_j(t) \stackrel{\text{def}}{=} y_j(t) \times \overline{y}(t)^{-1} \in (0, \infty).$$

Or, a researcher might be interested instead in $x_j(t)$ that is total factor productivity (TFP) in economy j at time t. Taking as given that TFP can be reliably estimated somehow, this would be appropriate when theories of TFP's evolution imply testable restrictions on its cross-section distribution (e.g., Aghion and Howitt (1998, Ch. 3) or Section 4 to follow). Moreover, certain studies (Easterly and Levine, 2001; Feyrer, 2001) have suggested TFP is the critical driver for the cross-country per capita income distribution, so that studying TFP's distribution dynamics then also carries independent interest.

Finally, we might also consider logarithmic or other mathematical transformations of the different possibilities for $x_j(t)$ just described. Measurement of the variable $x_j(t)$ need not, of course, be straightforward—depending on the intent of the study, for instance, measuring TFP can, in the literature, be involved and controversial. But so too is assessing the relevant purchasing power parity correction for per capita incomes across countries. Such concerns matter importantly but are more appropriate for studies with goals other than those here.

3.1.1 Distributions

Here I will assume that $x_j(t)$ are data given to the researcher. There can be no one $x_j(t)$ always appropriate to use independent of the current question of interest. But as long as a researcher keeps in mind what $x_j(t)$ is and interprets the findings accordingly, no confusion need arise.

Definition 3.1 *For fixed t, given data $\{x_j(t),\ j = 1, 2, \ldots, J\}$, the* **empirical distribution function at** t *is the mapping* $\widehat{F}_X : \mathbb{R}\times[0,\infty) \to [0,1]$ *defined by*

$$\widehat{F}_X(x,t) \stackrel{def}{=} \#\{j \ni x_j(t) \leq x\} \times J^{-1}, \qquad x \in (-\infty, \infty). \tag{3.1}$$

In words the empirical distribution function takes value at x that is the fraction of observations in the cross section bounded from above by x. With probability one the functional statistic $\widehat{F}_X$ is bounded, increasing, and right-continuous with left limits, and satisfies $\widehat{F}_X(-\infty) = 0$. (If a statement holds for all t or if the time t argument can be otherwise left implicit without ambiguity, then that argument will be omitted.) Wherever the derivative $\widehat{F}'_X = \partial\widehat{F}_X(x,t)/\partial x$ exists, it equals zero, again with probability one.[4]

To analyze the behavior of $\widehat{F}_X$, denote the indicator function

$$\delta(x) = \begin{cases} 1 & \text{if } 0 \leq x; \\ 0 & \text{otherwise.} \end{cases}$$

[4] Conventions in these and related definitions used later sometimes differ across authors. Here, we follow as far as possible Billingsley (1968, Ch. 3) and Chung (1974, Section 5.5).

Definition 3.1 then gives:

$$\widehat{F}_X(x) = J^{-1} \sum_{j=1}^{J} \delta(x - x_j), \qquad (3.2)$$

i.e., for each fixed x the value $\widehat{F}_X(x)$ is an average of $\{0, 1\}$-valued random variables. Obviously $\delta(x - x_j)$ is bounded. If $\{x_j\}$ are random variables iid with the common distribution F_X then $\{\delta(x - x_j)\}$ are also iid sharing an appropriate induced distribution. A standard strong law of large numbers (e.g., Kolmogorov's SLLN in Rao, 1973, p. 115) then gives pointwise almost sure convergence, i.e.,

$$\forall x: \quad \left| \widehat{F}_X(x) - \mathsf{F}_X(x) \right| \xrightarrow{\text{a.s.}} 0 \qquad \text{as } J \to \infty. \qquad (3.3)$$

The same properties also give a central limit theorem at each fixed x, i.e., for $\mathsf{F}_X(x) \notin \{0, 1\}$, an appropriately standardized separation of $\widehat{F}_X(x)$ from $\mathsf{F}_X(x)$ converges in distribution to the standard normal (e.g., Lindeberg-Lévy Theorem in Rao, 1973, p. 127):

$$\left[\frac{J}{[1 - \mathsf{F}_X(x)] \cdot \mathsf{F}_X(x)} \right]^{1/2} \times \left[\widehat{F}_X(x) - \mathsf{F}_X(x) \right] \xrightarrow{d} \mathsf{N}(0, 1) \qquad \text{as } J \to \infty. \qquad (3.4)$$

However, since the object $\widehat{F}_X$ is an entire function, we are interested not just in its behavior at distinct points but in its global properties as a function, i.e., in features such as the multi-modality, dispersion, shape, skewness, and the maxima and minima of the implied density. Not all such characterizations have yet been used explicitly in the growth empirics literature but the crucial ingredients for their study reside in *uniform* analogs of the law of large numbers and central limit theorem, (3.3) and (3.4), respectively.[5] These counterparts

[5] This somewhat exaggerates the importance of the uniform or L_∞ results to follow, subtle though those might already be. For instance, to assess shapes, ideally one wants information on all the derivatives

are, respectively, the Glivenko-Cantelli Theorem (e.g., Chung, 1974, p. 133) and a functional central limit theorem giving convergence in distribution to Brownian Bridge (e.g., Billingsley, 1968, Ch. 3).

The Glivenko-Cantelli Theorem or Uniform Law of Large Numbers states that if $\{x_j\}$ are iid with the common distribution F_X then

$$\sup_x \left| \widehat{F}_X(x) - \mathsf{F}_X(x) \right| \xrightarrow{\text{a.s.}} 0 \qquad \text{as } J \to \infty. \tag{3.5}$$

No additional assumptions have been imposed; the same conditions that previously delivered the relatively ordinary pointwise almost sure convergence also provide uniform almost sure convergence of the empirical distribution function. Conclusion (3.5) implies the earlier (3.3) but the converse is, in general, false.[6]

For the uniform analog to the central limit theorem in this application, recall that a continuous random function B° on $[0,1]$ that is Gaussian and has $EB^\circ(r) = 0$ and $\mathrm{Cov}(B^\circ(r), B^\circ(r')) = \min(r, r') -$

---

of the estimated function, not just the maximum separation over the function's domain, i.e., one seeks characterizations in Sobolev space rather than just in L_∞ using sup norm.

[6] Since the difference between pointwise and uniform convergence is unrelated to the probabilistic behavior in (3.3) and (3.5), the distinction can be usefully illustrated with the following deterministic example. For $J = 1, 2, \ldots$, let f_J and f be functions mapping the interval $[0,1]$ to itself, with $f_J(x) = x^J$, and $f(x) = 0$ for $x \in [0,1)$ and $f(1) = 1$. Then the sequence $\{f_J\}$ converges pointwise to f as $J \to \infty$ but $\sup_x |f_J(x) - f(x)| = 1$ for all J and thus $\{f_J\}$ fails to converge uniformly to f. Mathematical interest in uniform convergence lies in how properties such as continuity, differentiability, and Riemann integrability propagate to the limit function under uniform but not, in general, under pointwise convergence. An alternative description has it that under pointwise convergence the speed of convergence depends on where the functions are being evaluated but under uniform convergence that speed is invariant throughout the functions' domain.

rr' is known as a *Brownian Bridge*. Recall also that *standard Brownian Motion* B is a continuous random function on $[0, \infty)$ having independent increments and with $B(r) \sim \mathsf{N}(0, r)$. A Brownian Bridge can be constructed from standard Brownian Motion B by $B^\circ(r) \stackrel{\text{def}}{=} B(r) - rB(1)$, $r \in [0, 1]$ (e.g., Billingsley, 1968, Section 9). Just as, when properly standardized, $\widehat{F}_X(x)$ converges in distribution to the standard normal, so too the function $\widehat{F}_X$, appropriately normalized, converges in distribution to the Brownian Bridge. Alternative statements of this are possible (e.g., Billingsley, 1968, Sections 13 and 16, and in particular Theorems 13.1 and 16.4). Here, we seek only to convey a flavor of what is available without getting into an overly-detailed exposition.

The Lindeberg-Lévy Theorem (3.4) obtained its conclusion from standardizing $\widehat{F}_X(x)$ by first centering it on $\mathsf{F}_X(x)$ and then dividing the separation by $[1 - \mathsf{F}_X(x)]^{1/2} \times \mathsf{F}_X(x)^{1/2}$. For convergence to Brownian Bridge, begin instead by defining the transformed observations $z_j = \mathsf{F}_X(x_j) \in [0, 1]$. Construct the resulting empirical distribution function $\widehat{F}_Z$ for z_j, $j = 1, 2, \ldots, J$, as earlier done in equation (3.1) for $\widehat{F}_X$, but now using z_j instead. Next define the function $B_J : [0, 1] \to \mathbb{R}$ by

$$B_J(r) \stackrel{\text{def}}{=} J^{1/2} \left[\widehat{F}_Z(r) - r \right], \qquad r \in [0, 1]. \tag{3.6}$$

Then (Billingsley, 1968, Theorem 13.1 or Theorem 16.4)

$$B_J \xrightarrow{d} B^\circ \qquad \text{as } J \to \infty. \tag{3.7}$$

Obviously, the construction in (3.6) involves the unknown quantity F_X and so is not immediately useful. It is, however, comparable to the construction in the standard central limit theorem (3.4), as the latter also involves the unknown $\mathsf{F}_X(x)$, that has to be appropriately estimated through some other procedure.

Billingsley (1968, Section 22) discusses how the functional central limit theory in (3.7) can be extended to apply to dependent observations. This involves hypothesizing appropriate mixing and moment conditions on x_j. We do not discuss this further here but leave it for the interested reader to pursue.

3.1.2 Densities

The derivative of a distribution is the density function. But while the empirical distribution function $\widehat{F}_X$, as we have just seen, usefully estimates the underlying distribution, its derivative fails to do so for the corresponding density as that derivative whenever well-defined only takes value zero.[7]

Consider then a density estimator

$$\widehat{f}_{X,b}(x) \stackrel{\text{def}}{=} \frac{\widehat{F}_X(x+b) - \widehat{F}_X(x-b)}{2b}, \quad \text{for } b > 0. \tag{3.8}$$

Replacing $\widehat{F}_X$ on the right by F_X and letting $b \to 0$ would recover the underlying density $\mathsf{f}_X = \partial\mathsf{F}_X/\partial x$ whenever F_X is differentiable. The estimator (3.8) can thus be viewed as a difference ratio, taking a numerical two-sided approximate derivative, with the approximation indexed by the arc length h. Unlike the derivative $\partial\widehat{F}_X/\partial x$ itself the estimator $\widehat{f}_{X,b}$ in (3.8) is no longer trivially zero almost everywhere. To assess $\widehat{f}_{X,b}$'s properties and to motivate its extension to other useful estimators, define the rectangular weighting function

$$\mathsf{K}(x) = \begin{cases} \frac{1}{2} & \text{for } |x| \leq 1, \\ 0 & \text{otherwise;} \end{cases} \tag{3.9}$$

[7] This difficulty in estimating the density is profound and arises not just from looking at one particular estimator. For a large class of density functions, no reasonable estimator exists that is unbiased everywhere in the true density's domain (Prakasa Rao, 1983, Theorem 1.2.2 and Ch. 2.1). By contrast, (3.2) immediately implies that, besides being uniformly consistent, the empirical distribution function $\widehat{F}_X$ is also unbiased pointwise for the underlying limit.

and recognize that (3.8) can be rewritten

$$\widehat{f}_{X,b}(x) = \frac{1}{Jb}\sum_{j=1}^{J} \mathsf{K}\left(\frac{x - x_j}{b}\right). \tag{3.10}$$

Equation (3.8) says the estimator $\widehat{f}_{X,b}$ at x is the fraction normalized by $1/2b$ of observations x_j that fall within the interval $[-b, b]$ centered on point x. By contrast, equation (3.10) says $\widehat{f}_{X,b}$ at x is the sum normalized by Jb of $\left\{0, \frac{1}{2}\right\}$-values attached to each observation, varying with how distant that observation is from point x. This second expression (3.10) also shows that, as a function, $\widehat{f}_{X,b}$ is the sum of rectangles, and so now *its* derivative whenever defined is only ever 0, the same property as $\widehat{F}'_X$. Note, however, that this feature is inherited directly from the rectangular weighting function (3.9) and thus can be easily altered.

The weighting function (3.9) and associated density estimator (3.10) are usefully generalized as follows (Parzen, 1962; Rosenblatt, 1956).

Definition 3.2 *A non-negative function* $\mathsf{K} : \mathbb{R} \to \mathbb{R}_+$ *with*

$$\int_{\mathbb{R}} \mathsf{K}(x)\, dx = 1$$

is a **scalar kernel function** *or just a* **kernel**. *A kernel is* **symmetric** *if* $\mathsf{K}(-x) = \mathsf{K}(x)$ *for all* $x \in \mathbb{R}$. *For* K *an arbitrary kernel and* $b > 0$, *the function*

$$\widehat{f}_{X,b}(x) \stackrel{def}{=} \frac{1}{Jb}\sum_{j=1}^{J} \mathsf{K}\left(\frac{x - x_j}{b}\right) \tag{3.11}$$

is a **kernel density estimator with kernel** K *and* **bandwidth** b.

Every probability density function (pdf) is a kernel and every kernel a pdf.[8] The standard normal pdf is a symmetric kernel, as is the rectangular weighting kernel (3.9). For this last, the bandwidth h

[8] Since then $\widehat{f}_{X,b}$ in (3.11) is automatically also a pdf, this seems a

Table 3.1 Typical kernels used in kernel density estimation, with I denoting the indicator function.

Kernel	$\mathsf{K}(x)$
Normal	$\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$
Rectangular	$\frac{1}{2} \times I_{\{\lvert x\rvert \leqslant 1\}}(x)$
Triangular	$(1 - \lvert x\rvert) \times I_{\{\lvert x\rvert \leqslant 1\}}(x)$
Epanechnikov	$\left[\frac{3}{4}\left(1 - \frac{1}{5}x^2\right) 5^{-1/2}\right] \times I_{\{\lvert x\rvert \leqslant \sqrt{5}\}}(x)$

can also naturally be interpreted as a *window width*. With other kernels, such terminology is less obviously apposite but is nevertheless sometimes used. The estimator (3.8), motivated initially as a two-sided numerical derivative, is thus seen to be a special kind of kernel density estimator, where the kernel is symmetric and uniform, and the bandwidth is the arc length in the numerical derivative approximation. A list of kernels typically used in kernel density estimation appears in Table 3.1.

Comparing (3.9)–(3.10) with the generalized version (3.11) in Definition 3.2 we see that the kernel density estimator is the average of selected pdf's placed on the datapoints x_j. This gives a further useful interpretation for the kernel density estimator. Recall that if K is the pdf for a random variable η then $b^{-1}\mathsf{K}\left(\frac{x-x_j}{b}\right)$ is the pdf for $x_j + b\eta_j$, i.e., that induced random variable scaled by b and re-located to x_j. These $x_j + b\eta_j$ are simply the original observations perturbed or smoothed by adding in scaled versions of iid auxiliary disturbances with known pdf's. The kernel density estimator is thus the average

useful and natural restriction. However, circumstances exist where a kernel that goes negative, but still integrating to 1, might be useful: one example is bias reduction—see, e.g., Prakasa Rao (1983, Theorem 2.1.5, p. 42) or Silverman (1986, 3.6). To keep to the principal intent of the discussion here, we will not consider such kernels.

pdf of the smoothed observations $x_j + b\eta_j$, with b sometimes then also called a *smoothing parameter*. The difference ratio in (3.8), because it can be rewritten using the rectangular weighting function (3.9), has its auxiliary disturbances distributed uniformly on $[-1, 1]$. More generally, if b is small then so too are the perturbations; conversely, if b is large then the perturbations dominate the data x_j in forming the kernel density estimator.

To see the significance of alternative selections of kernel K and bandwidth b for estimating f_X, it is useful first to look at consistency and central limit properties available. Just as with empirical distribution function, pointwise and uniform versions are available.[9] All such results have the bandwidth b diminish as J grows but at a controlled rate. To flag this, we write b_J when useful.

The next two results provide pointwise and uniform consistency respectively.

Theorem 3.3 *Suppose* $\mathsf{f}_X = \partial \mathsf{F}_X / \partial x$ *is continuous. In Definition 3.2 let* K *be such that (a)* K *is bounded and (b)* $|x\mathsf{K}(x)| \to 0$ *as* $|x| \to \infty$*; and let* $b_J \to 0$ *and* $Jb_J \to \infty$ *as* $J \to \infty$*. Then*

$$\text{for each } x: \qquad \widehat{f}_{X,b}(x) - \mathsf{f}_X(x) \xrightarrow{\text{Pr}} 0 \quad \text{as } J \to \infty.$$

Theorem 3.3 is established in Prakasa Rao (1983, Theorem 2.1.2, p. 37); see also Silverman (1986, Section 3.7.1, pp. 71–72) and Prakasa Rao (1983, Theorem 2.1.1, p. 35–36).

Theorem 3.4 *In Definition 3.2 let* K *be such that (a)* K *is bounded; (b)* K *has bounded variation; and (c)* K *is continuous outside a set of Lebesgue measure zero. Then the two conditions*

[9] Modes of convergence other than the uniform are also studied in the statistical literature—as examples see the exhaustive study of L^1 convergence in Devroye and Györfi (1985) or of integrated mean square convergence in Bickel and Rosenblatt (1973). To maintain focus and keep within space restrictions, we do not consider those here.

(i) $\mathsf{f}_X = \partial \mathsf{F}_X/\partial x$ *is uniformly continuous; and*

(ii) $b_J \to 0$ *and* $(\log J)^{-1} J b_J \to \infty$ *as* $J \to \infty$

jointly are necessary and sufficient for

$$\sup_x |\widehat{f}_{X,b}(x) - \mathsf{f}_X(x)| \xrightarrow{a.s.} 0 \quad \text{as } J \to \infty.$$

(The reader might find it useful to refer to the Technical Appendix, section 9, for further details.) Apart from regularity conditions, the critical difference between pointwise and uniform convergence expressed in Theorems 3.3 and 3.4 is the strengthening of control on bandwidth variation, from $Jb_J \to \infty$ in Theorem 3.3 to $(\log J)^{-1} Jb_J \to \infty$ in Theorem 3.4. In words b_J must still vanish but Theorem 3.4 restrains it from doing so as quickly as does Theorem 3.3.

The conditions placed on the kernel K in both Theorems 3.3 and 3.4 are very weak. Typical kernels, including all those in Table 3.1, readily satisfy these restrictions.

The version of a uniform law of large numbers given in Theorem 3.4 is discussed in Silverman (1986, Section 3.7.1, pp. 71–72). For useful variations see Pagan and Ullah (1999, Theorem 2.8, pp. 36–39) and Prakasa Rao (1983, Theorem 2.1.3, pp. 37–38, and 2.1.15, p. 54).

Both pointwise and uniform central limit theorems are available.

Theorem 3.5 *Assume:*

(i) K *is bounded;*

(ii) K *is symmetric about* 0*, i.e.,* $\mathsf{K}(x) = \mathsf{K}(-x)$ *for all* x*;*

(iii) $\int x^2 \mathsf{K}(x)\, dx < \infty$*;*

(iv) f_X *has second derivative continuous and bounded;*

(v) $Jb_J \to \infty$ *and* $J^{1/5} b_J \to 0$ *as* $J \to \infty$.

Then at each $x^\dagger$ *with* $\mathsf{f}_X(x^\dagger) \neq 0$ *we have*

$$\frac{(Jb_J)^{1/2}}{\left(\mathsf{f}_X(x^\dagger)\int \mathsf{K}(x)^2\,dx\right)^{1/2}}\left[\widehat{f}_{X,b}(x^\dagger) - \mathsf{f}_X(x^\dagger)\right] \xrightarrow{d} \mathsf{N}(0,1) \qquad \textit{as } J \to \infty.$$

Since this result is likely one of the more directly useful for an applied researcher, its proof is given in section 9.

Notice that the bandwidth condition in (v) of Theorem 3.5 has $b_J = o(J^{-1/5})$, i.e., the bandwidth has to go to zero *faster* than $J^{-1/5}$, contradicting the fastest mean square error convergence rule of $b_J = O(J^{-1/5})$ (see the Technical Appendix). Using the latter, the limit distribution is not centered at 0, thus invalidating statistical inference. When $b_J = o(J^{-1/5})$, however, Theorem 3.5 implies that with probability approximately $1 - \alpha$ the underlying $\mathsf{f}_X(x^\dagger)$ lies in the interval having endpoints

$$\widehat{f}_{X,b}(x^\dagger) \pm (Jb_J)^{-1/2}\left[\widehat{f}_{X,b}(x^\dagger)\int \mathsf{K}(x)^2\,dx\right]^{1/2} z_{\alpha/2}, \qquad (3.12)$$

when $z_{\alpha/2}$ is the standard normal ($(100 \times \alpha/2)$-percentile) critical value, with accuracy improving as J grows.

The endpoints given in (3.12) indicate that for given bandwidth b_J the estimator variance depends on the kernel through $\int \mathsf{K}(x)^2 dx$. The second column in Table 3.2 calculates this integral for the kernels of Table 3.1. But while this information is useful for performing inference off (3.12), the researcher will typically set b_J and K simultaneously. Thus, the third column in Table 3.2 reports a measure of relative efficiency, based on minimizing integrated mean square error, where the bandwidth b_J is permitted to vary with the kernel K (see Theorem 9.6 in the Technical Appendix and Silverman (1986, Table 3.1, p. 43)). The Epanechnikov kernel turns out to be optimal but the other kernels achieve efficiencies remarkably close to it.[10]

[10] This brief discussion covers key ideas but, necessarily, cannot

Table 3.2 Variance factors and relative efficiencies for typical kernels used in kernel density estimation; see equation (3.12) following Theorem 3.5 and equation (9.19) following Theorem 9.6.

Kernel	$\int \mathsf{K}(x)^2\,dx$	J_{Ep}/J
Normal	$\frac{1}{2\sqrt{\pi}} \approx 0.282$	0.95
Rectangular	$\frac{1}{2} = 0.5$	0.93
Triangular	$\frac{2}{3} \approx 0.667$	0.99
Epanechnikov	$\frac{3}{5\sqrt{5}} \approx 0.268$	

Theorem 3.6 *Let [...].*

This result is established in Bickel and Rosenblatt (1973, p. 1072).

[...]

Many other methods are available to estimate densities, among them nearest-neighbor, orthogonal series, maximum penalized likelihood, and adaptive kernel density estimators (see, e.g., Devroye, 1987; Pagan and Ullah, 1999; Silverman, 1986). While each method has specific advantages and disadvantages, this section, to conserve space, has discussed only kernel density estimators.

Section 3.3 will again take up estimation, but we need first to develop some analytical concepts preliminary to that discussion. These concepts will be used also in Section 6, although then for analytical modelling rather than econometric estimation.

---

do justice to an extremely large literature on optimal bandwidth and kernel choice. The interested reader should refer to Pagan and Ullah (1999, Ch. 2), Silverman (1986, Ch. 3), and the references given there.

3.2 Stochastic kernels

The discussion thus far has provided, through equation (3.1), snapshots in time of the evolving cross-section distributions. Results (3.5) and (3.7) allow viewing what is calculated in (3.1) as reasonable descriptions of some underlying distribution F_X, while Theorems 3.3–3.6 achieve the same for $\widehat{f}_{X,b}$ as an estimator for the underlying density f_X.

To study dynamics, we need to consider explicit laws of motion in the cross-section distributions, not just snapshots in time. To that end, this article now draws on Markov process theory, as developed in, e.g., Gihman and Skorohod (1975), Karlin and Taylor (1981), or Rogers and Williams (1994). In this discussion the variable subscript (X or Z) is usefully left implicit.

Consider distributions $\{\mathsf{F}_t : t \geq 0\}$ evolving in continuous time, with each F_t defined on the real line $\mathbb{R}$. Let $\mathcal{T}^*_{t,s}$ be operators mapping $\mathbb{F}$ the space of such distributions into itself, and consider distribution dynamics in the transition equation:

$$\mathsf{F}_{t+s} = \mathcal{T}^*_{t,s}\mathsf{F}_t. \tag{3.13}$$

In stochastic process theory $\mathcal{T}^*_{t,s}$ typically operates on the underlying probability measures rather than on distributions as in (3.13); further, the operator would need to be defined on a space of bounded finitely-additive set functions rather than just probability measures.[11] However, the distinction is unimportant for the level of abstraction used here. With measures as operands, $\mathcal{T}^*$ also turns out to be the adjoint of operator $\mathcal{T}$ to be defined in (3.17), hence the * notation.

What does the $\mathcal{T}^*$ operation (3.13) mean? Denote by $\mathcal{R}$ the Borel-measurable subsets of the real numbers $\mathbb{R}$, and let $\mathcal{M}_{t,s}$ be mappings

[11] Durlauf and Quah (1999) and Quah (1997) discuss these technical points for studying the empirics of economic growth and convergence. Stokey and Lucas (1989, Ch.8) provides a rigorous development of the mathematics.

from $\mathbb{R} \times \mathcal{R}$ to $[0,1]$, where for each x we have $\mathcal{M}_{t,s}(x,\cdot)$ a probability measure. Associate with a given $\mathcal{T}^*_{t,s}$ some such $\mathcal{M}_{t,s}$ by

$$\forall\ x' : \qquad \mathsf{F}_{t+s}(x') = \int_{-\infty}^{\infty} \mathcal{M}_{t,s}\left(x, (-\infty, x']\right)\, d\mathsf{F}_t(x). \tag{3.14}$$

Equation (3.14) is the *stochastic kernel* representation of equation (3.13).

Whenever F_t admits a density, equation (3.14) becomes

$$\mathsf{f}_{t+s}(x') = \int_{-\infty}^{\infty} p_{t,s}(\,x, x'\,) \times \mathsf{f}_t(x)\, dx \tag{3.15}$$

for *transition density* $p_{t,s}$ satisfying

$$\forall x, x' \text{ in } \mathbb{R} : \quad \mathcal{M}_{t,s}(x, (-\infty, x']) = \int_{-\infty}^{x'} p_{t,s}(x, \zeta)\, d\zeta. \tag{3.16}$$

In words the transition density $p_{t,s}$ is the Radon-Nikodym derivative of the stochastic kernel $\mathcal{M}_{t,s}$ with respect to Lebesgue measure, just as the density f_t is the Radon-Nikodym derivative of the distribution F_t again with respect to Lebesgue measure. By construction then for each x we have $p_{t,s}(x,\cdot)$ a probability density function, i.e., $p_{t,s}(x, x') \geq 0$ for all x' and $\int_{-\infty}^{\infty} p_{t,s}(x, x')\, dx' = 1$.

Next define the operator $\mathcal{T}_{t,s}$ mapping bounded measurable functions to bounded measurable functions:

$$\left(\mathcal{T}_{t,s}\phi\right)(x) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} \phi(x')\mathcal{M}_{t,s}(x, dx'). \tag{3.17}$$

Call $\mathcal{T}_{t,s}$ the time t to time $t+s$ *transition operator*. If the transition density exists, equation (3.17) is

$$\left(\mathcal{T}_{t,s}\phi\right)(x) = \int_{-\infty}^{\infty} \phi(x') p_{t,s}(x, x')\, dx'.$$

To provide one interpretation to (3.17) let $\{\, X(t) : t \in \mathbb{R}_+ \,\}$ be a stochastic process whose marginal distributions for each $X(t)$ are given by F_t. Assume equation (3.13) but leave X's serial dependence properties otherwise unrestricted, i.e., assume that the timepath of F_t can be described by $\mathcal{T}^*_{t,s}$. Then (3.17) is the expectation of ϕ at time $t+s$ conditional on $X(t) = x$, i.e.,

$$(\mathcal{T}_{t,s}\phi)\,(x) = E\,[\,\phi(X(t+s)\,|\,X(t) = x\,]\,.$$

To emphasize, nothing has been assumed thus far on whether X is Markov or otherwise; there are no assumptions or restrictions to test. Instead, X is an artificial random variable that has simply been constructed in the course of our discussion; it need correspond to nothing observable in real data.

The stochastic kernel representation (3.14) and the transition operator definition (3.17) help clarify the notation and terminology; understanding the relation between them also aids intuition for the discussion to follow. Suppose $\mathcal{T}$ is an operator that maps a class of appropriate functions to itself—an example of such an operator would be $\mathcal{T}_{t,s}$. If $\mathcal{S}$ is an operator mapping distributions to distributions such that for ϕ an appropriate function and F a distribution, we have:

$$\int (\,\mathcal{T}\phi\,)(x)\;\mathsf{F}(dx) = \int \phi(x')\,(\,\mathcal{S}\mathsf{F}\,)(dx'), \quad \forall\ \phi \text{ and } \mathsf{F}, \tag{3.18}$$

then call $\mathcal{S}$ the *adjoint* to $\mathcal{T}$ and write $\mathcal{S} = \mathcal{T}^*$. Applying this to $\mathcal{S} = \mathcal{T}^*_{t,s}$ and $\mathsf{F} = \mathsf{F}_t$, calculate the right side of equation (3.18) as:

$$\begin{aligned}
\int \phi(x')\,(\,\mathcal{T}^*_{t,s}\mathsf{F}_t\,)(dx') &= \int \phi(x')\;\mathsf{F}_{t+s}(dx') && \text{(by (3.13))}\\
&= \int \phi(x')\left[\int \mathcal{M}_{t,s}(x,dx')\;\mathsf{F}_t(dx)\right] && \\
& && \text{(by (3.14))}\\
&= \int \left[\int \phi(x')\mathcal{M}_{t,s}(x,dx')\right]\;\mathsf{F}_t(dx) && \\
&= \int (\,\mathcal{T}_{t,s}\phi\,)(x)\;\mathsf{F}_t(dx). && \text{(by (3.17))}
\end{aligned}$$

Thus, $\mathcal{T}^*_{t,s}$ is indeed the adjoint to $\mathcal{T}_{t,s}$, and the notation is apt.

For analyzing observed income or consumption distribution dynamics, a researcher might well confine attention to $\mathcal{T}^*$—that alone would suffice to trace out the dynamics of the income distribution—and never need to consider explicitly the transition operator $\mathcal{T}$. However, economists will typically be interested not just in income or consumption alone, but in utility or poverty or yet other functionals of the income distribution. When applied to an appropriate function, the transition operator $\mathcal{T}$ allows conveniently characterizing those additional derived dynamics: an important example of this appears in Section 6, which uses the resolvent operator (Section 3.5) to describe economic welfare and the dynamics of poverty and income inequality.

The alternative descriptions ($\mathcal{T}^*$, $\mathcal{M}$, $\mathcal{T}$) are well-defined regardless of the exact dynamics and cross-correlation properties in $x_j(t)$, the same way that the projection of a scalar $X(t)$ on $X(t-1)$ is meaningful regardless of X's exact serial correlation properties (see Example 3.7). Equation (3.13) says only that operator $\mathcal{T}^*$ transforms one distribution into another; operator $\mathcal{T}^*$ does this whether or not those distributions come from a process that is generally serially dependent or Markov or, indeed, serially independent altogether. This same reasoning had been used previously, following (3.17), to place a conditional-expectations interpretation on $\mathcal{T}$.

Now to permit more precise statements, consider not just the timepath $\{\mathsf{F}_t : t \geq 0\}$ but also the associated conditional distributions. Recall that a stochastic process $\{X(t) : t \geq 0\}$ is said to be *Markov* when for all integer $n \geq 2$ and all $t_1, \ldots, t_n, t$ with $0 \leq t_1 < \cdots < t_n < t$, the conditional distributions satisfy

$$\begin{aligned} &\mathsf{F}\,(\,X(t) \leq x \mid X(t_1) = x_1, \ldots, X(t_n) = x_n\,) \\ &\qquad\qquad = \mathsf{F}\,(\,X(t) \leq x \mid X(t_n) = x_n\,) \qquad \forall x_1, \ldots, x_n, x. \end{aligned}$$

Whenever X is Markov and for all $t \geq 0$ has F_t as the distribution of $X(t)$, then say also that $\{\,\mathsf{F}_t : t \geq 0\,\}$ and the associated $\mathcal{T}^*$, $\mathcal{M}$, and $\mathcal{T}$ are Markov.

When $\mathcal{T}^*$ (or $\mathcal{M}$, p, or $\mathcal{T}$) allows its doubled time-subscript to collapse to just a single subscript denoting separation in time,

$$\mathcal{T}^*_{t,s} = \mathcal{T}^*_{0,s} = \mathcal{T}^*_s,$$

say that the process $\{\mathsf{F}_t : t \geq 0\}$ has *time-homogeneous transitions* or is *time-homogeneous*.

If $\{\, \mathsf{F}_t : t \geq 0 \,\}$ is Markov and has time-homogeneous transitions then $\{\, \mathcal{T}_t : t \geq 0 \,\}$ constitutes a semigroup of operators where

$$\mathcal{T}_{t+s} = \mathcal{T}_t \mathcal{T}_s = \mathcal{T}_s \mathcal{T}_t, \tag{3.19}$$

as for all bounded measurable ϕ

$$\begin{aligned}
(\mathcal{T}_{t+s}\phi)(x) &= E\,[\,\phi(X(t+s)) \mid X(0) = x\,] \\
&= E\,[\,E[\phi(X(t+s)) \mid X(t), X(0)] \mid X(0) = x\,] \\
&= E\,[\,E[\phi(X(t+s)) \mid X(t)] \mid X(0) = x\,] \\
&= E\,[\,(\mathcal{T}_s\phi)(X(t)) \mid X(0) = x\,] \\
&= (\,\mathcal{T}_t(\mathcal{T}_s\phi)\,)\,(x),
\end{aligned}$$

where the second equation uses the law of iterated expectations, the second the Markov property, and the third time-homogeneity.[12] It is here that the Markov assumption has become critical; the earlier manipulations of $(\mathcal{T}^*, \mathcal{M}, \mathcal{T})$ are all valid and well-defined even outside a Markov environment.[13]

Finally, Example 3.7 constructs an example of the distribution sequence $\{\, \mathsf{F}_t : t \,\}$ and the stochastic kernels for a discrete-time covariance stationary process. The example shows that when not Markov the transition operators, even if time-homogeneous, in general violate the semigroup property.

[12] The semigroup property is also known as the Chapman-Kolmogorov equation or the Fokker-Planck equation (Karlin and Taylor, 1981, pp. 285–286).

[13] Cox and Miller (1965, Ex. 3, p. 142) gives an example where the transition operator satisfies the semigroup property while the underlying stochastic process is not Markov.

Example 3.7 *Let*

$$\{X(t): \ t = 0, 1, 2, \ldots\}$$

be a sequence of random variables jointly normally distributed, and consider the associated distribution dynamics $\{F_t, \ t = 0, 1, 2, \ldots\}$. *Suppose the sequence* X *is covariance stationary with mean zero and covariogram:*

$$\{g(m) = E\left[X(m)X(0)\right], \ m = 0, \pm 1, \pm 2, \ldots\}$$

The projection of $X(t)$ *on* $X(t-1)$ *has residual*

$$\epsilon(t) \stackrel{def}{=} X(t) - \left[g(1)g(0)^{-1}\right] X(t-1)$$

normally distributed, mean zero, independent of $X(t-1)$, *and with variance*

$$\left[1 - g(1)^2 g(0)^{-2}\right] g(0)^2 > 0 \qquad \text{by Cauchy-Schwartz inequality.}$$

Because X *is normally distributed, the projection is also the conditional expectation, so that the distribution of* $X(t)$ *conditional on* $X(t-1)$ *is normal and can be written:*

$$X(t)|X(t-1) \sim \mathsf{N}(0, \left[1 - g(1)^2 g(0)^{-2}\right] g(0)^2).$$

Consequently, the stochastic difference equation

$$X(t) = \beta X(t-1) + \epsilon(t), \tag{3.20}$$

where $\beta = g(1)g(0)^{-1}$ *and* $\epsilon(t) \sim \mathsf{N}(0, [1-\beta^2]g(0))$ *independent of* $X(t-1) \sim \mathsf{N}(0, g(0))$, *is a valid representation of the distribution dynamics in* F_t, *regardless of the rest of the covariogram* $g(m)$. *Implicit in* (3.20) *is the operator* $\mathcal{T}^*_{t-1,1} = \mathcal{T}^*_1$ *of the distributional transition equation* (3.13). *But although the resulting* $\mathcal{T}^*$ *and thus* $\mathcal{M}$ *and* $\mathcal{T}$ *are time-homogeneous (by the stationarity of the sequence* X*) the semigroup property* (3.19) *fails in general. For instance, suppose* X *is a second-order autoregression,*

$$X(t) = \alpha_1 X(t-1) + \alpha_2 X(t-2) + \nu(t), \quad \nu(t) \sim iid\ \mathsf{N}(0, 1).$$

The still-valid conditional distribution equation (3.20) *has*

$$\beta = g(1)g(0)^{-1} = (1-\alpha_2)^{-1}\alpha_1$$
$$g(0) = \frac{1-\alpha_2}{\alpha_1}\left[(1-\alpha_1^2-\alpha_2^2)\frac{1-\alpha_2}{\alpha_1} - 2\alpha_1\alpha_2\right]^{-1}.$$

The 2-step transition $\mathcal{T}_2^*$ *is given by*

$$X(t) = \gamma X(t-2) + \eta(t),$$

where $\gamma = g(2)g(0)^{-1}$ *and* $\eta(t) \sim \mathsf{N}(0, [1-\gamma^2]g(0))$ *independent of* $X(t-2)$. *On the other hand, the convolution of the 1-step operators,* $\mathcal{T}_1^*\mathcal{T}_1^*$, *obeys*

$$\begin{aligned} X(t) &= \beta \times (\beta X(t-2)) + \{\epsilon(t) + \beta\epsilon(t-1)\} \\ &= \beta^2 X(t-2) + \tilde{\epsilon}(t), \end{aligned}$$

where $\beta^2 \neq \gamma$ *and* $\tilde{\epsilon}(t)$ *is not independent of* $X(t-2)$, *except if* $\alpha_2 = 0$, *i.e., if* X *is Markov. Thus, in general,* $\mathcal{T}_2^* \neq \mathcal{T}_1^*\mathcal{T}_1^*$*—the semigroup property fails—when* $\mathcal{T}_1^*$ *is calculated from the nonetheless-valid time-homogeneous equation* (3.20).

How is a result on violation of the semigroup property useful to the applied researcher? For studying growth and convergence Quah (1993a) compared transition probability descriptions of the cross-section of economy incomes over long horizons and short horizons. From the discussion in Example 3.7 we see that whether the long-horizon characterization matches the appropriate convolution of the short-horizon characterization is an examination of the Markov property. Conversely, modifications of the underlying statistical model—for instance, allowing certain kinds of mover-stayer mixtures (Blumen, Kogan, and McCarthy, 1955; Singer and Spilerman, 1976)—so long as they preserve the Markov property in F_t will, necessarily, not help explain deviation of the long-horizon description from the short-horizon ones.

Section 3.5 continues this general discussion to consider resolvent operators. Before proceeding to that, however, Section 3.3 considers estimation of these relatively abstract objects just developed. Then Section 3.4 treats Markov chains and shows how the abstract concepts can be given concrete and explicit form in the discrete-state special case.

3.3 Estimating stochastic kernels

For the remainder of this section we can confine discussion to the stochastic kernel $\mathcal{M}$ or, in effect, its derivative the transition probability p_t. To estimate the stochastic kernel notice that equation (3.15) is solved for $p_{t,s}$ by

$$p_{t,s}(x,x') = \begin{cases} \mathsf{f}_{X,(t,t+s)}(x,x')\,\mathsf{f}_{X,t}(x)^{-1} & \text{when } \mathsf{f}_{X,t}(x) \neq 0 \\ 0 & \text{otherwise,} \end{cases} \tag{3.21}$$

where $\mathsf{f}_{X,(t,t+s)}(x,x')$ denotes the joint probability density function at $x_j(t) = x$ and $x_j(t+s) = x'$.

From the reasoning leading up to the kernel density estimator (3.11) for f_X, we can similarly define a kernel density estimator for $\mathsf{f}_{X,(t,t+s)}$.

Definition 3.8 *A non-negative function* $\mathsf{K} : \mathbb{R}^2 \to \mathbb{R}_+$ *with*

$$\int_{\mathbb{R}^2} \mathsf{K}(x)\,dx = 1$$

is a **two-dimensional kernel function** *or just a* **kernel**. *A kernel is* **radially symmetric** *if* $\mathsf{K}(Sx) = \mathsf{K}(x)$ *for all* $x \in \mathbb{R}^2$ *and all* S *such that* $SS' = I$. *Fixing* t *and* s, *let* $x_j = (\,x_j(t), x_j(t+s)\,)'$. *A kernel is a* **product kernel** *if it can be factored into a product of scalar kernel functions, i.e., for* $x = (x_1, x_2)'$, $\mathsf{K}(x) = \mathsf{K}_1(x_1)\mathsf{K}_2(x_2)$. *For* K *a two-dimensional kernel function and* $b > 0$, *the function*

$$\widehat{\mathbf{f}}_{\mathbf{X},b}(x) \stackrel{def}{=} \frac{1}{Jb^2}\sum_{j=1}^{J} \mathsf{K}\left([x - x_j]\,b^{-1}\right) \tag{3.22}$$

is a **kernel density estimator with kernel** K *and* **bandwidth** b.

This definition parallels Definition 3.2 with the obvious change from a scalar x to the bivariate x and from a normalizing b to b^2.

> *[[Product kernels as mutually independent perturbations, following noise-interpretation in Definition 3.2 of kernel as noise-augmentation. Table some typical multivariate kernels.]]*

The estimator for the stochastic kernel simply puts together Definitions 3.2 and 3.8.

Definition 3.9 *Let* K_1 *and* K_0 *be two-dimensional and scalar kernel functions, respectively; and let* $b_1, b_0 > 0$. *Given* t *and* $s > 0$ *let* $\widehat{\mathbf{f}}_{\mathbf{X},b}$ *be the kernel density estimator with kernel* K_1 *and bandwidth* b_1 *from* $\{\, x_j(t, t+s) = (x_j(t), x_j(t+s))' : j = 1, 2, \ldots, J \,\}$*; and let* $\widehat{f}_{X,b}$ *be the kernel density estimator with kernel* K_0 *and bandwidth* b_0 *from* $\{\, x_j(t) : j = 1, 2, \ldots, J \,\}$.

Consistency and asymptotic distribution properties are available for the stochastic kernel estimator.

Theorem 3.10 *Let [. . .]. (Stochastic kernel LLN)*

Theorem 3.11 *Let [. . .]. (Stochastic kernel CLT)*

> *[[In Theorems 3.10 and 3.11 describe the similarities with and differences from standard results in Cacoullos (1966), Chen, Linton, and Robinson (2001), and Roussas (1969). Combine them. Describe estimated stochastic kernels in Figure 5.1 and Figure 5.2.]]*

3.4 Markov chains

Empirical researchers often consider two different discrete models that lead to *Markov chains* and *transition probability matrices*. A Markov chain is a Markov process taking values only on a discrete set; a transition probability matrix is the counterpart, for Markov chains, of a transition density (3.15)–(3.16).

Of these two discrete models, one will be implied by equation (3.15) and therefore is a special case of the analysis where densities are available. But the other is not. While their common form allows a unified presentation, understanding that they differ helps sharpen the empirical modelling.

Of course, many excellent presentations of Markov chains already exist in the literature. What is distinctive here is that the presentation begins from where a researcher observes a set of data (Definition 3.1) and has access (Theorems 3.3–3.6, Theorems 3.10–3.11) to the objects estimable from those data. The development then builds on these data-derived constructions, rather than, say, hypothesizing discrete probability vectors and transition probability matrices a priori. This approach makes transparent the relation between, among other objects previously discussed, transition probability matrices and transition operators; it shows how a researcher calculates one from another and what information might be lost in one but preserved in yet others. Such matters are (rightly) never considered in rigorous theoretical development of Markov chain analysis. But, on the other hand, they constitute part of what applied researchers need to know.

For the first model, take F_t continuous but piecewise linear; for the second, take F_t discontinuous but piecewise constant (see Figures 3.1 and 3.2 respectively). Call *interval-discrete* the first case—that in Figure 3.1; call *point-discrete* the second case—that in Figure 3.2. An interval-discrete model continues to have density with respect to Lebesgue measure, as shown in the graph for f in Figure 3.1. Following our usual convention, take the density to be right continuous with

Figure 3.1 Distribution F and implied density f for a discrete model; F is continuous but piecewise linear.

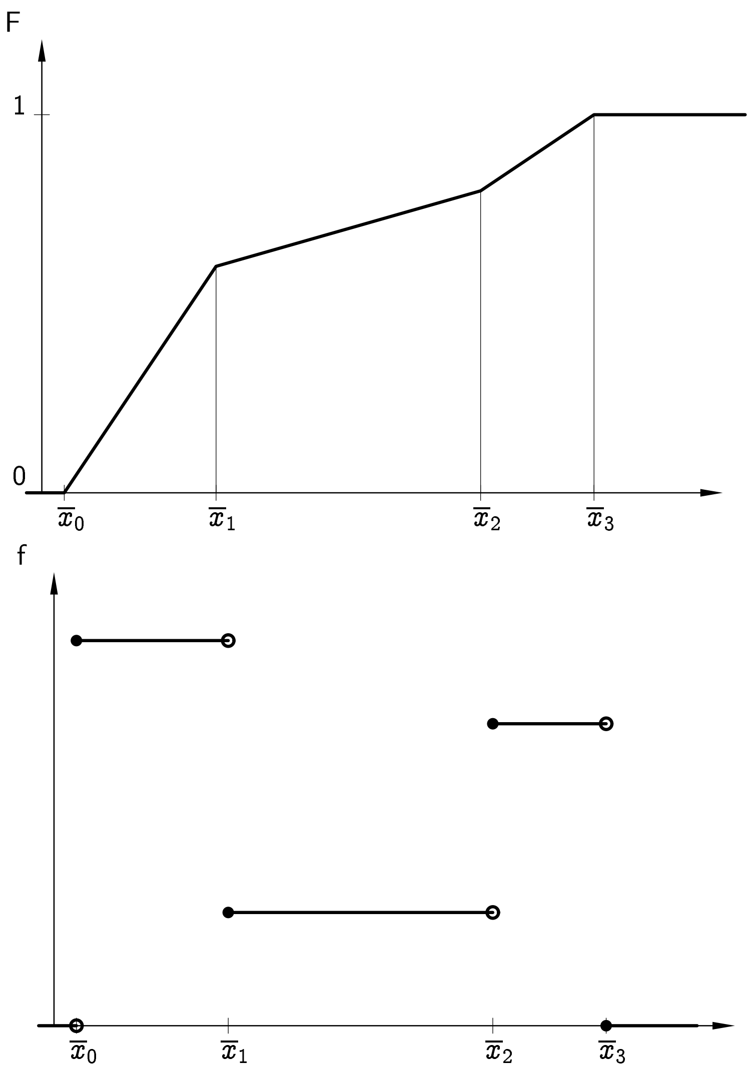
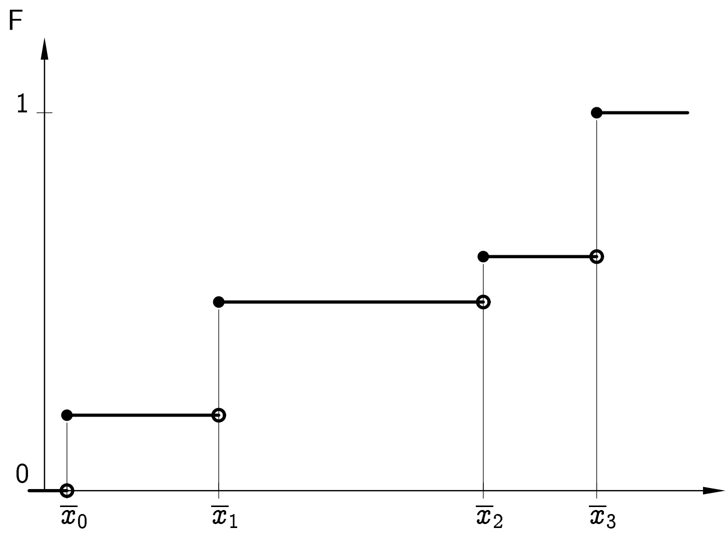


Figure 3.2 Distribution F for a discrete model; F is discontinuous but piecewise constant.



left limits. By contrast, a point-discrete model has no such density.

Interval-discrete and point-discrete models obviously don't exhaust the class of Markov chain models in general, where the discrete states might be abstract sets, not just connected intervals or points on the real line. However, these two do constitute all the useful discrete models that can be derived from framework in section 3.2, where the initial underlying state space is naturally taken to be the set of real numbers.

Both interval-discrete and point-discrete models define for each t a finite or at most countably infinite set

$$\{\,\overline{x}_0, \overline{x}_1, \overline{x}_2, \ldots,\,\} \subset \mathbb{R}, \quad \overline{x}_m < \overline{x}_{m+1}\ \forall\ m = 0, 1, 2, \ldots,$$

made up of either the points of non-differentiability in F_t for the interval-discrete model or the points of discontinuity in F_t for the point-discrete model. Assume this set is invariant in time. This restriction can be relaxed with no conceptual difficulty, but at the cost of cumbersome notation.

Identify the distinguished points with a set of discrete states, labelled $\{\,0, 1, 2, \ldots\,\}$. The number of states M can be finite or countably infinite, where for the interval-discrete case

$$M = \inf\{\,m : \mathsf{F}(\overline{x}_m) = 1\,\}$$

and for the point-discrete case

$$M = \inf\{\,m : \mathsf{F}(\overline{x}_m) = 1\,\} + 1.$$

For both models associate to each discrete state $m = 0, 1, \ldots, M-1$ a probability

$$\overline{\mathsf{f}}(m) = \int_{\overline{x}_m}^{\overline{x}_{m+1}} d\mathsf{F}(x) = \left[\lim_{x \uparrow \overline{x}_{m+1}} \mathsf{F}(x)\right] - \left[\lim_{x \uparrow \overline{x}_m} \mathsf{F}(x)\right]. \qquad (3.23)$$

Because the interval-discrete case has F continuous and piecewise linear, equation (3.23) gives

$$\begin{aligned}\overline{\mathsf{f}}(m) = \mathsf{F}(\overline{x}_{m+1}) - \mathsf{F}(\overline{x}_m) &= \int_{\overline{x}_m}^{\overline{x}_{m+1}} \mathsf{f}(x)\,dx \\ &= \mathsf{f}(\overline{x}_m) \times (\,\overline{x}_{m+1} - \overline{x}_m\,)\,. \end{aligned} \tag{3.24}$$

Similarly, our convention on right-continuity implies that the point-discrete case has (3.23) as

$$\overline{\mathsf{f}}(m) = \mathsf{F}(\overline{x}_m) - \lim_{x\uparrow\overline{x}_m} \mathsf{F}(x). \tag{3.25}$$

The general definition (3.23) shows the probabilities $\overline{\mathsf{f}}$ to be changes in the values of the distribution F. Special cases (3.24) and (3.25), respectively, show $\overline{\mathsf{f}}$ as areas under the density and as the sizes of jumps in the distribution at discontinuity points.

Reintroduce the time index t for $\overline{\mathsf{f}}_t(m)$ and form the vector

$$\overline{\mathsf{f}}_t = \begin{pmatrix} \overline{\mathsf{f}}_t(0) & \overline{\mathsf{f}}_t(1) & \cdots & \overline{\mathsf{f}}_t(M-1) \end{pmatrix}^{\top},$$

where $^{\top}$ denotes vector and matrix transposition.[14] For both interval-discrete and point-discrete models the version of the transition equation (3.13) used here will be:

$$\overline{\mathsf{f}}_{t+s} = P_{t,s}^{\top} \times \overline{\mathsf{f}}_t, \tag{3.26}$$

where the $M \times M$ *transition probability matrix*

$$P_{t,s}(m, m'), \quad m, m' = 0, 1, \ldots, M-1,$$

characterizes transitions from time t to time $t + s$. Even though the transition probability matrix equation (3.26) for Markov chains

[14] This monograph adopts the convention that vectors and matrices begin their indexes at 0 rather than 1. The development of the material in the text makes that convenient and natural. This zero-offset convention is also used in some linear algebra texts and in many programming languages.

applies in all cases of interest, because F_t will not always admit a density, equation (3.26) is not a direct counterpart to the transition density equation (3.15). Nor is it just a special case of the transition equation (3.13), which describes dynamics in the distributions, not in their Radon-Nikodym derivatives.

In the interval-discrete case, however, the transition density does exist and can be usefully and intuitively related to the transition probability matrix.

Proposition 3.12 *For the interval-discrete case:*

(i) Matrix $P_{t,s}$ in equation (3.26) has

$$P_{t,s}(m, m') = \\ (\overline{x}_{m+1} - \overline{x}_m)^{-1} \times \int_{\overline{x}_{m'}}^{\overline{x}_{m'+1}} \int_{\overline{x}_m}^{\overline{x}_{m+1}} p_{t,s}(x, x')\, dx\, dx'. \quad (3.27)$$

(ii) For a function ϕ piecewise constant on $[\overline{x}_m, \overline{x}_{m+1})$ define the vector $\overline{\phi}$ where $\overline{\phi}(m) = \phi(x)$, $x \in [\overline{x}_m, \overline{x}_{m+1})$. Then the image of ϕ under transition operator $\mathcal{T}_{t,s}$ has its interval-averaged version given by $P_{t,s}\overline{\phi}$, i.e.,

$$\left(P_{t,s}\overline{\phi} \right)(m) = (\overline{x}_{m+1} - \overline{x}_m)^{-1} \int_{\overline{x}_m}^{\overline{x}_{m+1}} (\mathcal{T}_{t,s}\phi)(x)\, dx. \quad (3.28)$$

If further for each x' the transition density function $p_{t,s}(\cdot, x')$ is piecewise constant on $[\overline{x}_m, \overline{x}_{m+1})$ then

$$\left(P_{t,s}\overline{\phi} \right)(m) = (\mathcal{T}_{t,s}\phi)(\overline{x}_m).$$

How is Prop. 3.12, or the framework here more generally, useful to an empirical researcher? Recall that theoretical analyses typically take as given the set of states for a Markov chain. In applied empirical work, however, a researcher might need to decide which discretization or finite set of states to use, calculated off of observations on the income distribution.[15] Since incomes data are continuous variables the researcher must have hypothesized a particular piecewise linear distribution or interval-discrete model, thereby dividing the income observations into a partition of discrete sets. Prop. 3.12 makes precise what it means to represent the dynamics of the income distribution in a transition probability matrix.[16]

Equation (3.27) says that the transition probability matrix P is an integral or average of the underlying transition density. Different transition densities p therefore imply the same transition probability matrix P. The underlying transition density need not and generally will not display the same piecewise linearity as the distributions F. When a researcher estimates a transition probability matrix following a particular discretization of the space of incomes—hypothesizes a particular Figure 3.1, even when entirely accurate—the underlying dynamics can be identified only up to some equivalence class. The underlying transition density, and thus $\mathcal{T}$, $\mathcal{T}^*$, and $\mathcal{M}$ still contain relevant information additional to that estimated by the researcher.

This can be seen in part (ii) of Prop. 3.12: vector $P\overline{\phi}$ contains only smoothed or averaged versions of the conditional expectation $\mathcal{T}\phi$; it does not recover that conditional expectation itself, even when ϕ is piecewise constant. Only when the transition density is also

[15] Examples include Epstein, Howlett, and Schulze (2003), Quah (1993a), and ... [[put others]].

[16] To be clear, Prop. 3.12 does not claim that every continuous-state Markov process with a transition density can be accurately approximated by a discretization. From Chung (1967) we know that a continuous-state Markov chain when inappropriately discretized will no longer be Markov; however, see Bulli (2001) for an interesting attempt to deal with this confounding of the underlying law of motion. By contrast, in the text the interval-discrete case *begins* with underlying distribution F in Figure 3.1. The discussion here hypothesizes an appropriate discretization and seeks the features of the transition dynamics consistent with that discretization.

piecewise constant, do $P\overline{\phi}$ and $\mathcal{T}\phi$ coincide. Variation in the first argument of the transition density function allows dynamics strictly richer than that recoverable in the discretization, regardless of that discretization's validity and accuracy. Similarly, the conclusion in (3.28) fails in general if the function ϕ is not piecewise constant, even with $\overline{\phi}$ replaced by, say, fixed averages of ϕ over the appropriate $[\overline{x}_m, \overline{x}_{m+1})$ intervals.

More positively, however, by Prop. 3.12 we have:

$$P_{t,s}(m, m') \geq 0, \quad \text{all } m,\, m'. \tag{3.29}$$

Moreover,

$$\sum_{m=0}^{M-1} \int_{\overline{x}_{m'}}^{\overline{x}_{m'+1}} p_{t,s}(x, x')\, dx' = \int_{-\infty}^{\infty} p_{t,s}(x, x')\, dx' = 1$$

implies

$$\sum_{m'=0}^{M-1} P_{t,s}(m, m') = (\overline{x}_{m+1} - \overline{x}_m)^{-1} \int_{\overline{x}_m}^{\overline{x}_{m+1}} 1\, dx = 1. \tag{3.30}$$

In words (3.29) states non-negativity and (3.30) row sums equalling unity.[17]

For the point-discrete case described in Figure 3.2 part (i) of Prop. 3.12 no longer applies. Instead, by direct inspection, we can set

$$P_{t,s}(m, m') = \mathcal{M}_{t,s}(\, \overline{x}_m, (-\infty, \overline{x}_{m'}]\,) - \lim_{x \uparrow \overline{x}_{m'}} \mathcal{M}_{t,s}(\, \overline{x}_m, (-\infty, x]\,). \tag{3.31}$$

[17] A matrix satisfying (3.29) and (3.30) is often referred to as a *Markov matrix* or a *stochastic matrix*. Although the logic for this terminology is long-established, the language can nonetheless be confusing to non-specialists, as obviously such a matrix is Markov or stochastic in a way different from how a stochastic process is Markov or a random variable is stochastic.

To understand this, notice that when F_t has the point-discrete property of Figure 3.2 the integral on the right side of (3.14) selects values $\mathcal{M}_{t,s}(x,\cdot)$ only at $x = \overline{x}_m$. Then that the left side F_{t+s} of (3.14) has the point-discrete property means $\mathcal{M}_{t,s}(\overline{x}_m, (-\infty, \cdot])$ must do so as well. Here too the transition probability matrix P fails to contain all the features in the underlying stochastic kernel $\mathcal{M}$; however, unlike in the interval-discrete case, any possible aliasing occurs only on zero-probability events.

The conditions (3.29) and (3.30) follow immediately from (3.31). This point-discrete case is not, however, typically observed directly in actual incomes data and therefore is not generally useful for empirical work. On the other hand, it turns out to be well-suited to analytical modelling, e.g., in Section 6.2 to follow. For the point-discrete case, there is no longer an entire range $[\overline{x}_m, \overline{x}_{m+1})$ associated with each state m; instead, only the value $\overline{x}_m$ matters. Thus, although F_t here admits no density in the usual sense, the analysis actually simplifies, relative to the interval-discrete model. The transition operator expression, analogous to (3.28), no longer involves an averaging but is always exact. Letting ϕ be any function, no longer necessarily pointwise constant, define $\overline{\phi}$ by $\overline{\phi}(m) = \phi(\overline{x}_m)$. Using (3.31),

$$
\begin{aligned}
&\left(P_{t,s}\overline{\phi} \right)(m) \\
&= \sum_{m'=0}^{M-1} P_{t,s}(m,m')\overline{\phi}(m') = \sum_{m'=0}^{M-1} P_{t,s}(m,m')\phi(\overline{x}_{m'}) \\
&= \phi(\overline{x}_0)\mathcal{M}_{t,s}(\,\overline{x}_m, (-\infty, \overline{x}_0]\,) + \sum_{m'=1}^{M-1} \int_{\overline{x}_{m'-1}}^{\overline{x}_{m'}} \phi(x')\mathcal{M}_{t,s}(\,\overline{x}_m, dx'\,) \\
&= \int_{-\infty}^{\infty} \phi(x')\mathcal{M}_{t,s}(\,\overline{x}_m, dx'\,) = (\mathcal{T}\phi)(\overline{x}_m).
\end{aligned}
$$

The expression after the third equals sign follows from noting that the right side of (3.31) equals $\mathcal{M}_{t,s}(\,\overline{x}_m, \overline{x}_0\,)$ for $m' = 0$ whereas for

$m' = 1, 2, \ldots$ the right side satisfies

$$\begin{aligned}
&\mathcal{M}_{t,s}(\,\overline{x}_m, (-\infty, \overline{x}_{m'}]\,) - \lim_{x \uparrow \overline{x}_{m'}} \mathcal{M}_{t,s}(\,\overline{x}_m, (-\infty, x]\,) \\
&\quad = \mathcal{M}_{t,s}(\,\overline{x}_m, (-\infty, \overline{x}_{m'}]\,) - \mathcal{M}_{t,s}(\,\overline{x}_m, (-\infty, x]\,) \;\forall x \in [\overline{x}_{m'-1}, \overline{x}_{m'}) \\
&\quad = \int_{\overline{x}_{m'-1}}^{\overline{x}_{m'}} \mathcal{M}_{t,s}(\,\overline{x}_m, dx'\,).
\end{aligned}$$

For all Markov chains—whether interval- or point-discrete—each row of the matrix $P_{t,s}$ is a vector of probabilities. Therefore, the operation on the right side of equation (3.26) finds the $\overline{\mathsf{f}}_t$-probability mixture of a collection of probability vectors, and thus generates yet another probability vector. Again, for all Markov chains—interval- or point-discrete—the adjoint equation (3.18) is immediate:

$$\begin{aligned}
\sum_{m=0}^{M-1} \left(P_{t,s}\overline{\phi} \right)(m) \times \overline{\mathsf{f}}_t(m) &= \left(P_{t,s}\overline{\phi} \right)^{\top} \overline{\mathsf{f}}_t = \overline{\phi}^{\top} (\, P_{t,s}^{\top} \overline{\mathsf{f}}_t \,) \\
&= \sum_{m'=0}^{M-1} \overline{\phi}(m') \times (\, P_{t,s}^{\top} \overline{\mathsf{f}}_t \,)(m'),
\end{aligned}$$

so that we can identify $P_{t,s}$ with $\mathcal{T}_{t,s}$ and the adjoint with the matrix transpose.

Summarize the notation for describing a Markov chain. Given M the number of discrete states, call $\overline{x}$ the vector $(\,\overline{x}_0, \overline{x}_1, \ldots \overline{x}_{M-1}\,)^{\top}$ of possible realizations and P the matrix function

$$\{\, P_{t,s}(m, m'), \quad m, m' = 0, 1, \ldots, M-1;\; t, s \geq 0 \,\}.$$

Let $\overline{\mathsf{f}}_0$ be a probability vector, taken to describe an initial distribution. A Markov chain is then the 4-tuple $(M, \overline{x}, P, \overline{\mathsf{f}}_0)$.

If P is time-homogeneous, i.e., $P_{t,s} = P_{0,s}$ for all $t,\ s \geq 0$, and P satisfies the semigroup property (3.19), call $P_s = P_{0,s}$. Then in a direct extension of the property for scalar real-valued functions $P_t = e^{\mathbf{G}t}$ for some fixed $M \times M$ matrix $\mathbf{G}$. The matrix $\mathbf{G}$ is called the *infinitesimal generator* of the semigroup P and can be obtained by matrix differentiation:

$$\mathbf{G} = \lim_{t \downarrow 0} \frac{P_t - I}{t}.$$

(These will be reintroduced for the general case in Definition 3.15.) Writing out the entries explicitly as

$$\mathbf{G} = \{ \mathbf{G}(m, m'),\ m, m' = 0, 1, \ldots, M-1, \}$$

it is immediate that $\sum_{m'} \mathbf{G}(\cdot, m') = 0$ and $\mathbf{G}(m, m) \leq 0$.

3.5 Resolvent operators

Time-homogeneity allows two further tools that will be used repeatedly later (section 6.2) to analyze growth and distribution.

Definition 3.13 *For X Markov with time-homogeneous transitions $\mathcal{T}$, define $\mathbf{R}_\lambda$ $(\lambda > 0)$, the* **resolvent operator** *or simply* **resolvent** *of X, to be the Laplace transform of the semigroup $\{ \mathcal{T}_t : t \geq 0 \}$, i.e.,*

$$\mathbf{R}_\lambda = \int_0^\infty e^{-\lambda t}\, \mathcal{T}_t \, dt, \quad \lambda > 0.$$

Resolvent operators map bounded measurable functions to bounded measurable functions by

$$\begin{aligned}
(\mathbf{R}_\lambda \phi)(x) &= \int_0^\infty e^{-\lambda t}\, (\mathcal{T}_t \phi)(x)\, dt \\
&= \int_0^\infty e^{-\lambda t}\, E\,[\, \phi(X(t)) \mid X(0) = x \,]\; dt \\
&= \psi(x).
\end{aligned}$$

Moreover, the resolvent has a kernel representation useful for calculations. To obtain it, use (3.17) and interchange the integration order in

$$
\begin{aligned}
(\mathbf{R}_\lambda \phi)(x') &= \int_0^\infty e^{-\lambda t}\, (\mathcal{T}_t \phi)(x')\, dt \\
&= \int_0^\infty e^{-\lambda t}\, E\left[\, \phi(X(t)) \mid X(0) = x' \,\right]\, dt \\
&= \int_0^\infty e^{-\lambda t} \left[\int_{-\infty}^\infty \phi(x) \mathcal{M}_t(x', dx) \right] dt \\
&= \int_{-\infty}^\infty \phi(x) \left[\int_0^\infty e^{-\lambda t}\, \mathcal{M}_t(x', dx)\, dt \right]
\end{aligned}
$$

so that

$$(\mathbf{R}_\lambda \phi)(x') = \int_{-\infty}^\infty \phi(x)\, G_\lambda(x', x)\, dx, \tag{3.32}$$

where the **resolvent kernel** satisfies[18]

$$G_\lambda(x', x)\, dx = \int_0^\infty e^{-\lambda t} \mathcal{M}_t(x', dx)\, dt. \tag{3.33}$$

When the stochastic kernel $\mathcal{M}$ admits a density, as in equation (3.16), then the resolvent kernel equals

$$G_\lambda(x', x) = \int_0^\infty e^{-\lambda t} p_t(x', x)\, dt,$$

i.e., the resolvent kernel is the Laplace transform of the transition density.

Record the transition density and resolvent kernel for Brownian Motion, for use in Section 6.4:

Theorem 3.14 *For $X = \sigma B$, where B is standard Brownian Motion, the transition density is*

$$p_t(x', x) = (2\pi\sigma^2 t)^{-1/2} \exp\left(-\frac{(x - x')^2}{2\sigma^2 t} \right); \tag{3.34}$$

[18] This notation follows Karlin and Taylor (1981, Ch. 15.11, p. 287) but differs from, say, Gihman and Skorohod (1975). It is motivated by G_λ being related to so-called Green functions for boundary value problems in differential equations.

and the resolvent kernel is

$$G_\lambda(x', x) = \sigma^{-1} \frac{1}{\sqrt{2\lambda}} e^{-|x-x'|\sqrt{2\lambda}/\sigma} . \tag{3.35}$$

Finally, we have

Definition 3.15 *For X Markov with time-homogeneous transitions $\mathcal{T}$, define $\mathbf{G}$ the* **infinitesimal generator** *of X as the derivative of the semigroup $\{\mathcal{T}_t : t \geq 0\}$ at $t = 0$, i.e.,*

$$\mathbf{G} = \lim_{\Delta t \downarrow 0} \left[\frac{\mathcal{T}_{\Delta t} - I}{\Delta t} \right] ,$$

where I denotes the identity operator.

The resolvent operator and infinitesimal generator can be related to each other (e.g., Gihman and Skorohod, 1975, Theorem 1, p. 97):

Theorem 3.16 *An X Markov with time-homogeneous transitions $\mathcal{T}$ has its resolvent operator $\mathbf{R}_\lambda$ and infinitesimal generator $\mathbf{G}$ satisfy:*

$$\mathbf{R}_\lambda = (\lambda - \mathbf{G})^{-1}, \quad \forall \lambda > 0. \tag{3.36}$$

By definitions 3.13 and 3.15, respectively, in equation (3.36) the left side, $\mathbf{R}_\lambda$, is an integral taking into account all $\mathcal{T}_t$ ($t > 0$) whereas the right side, $(\lambda - \mathbf{G})^{-1}$, involves only the derivative of $\mathcal{T}_t$ at a single point $t = 0$. That the two sides of (3.36) are equal to each other is remarkable, and shows the power of the time-homogeneity restriction in semigroups.

Some further insight into this is provided by the discrete-state Markov chain (section 3.4) case. There the resolvent operator can be explicitly calculated:

$$\begin{aligned} \mathbf{R}_\lambda &= \int_0^\infty e^{-\lambda t} P_t \, dt = \int_0^\infty e^{-(\lambda I - \mathbf{G})t} \, dt \\ &= (\lambda - \mathbf{G})^{-1}, \end{aligned}$$

exactly as in equation (3.36).

To anticipate the use of equations (3.33) and (3.36) in section 6.2, notice that when the resolvent is evaluated at a utility function, the result is the expected present discounted value of the flow of utility.[19] When consumption, the argument of the utility function, is a diffusion process—a Markov process with continuous timepaths—then its infinitesimal generator can be described by a second-order differential operation. If, however, consumption is a Markov chain—with jumps representing sudden discrete changes—then its infinitesimal generator can be described by a matrix. Either way, equation (3.36) allows explicitly calculating the present discounted utility value of consumption under alternative scenarios for growth and inequality.

3.6 Long run distributions and other characteristics

We can analyze the distributions and stochastic kernel in (3.14), having obtained their empirical counterparts using the results in 3.1.1–3.2.

Durlauf and Quah (1999, Appendix A) describes ....

> *Limiting behavior. Doblin's condition. Ergodicity. First-passage times. Doob (1953). Futia (1982). Quah (1996a). Stokey and Lucas (1989).*

For a Markov chain $(M, \overline{x}, P, \overline{\mathrm{f}}_0)$ having time-homogeneous transition probability matrix, a *stationary* or *invariant probability (vector)* is any vector that can be formed as a convex combination of the rows of

$$P^{(\infty)} \stackrel{\text{def}}{=} \lim_{t\to\infty} P_t,$$

provided the limit on the right exists.

[19] Rogers (1997) used $\cdots$. Duffie and Glynn (2004) applied $\cdots$.

To see why this is reasonable suppose $\bar{f}_0$ is an invariant probability. Because $\bar{f}_0$ is a convex combination of the rows of $P^{(\infty)}$, a probability vector $\overline{\phi}$ exists so that

$$\bar{f}_0 = \left[P^{(\infty)}\right]^{\top} \overline{\phi}.$$

Then for all $t > 0$

$$\begin{aligned}\bar{f}_t &= P_t^{\top}\bar{f}_0 = P_t^{\top}\left[P^{(\infty)}\right]^{\top}\overline{\phi} = \left[P^{(\infty)}P_t\right]^{\top}\overline{\phi} \\ &= \left[\lim_{s\to\infty} P_{s+t}\right]^{\top}\overline{\phi} = \left[P^{(\infty)}\right]^{\top}\overline{\phi} = \bar{f}_0,\end{aligned}$$

since P satisfies the semigroup property (3.19). In words, an invariant probability vector is preserved through time. An invariant probability corresponds to a long-run steady state of the distributional dynamics; or, it is a fixed point of the associated transition operator.

If $P^{(\infty)}$ has all its rows identical, then every convex combination of those rows is just the first row: the stationary probability vector is therefore unique and equals that first row. But the converse also holds. If the stationary probability vector is unique, then all rows of $P^{(\infty)}$ must equal that stationary probability, for otherwise some convex combination of the rows would be available as a yet different stationary probability.

If P_1 has all rows identical then the Markov process is, when observed in discrete time, serially independent. From any initial distribution the process attains its invariant distribution equal to the first row of P_1 in a single time period, and it remains there forever after. Thus, the limit $P^{(\infty)}$ is well-defined with P_1 the unique limit point.

4 Catch-up models

To see how growth models restrict distribution dynamics, we begin with the simplest version of the Solow (1956) growth model. Let Y

be total output, N be the workforce, and K be the total capital stock. Denote per worker quantities in lower case:

$$y \stackrel{\text{def}}{=} Y/N \qquad k \stackrel{\text{def}}{=} K/N. \tag{4.1}$$

Output depends on K, N, and technology A through a standard smooth neoclassical production function. Assume technology A enters the production function multiplicatively in N so that output per worker can then be written as

$$y = Af(k/A), \quad f' > 0,\ f'' < 0,\ \lim_{k\to\infty} f(k)k^{-1} = 0. \tag{4.2}$$

Technology and the workforce evolve exogenously at constant growth rates

$$\dot{A}/A = \xi \geq 0, \quad A(0) > 0 \tag{4.3}$$

$$\dot{N}/N = \nu \geq 0. \quad N(0) > 0 \tag{4.4}$$

Capital depreciates at a constant rate δ and accumulates through savings equal to fraction τ of total income Y:

$$\dot{K} = \tau Y - \delta K, \quad \tau \text{ in } (0,1) \text{ and } \delta > 0. \tag{4.5}$$

Combining (4.1) through (4.5) gives the dynamic equation for capital per worker:

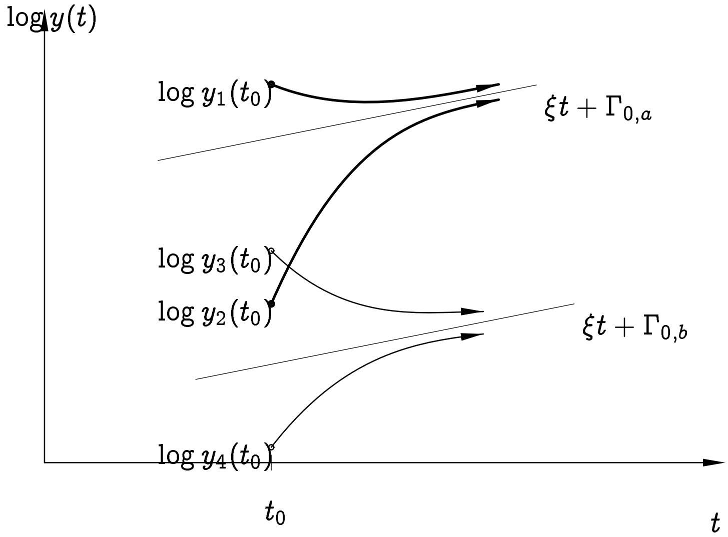
$$\dot{k}/k - \dot{A}/A = \tau\frac{f(k/A)}{k/A} - (\delta + \nu + \xi) \tag{4.6}$$

Under the standard curvature assumptions on f given in (4.2), equation (4.6) has a unique steady-state value $[k/A]^*$.

Taking together equations (4.2), (4.3), and (4.6) then gives observable dynamics for labor productivity:

$$\log y(t) = \Gamma_0 + \xi \cdot t + [\log y(0) - \Gamma_0]e^{\lambda t}, \tag{4.7}$$

Figure 4.1 A cross section of economies, each having different possible steady-state paths varying with Γ_0.



where

$$\begin{aligned}\Gamma_0 &= \log f\left([k/A]^*\right) + \log A(0) \\ &= g\left(\left[\delta + \nu + \xi\right]^{-1}\tau\right) + \log A(0), \quad \text{with } g' > 0, \\ \text{and } \lambda &= \lambda(f, (\delta + \nu + \xi), \tau) < 0.\end{aligned}$$

These dynamics are illustrated in Figure 4.1. For any one economy, say with output per worker y_1, economic history is the transition from its initial level to a specific steady-state path. However, the Figure also shows that the cross section of economies, having different underlying steady-state paths varying with Γ_0 and different initial incomes, displays a wide range of possible behaviors. Economies 2 and 3 diverge away from each other, criss-crossing along the way although they began close together at a middle-income level. Economies like 1 begin and remain rich; those like 4 begin and remain poor.

A stochastic kernel $\mathcal{M}$ for this range of possibilities would show simultaneously convergence (economies 1 and 2 approaching each other, and 3 and 4 similarly); divergence (economies 2 and 3 starting out together but then growing apart); and persistence at both high and low ends of the income range (economies 1 and 4 remaining, respectively, rich and poor).

All these different behaviors can be seen in the data (Quah, 2001b, 1993a, 1997). The analytically-derived Figure 4.1

Thus, while divergence might be an important feature of economic growth across countries (Pritchett, 1997), it is not the only prominent feature of distribution dynamics overall. Similarly, convergence (Barro and Sala-i-Martin, 1992). Theorizing on the convergence and divergence properties in economic growth should try to incorporate both features simultaneously, rather than exclusively one or the other.[20]

---

[20] Quah (1996a) has interpreted the model in Galor and Zeira (1993) as displaying exactly such diverse cross-section dynamics.

Quantify physical and human capital contributions
in pictures like Jones (1998, Figs. 3.1–3.2).
Importance of A, regardless whether exogenous or endogenous.

Jones (1997)
Mankiw, Romer, and Weil (1992)
Romer (1990)

(i) Importance of technology, or something like it to segue into subsection 4.2.
(ii) Panel data empirical methods can be misleading for question of interest

---

4.1 Theoretical models for cross-section distribution dynamics: Technology diffusion across countries

One convenient framework to model interacting economies, evolving in a rich cross section through time, is the following. Let $\mathcal{J}$ denote the given set of economies, j the typical element in $\mathcal{J}$, and $\mathcal{C} \subseteq \mathcal{J}$ a subset of economies. Thus,

$$A_j = \{A_j(t) : t \geq 0\}$$

is the timepath for the single, isolated economy j whereas

$$A_{\mathcal{C}} = \{A_j(t) : t \geq 0,\ j \in \mathcal{C}\}$$

is the collection of timepaths for economies in the group $\mathcal{C}$.

Applied to distribution dynamics, many models of technology and knowledge diffusion can be summarized as a *diffusion scheme*, i.e., a pair $(A_*, \mathcal{D})$, with A_* the *target* and $\mathcal{D}$ the *dynamic catch-up equation*. Examples of A_* include:

(i) $\overline{A}$ the average, either across a subgroup $\mathcal{C}$ or the entire set of economies $\mathcal{J}$;
(ii) A_0 some distinguished, fixed-identity leader economy;
(iii) $A_{\max}$ a specific leader, but whose identity varies through time.

Target A_* will also, in general, evolve through time in some way to be specified.

Dynamics $\mathcal{D}$ can be divided conceptually into two classes: First, where the resulting timepaths are continuous; second, where they are discontinuous, i.e., can take discrete jumps. Since we are modelling A, no physical limits prevent the second—indeed, certain distribution dynamics naturally call for jumps in the timepaths of specific countries. Examples of the first class of dynamics $\mathcal{D}$ include, for $\xi_j,\ \beta_j \geq 0$,

$$\dot{A}_j = \left[A_j^{1-\beta_j} A_*^{\beta_j}\right] \xi_j \tag{C1}$$

and

$$\dot{A}_j/A_j = \xi_j + \left(1 - \frac{A_j}{A_*}\right) \beta_j \tag{C2}$$

as distinct possibilities. In all cases, β_j can be interpreted as implying a *growth premium*: it is a "catch up to the leader" coefficient—the higher is β_j, the more rapidly A_j grows whenever $A_j < A_*$. By contrast, ξ_j can be viewed as the natural growth rate in A_j, distinct from catch-up tendencies: it is the growth rate in A_j whenever β_j equals zero. With β_j positive, the further A_j is below A_*, the faster does the economy grow, relative to the underlying natural growth rate ξ_j. Finally, notice that in (C1)–(C2) the target A_* behaves like just any other A_j.

Models with discrete jumps can be described as follows. For each $j \in \mathcal{J}$, define $t_0 = 0$, and for $k \geq 1$ let $t_k - t_{k-1}$ be continuous, positive-valued random variables, with distributions to be specified, varying with j.

Following Feller (1968, p. 73), call the (random) time points t_k *epochs*. If, for example, the j-th economy has its inter-epoch times distributed iid exponentially, then we can call epochs t_k the arrival times for a Poisson process. More generally, denote by P the distribution of time from one epoch until the next; and let p be the resulting probability density,

$$\mathsf{p}(t) = \frac{d\mathsf{P}}{dt}(t).$$

Then define the *hazard rate*

$$\mathsf{h}(t) = -\frac{d}{dt}\log\left(1 - \mathsf{P}(t)\right) = \frac{\mathsf{p}(t)}{1 - \mathsf{P}(t)} \qquad t \geq 0,$$

the expected arrival rate of events. The implied random process is a *point process*. Like the distribution P, the hazard h is fully informative on the properties of the point process, as we can always calculate:

$$\mathsf{P}(t) = 1 - e^{-\int_0^t \mathsf{h}(s)\,ds} \qquad t \geq 0. \tag{4.8}$$

Returning to the cross section of economies, assume technology levels change only at epochs, i.e., for $t \neq t_k$, let $\dot{A}_j(t) = 0$, but for $t = t_k$, one of the following holds:

$$A_j(t) = A_{\max}(t) \text{ for some timepath } \{A_{\max}(t) : t \geq 0\} \tag{D1}$$

or

$$A_j(t) = (1 + \xi_j) \lim_{s \uparrow t} A_j(s). \tag{D2}$$

Equation (D1) describes a jump to the frontier, regardless how far advanced the frontier might be ahead of A_j before j's epochal event, while equation (D2) is a discrete jump of height ξ_j, independent of where the frontier currently is. In (D1) the frontier $A_{\max}$ evolves in a way that varies across models. The average or expected growth rate of an economy j depends on ξ_j, $A_{\max}$, and the epochal time distributions.

Equations (C1) and (C2) represent gradual adjustment towards a target level A_*, in turn evolving through time. Equations (D1) and (D2), on the other hand, represent discontinuous, abrupt adjustment. What happens at the frontier $A_{\max}$ in (D1) matters importantly for the distribution dynamics: One typical assumption (e.g., Aghion and Howitt, 1998; Howitt, 2000) is that $A_{\max}$ evolves in a way related to but different from how A_j's not yet at the frontier do so. We return to this in section 4.2.

We will also consider hybrid schemes that mix continuous and discontinuous catchup dynamics.

4.2 Explicit diffusion schemes $(A_*, \mathcal{D})$: Examples

Many studies in the literature can be usefully organized into alternative diffusion schemes. To see this, first define the relative technology ratio $a_j = A_j/A_*$. Consider continuous and discontinuous $\mathcal{D}$ in turn.

Continuous $\mathcal{D}$

One diffusion scheme commonly used in empirical studies sets the target equal to a permanent leader economy, $A_* = A_0$, and the catchup dynamics to be gradual, either $\mathcal{D} =$ (C2) or $\mathcal{D} =$ (C1). Then,

$$\dot{A}_*/A_* = \dot{A}_0/A_0 = \xi_0.$$

Defining $\overline{a}$ to be the steady-state counterpart to the relative technology ratio a, either:

$$\overline{a}_j \stackrel{\text{def}}{=} (\xi_j/\xi_0)^{1/\beta_j}$$
$$\dot{a}_j/a_j = \left[(\overline{a}_j/a_j)^{\beta_j} - 1\right]\xi_j$$

for (C1), or

$$\overline{a}_j \stackrel{\text{def}}{=} 1 + (\xi_j - \xi_0)/\beta_j^{-1}$$
$$\dot{a}_j/a_j = -(a_j - \overline{a}_j) \times \beta_j$$

for (C2). That $\overline{a}_j$ differs from 1 implies a permanent cross-country difference in steady state equilibrium, and thus a non-degenerate invariant cross-section distribution, depending in general on intrinsic underlying growth rates ξ_j as well as catch-up coefficients β_j. If, however, $\xi_j = \xi_0$, then $a_j(t) \to 1$ as $t \to \infty$, independent of β_j; the invariant distribution is then a degenerate point mass.

Empirical examples of the special case of (C2) above include, among others, Bernard and Jones (1996) and Cameron, Proudman, and Redding (1998). Illustrative of the kinds of results available in this approach, Cameron, Proudman, and Redding (1998) find that compared to the US, OECD economies display long-run steady-state technology ratios a_j between 53% and 92%.

Lucas (1993) is a different variant on this class of examples, setting $A_* = \overline{A}$ and $\mathcal{D} =$ (C1). In a slight distortion of language, catch up here is not to a leader economy, but to the *average* of the cross section distribution. Economies that are advanced relative to the average slow down, while those that are behind speed up. When $\xi_j = \xi_0$, then again $a_j(t) \to 1$ as $t \to \infty$, so that in steady state, $A_j = A_0 = \overline{A}$: Convergence of the cross-section distribution is to a degenerate point mass.

Why might a target equalling the cross-section average, $A_* = \overline{A}$, make sense in an economic model? Lucas (1993, p. 255) refers to externalities in accumulating human capital, but suggests the specification is just a convenience in modelling as nothing essential changes if we replace $A_* = \overline{A}$ by the maximum $A_* = A_{\max}$ or the permanent leader $A_* = A_0$. Outside the present predictive analysis, however, this "dependence on the cross-section mean" externality specification might well bear substantive implications.

In all cases above, no intra-distribution crossings occur. Economy j never surpasses a j' previously better off than itself; at best, j becomes as rich as j', and then only asymptotically. Convergence of the cross-section distribution is monotone, and, apart from exogenous heterogeneity $\xi_j \neq \xi_0$, is always to a degenerate point mass. Any sensible measure of cross-section inequality or of the spread in the cross-section distribution will constantly diminish towards zero. Cross-economy dependence is straightforward: Dynamics in economy

j depend on one simple characteristic of the cross-section distribution, namely one of A_0, $A_{\max}$, or $\overline{A}$.

Extensions of this approach can proceed in a number of different directions, modifying the conclusions just given. One possibility is that target A_* might differ across distinct subsets of the cross section $\mathcal{J}$. Thus, denote groupings of economies,

$$\mathcal{C}_0,\ \mathcal{C}_1, \cdots \subsetneq \mathcal{J},$$

and suppose that each $\mathcal{C}$ bears a specific target A_* for all economies in the grouping. Each $\mathcal{C}$ is a *convergence club*, in the language used in Abramowitz (1986) or Baumol (1986). The analysis for each subset $\mathcal{C}$ proceeds exactly as above. However, combining analyses across subsets, we see that criss-crossings and overtakings can now potentially occur; and convergence of the distribution is, in general, to a multi-peaked long-run distribution. The economic question to analyze here becomes, Which j's *join* which groupings $\mathcal{C}$? What forces determine the membership of any given $\mathcal{C}$? Keely (1999) and Quah (2001b) have applied ideas from the analysis of endogenous coalition formation (Moldovanu and Winter, 1995; Ray and Vohra, 1999; Yi, 1997) to study these questions. Such work complements the standard convergence analysis that consider only (exogenously-specified) spillovers or externalities, and is an alternative description of the process of technology dissemination.

Hybrid $\mathcal{D}$

Lucas (2000) provides a cross-economy growth model combining the continuous timepath analyses described earlier with the discontinuous ones to follow. The model can also be viewed as one that places a probability structure on the endogenous coalition formation just described. Thus, it is a different representation of the idea that not all economies are simultaneously in the same convergence club.

The economics in the model turns on human capital development and knowledge dissemination, along lines studied in Tamura (1996).

Identify technology levels interchangeably with incomes, and suppose for simplicity that all economies are equal-sized with constant populations, normalized to unity. Economies fall, endogenously, into two categories: they are stagnant, or they have taken off and are on a development path converging towards the leading economy. Take-offs happen probabilistically and at varying calendar times for different individual economies.

How any given economy evolves, however, cannot be determined without knowing how the entire cross-section distribution evolves. Mechanically, the converse holds as well—the cross-section distribution is transformed along with the economies in it. In the model the cross section influences any given economy in two distinct ways: First, it affects the probability of a stagnant economy taking off. The more developed the world is, the more likely a currently-stagnant economy will launch itself on the path to development. Second, how developed the rest of the world is (or, in particular, the leading economy) determines how fast a newly-developing economy grows—a catch-up premium speeds up the growth of an economy that starts far behind the world's leaders.

The key results from the model are two-fold. The world develops through initially slow diffusion and then acceleration—as the world becomes richer both on average and cross-sectionally, more economies latch on yet more rapidly onto the path to development. This acceleration would not occur without *some* positive feedback from the developed world to the possibilities for takeoff among the world's stagnant economies. The model generates first increasing and then decreasing cross-country inequality, and displays a emerging but eventually only transient bimodality in the cross-section distribution. The early part of the transition, therefore, shows both divergence and twin-peakedness of the kind documented in Jones (1997), Pritchett (1997), and Quah (1997).

The model works as follows. Up through time 0 all J economies in the cross section $\mathcal{J}$ remain stagnant,

$$y_j(t) = \underline{y} > 0, \quad \text{for all } t \leq 0.$$

At $t = 0$ the incipient leader economy, $j = 1$, starts growing:

$$t \geq 0: \qquad \dot{y}_1(t)/y_1(t) = \xi > 0 \implies y_1(t) = \underline{y}e^{\xi t}. \qquad (4.9)$$

(In Lucas's interpretation, economy 1 at time 0 can be thought of as the UK at the beginning of the Industrial Revolution.) All other economies stagnate with incomes at $\underline{y}$ until, individually, they experience a take-off event, whereupon that now-developing economy converges upwards towards the leader economy. As time proceeds, like marathon runners at the starting line, more and more economies launch onto their respective development paths and race towards the leader.

For $j \in \mathcal{J}$, $j \neq 1$, call the random variable $t_j \geq 0$ the epoch of j's takeoff; write $t_1 = 0$. Conditional on history up through time t, call $\mathcal{C}_0(t)$ the collection of still-stagnant economies and call $\mathcal{C}_1(t)$ the complement, $\mathcal{J}\backslash\mathcal{C}_0(t)$, i.e., $\mathcal{C}_1(t) = \{j \in \mathcal{J} : t_j \leq t\}$. For any positive t, economy 1 is always in $\mathcal{C}_1$ and, probabilistically, others might be as well. Denote by J_0 the number of economies in $\mathcal{C}_0$ and J_1 the number in $\mathcal{C}_1$. Independently and identically distributed across the still-stagnant economies, take-off events occur over the time interval $[t, t + dt]$ with probability $\mathsf{h}(t)\, dt$. Thus, the hazard rate h probabilistically splits the world into two groups, the stagnant $\mathcal{C}_0$, where $y_j(t) = \underline{y}$ for $j \in \mathcal{C}_0$; and the developing $\mathcal{C}_1$, comprised of economy 1 and $j \neq 1$ for which we assume $y_j(t)$ grows as

$$1 - \frac{y_j(t)}{y_1(t)} = \left[1 - \frac{y_j(t_j)}{y_1(t_j)}\right] e^{-(t-t_j)\beta}, \qquad \beta > 0,\ t \geq t_j. \qquad (4.10)$$

Equation (4.10) says that eventually $y_j(t)$ catches up to $y_1(t)$: conditional on t_j, the right side of (4.10) declines exponentially to zero at rate β as t grows, so that $y_j(t) \nearrow y_1(t)$. The β coefficient in (4.10) properly describes convergence for economy j to the leader, not simply an economy's growth rate slowing down as that economy approaches its underlying steady-state growth path. This can be rewritten to say that later-developing economies enjoy a growth premium as (4.10) is also:

$$y_j(t) = y_1(t) \times \left[1 - \left(1 - \frac{y_j(t_j)}{y_1(t_j)}\right) e^{-(t-t_j)\beta}\right] \tag{4.11}$$

and log-differentiating with respect to time gives:

$$\frac{\dot{y}_j(t)}{\dot{y}_j(t)} = \xi + \beta \times \left[\frac{\left(1 - \frac{y_j(t_j)}{y_1(t_j)}\right) e^{-(t-t_j)\beta}}{1 - \left(1 - \frac{y_j(t_j)}{y_1(t_j)}\right) e^{-(t-t_j)\beta}}\right].$$

Developing economies grow nonlinearly. Their growth rates equal that of the leader's, ξ, plus a second convergence term that is larger the greater the distance from the leader economy at the takeoff date, but that then falls to zero over time. The later an economy begins to develop, the faster it grows ($\underline{y}/y_1(t_j)$ is lower), but as that economy matures, its growth rate slows, although remaining sufficiently high so that eventually it catches up to the leader.

As history unfolds, not only do the incomes of the economies in C_1 grow, but the size and membership of C_0 and C_1 also change. World income is the sum of group incomes across stagnant and developing economies, i.e., at time t,

$$\sum_{j \in C_0(t)} y_j(t) = J_0(t) \times \underline{y} \tag{4.12}$$

and

$$\sum_{j \in C_1(t)} y_j(t) = \sum_{\{j \ni t_j \leq t\}} y_j(t). \tag{4.13}$$

Provided J is sufficiently large that laws of large numbers apply to averages, we also have

$$\frac{J_0(t)}{J} = e^{-\int_0^t \mathsf{h}(s)\, ds} \tag{4.14}$$

and for (4.13)

$$J^{-1} \sum_{j \in C_1(t)} y_j(t) = \int_0^t \frac{J_0(s)}{J} \mathsf{h}(s) \cdot \left\{ y_1(t) \times \left[1 - \left(1 - \frac{y}{y_1(s)} \right) e^{-(t-s)\beta} \right] \right\} ds. \quad (4.15)$$

Using (4.11) the term in braces in equation (4.15) is the income at time t of an economy that began growing at epoch $t_j = s$. The multiplier $J^{-1}J_0(s)\mathsf{h}(s)\,ds$ is the ex-ante expected fraction of economies experiencing a takeoff event in an infinitesimal time interval at s. The integral then takes the total over all past epochs $s \in [0, t]$.

Specifying how h behaves will now allow the model to produce distribution dynamics for the cross section of economies through equations (4.9), (4.11), (4.13), (4.14), and (4.15). As Lucas (2000) observes, a constant hazard $\mathsf{h}(t) = \mathsf{h}(0)$ cannot adequately capture how as world development has proceeded in the 20th century, we observe more and more economies take off more and more rapidly. To allow this the hazard rate might be chosen to vary with the size of C_1 or with average world income (which, through (4.12), (4.14), and (4.15), varies positively with C_1). Lucas (2000) shows that selecting the latter allows a better fit to the data.

To summarize, the model displays two clear successes. First, allowing individual-economy takeoffs to depend on world development, the model explains the accelerating pace of development cross-sectionally. Second, if we identify the economic history analyzed econometrically in Jones (1997), Pritchett (1997), and Quah (1997) with the early part of the model's distributional transition, the model explains divergence and emerging twin-peakedness.

But note that however h might evolve through time, provided only that it is bounded from below by a positive constant, equation (4.14) has its right side converging to zero. More precisely, the discussion surrounding equation (4.8) implies that in this model all economies almost surely enter C_1 in finite time. After that time, the cross section of economies behaves as discussed earlier for (C1) or (C2): inequality decreases towards zero. Since initial inequality is also zero the transition path therefore has first increasing and then decreasing cross-section inequality, with a bimodality—the two clusters $\mathcal{C}_0$ and $\mathcal{C}_1$—emerging, becoming more pronounced, and then disappearing in the cross-section distribution as all economies get absorbed into $\mathcal{C}_1$.

Because the model treats takeoffs into development as a one-way transition—$\mathcal{C}_1$ membership is an absorbing state—the framework disallows the possibility that world development might show persistent divergence and twin-peakedness. This suggests that a useful feature to explore in the model might be *decay* events, the opposite to takeoff events. This terminology is intended to recall the title in Tamura (1996) and to suggest that transitions might be two-way, rather than simply uni-directional. Just as the hazard function h parametrizes the probabilistic beginning of industrialization, so too there might be another describing the probabilistic decay of a once-successful economy. Specifying it at the same level of theoretical transparency as done for h in Lucas (2000) allows a first step in the analysis but then investigating the structure of both more deeply would provide insight as well.

However the two hazard rates vary, as long as each is bounded from below by a positive constant, the cross section distribution of economies will settle down to a long-run steady state that is bimodal, although with continually ongoing simultaneous upwards and downwards transitions. Indeed, it is the balancing of those transitions that will determine a nondegenerate configuration in the long-run steady state distribution.

Neither economic logic nor economic history rules out such transitions to decay. The UK, despite its early lead with the Industrial Revolution, relinquished economic dominance to the US in the early 20th century. China's advanced state of technology in the 14th century underwent not just relative but absolute economic decline over the subsequent five millenia (Jones, 1988; Landes, 1998; Mokyr, 1990; Pomeranz, 2000; Quah, 2001a). After the flooding of the Bass Strait following the end of the last Ice Age 10,000 years ago, Aboriginal Tasmanians grew culturally isolated from mainland Australians and subsequently abandoned bone tools and fishing—relatively advanced technologies—and reverted to primitive hunting-gathering, where they remained up through their first encounter with modern Europeans in 1642 (Diamond, 1997).

A second line of potential further work begins by noting that spillovers in the model are global. Spillovers run along two distinct channels: First, from the leading economy $j = 1$ to every other economy that has already begun its takeoff and second, from the state of the world—the distribution of world incomes, through the hazard rate h—to takeoff possibilities for all the remaining, still-stagnant economies. Useful extensions to the framework might thus study if spillovers have more restricted scope and if so, what determinants might be important for them. If the spillovers are primarily in knowledge, do geography and physical distance matter? How do technologies for knowledge transmission alter patterns of world development? Or, does the cross-economy spread of knowledge occur through trade in ordinary merchandise, rather than more abstract trade in knowledge? If knowledge disseminates through learning of some kind, does that mean university training, or seeing how production and management take place in factories, or accessing engineering blueprints and scientific formulas? If it's the last of these that matter, how do regimes of intellectual property rights distort the dissemination of economic growth across countries? How do domestic attitudes and mechanisms for political control influence the acceptance of new ideas? Does the learning that matters come about through moving across economies people rather than ideas; is one mechanism more reliable than another?

Having some answers to these questions might lead in the model's formalization to the endogenous emergence of distinct or perhaps overlapping clusters C_0, C_1, ..., that are, again, convergence clubs.

Both this and the decay-events mechanism previously described will also produce over-taking and leapfrogging in across economies, thus far absent in the framework.

As Lucas (2000) said, "Gaining a quantitative understanding of all of these forces for diffusion—and, as Parente and Prescott (1994) would stress, the forces that oppose them—is the central question of the theory of economic growth and development." A limited selection of studies that have studied such questions explicitly for cross-economy patterns of growth includes Coe and Helpman (1995); Eaton and Kortum (1999); Epstein, Howlett, and Schulze (2003); Helpman (1993); Keller (2004, 1998); Matsuyama (2000); Parente and Prescott (2000, 1994); Quah (2001a, 1997).

Discontinuous $\mathcal{D}$

The leapfrogging conspicuously absent in the model just discussed appears instead as a leading feature in models of creative destruction (Aghion and Howitt, 1992, 1998). These models specify follower-economy technologies that leapfrog to the frontier whenever a specific event occurs (equation (D1)). R&D-driven models associate those events with fundamental discoveries or innovations. At such an epoch, the leader is deposed, and a different economy, that for whom the innovation has just realized, takes over as the leading economy. Criss-crossings obviously occur in such a specification, always to the frontier technology level.[21] The model is presented here in a way to make it directly comparable with that of Lucas's (Lucas, 2000) just described.

The leading economy is no longer fixed at $j = 1$ but has its identity vary across j's, depending on the unfolding of economic history.

[21] Typically, such a model is used to analyze economic performance across different industries within a single economy, although Howitt (2000) has also applied it to examine multiple industries across different economies.

But while its identity can vary, the growth rate of income at the frontier does not depend on which economy it is that's there; instead that growth rate depends on the global resources devoted to improving productivity—this might occur from research and development or learning by doing or accumulating human capital. Thus, across all economies spillover is extreme in pushing back the global frontier. Next, economies can jump, discontinuously, to the frontier from wherever they might be in the distribution, leap-frogging over all those initially ahead of them. This transition occurs when a discovery or revelation event suddenly occurs. Between such events, however, a given economy's productivity simply remains stagnant, thereby having the frontier move ever distant as well as yet other economies overtaking it in turn.

While significant intra-distributional churning occurs and growth at the frontier is ongoing, nonetheless a steady-state long-run distribution obtains. Under certain assumptions, moreover, that long-run cross-economy income distribution takes a power-law form.

Although all economies, typically, have different incomes and productivity levels at any time t, suppose that the returns to improving productivity equalize across economies. Each economy devotes resources so that its likelihood of transiting to the frontier is the same as as that for all others. Such expenditure will appear to have no discernible impact on most economies most of the time—most economies do not successfully jump to the frontier but instead have their income and productivity only remain constant. However, those expenditures do collectively contribute to pushing the frontier ever outwards—for instance, all research contributes to the global stock of knowledge—and, for that economy that happens to be successful, the steps it undertook to improve productivity appear to pay off out of all proportion.

A simple formalization of this dynamic mechanism is as follows. When economy j devotes resources $Res(t)$ at time t assume the hazard rate on that economy's leap-frogging to the frontier is $\mathsf{h}(t) =$

$Res(t)$ (by normalizing Res to appropriate units). Summing $Res(t)$ across economies, suppose that the global frontier per capita income $y_{\max}$ evolves as:

$$\dot{y}_{\max}(t)/y_{\max}(t) = \zeta \times Res(t), \qquad \zeta > 0,$$

with $\zeta \gtreqless 1$ depending on how well or how badly incipient individual productivity spills over towards the global frontier. Denote per capita income relative to the leading economy $x = y/y_{\max}$.

Theorem 4.1 *Suppose J is so large that laws of large numbers apply to averages. When the intra-distribution hazard is proportional to growth at the frontier, i.e.,*

$$\mathsf{h}(t) = \dot{y}_{max}(t)/y_{max}(t) \times \zeta^{-1}, \qquad \zeta > 0, \tag{4.16}$$

then the steady-state cross-country distribution F_X of relative incomes exists, is unique, and has the power-law form:

$$x \in [0,1]: \quad \mathsf{F}_X(x) = x^{1/\zeta} \implies \mathsf{f}_X(x) = \zeta^{-1}x^{\zeta^{-1}-1},$$

where the density f_X is rising or falling depending on $\zeta \lesseqgtr 1$.

The result on the shape of the long-run distribution is intuitive. When ζ exceeds 1, the frontier is growing rapidly relative to the intra-distribution leapfrogging. The high tail then grows thin and the pileup in the cross section then occurs at the other end of the distribution, where relative incomes are low. Conversely, when ζ is lower than 1, the frontier is growing relatively slowly, and economies leapfrogging to the high-income frontier, probabilistically, pile up there.

The cross-economy income distribution overall smoothly grows towards ever higher levels of income. Nonetheless, incomes relative to the frontier have an invariant long-run distribution. That long-run distribution displays ongoing churning, as within it economies leap-frog and criss-cross stochastically over time. The model displays unimodality in the cross-economy long-run distribution, but where the mode is at one or the other extreme. Allowing for two subpopulations—economies could stochastically switch between them—where in one group ζ exceeds 1, but in the other this inequality is reversed would produce a bimodal distribution, where the two modes appear at relative income extremes.

4.3 Underlying causes

Additional models.

> *Fiaschi and Lavezzi (2003) Feyrer (2001) Bloom, Canning, and Sevilla (2003) Azariadis and Drazen (1990) Azariadis and Stachurski (2003). Paul Johnson. Stachurski. Conditioning Quah (1997). Epstein, Howlett, and Schulze (2003). Solow (1997)*

5 Polarization and twin peaks

An important finding from the distribution-dynamics approach is *twin peaks* emergence (Jones, 1997; Pritchett, 1997; Quah, 1993a). Quah (1993b, 1996b) has argued that this empirical characterization is typically unavailable to standard cross-section or panel data regression although its critical features are certainly visible in regime-sensitive regression analyses such as in Durlauf and Johnson (1995). While the robustness of the twin-peaks finding has been criticized in, e.g., Kremer, Onatski, and Stock (2001) the alternative empirical analyses in Bianchi (1997) and Paap and van Dijk (1998) have confirmed its essential insights.

> *This section is empirical, using latest data from (Heston, Summers, and Aten, 2002). Show beginning and end point distributions. Show twinpeaks emergence from the stochastic kernel Figure 5.1. Polarization calculations.*

Figure 5.1 (Update with latest data) Relative income dynamics across 105 countries: Stochastic kernel

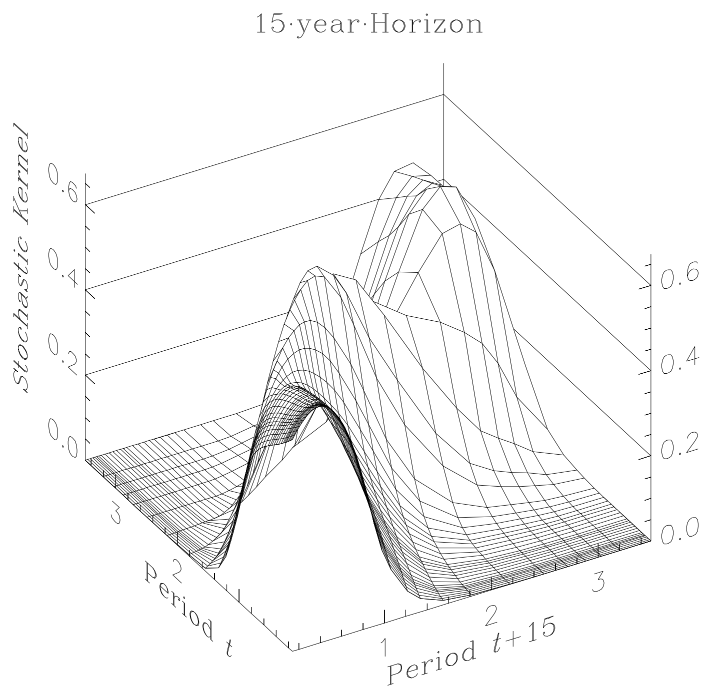
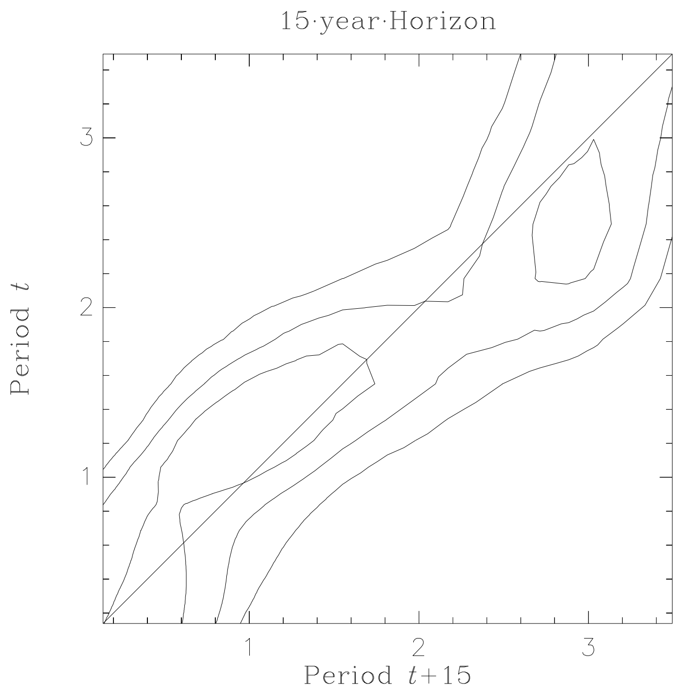


Figure 5.2 (Update with latest data) Relative income dynamics across 105 countries: Contour plot



Acceleration in spread of growth (Lucas, 2000 - Hybrid in Section 4.2)? Mobility - leapfrogging and overtaking

Some of this graphical and statistical evidence can also be re-expressed directly:

(i) Relative to the world average, the very poorest economies ...

(ii) Relative to the world average, the very richest economies ...

(iii) Relative to the world average, ...

(iv) Growth miracles ...

(v) Growth disasters ...

(vi) Africa ...

Bianchi (1997); Durlauf and Johnson (1995); Paap and van Dijk (1998) Zhu (2003) Pritchett (1997). Jones (1997) Jones, Pritchett. Esteban and Ray. Wolfson. Paul Johnson US states. Shorrocks. Geweke++

Is the country the right unit of observation? Re-draw state boundaries? The state of California; provinces of China. People? (But then income distribution across people too, so return in Section 6.)

6 Extensions of cross-section distribution dynamics

We have seen that models of distribution dynamics allow analyzing economic growth, convergence, and polarization in an integrative framework. But these models extend naturally to other domains as well, to wherever questions arise regarding cross-section economic performance. The relation between aggregate economic growth and cross-inhabitant inequality is one leading such application. We turn to this first application in 6.1–6.7. Another is spatial clustering: Economic inequality across geography is as profound as it is across people. What can we learn from models of explicit spatial distribution dynamics? We consider this second application in 6.7.

6.1 Inequality and growth

A putative tradeoff between equality and efficiency (Okun, 1975) is the kind of issue in economics so enduring that it gets discussed even in introductory textbooks. The view is that reducing inequality might come only at a cost of impairing efficiency; if so then the socially-optimal point might involve operating below economic efficiency. A related argument manifests in the conjecture that opening up economies to trade might well improve economic growth but at the cost of worsening inequality.

Banerjee and Duflo (2003). Durlauf (1996). Dollar and Kraay (2002). Son (2004).

Kuznets, Barro. Martin Ravallion Bertola monograph

Kuznets (1955) early on documented how inequality first rises and then falls across economies at progressively higher levels of income and development. Since then, theoretical and empirical research has sought to explain, confirm, or reject this so-called *Kuznets curve*. Researchers have attempted to identify causal mechanisms, either running from individual inequality to aggregate economic growth, from economic growth to inequality, or simultaneously in both directions.

The literature is large. The important contributions of Aghion, Caroli, and García-Peñalosa (1999) and Bénabou (1996) survey many of the central issues and key models. Different political economy mechanisms for the relation between inequality and growth are developed in Alesina and Rodrik (1994) and Persson and Tabellini (1994).

Galor and Zeira (1993) study capital market imperfections linking inequality and economic growth. Empirical studies on causal mechanisms between growth and inequality include Banerjee and Duflo (2003), Barro (2000), Deininger and Squire (1998), Forbes (2000), Li, Squire, and Zou (1998), Perotti (1996), among others.

6.2 Value and distribution

No single clear picture emerges from the many empirical studies. Had one done so then we might be more confident on the causal relationship between growth and inequality. But, regardless, assessing more generally the quantitative significance of these studies is not always transparent or direct. For example, if inequality were truly an important driver of economic growth, it would necessarily be only one of many. Or, conversely, if aggregate economic growth did fundamentally worsen income inequality, many other things likely do so as well. No matter how strong a link one is able to show between inequality and growth, a complete picture remains to be drawn of how the joint dynamics in these two variables matter for economic welfare more generally.

One way to flesh out such a picture is to explicitly trace out the dynamics of the (within-economy) income distribution. Doing so allows quantifying a range of key characteristics of growing unequal economies. As we will see the framework extends generally, but in this section we confine attention to two such characteristics. First, the dynamics of poverty, as also previously considered in Quah (2003); Sala-i-Martin (2002b) and second, the dynamics of utility.

Take as given an unequal but evolving economy $\{ \mathsf{F}_t : t \geq 0 \}$, where F_t is the distribution of income or consumption or their logs across the people living there. In practice, whether a researcher uses income or consumption in the calculation matters importantly (e.g., Blundell and Preston, 1998), but in the analytical description it is convenient to be able to refer to either one interchangeably. Since the economy is held fixed throughout this discussion, j now indexes individual agents where explicitly needed. To save notation, omit the t subscript unless again explicitly needed.

Whereas previously we had used y to denote per capita income, to emphasize the change in focus from economy-wide to individual characteristics, write per capita income (or consumption) now as $\mathcal{E}$ instead. This is standard notation for the *mean* of a distribution

$$\mathcal{E} = \int x \, d\mathsf{F}(x).$$

Hereafter, F is always taken to have finite and positive mean $\mathcal{E} \in (0, \infty)$. Record also the expectation of the logarithm,

$$\mathcal{E}_\ell \stackrel{\text{def}}{=} \int \log x \, d\mathsf{F}(x).$$

Growth is now $\dot{\mathcal{E}}/\mathcal{E}$.

Let $\mathcal{I}$ denote a vector of inequality measures. This vector might include the Gini coefficient

$$\mathcal{I}_G = \frac{1}{2}\,\mathcal{E}^{-1} \times \int_0^\infty \int_0^\infty |x - x'| \, d\mathsf{F}(x') d\mathsf{F}(x), \tag{6.1}$$

the Theil index

$$\mathcal{I}_{\mathrm{Th}} = \mathcal{E}^{-1} \int (\log x - \log \mathcal{E}) \, x \, d\mathsf{F}(x),$$

the Generalized Entropy index

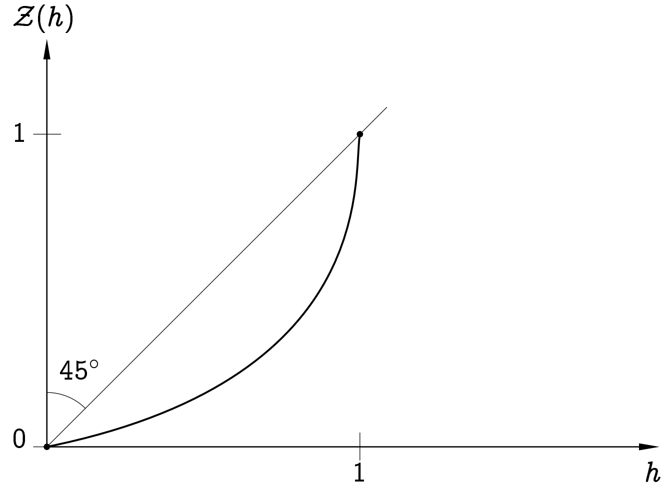
$$\mathcal{I}_{\mathrm{GE},\alpha} = (\alpha^2 - \alpha)^{-1} \int [\,(x/\mathcal{E})^\alpha - 1\,] \; d\mathsf{F}(x), \quad \alpha \in \mathbb{R},$$

the log standard deviation

$$\mathcal{I}_\ell = \left[\int (\log x - \mathcal{E}_\ell)^2 \; d\mathsf{F}(x) \right]^{1/2},$$

the mean-median ratio, the interquartile range, the ratio between the 90th and 10th percentiles, and so on. Although their details differ

Figure 6.1 The Lorenz curve plots $\mathcal{Z}$—the share of total income or consumption—against the population share.



and a large literature surrounds them (see, e.g., the masterful survey in Cowell, 2000), here it suffices to note that all inequality indexes assess the spread of the cross-section distribution. Some indexes take values bounded between 0 and 1; yet others need not.

The Gini coefficient $\mathcal{I}_G$ is likely the inequality index most used in the empirical literature. From the definition (6.1), $\mathcal{I}_G$ is the mean income gap in the population, normalized by mean income.

A different description of inequality is given in the Lorenz curve. As the name suggests the Lorenz curve does not seek to collapse all the information in an income distribution into just a single, scalar characteristic. Instead, the *Lorenz curve* $\mathcal{Z}$ is a non-decreasing function taking values between 0 and 1, that is the following transformation of an income or consumption distribution:

$$\mathcal{Z}(h) = \mathcal{E}^{-1}\left[\sup_{\mathrm{F}(x^{\dagger})\leq h}\int_0^{x^{\dagger}} x\, d\mathrm{F}(x)\right], \quad h \in [0,1]. \tag{6.2}$$

In words, the Lorenz curve is the share $\mathcal{Z}(h)$ of total income or consumption going to the poorest share h of the population. Thus its graph (e.g., Figure 6.1) is always bounded between the horizontal axis and the 45-degree line,

$$0 \leq \mathcal{Z}(h) \leq h, \qquad \forall h \in [0, 1].$$

As a function the Lorenz curve can also be regarded (Stuart and Ord, 1993, 2.25) as the *incomplete first moment* or the *first moment distribution* of F.

Theorem 6.1 *If* F *on* $[0, \infty)$ *admits a positive finite mean* $\mathcal{E}$ *and a density that is always positive then the Lorenz curve* $\mathcal{Z}(h)$, $h \in [0, 1]$,

(i) is strictly convex from below;

and at $h = \mathsf{F}(\mathcal{E})$*:*

(ii) reaches its maximum horizontal separation from the 45-degree line through the origin; and

(iii) has $\mathcal{Z}'(h) = 1$.

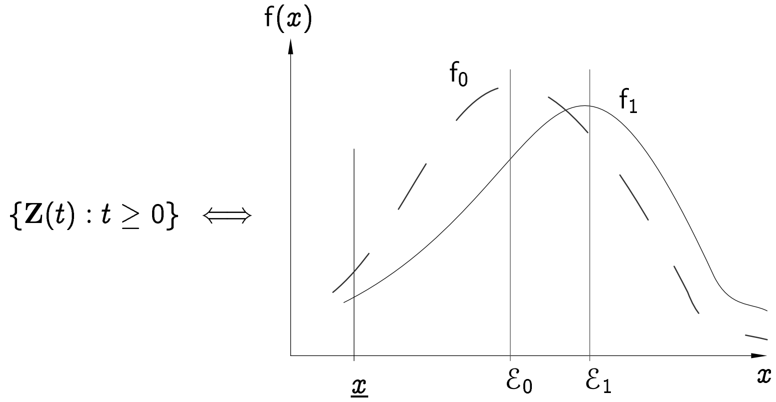
When distribution F divides society's income equally across the population then $\mathcal{Z}$ is the 45-degree line through the orign. When, in contrast, most of society's income is concentrated in only a small part of the population, the opposite occurs and $\mathcal{Z}$ strays further from the 45-degree line.

Theorem 6.2 *The Gini coefficient* $\mathcal{I}_G$ *can be found from the Lorenz curve* $\mathcal{Z}$ *by*

$$\mathcal{I}_G = 2 \times \int_0^1 [h - \mathcal{Z}(h)]\, dh. \tag{6.3}$$

Whereas equation (6.1) describes the Gini coefficient as a normalized mean income gap, Theorem 6.2 gives the Gini coefficient as twice the area between the 45-degree line and the Lorenz curve.

Figure 6.2 The dynamics in $\mathbf{Z}$ map into those in the distribution F_t, represented here by the evolving density f shown at two epochs t_0 and t_1. The horizontal axis shows mean incomes (or consumption) $\mathcal{E}_0$ and $\mathcal{E}_1$ for the densities f_0 and f_1 respectively, together with a fixed threshold income or consumption level $\underline{x}$, to be used later.



Research analyzing hypothesized causal mechanisms between inequality and growth seeks to uncover and place structural interpretation on functions ϕ, ψ, where

$$\frac{\dot{\mathcal{E}}}{\mathcal{E}} = \phi(\mathcal{I}) \quad \text{or} \quad \left(\mathcal{I}\right) = \psi\left(\dot{\mathcal{E}}/\mathcal{E}\right).$$

These would be read, respectively, as inequality causing economic growth or vice versa, conditional on auxiliary variables. (The dates on the different variables would be whatever are appropriate for a given economic model or hypothesis.)

Stack $\mathcal{E}$, $\dot{\mathcal{E}}/\mathcal{E}$, $\mathcal{I}$, and other relevant variables $\mathbf{Z}_0'$ into a vector $\mathbf{Z}$. The dynamics of the distribution F_t implies dynamics in the vector $\mathbf{Z}$. Conversely, when $\mathbf{Z}$ is sufficiently informative, then its dynamics in turn implies F_t-distribution dynamics. Figure 6.2 shows the logic generally, and Example 6.3 describes an explicit special case.

Example 6.3 *If distribution* F *is Pareto Type-1 with parameters* $\theta = (\theta_1, \theta_2)$*, i.e.,*

$$\mathsf{F}(x) = 1 - (\theta_1 x^{-1})^{\theta_2}, \quad \theta_1 > 0,\ x \geq \theta_1,\ \theta_2 > 1,$$

then

$$\begin{aligned} \mathcal{E} &= (\theta_2 - 1)^{-1}\theta_2\theta_1, \\ \mathcal{I}_G &= (2\theta_2 - 1)^{-1} \end{aligned}$$

(having additional inequality indexes here would over-identify the model). Inverting these equations allows constructing the mapping from $\mathbf{Z}$ *to* F*,*

$$\begin{aligned} \theta_2 &= (1 + \mathcal{I}_G^{-1})/2, \\ \theta_1 &= (1 - \hat{\theta}_2^{-1})\mathcal{E}, \end{aligned}$$

with dynamics

$$\begin{aligned} \dot{\mathcal{E}}/\mathcal{E} &= \frac{\dot{\theta}_1}{\theta_1} - (\theta_2 - 1)^{-1}\frac{\dot{\theta}_2}{\theta_2}, \\ \dot{\mathcal{I}}_G/\mathcal{I}_G &= \left(\frac{-2\theta_2}{2\theta_2 - 1}\right)\frac{\dot{\theta}_2}{\theta_2}. \end{aligned}$$

These dynamics, however, do not give stochastic kernels: the latter contain information on mobility within the distribution, not just on changes in the distribution's external features as described by the mean and the Gini inequality index.

Whatever the causal mechanism relating aggregate growth and cross-sectional inequality, calibrating the dynamics in F_t to historical patterns, as depicted in Figure 6.2 can shed light on how growth and distribution matter. If, for instance, higher growth is possible only with greater inequality, we can ask, How much increase in inequality is tolerable for a given increase in the rate of economic growth? How have growth and inequality dynamics affected the number of people who are very poor? How have growth and inequality dynamics affected the utility of agents at different parts of the income distribution?

Such questions can be addressed without restricting the discussion to any one specific mechanism relating growth and inequality. The general framework here can thus be usefully added to any theoretical or empirical analysis of a hypothesized causal mechanism.

6.3 Dynamics of Poverty

Growth and inequality differently affect the fraction of the population living below a given threshold income level. Given a fixed shape in the cross-section income distribution, growth on average raises everyone's income and thus reduces the fraction of the population living in poverty. By contrast, an increase in the spread of the distribution while the average is unchanged—a rise in inequality with no economic growth—increases the proportion of individuals in poverty. Holding fixed both growth and inequality but increasing the population size raises proportionally the number of the poor.

Thus, a joint law of motion in the population and distribution dynamics will also mechanically give the dynamics of poverty. The impact of growth, inequality, or population shocks on poverty is a statement on the elasticity of poverty with respect to those variables.

Atkinson (1987).

Before investigating further that law of motion, Table 6.1 provides some data on the numbers and geographical distribution of the poor worldwide.[22] The first column of numbers shows the number of people, across continental regions, living on less than US$1 a day

[22] Such numbers are subject to considerable debate and controversy. See, e.g., Bourguignon and Morrisson (2002) and Deaton (2005). Nonetheless, *some* numbers need to be first put down to clarify what's at issue, even or especially if they are not to be taken as the final word on the matter. Such figures can then be subsequently refined and improved as better information becomes available.

Table 6.1 Geographical distribution of the world's poor, 2001. Source: Chen and Ravallion (2004a) and author's calculations

	$< \$1$/day, 10^6/%	Popl., 10^6/%	Conc.
South Asia	**428/39.2**	**1369/26.5**	**1.5**
excl. India	70/6.4	335/6.5	1.0
India	359/32.8	1033/20.0	1.6
Sub-Sah. Africa	**316/28.9**	**673/13.1**	**2.2**
East Asia	**271/24.8**	**1821/35.3**	**0.7**
excl. China	60/5.5	546/10.6	0.5
China	212/19.4	1275/24.7	0.8
Latin America	**50/4.6**	**524/10.2**	**0.4**
EECA	**18/1.6**	**476/9.2**	**0.2**
MENA	**7/0.6**	**296/5.7**	**0.1**
Total:	1093/	5159/	

[...]

To formalize this, let $\underline{x}$ denote, as in Figure 6.2, a fixed threshold level of income or consumption. Call the poor those with x at level $\underline{x}$ or lower. Define, for just this discussion, the indicator function

$$\phi(x) = \begin{cases} 1 & \text{if } x \leq \underline{x}; \\ 0 & \text{otherwise.} \end{cases}$$

In the inequality literature the fraction of the population that is poor given a particular $\underline{x}$ is sometimes known as the *Poverty Headcount Index*. Denote this as

$$HC_{\underline{x}} \stackrel{\text{def}}{=} \int_0^{\underline{x}} d\mathsf{F}(x) = \int_0^{\infty} \phi(x)\, d\mathsf{F}(x). \tag{6.4}$$

Combining the stochastic kernel and definition (6.4), the expected Poverty Headcount Index at time $t+s$, conditional on time t, is

$$E\left(HC_{\underline{x}}(t+s) \mid \mathsf{F}_t\right) = \int_0^{\infty} \left(\mathfrak{T}_{t,s}\phi\right)(x)\, d\mathsf{F}_t(x). \tag{6.5}$$

How growth and inequality dynamics $(\mathcal{T}_{t,s}, \mathsf{F}_t)$ affect the poverty headcount can be explicitly traced through equation (6.5).

Although Example 6.4 is a special case of Example 6.5 to follow, it illustrates the ideas being considered especially transparently, and so is useful to go through.

Example 6.4 *Suppose the cross-section distribution has timepath*

$$\{\, \mathsf{F}_t : t \geq 0 \,\}$$

matching that of the exponential of Brownian motion, σB, where σ is a positive constant. That is,

$$\mathsf{F}_t(x) = \text{Prob}\,\{\, \exp\,(\, Z(t)\,) = X(t) \leq x \,\}$$

with $Z(t) = \sigma B(t)$. Hypothesizing some initial condition $X(0) = x = \exp(z)$, the distribution F_t is lognormal with

$$\begin{aligned} \mathcal{I}_\ell(\mathsf{F}_t) &= \sigma\sqrt{t}; \\ \mathcal{E}(\mathsf{F}_t) &= \exp\left[\log x + \mathcal{I}_\ell^2/2\right]; \\ \mathcal{I}_G(\mathsf{F}_t) &= 2 \times \mathsf{F}_{\mathsf{N}(0,1)}\left(\mathcal{I}_\ell/\sqrt{2}\right) - 1\,; \end{aligned}$$

where $\mathsf{F}_{\mathsf{N}(0,1)}$ denotes the standard normal cdf. The transition density function, from equation (3.15), is

$$p_t\left(x, x^\dagger\right) = \left(2\pi\,\mathcal{I}_\ell^2\right)^{-1/2} \times \exp\left[-\frac{1}{2\,\mathcal{I}_\ell^2}\left(\log x^\dagger - \log x\right)^2\right].$$

Growth is

$$\dot{\mathcal{E}}/\mathcal{E} = \mathcal{I}_\ell \frac{d\mathcal{I}_\ell}{dt} = \frac{1}{2}\sigma^2 > 0,$$

where, from Jensen's inequality, the growth rate of the average $\mathcal{E}$ differs from 0, the average of the individual growth rates. Over time, as $t \to \infty$, inequality rises as both $\mathcal{I}_\ell \uparrow \infty$ and $\mathcal{I}_G \uparrow 1$. The expected

Poverty Headcount Index, equation (6.5), *is*

$$\begin{aligned} E\left(HC_{\underline{x}}(t) \mid \mathsf{F}_0\right) &= \int_0^\infty (\mathcal{T}_t\phi)(x)\, d\mathsf{F}_0(x) \\ &= \mathsf{F}_{\mathsf{N}(0,1)}\left(\frac{\log \underline{x} - \log X(0)}{\sigma\sqrt{t}}\right) \\ &\to \frac{1}{2} \quad \text{as } t \to \infty. \end{aligned}$$

Even though measured growth is positive, the increase in inequality is so rapid that the numbers in poverty are always expected to remain substantial.

Example 6.4 shows a situation where, despite positive growth, a large fraction of the population always remains poor. In general, however, a nontrivial tradeoff between growth and inequality allows precise statements about the Poverty Headcount only after quantifying the different, opposing effects.

Example 6.5 *Suppose the cross-section distribution has timepath*

$$\{\,\mathsf{F}_t : t \geq 0\,\}$$

matching that of the exponential of a drifted Ornstein-Uhlenbeck process. That is,

$$\mathsf{F}_t(x) = \text{Prob}\,\{\,\exp\left(\,Z(t) + \xi t\,\right) = X(t) \leq x\,\} \tag{6.6}$$

with

$$dZ = -\mu Z\, dt + \sigma\, dB, \tag{6.7}$$

where, as in section 3.1, the process B is standard Brownian motion; (ξ, μ, σ) are constants satisfying $\mu > 0$, $\sigma > 0$; and equation (6.7) *is a stochastic differential equation whose solution $\{\,Z(t) : t \geq 0\,\}$ is an Itô integral (e.g., Karlin and Taylor, 1981, Ch. 15, pp. 340–377). Hypothesizing some initial condition $X(0) = x = \exp(z)$ the solution* *to* (6.7) *implies that distribution* F_t *is lognormal with*

$$\begin{aligned}
\mathcal{I}_\ell(\mathsf{F}_t) &= \left[\left(1 - e^{-2\mu t} \right) \frac{\sigma^2}{2\mu} \right]^{1/2} ; \\
\mathcal{E}(\mathsf{F}_t) &= \exp\left[e^{-\mu t} \log X(0) + \mathcal{I}_\ell^2/2 \right] \times e^{\xi t} ; \\
\mathcal{I}_G(\mathsf{F}_t) &= 2 \times \mathsf{F}_{\mathsf{N}(0,1)} \left(\mathcal{I}_\ell/\sqrt{2} \right) - 1 .
\end{aligned}$$

The transition density function, from equation (3.15), *is*

$$\begin{aligned}
p_t\left(x, x^\dagger\right) = \left(2\pi\, \mathcal{I}_\ell^2\right)^{-1/2} &\times \frac{e^{\xi t}}{x^\dagger} \\
&\times \exp\left[-\frac{1}{2\,\mathcal{I}_\ell^2} \left(\log x^\dagger - \xi t - e^{-\mu t} \log x \right)^2 \right] .
\end{aligned}$$

Thus, growth is

$$\dot{\mathcal{E}}/\mathcal{E} = \xi + \left[-\mu e^{-\mu t} \log x + e^{-2\mu t} \sigma^2/2 \right] ,$$

where, again because of Jensen's inequality, the growth rate of the average $\mathcal{E}$ *differs from* ξ *the average of the growth rates. As* $t \to \infty$ *we have long-run growth and inequality described by:*

$$\begin{gathered}
\mathcal{I}_\ell \to \sqrt{\sigma^2/2\mu} ; \qquad \mathcal{I}_G \to 2 \times \mathsf{F}_{\mathsf{N}(0,1)} \left(\frac{1}{2} \sqrt{\sigma^2/\mu} \right) - 1 ; \\
\dot{\mathcal{E}}/\mathcal{E} \to \xi ; \quad \text{and} \quad \left| \mathcal{E} - \frac{\sigma^2}{4\mu} e^{\xi t} \right| \to 0 .
\end{gathered}$$

When scaled by $e^{-\xi t}$ *the limit distribution remains log-normal. The dynamics of the expected Poverty Headcount Index, equation* (6.5), *can then be explicitly given:*

$$\begin{aligned}
E\left(HC_{\underline{x}}(t) \mid \mathsf{F}_0\right) &= \int_0^\infty \left(\mathcal{T}_t \phi\right)(x)\, d\mathsf{F}_0(x) \\
&= \mathsf{F}_{\mathsf{N}(0,1)} \left(\frac{\log \underline{x} - \xi t - e^{-\mu t} \log X(0)}{\sqrt{(1 - e^{-2\mu t}) \sigma^2/2\mu}} \right) .
\end{aligned}$$

Depending on the values of ξ and μ, among others, the Poverty Headcount can be expected to rise or fall in the short run. However, provided $\xi > 0$, i.e., provided growth occurs in the long run, eventually all else gets swamped, and economic growth alone drives the poor to an ever smaller segment of the population.

In Example 6.5 average income can rise or fall in the short-term but is expected eventually to grow at a constant rate ξ. However, inequality—whether measured by $\mathcal{I}_G$ or $\mathcal{I}_\ell$—only ever rises, independent of the initial condition $X(0)$. Even though the Poverty Headcount remains free to rise or fall, the feature that inequality always increases is obviously overly restrictive,

A useful way to enrich the model, while preserving the tractability of the Ornstein-Uhlenbeck mechanism, is to allow the evolving distribution F_t not to be simply the distributions of a scalar (6.6)–(6.7), but to be instead a convolution of a mixture distribution (across $X(0)$'s, say) together with that in (6.6)–(6.7). A special case of this preserves the lognormality in F_t but frees up its mean and variance parameters so that these latter evolve through time less restrainedly:

Example 6.6 *Suppose that* $\{\theta_1(t), \theta_2(t) : t \geq 0\}$ *has* $\theta_2(t) > 0$, *and each* F_t *has density*

$$\mathsf{f}_t(x) = (2\pi\theta_2(t))^{-1/2} \cdot x^{-1} \times \exp\left\{ \frac{-(\log x - \theta_1(t))^2}{2\theta_2(t)} \right\}, \quad \textit{for } x > 0.$$

Then

$$\begin{aligned} \mathcal{I}_\ell(\mathsf{F}_t) &= \theta_2(t)^{1/2}; \\ \mathcal{E}(\mathsf{F}_t) &= \exp(\theta_1(t) + \theta_2(t)/2); \\ \mathcal{I}_G(\mathsf{F}_t) &= 2 \times \mathsf{F}_{\mathsf{N}(0,1)}\left(\theta_2(t)^{1/2}/\sqrt{2}\right) - 1. \end{aligned}$$

The expected Poverty Headcount Index is

$$E\left[HC_{\underline{x}}(t) \mid \mathsf{F}_0\right] = E\left[\mathsf{F}_{\mathsf{N}(0,1)}\left((\log \underline{x} - \theta_1(t))/\sqrt{\theta_2(t)} \right) \mid \mathsf{F}_0\right].$$

Inequality, whether measured by $\mathcal{I}_\ell$ or $\mathcal{I}_G$ can now rise or fall through time; so too the Poverty Headcount. Depending on the relative magnitudes of the underlying changes in θ_1 and θ_2, a wide range of behavior in growth and inequality can be captured. By contrast, the Ornstein-Uhlenbeck model of Example 6.5 restricted $\theta_1(t) = e^{-\mu t}\log X(0)$ and $\theta_2(t) = (1 - e^{-2\mu t})\sigma^2/2\mu$.

In Example 6.6 a law of motion for the hyper-parameters $\theta(t)$ would imply restrictions on the transition operator for $\{\mathsf{F}_t : t \geq 0\}$, but would typically not uniquely identify that operator. Despite this lack of identification, the Poverty Headcount Index can, obviously, be computed straightforwardly. The following Example illustrates the same point but is also interesting in its own right.

Example 6.7 *Suppose that $\{\theta_1(t), \theta_2(t) : t \geq 0\}$ has $\theta_1(t) > 0$ and $\theta_2(t) > 1$ and*

$$\mathsf{F}_t(x) = \begin{cases} 1 - (\theta_1(t)/x)^{\theta_2(t)}, & \text{for } x \geq \theta_1(t); \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\begin{aligned} \mathcal{E}(\mathsf{F}_t) &= (\theta_2(t) - 1)^{-1}\,\theta_2(t)\theta_1(t); \\ \mathcal{I}_G(\mathsf{F}_t) &= (2\theta_2(t) - 1)^{-1}. \end{aligned}$$

Measured growth is

$$\dot{\mathcal{E}}/\mathcal{E} = \frac{\dot{\theta}_1}{\theta_1} - (\theta_2 - 1)^{-1}\,\frac{\dot{\theta}_2}{\theta_2},$$

while the expected Poverty Headcount Index is

$$E\left[HC_{\underline{x}}(t) \mid \mathsf{F}_0\right] = E\left[1 - (\theta_1(t)/\underline{x})^{\theta_2(t)} \mid \theta_1(0), \theta_2(0)\right].$$

Using growth and distribution data, Quah (2003) has estimated versions of Examples 6.6 and 6.7 for China and India. He concludes that given the historical experience over 1980–1992 it is growth that has dominated the evolution of Poverty Headcounts, even with the rapid rise of inequality in China over this period. For instance, ...

The advantage of ....

Using related ideas Sala-i-Martin (2002a,b) provide .... Deaton (2005). Chen and Ravallion (2001, 2004b).

To continue this discussion we broaden the scope of how growth and inequality can affect a population. Instead of focusing on only the fraction of the population that remaining poor, we ask instead how welfare varies over time and across the cross section. To do this we hypothesize dynamic utility functions for each person, and then confront that individual with different dynamically evolving distribution processes in income or consumption. We can then calibrate the relative importance of growth and inequality components in affecting welfare (Quah, 2007).

6.4 Dynamics of Utility

Assume time is continuous and agents live for t in $[0, \infty)$. (The previous sections have developed a continuous-time stochastic process machinery so it is convenient to preserve that structure here. But a further reason for a continuous-time analysis is that in discrete time, sudden large changes in consumption—disasters or takeoffs—are subtle to describe, whereas in continuous time they are simply discontinuities in timepaths.)

Write $C_j(t)$ for the consumption of individual j at time t. Suppose that the long-run conditional expectation

$$\xi = \lim_{t \to \infty} E\left[\log(C_j(t)/C_j(0)) \times t^{-1} \mid C_j(0) \right] \tag{6.8}$$

is invariant across j, i.e., that the economy has some constant long-run proportional growth rate in consumption, the same for everyone. For example, in balanced growth steady state for a Solow growth model this constant would be the rate of technical progress, hence the notation ξ on the left side of (6.8). More generally, ξ is the aggregate growth rate of the economy.

Write individual consumption fluctuations as j-indexed stochastic processes:

$$Z_j(t) \stackrel{\text{def}}{=} \log\left[C_j(t)e^{-\xi t}\right] \tag{6.9}$$

so that we can write

$$C_j(t) = e^{Z_j(t)}e^{\xi t}.$$

Collection Z describes consumption differences across people. Thus ξ represents aggregate economic growth and Z inequality. Strictly, the collection Z of course characterizes everything about the cross-section distribution but referring to it as simply inequality denotes better the analysis's intent. If doing so misleads, at least it does so in an innocuous way.

Equations (6.8)–(6.9) define Z by using ξ to transform C. They do not assume inequality and growth are functionally independent—both Z and ξ are induced from a single underlying mechanism for C. Moreover, while these equations restrict Z's trend, they do not require that Z be stationary with bounded variance. For instance, Z might be Brownian motion independent across j, and inequality would then be always increasing through time (as in Example 6.4) even as consumption obeys the trend growth in equation (6.8).

The utility experienced by an agent in a growing but unequal society varies with that agent's consumption timepath, but not that of anyone else. Thus, an agent's utility does not depend on aggregate social indicators such as the shape, dispersion, or other characterisitics of the economy-wide consumption distribution.[23]

For concreteness assume that the agent's utility, conditional on

[23] This individualistic restriction is consistent with the *social welfare function* approach to income inequality—see, e.g., Cowell (1995, Ch. 3).

time-t information, is

$$W_j(t) = E_t \left[\int_t^\infty e^{-(s-t)\rho} U(C_j(s))\, ds \right], \qquad \rho > 0, \qquad (6.10)$$

where the positive discount rate ρ will be required, in Theorem 6.8, to be sufficiently large. Say that the instantaneous utility function U shows constant relative risk aversion (CRRA) when

$$U(c) = \frac{c^{1-R} - 1}{1 - R}, \qquad R > 0. \qquad (6.11)$$

In social welfare analysis of income inequality the coefficient R is known as the *inequality aversion parameter* (see, e.g., Cowell, 1995, Ch. 3).

We seek to quantify how different patterns of growth and inequality (ξ, Z) affect welfare (6.10). The calculation is similar to that in the analysis of the relative importance of growth and business cycles in Lucas (1981, Ch. 3) but with two significant differences. One, the economic interpretation: we treat Z as inequality rather than business cycles. Two, the model here evolves in continuous rather than discrete time. While this second feature might seem merely technical—drawing on the resolvent operators of Section 3.5—it allows explicitly modelling events such as sharp discrete takeoffs and growth disasters. Such behaviour manifest as discontinuities in the timepath whereas they would be unavailable in a discrete-time framework. Features like these matter more for growth and inequality than for business cycles.

Since C_j is Markov, so too is Z_j. Hereafter, except where doing so causes ambiguity, the j subscript will be suppressed. Assume Z has time-homogeneous transitions, with the associated transition semigroup $\{ \mathcal{T}_t \}$ and resolvent $\mathbf{R}_\lambda$ (Section 3.5). Lifetime utility can then be expressed in the following convenient form:

Theorem 6.8 *Let consumption*

$$\{ C(t) : t \geq 0 \},$$

obeying (6.8), *be Markov with time-homogeneous transitions, and let lifetime utility* (6.10) *have* U *CRRA, as in* (6.11). *Assume* $\rho > \max\{0, (1-R)\xi\}$. *Define*

$$\widetilde{\rho} \stackrel{\text{def}}{=} \rho - (1-R)\xi;$$

$$\widetilde{U}(z) \stackrel{\text{def}}{=} \begin{cases} (1-R)^{-1}e^{(1-R)z} & \text{for } R \neq 1; \\ z & \text{otherwise;} \end{cases}$$

and denote by $\mathbf{R}_\lambda$ *the resolvent for the inequality process* $Z(t) = C(t)e^{-\xi t}$. *Then conditional on* $C(0)$,

$$W(0) = (\mathbf{R}_{\widetilde{\rho}}\widetilde{U})(\log C(0)) + \begin{cases} -(1-R)^{-1}\rho^{-1} & \text{for } R \neq 1; \\ \xi\rho^{-2} & \text{otherwise.} \end{cases}$$

In words, up to an additive constant, lifetime utility equals the resolvent operator for growth-discounted consumption $Z(t) = C(t)e^{-\xi t}$, evaluated at a modified instantaneous utility function $\widetilde{U}$ and for the resolvent λ set to a suitably adjusted discount rate $\widetilde{\rho}$. The resolvent expression $\mathbf{R}_{\widetilde{\rho}}\widetilde{U}$ can be either explicitly calculated or conveniently characterized for a range of inequality processes.

Theorem 6.9 *Assume the hypotheses of Theorem 6.8 and let inequality* Z *be Brownian Motion with variance parameter* $\sigma^2 > 0$, *i.e.,* $Z(t) = \sigma B(t)$. *Define the function* $r : \mathbb{R}^4 \to \mathbb{R}$ *by*

$$r(R, \rho, \xi, \sigma) = (1-R) \times \left[\rho - (1-R)\xi - \frac{1}{2}(1-R)^2\sigma^2\right]. \quad (6.12)$$

Provided $\widetilde{\rho} > \frac{1}{2}(1-R)^2\sigma^2$ *then*

$$(\mathbf{R}_{\widetilde{\rho}}\widetilde{U})(z) = \begin{cases} r(R, \rho, \xi, \sigma)^{-1} \times e^{(1-R)z} & \text{for } R \neq 1; \\ \rho^{-1}z & \text{otherwise.} \end{cases}$$

Discontinuities in the consumption timepath, to capture sudden discrete upwards jumps or conversely economic disasters, can be modelled using Markov chains. The specialization of Theorem 6.8 can be stated as the following.

Theorem 6.10 *Assume the hypotheses of Theorem 6.8 and let inequality Z be a point-discrete Markov chain $(M, \overline{z}, P, \overline{\mathfrak{f}}_0)$ with states*

$$\overline{z}_0, \overline{z}_1, \overline{z}_3, \ldots, \overline{z}_{M-1}, \quad \overline{z}_m < \overline{z}_{m+1};$$

and time-homogeneous transition probability matrices

$$\{ P_{0,t} = P_t : t \geq 0 \} .$$

Let $\mathbf{G}$ be the infinitesimal generator of the semigroup P,

$$\mathbf{G} = \lim_{t \downarrow 0} \frac{P_t - I}{t},$$

and define the vector

$$\overline{U} = \left(\widetilde{U}(\overline{z}_0) \quad \widetilde{U}(\overline{z}_1) \quad \cdots \quad \widetilde{U}(\overline{z}_{M-1}) \right)^{\top} .$$

Then

$$(\mathbf{R}_{\tilde{\rho}} \widetilde{U})(\overline{z}_m) = \left[(\tilde{\rho} - \mathbf{G})^{-1} \overline{U} \right] (m), \quad m = 0, 1, \ldots, M-1.$$

Theorem 6.10 allows calculating the utility for an agent facing a Markov chain process in inequality using nothing more than matrix inversion.

Combining Theorems 6.8–6.9 give that when Z is Brownian motion and $\log C(0) = Z(0) = z$ then

$$W(0) = \begin{cases} r(R, \rho, \xi, \sigma)^{-1} \times e^{(1-R)z} - (1-R)^{-1} \rho^{-1} & \text{for } R \neq 1; \\ \rho^{-1} z + \xi \rho^{-2} & \text{otherwise} \end{cases} \tag{6.13}$$

with multiplier r from (6.12).

(For completeness and to aid intuition, an appendix, Section 8, contains counterpart calculations for when time is discrete.)

The Table $\cdots$.

Benabou, mobility.

6.5 Dynamics of World Income Inequality

Although the bias implicit in our analysis has been to work with the distributions directly, the discussion in this section connects also with the extremely large literature on within-country income inequality.

The question considered here is, How much of world income inequality is due to inequality across countries, and how much to inequality across people within a given country?

Let Y denote

Decomposition: Inequality between and within countries. (Rewrite extant decompositions in Lebesgue-Stieltjes integral form.) Generalized Entropy indexes. Bourguignon (1979). Shorrocks (1980). Cowell (1995). Sala-i-Martin (2002a,b).

6.6 Polarization

The.

Anderson (2004)

Esteban and Ray (1994), Wolfson (1994).

Decomposition?

6.7 Spatial inequality and distribution

Spatial hazard models. Quah (2002). Quah and Simpson (2003). Duranton and Overman (2002). Duranton and Puga. Ioannides.

7 Conclusion

This.

8 Appendix: Discrete-time dynamics of utility

Discrete-time counterparts to the Section 6.4 calculations on welfare provide a direct check and further intuition to those continuous-time results.

For this section only, suppose time is discrete, $t = 0, 1, 2, \ldots$, and that preferences for agent j at time t are described by:

$$W_j(t) = E_t\left[\sum_{s=0}^{\infty} \delta^s U(C_j(s))\right], \qquad \delta \in (0,1), \tag{8.1}$$

with instantaneous utility U again CRRA as in (6.11),

$$U(c) = \frac{c^{1-R} - 1}{1 - R}, \qquad R > 0.$$

Suppose consumption is:

$$C_j(t) = Z_j(t)\xi^t, \qquad \xi \geq 1, \tag{8.2}$$

for each j, i.e., a trending, lognormally-distributed first-order autoregression, with the underlying growth rate ξ common across the population.

Let the cross-sectional differences $Z_j(t)$ satisfy:

$$Z_j(t) = \overline{z}_j \epsilon_j(t),$$

so that initial consumption conditional on the stochastic disturbance $\epsilon_j(0)$ is

$$C_j(0) = \overline{z}_j \epsilon_j(0).$$

Let ϵ be stationary and follow:

$$\log \epsilon_j(t) = -\frac{1}{2}(1+\alpha)^{-1}\sigma_j^2 + \alpha \log \epsilon_j(t-1) + \nu_j(t), \tag{8.3}$$

with

$$|\alpha| < 1 \quad \text{and} \quad \nu_j(t) \sim \text{iid } \mathrm{N}(0, \sigma_j^2). \tag{8.4}$$

Equations (8.3)–(8.4) imply ϵ_j is Markov. In the unconditional distribution

$$E \log \epsilon_j(t) = -\frac{1}{2}(1-\alpha^2)^{-1}\sigma_j^2$$

and

$$\text{Var} \log \epsilon_j(t) = (1-\alpha^2)^{-1}\sigma_j^2,$$

implying

$$\begin{aligned} E\epsilon_j(t) &= \exp\left[E \log \epsilon_j(t) + \frac{1}{2}\text{Var} \log \epsilon_j(t)\right] \\ &= \exp\left[-\frac{1}{2}(1-\alpha^2)^{-1}\sigma_j^2 + \frac{1}{2}(1-\alpha^2)^{-1}\sigma_j^2\right] = 1 \end{aligned}$$

and

$$\text{Var}\, \epsilon_j(t) = \exp\left[(1-\alpha^2)^{-1}\sigma_j^2\right] - 1,$$

increasing in σ_j^2.

Equation (8.2) then gives the unconditional expectation

$$EC_j(t) = \overline{z}_j \xi^t, \qquad t \geq 0,$$

i.e., $\overline{z}_j$ parametrizes the (unconditional) expected level of j's consumption profile, with the initial value $EC_j(0) = \overline{z}_j$.

More generally, $\epsilon_j(t)$ is conditionally lognormal:

$$\begin{aligned} &\log \epsilon_j(t_0+t) \mid \epsilon_j(t_0) \\ &\quad \sim \mathsf{N}\left(\alpha^t \log \epsilon_j(t_0) - \frac{1-\alpha^t}{1-\alpha^2}\frac{\sigma_j^2}{2},\ \frac{1-\alpha^t}{1-\alpha^2}\sigma_j^2\right) \text{ for any } t > 0. \end{aligned} \tag{8.5}$$

The value to (8.1) can then be directly calculated:

Theorem 8.1 *Assume utility has the form* (6.11) *and that consumption evolves from the combination of growth and inequality given in* (8.2)–(8.4). *There are two cases:*

(i) When $R = 1$*,*

$$W_j(0; \epsilon_j(0), \xi, \overline{z}_j, \sigma_j^2, \alpha) = (1 - \delta\alpha)^{-1} \log \epsilon_j(0)$$
$$+ \frac{\delta}{(1-\delta)^2} \log \xi + \frac{1}{1-\delta} \log \overline{z}_j - \frac{\delta}{(1-\delta)(1-\delta\alpha)(1+\alpha)} \times \frac{\sigma_j^2}{2}.$$

(ii) Otherwise,

$$W_j(0; \epsilon_j(0), \xi, \overline{z}_j, \sigma_j^2, \alpha) = (1 - R)^{-1} \left[-\frac{1}{1-\delta} + (\overline{z}_j)^{1-R} \times \right.$$
$$\left\{ \epsilon_j(0)^{1-R} + e^{-(1-R)R(1-\alpha^2)^{-1}\sigma_j^2/2} \times \right.$$
$$\left. \left. \sum_{t=1}^{\infty} \delta^t \xi^{-(R-1)t} e^{D_1 \alpha^t} e^{-D_2 \alpha^{2t}} \right\} \right],$$

where

$$D_1 = (1 - R) \left[\log \epsilon_j(0) + (1 - \alpha^2)^{-1} \frac{\sigma_j^2}{2} \right] \text{ and}$$
$$D_2 = (1 - R)^2 (1 - \alpha^2)^{-1} \frac{\sigma_j^2}{2} > 0.$$

(As with the other results, the proof to Theorem 8.1 is in Section 9.)

9 Technical Appendix

This Technical Appendix has two purposes: First, provides a brief reminder of mathematical details used in the text. Second, it contains proofs to all the results developed in the text. On the technical details, analysis and probability monographs such as Ash (1972) provide further background if the reader wishes to investigate more deeply.

Recall that a function ϕ has *variation on interval* $[a, b]$ defined to be $\sup_{\mathcal{P}} \sum_i |\phi(x_{i+1}) - \phi(x_i)|$ for $\mathcal{P}$ that are partitions of $[a, b]$ given by $\{ x_i : \text{integer } i \}$ where $x_i < x_{i+1}$. Variation assesses choppiness.

A function has *bounded variation* when it has finite variation on its domain.

A function ϕ is *absolutely continuous on* $[a, b]$ when for any positive ϵ there is δ such that for all positive integer M and all families (a_1, b_1), (a_2, b_2), $\ldots$, (a_M, b_M) of open disjoint subintervals of $[a, b]$, we have

$$\sum_{i=1}^{M} |b_i - a_i| < \delta \implies \sum_{i=1}^{M} |\phi(b_i) - \phi(a_i)| < \epsilon.$$

If a function is absolutely continuous, then it has bounded variation. Distinguish absolute continuity from uniform continuity, also used in Theorem 3.4 of Section 3.1.2 in the text. Function ϕ is *uniformly continuous* if for any positive ϵ there is δ such that

$$\forall x', x : \quad |x' - x| < \delta \implies |\phi(x') - \phi(x)| < \epsilon.$$

The crucial distinction in uniform continuity from (ordinary) continuity is that δ can be chosen independent of the particular x in ϕ's domain, i.e., the criterion above applies uniformly, for all x.

Some probability results useful for understanding the discussion of bandwidth choice surrounding Theorems 3.5 and 3.6 are routinely used in econometric theory but are perhaps less familiar in macroeconomics discussion. For completeness, I state them here.

Lemma 9.1 (c_r Inequality) *If for r positive the random variables Z_0 and Z_1 have finite r-th absolute moments, then*

$$E|Z_0 + Z_1|^r \leq c_r \times (E|Z_0|^r + E|Z_1|^r)$$

where c_r is 1 if $r \in (0, 1)$ and 2^{r-1} if $r \geq 1$.

The c_r inequality is discussed in Chung (1974, Ex. 12, p. 48) and White (1984, Prop. 3.8, p. 33).

Lemma 9.2 (Hölder Inequality) *If Z_0 and Z_1 are random variables and $p,\ q > 1$ such that $p^{-1} + q^{-1} = 1$, then*

$$|EZ_0 Z_1| \leq E|Z_0 Z_1| \leq (\,E|Z_0|^p\,)^{1/p}\,(\,E|Z_1|^q\,)^{1/q}\,.$$

See Chung (1974, p. 47) and White (1984, Prop. 3.4, p. 30). Setting $Z_1 = 1$, $|Z_0| = |Z|^r$ for $r > 0$, and $r' = rp$, the Hölder Inequality gives

$$(\, E|Z|^r \,)^{1/r} \leq (\, E|Z^{r'} \,)^{1/r'}, \qquad \text{for } 0 < r < r' < \infty$$

(a special case that is also known as the *Liapounov Inequality*).

The limiting distribution result Theorem 3.5 draws on a central limit theorem for doubly-indexed random variables:

Theorem 9.3 (Liapounov CLT) *Let*

$$\{\, Z_{Jj} : j = 1, \ldots, J; J = 1, 2, 3, \ldots \,\}$$

be a doubly-indexed array of random variables where

(i) for each J the random variables $\{\, Z_{Jj} : j = 1, \ldots, J \,\}$ are mutually independent;

(ii) $\mathrm{Var}(Z_{Jj}) \neq 0$*;*

(iii) for some $r > 2$,

$$\frac{\left(\sum_{j=1}^{J} E|Z_{Jj} - EZ_{Jj}|^r\right)^{1/r}}{\mathrm{Var}^{1/2}\left(\sum_{j=1}^{J} Z_{Jj}\right)} \to 0 \qquad \textit{as } J \to \infty. \tag{9.1}$$

Then

$$\frac{J^{-1/2}\left(\sum_{j=1}^{J}[Z_{Jj} - EZ_{Jj}]\right)}{\mathrm{Var}^{1/2}\left(J^{-1/2}\sum_{j=1}^{J} Z_{Jj}\right)} \xrightarrow{d} \mathrm{N}(0,1) \qquad \textit{as } J \to \infty.$$

This Liapounov CLT for doubly-indexed random variables is given in Chung (1974, Sections 7.1–7.2, p. 209 in particular), Loeve (1977, Section 21.1, pp. 286–290) and White (1984, Theorem 5.11, p. 113). Relation (9.1) is often known as the *Liapounov condition*: it asserts, roughly, that some moment order higher than 2 must grow more slowly than that of order 2. Checking this condition will constitute an important part of establishing the limiting properties for the kernel density estimator.

Also used repeatedly in the calculations to follow is this:

Lemma 9.4 *Suppose kernel* K *in Definition 3.2 is bounded and*

$$\int \mathsf{K}(x)x^2\,dx < \infty.$$

Then for all $q \geq 1$*:*

$$\int \mathsf{K}(x)^q x^2\,dx < \infty \quad \textit{and} \quad \int \mathsf{K}(x)^q\,dx < \infty.$$

Proof of Lemma 9.4 *Write*

$$\begin{aligned}\int \mathsf{K}(x)^q x^2\,dx &= \int \mathsf{K}(x)\mathsf{K}(x)^{r-1}x^2\,dx\\ &\leq \left[\sup_x \mathsf{K}(x)^{r-1}\right] \times \int \mathsf{K}(x)x^2\,dx\\ &< \infty.\end{aligned}$$

Use this in

$$\begin{aligned}\int \mathsf{K}(x)^q\,dx &= \int_{|x|\leq 1} \mathsf{K}(x)^q\,dx + \int_{|x|>1} \mathsf{K}(x)^q\,dx\\ &\leq [\sup_x \mathsf{K}(x)^q\,] \int_{|x|\leq 1} dx + \int_{|x|>1} \mathsf{K}(x)^q x^2\,dx < \infty.\end{aligned}$$

Q.E.D.

The pointwise central limit result Theorem 3.5 is, interestingly, not given in standard texts such as Silverman (1986). With some slight variation, it is provided in, e.g., Pagan and Ullah (1999, Theorem 2.10, p. 41) or Prakasa Rao (1983, pp. 61–62), but where it appears as a minor consequence of deeper, more abstract results, or is proven only in tandem with many other propositions, so that an applied researcher might sometimes face difficulty in assessing what is critical for the result's validity. For completeness, therefore, a relatively self-contained proof is given here.

The proof consists of two principal parts. First, verify the Liapounov condition (9.1) on the deviation of the kernel estimator from its expectation; then, correct that expectation for its asymptotic bias. The second of these requires bandwidth $b_J = o(J^{-1/5})$.

Proof of Theorem 3.5 *Fix a point* $x^\dagger$ *with* $\mathsf{f}_X(x^\dagger) \neq 0$ *and for*

$$\widehat{f}_{X,b}(x^\dagger) = J^{-1} \sum_{j=1}^{J} b_J^{-1} \mathsf{K}\left(\frac{x^\dagger - x_j}{b_J}\right)$$

write the right-side summands

$$Z_{Jj} \stackrel{\text{def}}{=} b_J^{-1} \mathsf{K}\left(\frac{x^\dagger - x_j}{b_J}\right).$$

From (iv) write the second-order Taylor's series expansion of $\mathsf{f}_X(x^\dagger - xb_J)$ *about* $b_J = 0$ *as*

$$\mathsf{f}_X(x^\dagger - xb_J) = \mathsf{f}_X(x^\dagger) - \mathsf{f}'_X(x^\dagger)xb_J + \frac{1}{2}\mathsf{f}''_X(x^\dagger)x^2 b_J^2 + o(b_J^2).$$

Using this calculate for $q > 0$,

$$\begin{aligned}
EZ_{Jj}^q &= \int b_J^{-q} \mathsf{K}\left(\frac{x^\dagger - x}{b_J}\right)^q \mathsf{f}_X(x)\, dx \\
&= b_J^{1-q} \int \mathsf{K}(x)^q\, \mathsf{f}_X(x^\dagger - xb_J)\, dx \\
&= b_J^{1-q} \left[\mathsf{f}_X(x^\dagger) \int \mathsf{K}(x)^q\, dx - \mathsf{f}'_X(x^\dagger) b_J \int \mathsf{K}(x)^q x\, dx \right. \\
&\qquad \left. + \frac{1}{2}\mathsf{f}''_X(x^\dagger) b_J^2 \int \mathsf{K}(x)^q x^2\, dx + o(b_J^2) \right] \\
&= b_J^{1-q} \left[\mathsf{f}_X(x^\dagger) \int \mathsf{K}(x)^q\, dx + \frac{1}{2}\mathsf{f}''_X(x^\dagger) b_J^2 \int \mathsf{K}(x)^q x^2\, dx \right. \\
&\qquad \left. + o(b_J^2) \right],
\end{aligned} \tag{9.2}$$

where the second equality applies a change of variables, the third uses the second-order Taylor's series expansion, and the fourth uses (ii) K *symmetric implying* $\int \mathsf{K}(x)^q x\, dx = 0$. *From (i), (iii), and Lemma 9.4, we have that for all* $q \geq 1$, $\int \mathsf{K}(x)^q x^2\, dx < 0$ *and* $\int \mathsf{K}(x)^q\, dx < 0$. *Then* (9.2) *has for* $q = 1$, 2, *and* r, *respectively,*

$$EZ_{Jj} = \mathsf{f}_X(x^\dagger) + O(b_J^2); \tag{9.3}$$

$$EZ_{Jj}^2 = b_J^{-1}\mathsf{f}_X(x^\dagger)\int \mathsf{K}(x)^2\, dx + O(b_J^2); \tag{9.4}$$

$$\begin{aligned} EZ_{Jj}^r &= b_J^{1-r}\left[\mathsf{f}_X(x^\dagger)\int \mathsf{K}(x)^r\, dx + O(b_J^2)\right] \\ &= b_J^{1-r}\mathsf{f}_X(x^\dagger)\int \mathsf{K}(x)^r\, dx + O(b_J^{3-r}). \end{aligned} \tag{9.5}$$

Using (9.3) *and* (9.4) *gives*

$$\begin{aligned} \mathrm{Var}\,(\, Z_{Jj}\,) &= EZ_{Jj}^2 - (\, EZ_{Jj}\,)^2 \\ &= b_J^{-1}\mathsf{f}_X(x^\dagger)\int \mathsf{K}(x)^2\, dx - \mathsf{f}_X(x^\dagger)^2 + O(b_J) \end{aligned}$$

$$\Longrightarrow \mathrm{Var}\left(\sum_{j=1}^{J} Z_{Jj}\right) = Jb_J^{-1}\mathsf{f}_X(x^\dagger)\int \mathsf{K}(x)^2\, dx + O(J). \tag{9.6}$$

Also, by c_r *and Hölder Inequalities, Lemma 9.1 and 9.2, and* (9.5),

$$\begin{aligned} &\left(\sum_{j=1}^{J} E|Z_{Jj} - EZ_{Jj}|^r\right) \\ &\qquad \leq 2^r \sum_{j=1}^{J} E|Z_{Jj}|^r \\ &\qquad = 2^r\left(Jb_J^{1-r}\mathsf{f}_X(x^\dagger)\int \mathsf{K}(x)^r\, dx + O(Jb_J^{3-r})\right). \end{aligned}$$

Liapounov condition (9.1) *then follows from*

$$\frac{\left(\sum_{j=1}^{J} E|Z_{Jj} - EZ_{Jj}|^r\right)^{1/r}}{\mathrm{Var}^{1/2}\left(\sum_{j=1}^{J} Z_{Jj}\right)} = \frac{O(J^{1/r} b_J^{1/r-1})}{O(J^{1/2} b_J^{-1/2})}$$
$$= O\left((Jb_J)^{r^{-1}-2^{-1}}\right) \to 0 \quad \textit{since } Jb_J \to \infty \textit{ as } J \to \infty. \quad (9.7)$$

Using

$$\mathrm{Var}^{1/2}\left(J^{-1/2}\sum_{j=1}^{J} Z_{Jj}\right) = \left[b_J^{-1}\mathsf{f}_X(x^\dagger)\int \mathsf{K}(x)^2\,dx + O(1)\right]^{1/2}$$

$$\Longrightarrow\ b_J^{1/2}\,\mathrm{Var}^{1/2}\left(J^{-1/2}\sum_{j=1}^{J} Z_{Jj}\right) = \left[\mathsf{f}_X(x^\dagger)\int \mathsf{K}(x)^2\,dx\right]^{1/2} + o(1)$$

and

$$b_J^{1/2} J^{-1/2}\sum_{j=1}^{J}(Z_{Jj} - EZ_{Jj}) = (Jb_J)^{1/2}\left[\widehat{f}_{X,b}(x^\dagger) - E\widehat{f}_{X,b}(x^\dagger)\right],$$

the Liapounov CLT Theorem 9.3 gives

$$\frac{(Jb_J)^{1/2}\left[\widehat{f}_{X,b}(x^\dagger) - E\widehat{f}_{X,b}(x^\dagger)\right]}{\left[\mathsf{f}_X(x^\dagger)\int \mathsf{K}(x)^2\,dx\right]^{1/2}} \xrightarrow{d} \mathsf{N}(0,1). \quad (9.8)$$

To obtain the conclusion we need to establish the asymptotic bias vanishes relative to the normalizing $(Jb_J)^{1/2}$ *in the Liapounov CLT, i.e., that*

$$\frac{(Jb_J)^{1/2}\left[E\widehat{f}_{X,b}(x^\dagger) - \mathsf{f}_X(x^\dagger)\right]}{\left[\mathsf{f}_X(x^\dagger)\int \mathsf{K}(x)^2\,dx\right]^{1/2}} \to 0 \qquad \textit{as } J \to \infty. \quad (9.9)$$

Since (9.3) *gives*

$$\left[E\widehat{f}_{X,b}(x^\dagger) - \mathsf{f}_X(x^\dagger) \right] = O(b_J^2),$$

condition (v) $J^{1/5} b_J \to 0$ *as* $J \to \infty$ *implies* (9.9). *Combining* (9.8) *and* (9.9),

$$\frac{(Jb_J)^{1/2}\left[\widehat{f}_{X,b}(x^\dagger) - \mathsf{f}_X(x^\dagger) \right]}{\left[\mathsf{f}_X(x^\dagger) \int \mathsf{K}(x)^2\, dx \right]^{1/2}} = \frac{(Jb_J)^{1/2}\left[\widehat{f}_{X,b}(x^\dagger) - E\widehat{f}_{X,b}(x^\dagger) \right]}{\left[\mathsf{f}_X(x^\dagger) \int \mathsf{K}(x)^2\, dx \right]^{1/2}} + o(1) \xrightarrow{d} \mathsf{N}(0,1).$$

Q.E.D.

The Liapounov limiting normal distribution (9.8) obtains with nothing more than $Jb_J \to \infty$ in (9.7). However, that limit distribution is centered about the the expectation $E\widehat{f}_{X,b}(x^\dagger)$ rather than the underlying population quantity $\mathsf{f}_X(x^\dagger)$. That $J^{1/5} b_J \to 0$ guarantees this asymptotic bias converges to zero.

Embedded in the proof is also the result that the bandwidth choice that most rapidly forces to zero the mean square error of the estimator is $b_J = O(J^{-1/5})$, i.e., slower than the rate that gives a properly centered limit normal distribution. To see this, notice (9.3) and (9.4) give

$$\left| E\widehat{f}_{X,b}(x^\dagger) - \mathsf{f}_X(x^\dagger) \right|^2 = O(b_J^4) \tag{9.10}$$

$$\operatorname{Var}\left(\widehat{f}_{X,b}(x^\dagger) \right) = (\, Jb_J \,)^{-1} \mathsf{f}_X(x^\dagger) \int \mathsf{K}(x)^2\, dx + o(\left(\, Jb_J \,\right)^{-1}), \tag{9.11}$$

so that mean square error is $O(b_J^4) + O((Jb_J)^{-1})$. The faster b_J converges to zero, so too the first component $O(b_J^4)$, but at the cost of slowing the convergence rate in the second, $O((Jb_J)^{-1})$. Since the less rapid rate dominates, the optimal mean square error convergence rate equates the two, by setting $b_J = O(J^{-1/5})$, so that mean square error then converges at its fastest rate, as $O(J^{-4/5})$. But with this bandwidth,

$$\frac{(Jb_J)^{1/2}\left[\widehat{f}_{X,b}(x^\dagger) - \mathsf{f}_X(x^\dagger)\right]}{\left[\mathsf{f}_X(x^\dagger)\int \mathsf{K}(x)^2\,dx\right]^{1/2}} = \frac{(Jb_J)^{1/2}\left[\widehat{f}_{X,b}(x^\dagger) - E\widehat{f}_{X,b}(x^\dagger)\right]}{\left[\mathsf{f}_X(x^\dagger)\int \mathsf{K}(x)^2\,dx\right]^{1/2}} + \frac{(Jb_J)^{1/2}\left[E\widehat{f}_{X,b}(x^\dagger) - \mathsf{f}_X(x^\dagger)\right]}{\left[\mathsf{f}_X(x^\dagger)\int \mathsf{K}(x)^2\,dx\right]^{1/2}}$$

has the second term on the right $O(1)$ and the first term converging in distribution to $\mathsf{N}(0,1)$, so that the resulting limit distribution can no longer be used for inference.

Despite this, a minimum mean square error criterion is often used to justify bandwidth and kernel choice. Developing this explicitly requires more detail than available in (9.10) and (9.11). Under hypotheses (i)–(iv) of Theorem 3.5, equation (9.2) with $q = 1$ gives

$$E\widehat{f}_{X,b}(x^\dagger) - \mathsf{f}_X(x^\dagger) = \frac{1}{2}\mathsf{f}''_X(x^\dagger)b_J^2\int \mathsf{K}(x)x^2\,dx + o(b_J^2)$$

$$\Longrightarrow \left|E\widehat{f}_{X,b}(x^\dagger) - \mathsf{f}_X(x^\dagger)\right|^2 = \left[\frac{1}{2}b_J^2\int \mathsf{K}(x)x^2\,dx\right]^2 \mathsf{f}''_X(x^\dagger)^2 + o(b_J^4),$$

while from (9.6),

$$\mathrm{Var}\left(\widehat{f}_{X,b}(x^\dagger)\right) = (Jb_J)^{-1}\left[\int \mathsf{K}(x)^2\,dx\right]\mathsf{f}_X(x^\dagger) + O(J^{-1}).$$

Mean integrated square error is then

$$\begin{aligned} MISE &\stackrel{\text{def}}{=} \int \left| E\widehat{f}_{X,b}(x^\dagger) - \mathsf{f}_X(x^\dagger)\right|^2 dx^\dagger + \int \mathrm{Var}\left(\widehat{f}_{X,b}(x^\dagger)\right) dx^\dagger \\ &= \left[\frac{1}{2} b_J^2 \int \mathsf{K}(x)x^2\,dx\right]^2 \int \mathsf{f}_X''(x)^2\,dx \\ &\qquad + (Jb_J)^{-1}\int \mathsf{K}(x)^2\,dx + o(b_J^4) + O(J^{-1}). \end{aligned}$$

By the reasoning following (9.10) and (9.11) we can ignore the trailing order terms on the right side provided $b_J \to 0$ and $J \to \infty$.

To give the optimal kernel and bandwidth minimizing *MISE*, it is necessary to provide first an analytical characterization of the Epanechnikov kernel in Table 3.1. Do so in Lemma 9.5. That Lemma and Theorem 9.6 are not used elsewhere in the paper. They are therefore stated and proven in this Technical Appendix and numbered accordingly.

Lemma 9.5 *The problem*

$$\inf_{\mathsf{K}} \left(\int_0^\infty \mathsf{K}(x)x^2\,dx\right)\left(\int_0^\infty \mathsf{K}(x)^2\,dx\right)^2 \tag{9.12}$$

across K*'s that are symmetric kernel densities is solved by parabolic densities of the form*

$$\beta > 0: \qquad \mathsf{K}(x) = \begin{cases} \frac{3}{4}\beta^{-1}\left[1-\beta^{-2}x^2\right] & \textit{for } |x| \le \beta, \\ 0 & \textit{otherwise;} \end{cases} \tag{9.13}$$

i.e., all densities (9.13) *achieve the same minimized value for the objective function in* (9.12)*, independent of* β*. If, further,*

$$\int_{-\infty}^\infty \mathsf{K}(x)x^2\,dx = 1$$

is imposed then the unique solution to the minimization problem is

$$\mathsf{K}(x) = \left[\frac{3}{4}\left(1-\frac{1}{5}x^2\right)5^{-1/2}\right] \times I_{\{|x|\leqslant\sqrt{5}\}}(x). \tag{9.14}$$

Proof of Lemma 9.5 *Since* $\int_0^\infty \mathsf{K}(x)\,dx = \frac{1}{2}$ *construct*

$$\mathcal{L} = \left(\int_0^\infty \mathsf{K}(x)x^2\,dx\right)\left(\int_0^\infty \mathsf{K}(x)^2\,dx\right)^2 - \lambda\left[\int_0^\infty \mathsf{K}(x)\,dx - \frac{1}{2}\right], \quad \lambda \in \mathbb{R}.$$

Consider perturbations in K *of the form*

$$\widetilde{\mathsf{K}}_\alpha(x) = \mathsf{K}(x) + \alpha\mathsf{W}(x), \quad \alpha \in \mathbb{R},$$

where the test function W *is symmetric and integrable but otherwise unrestricted, and the scalar* α *can be chosen sufficiently small so that* $\widetilde{\mathsf{K}}_\alpha$ *remains non-negative. The Lagrangean function*

$$\mathcal{L}(\alpha;\lambda) = \left(\int_0^\infty \widetilde{\mathsf{K}}_\alpha(x)x^2\,dx\right)\left(\int_0^\infty \widetilde{\mathsf{K}}_\alpha(x)^2\,dx\right)^2 - \lambda\left[\int_0^\infty \widetilde{\mathsf{K}}_\alpha(x)\,dx - \frac{1}{2}\right]$$

implies

$$\frac{\partial\mathcal{L}}{\partial\alpha} = \left(\int_0^\infty \mathsf{W}(x)x^2\,dx\right)\left(\int_0^\infty \widetilde{\mathsf{K}}_\alpha(x)^2\,dx\right)^2 + 4\left(\int_0^\infty \widetilde{\mathsf{K}}_\alpha(x)x^2\,dx\right)\left(\int_0^\infty \widetilde{\mathsf{K}}_\alpha(x)^2\,dx\right)\int_0^\infty \widetilde{\mathsf{K}}_\alpha(x)\mathsf{W}(x)\,dx - \lambda\left[\int_0^\infty \mathsf{W}(x)\,dx\right].$$

If an optimum K *exists (to be verified by calculating one explicitly) denote*

$$\beta_0 = \int_0^\infty \mathsf{K}(x)^2\,dx \quad \text{and} \quad \beta_1 = \int_0^\infty \mathsf{K}(x)x^2\,dx$$

and notice that at an optimum, $\partial\mathcal{L}/\partial\alpha = 0$ *at* $\alpha = 0$, *i.e.,*

$$\left.\frac{\partial\mathcal{L}}{\partial\alpha}\right|_{\alpha=0} = \int_0^\infty \left[\beta_0^2x^2 + 4\beta_0\beta_1\mathsf{K}(x) - \lambda\right]\mathsf{W}(x)\,dx = 0.$$

But since W *is arbitrary and* $\mathsf{K}(x) \geq 0$ *we must have*

$$\beta_0^2 x^2 + 4\beta_0\beta_1 \mathsf{K}(x) - \lambda = 0 \implies$$

$$\mathsf{K}(x) = \begin{cases} (4\beta_0\beta_1)^{-1}\,[\,\lambda - \beta_0^2 x^2\,] & \textit{for } |x| \leq \lambda^{1/2}\beta_0^{-1}, \\ 0 & \textit{otherwise,} \end{cases} \tag{9.15}$$

with $\lambda > 0$. *The form* (9.15) *gives*

$$\begin{aligned} \frac{1}{2} &= \int_0^\infty \mathsf{K}(x)\,dx = (4\beta_0\beta_1)^{-1} \int_0^{\lambda^{1/2}\beta_0^{-1}} \left[\lambda - \beta_0^2 x\right] dx \\ &= (4\beta_0\beta_1)^{-1} \left[\left(\lambda - \frac{1}{3}\beta_0^2 x^2\right) x^2 \,\Big|_0^{\lambda^{1/2}\beta_0^{-1}} \right] \\ &= \frac{1}{2}(3\beta_0\beta_1)^{-1}\lambda^{3/2}\beta_0 \implies \lambda = \left(3\beta_0^2\beta_1\right)^{2/3} \end{aligned}$$

and

$$\begin{aligned} \beta_0 &= \int_0^\infty \mathsf{K}(x)^2\,dx = (4\beta_0\beta_1)^{-2} \int_0^{\lambda^{1/2}\beta_0^{-1}} \left[\lambda - \beta_0^2 x\right]^2 dx \\ &= (4\beta_0\beta_1)^{-2} \left[\left(\frac{\beta_0^4}{5} x^4 - \frac{2}{3}\lambda\beta_0^2 x^2 + \lambda^2 \right) x \,\Big|_0^{\lambda^{1/2}\beta_0^{-1}} \right] \\ &= \frac{3^{2/3}}{10}(\beta_0/\beta_1)^{1/3} \implies \beta_0 = \frac{3}{10^{3/2}}\beta_1^{-1/2}. \end{aligned}$$

Together these imply $\lambda = (3/10)^2$ *and that the objective function in* (9.12) *achieves value* $\beta_1\beta_0^2 = 3^2/10^3$, *independent of* β_1. *Substituting for* λ *and* β_0 *in* (9.15) *and letting* $\beta = (10\beta_1)^{1/2} > 0$ *verifies* (9.13). *Finally, imposing* $\int_{-\infty}^\infty \mathsf{K}(x)x^2\,dx = 1$ *gives* $\beta_1 = \int_0^\infty \mathsf{K}(x)x^2\,dx = 1/2$ *and thus* $\beta = \sqrt{5}$, *verifying* (9.14). *Q.E.D.*

If, contrary to the last part of Lemma 9.5, $\int_{-\infty}^{\infty}, \mathsf{K}(x)x^2\,dx = 1$ is violated then solutions such as

$$\mathsf{K}(x) = \frac{3}{4}\left(1 - x^2\right) \times I_{\{|x|\leqslant 1\}}(x),$$

different from the kernel in (9.14), become available; see, e.g., Hardle and Linton (1994, Table 1, p. 2303) and Pagan and Ullah (1999, p. 28).

The objective function in (9.12), and thus Lemma 9.5, will play a central role in analyzing *MISE*.

Theorem 9.6 *Assume hypotheses (i)–(iv) of Theorem 3.5 and $b_J \to 0$ as $J \to \infty$. Then for J large, the (K, b_J) minimizing MISE can be taken to be*

$$\mathsf{K}(x) = \left[\frac{3}{4}\left(1 - \frac{1}{5}x^2\right)5^{-1/2}\right] \times I_{\{|x| \leqslant \sqrt{5}\}}(x)$$

(i.e., equation (9.14) of Lemma 9.5) and

$$\begin{aligned} b_J &= J^{-1/5}\frac{\left(\int \mathsf{K}(x)^2\,dx\right)^{1/5}}{\left(\int \mathsf{K}(x)x^2\,dx\right)^{2/5}\left(\int \mathsf{f}_X''(x)^2\,dx\right)^{1/5}} \\ &= J^{-1/5}\frac{\left(3 \times 5^{-3/2}\right)^{1/5}}{\left(\int \mathsf{f}_X''(x)^2\,dx\right)^{1/5}}\,. \end{aligned} \tag{9.16}$$

Defining

$$\mu(\mathsf{K}) \overset{def}{=} \left(\int \mathsf{K}(x)x^2\,dx\right)^{2/5}\left(\int \mathsf{K}(x)^2\,dx\right)^{4/5},$$

the minimized MISE is

$$\begin{aligned} &\mu(\mathsf{K})J^{-4/5} \times \left\{\frac{5}{4}\left(\int \mathsf{f}_X''(x)^2\,dx\right)^{1/5}\right\} \\ &\qquad = \left(3 \times 5^{-3/2}\right)^{4/5} J^{-4/5} \times \left\{\frac{5}{4}\left(\int \mathsf{f}_X''(x)^2\,dx\right)^{1/5}\right\}. \end{aligned} \tag{9.17}$$

Proof of Theorem 9.6 *For sufficiently large J, MISE is*

$$\left[\frac{1}{2}b_J^2 \int \mathsf{K}(x)x^2\,dx\right]^2 \int \mathsf{f}_X''(x)^2\,dx + (Jb_J)^{-1}\int \mathsf{K}(x)^2\,dx. \tag{9.18}$$

Minimize this first over b_J *and then over* K*. Setting to zero the first derivative of* (9.18) *with respect to* b_J *gives*

$$b_J^5 = J^{-1}\frac{\int \mathsf{K}(x)^2\,dx}{\left(\int \mathsf{K}(x)x^2\,dx\right)^2 \int \mathsf{f}_X''(x)^2\,dx},$$

verifying (9.16)*. Substituting this into* (9.18) *gives*

$$\begin{aligned}
&\frac{1}{4}\left(\int \mathsf{K}(x)x^2\,dx\right)^2\left(\int \mathsf{f}_X''(x)^2\,dx\right)J^{-4/5} \\
&\qquad\times\frac{\left(\int \mathsf{K}(x)^2\,dx\right)^{4/5}}{\left(\int \mathsf{K}(x)x^2\,dx\right)^{8/5}\left(\int \mathsf{f}_X''(x)^2\,dx\right)^{4/5}} \\
&\quad+\left(\int \mathsf{K}(x)^2\,dx\right)J^{-4/5}\frac{\left(\int \mathsf{K}(x)^2\,dx\right)^{-1/5}}{\left(\int \mathsf{K}(x)x^2\,dx\right)^{-2/5}\left(\int \mathsf{f}_X''(x)^2\,dx\right)^{-1/5}} \\
&\quad=\frac{5}{4}\left(\int \mathsf{K}(x)x^2\,dx\right)^{2/5}\left(\int \mathsf{f}_X''(x)^2\,dx\right)^{1/5}\left(\int \mathsf{K}(x)^2\,dx\right)^{4/5} \\
&\qquad\times J^{-4/5} \\
&\quad=\frac{5}{4}\mu(\mathsf{K})J^{-4/5}\left(\int \mathsf{f}_X''(x)^2\,dx\right)^{1/5},
\end{aligned}$$

verifying the left side of (9.17)*. Minimizing MISE in* (9.18) *therefore reduces to minimizing* $\mu(\mathsf{K})$ *over choice of* K*. But by Lemma 9.5, the latter minimization subject to (i)–(iii) of Theorem 3.5 is achieved by selecting* K *in* (9.14)*. Substituting this kernel into the right-side of equation* (9.16) *recovers the expression following the latter, and into the left side of equation* (9.17) *recovers the right side.* *Q.E.D.*

The left side of equation (9.17) allows a natural comparison across kernels. Using an arbitrary K on sample size J achieves only the same *MISE* as using the optimal kernel K_{Ep} in (9.14) on a hypothetical (smaller) sample size J_{Ep}, where

$$\mu(\mathsf{K})J^{-4/5} = \mu(\mathsf{K}_{\mathrm{Ep}})J_{\mathrm{Ep}}^{-4/5}.$$

The ratio

$$J_{\mathrm{Ep}}/J = [\,\mu(\mathsf{K}_{\mathrm{Ep}})/\mu(\mathsf{K})\,]^{5/4}$$
$$= \frac{3}{5\sqrt{5}}\left(\int \mathsf{K}(x)x^2\,dx\right)^{-1/2}\left(\int \mathsf{K}(x)^2\,dx\right)^{-1} \qquad (9.19)$$

can then be defined to be the *relative efficiency* of kernel K. Values of this for several commonly-used kernels are given in Table 3.2.

Continue from here with proofs of results in the body of the paper.

Proof of Proposition 3.12 *To prove (i) notice that by equation* (3.24) *the typical entry in* $\bar{\mathsf{f}}_{t+s}$ *is*

$$\bar{\mathsf{f}}_{t+s}(m') = \int_{\overline{x}_{m'}}^{\overline{x}_{m'+1}} \mathsf{f}_{t+s}(x')\,dx'$$
$$= \int_{\overline{x}_{m'}}^{\overline{x}_{m'+1}} \left[\int_{-\infty}^{\infty} p_{t,s}(x,x') \times \mathsf{f}_t(x)\,dx\right] dx'.$$

Since F_t *is piecewise linear, the integrand can be re-written:*

$$\int_{-\infty}^{\infty} p_{t,s}(x,x') \times \mathsf{f}_t(x)dx = \sum_{m=0}^{M-1} \int_{\overline{x}_m}^{\overline{x}_{m+1}} p_{t,s}(x,x') \times \mathsf{f}_t(x)\,dx$$
$$= \sum_{m=0}^{M-1} \left[\int_{\overline{x}_m}^{\overline{x}_{m+1}} p_{t,s}(x,x')\,dx\right] \times \mathsf{f}_t(\overline{x}_m)$$

so that

$$\bar{\mathsf{f}}_{t+s}(m') = \int_{\overline{x}_{m'}}^{\overline{x}_{m'+1}} \sum_{m=0}^{M-1} \left[\int_{\overline{x}_m}^{\overline{x}_{m+1}} p_{t,s}(x,x')\,dx\right] \times \mathsf{f}_t(\overline{x}_m)\,dx'$$
$$= \sum_{m=0}^{M-1} \left[\int_{\overline{x}_{m'}}^{\overline{x}_{m'+1}} \int_{\overline{x}_m}^{\overline{x}_{m+1}} p_{t,s}(x,x')\,dx\,dx'\right] \times \mathsf{f}_t(\overline{x}_m).$$

Using again equation (3.24)

$$\bar{\mathsf{f}}_t(m) = \mathsf{f}_t(\overline{x}_m) \times [\,\overline{x}_{m+1} - \overline{x}_m\,]$$

gives

$$
\begin{aligned}
&\overline{\mathsf{f}}_{t+s}(m') \\
&= \sum_{m=0}^{M-1} \left[(\overline{x}_{m+1} - \overline{x}_m)^{-1} \times \int_{\overline{x}_{m'}}^{\overline{x}_{m'+1}} \int_{\overline{x}_m}^{\overline{x}_{m+1}} p_{t,s}(x, x')\, dx\, dx' \right] \times \overline{\mathsf{f}}_t(m) \\
&= \sum_{m=0}^{M-1} P_{t,s}(m, m') \times \overline{\mathsf{f}}_t(m)
\end{aligned}
$$

so that $\overline{\mathsf{f}}_{t+s} = P_{t,s}^{\top} \times \overline{\mathsf{f}}_t$ *with* $P_{t,s}$ *given in equation* (3.27). *To prove (ii) write*

$$
\begin{aligned}
&\left(P_{t,s}\overline{\phi} \right)(m) \\
&= (\overline{x}_{m+1} - \overline{x}_m)^{-1} \int_{\overline{x}_m}^{\overline{x}_{m+1}} \sum_{m'=0}^{M-1} \int_{\overline{x}_{m'}}^{\overline{x}_{m'+1}} p_{t,s}(x, x')\, dx'\, dx \times \overline{\phi}(m') \\
&= (\overline{x}_{m+1} - \overline{x}_m)^{-1} \int_{\overline{x}_m}^{\overline{x}_{m+1}} \sum_{m'=0}^{M-1} \int_{\overline{x}_{m'}}^{\overline{x}_{m'+1}} p_{t,s}(x, x')\phi(x')\, dx'\, dx \\
&= (\overline{x}_{m+1} - \overline{x}_m)^{-1} \int_{\overline{x}_m}^{\overline{x}_{m+1}} \int_{-\infty}^{\infty} p_{t,s}(x, x')\phi(x')\, dx'\, dx \\
&= (\overline{x}_{m+1} - \overline{x}_m)^{-1} \int_{\overline{x}_m}^{\overline{x}_{m+1}} (\mathcal{T}_{t,s}\phi)(x)\, dx,
\end{aligned}
$$

establishing equation (3.28). *If for each* x' *the function* $p_{t,s}(\cdot, x')$ *is piecewise constant on* $[\overline{x}_m, \overline{x}_{m+1})$ *then*

$$
(\mathcal{T}_{t,s}\phi)(x) = \int_{-\infty}^{\infty} p_{t,s}(x, x')\phi(x')\, dx'
$$

is similarly piecewise constant on $[\overline{x}_m, \overline{x}_{m+1})$ *so that the interval average in equation* (3.28) *becomes just* $(\mathcal{T}_{t,s}\phi)(\overline{x}_m)$. *Q.E.D.*

Proof of Theorem 3.14 *The transition density* $p_t(x^{\dagger}, x)$ *of Brownian Motion* σB *is, by definition, the pdf of a* $\mathsf{N}(x, \sigma^2 t)$ *random variable. But the right side of equation* (3.34) *is exactly that pdf, thereby* *confirming the first part of the Theorem. To obtain the resolvent kernel* (3.35) *begin by calculating that for standard Brownian Motion. When* $\sigma = 1$, *the resolvent kernel is*

$$\int_0^\infty e^{-\lambda t}(2\pi t)^{-1/2} e^{-(x-x^\dagger)^2/2t}\,dt$$
$$= (2/\pi)^{-1/2}\int_0^\infty e^{-[\lambda t^2+(x-x^\dagger)^2/2t^2]}\,dt, \qquad (9.20)$$

where the expression on the right follows from applying the change of variable $t' = t^{1/2}$ *and then relabelling. Define* $\nu = |x - x^\dagger|\sqrt{\lambda}/\sqrt{2}$ *and* $s = (\lambda/\nu)^{1/2} \times t$, *and rescale and complete the square in the exponent on the right side of* (9.20)*:*

$$\lambda t^2 + \frac{(x-x^\dagger)^2}{2t^2} = \nu\left[\frac{\lambda}{\nu}t^2 + \frac{\nu}{\lambda}t^{-2}\right]$$
$$= \nu\left[s^2 + s^{-2}\right] = \nu\left[(s - s^{-1})^2 + 2\right].$$

Since $dt = (\nu/\lambda)^{1/2}\,ds$ *the right side of* (9.20) *becomes*

$$(2/\pi)^{-1/2}(\nu/\lambda)^{1/2}e^{-2\nu}\int_0^\infty e^{-(s-1/s)^2\nu}\,ds.$$

But

$$\begin{aligned}\int_0^\infty e^{-(s-1/s)^2\nu}\,ds &= \int_0^1 e^{-(s-1/s)^2\nu}\,ds + \int_1^\infty e^{-(s-1/s)^2\nu}\,ds\\ &= \int_0^1 e^{-(s-1/s)^2\nu}\,ds + \int_0^1 e^{-(s-1/s)^2\nu}\times s^{-2}\,ds\\ &= \int_0^1 e^{-(s-1/s)^2\nu}\times(1+s^{-2})\,ds\\ &= \int_0^\infty e^{-t^2\nu}\,dt = (\pi/2)^{1/2}(2\nu)^{-1/2}.\end{aligned}$$

Therefore,

$$\int_0^\infty e^{-\lambda t}(2\pi t)^{-1/2}e^{-(x-x^\dagger)^2/2t}\,dt$$
$$= e^{-2\nu}\frac{1}{\sqrt{2\lambda}} = e^{-(2\lambda)^{1/2}|x-x^\dagger|}\times\frac{1}{\sqrt{2\lambda}}.$$

The resolvent kernel for $X = \sigma B$ can then be calculated:

$$
\begin{aligned}
G_\lambda(x^\dagger, x) &= \int_0^\infty e^{-\lambda t} p_t(x^\dagger, x)\, dt \\
&= \int_0^\infty e^{-\lambda t} (2\pi\sigma^2 t)^{-1/2} \exp\left(-\frac{(x - x^\dagger)^2}{2\sigma^2 t} \right) dt \\
&= \int_0^\infty e^{-(\lambda/\sigma^2)s} (2\pi s)^{-1/2} \exp\left(-\frac{(x - x^\dagger)^2}{2s} \right) \frac{1}{\sigma^2}\, ds \\
&= \frac{1}{\sigma^2} \left[\frac{\sigma}{\sqrt{2\lambda}} e^{-|x - x^\dagger|\sqrt{2\lambda}/\sigma} \right] \\
&= \sigma^{-1} \frac{1}{\sqrt{2\lambda}} e^{-|x - x^\dagger|\sqrt{2\lambda}/\sigma} ,
\end{aligned}
$$

verifying (3.35). *Q.E.D.*

Proof of Theorem 3.16 *For small $\Delta t > 0$ and bounded measurable function ϕ, calculate:*

$$
\begin{aligned}
(\mathcal{T}_{\Delta t}(\mathbf{R}_\lambda \phi)) &= \mathcal{T}_{\Delta t} \left[\int_0^\infty e^{-t\lambda} (\mathcal{T}_t \phi)\, dt \right] \\
&= \int_0^\infty e^{-t\lambda} (\mathcal{T}_{t+\Delta t} \phi)\, dt \\
&= e^{\lambda \Delta t} \times \int_0^\infty e^{-(t+\Delta t)\lambda} (\mathcal{T}_{t+\Delta t} \phi)\, dt \\
&= e^{\lambda \Delta t} \times \left[(\mathbf{R}_\lambda \phi) - \int_0^{\Delta t} e^{-t\lambda} (\mathcal{T}_t \phi)\, dt \right] .
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\left(\frac{\mathcal{T}_{\Delta t} - I}{\Delta t} \right) (\mathbf{R}_\lambda \phi) \\
&\quad = \frac{e^{\lambda \Delta t} (\mathbf{R}_\lambda \phi) - (\mathbf{R}_\lambda \phi)}{\Delta t} - \frac{e^{\lambda \Delta t} \int_0^{\Delta t} e^{-t\lambda} (\mathcal{T}_t \phi)\, dt}{\Delta t} \\
&\quad = \left(\frac{e^{\lambda \Delta t} - 1}{\Delta t} \right) (\mathbf{R}_\lambda \phi) - \frac{e^{\lambda \Delta t} \int_0^{\Delta t} e^{-t\lambda} (\mathcal{T}_t \phi)\, dt}{\Delta t} . \qquad (9.21)
\end{aligned}
$$

As $\Delta t \to 0$

$$\frac{e^{\lambda \Delta t} - 1}{\Delta t} \to \lambda$$

and

$$\frac{e^{\lambda \Delta t} \int_0^{\Delta t} e^{-t\lambda} \, (\mathcal{T}_t \, \phi \,) \; dt}{\Delta t}$$
$$\to \lambda e^{\lambda \Delta t} \int_0^{\Delta t} e^{-t\lambda} \, (\mathcal{T}_t \, \phi \,) \; dt + e^{\lambda \Delta t} e^{-\lambda \Delta t} \, (\mathcal{T}_{\Delta t} \phi \,) \to \phi,$$

so that equation (9.21) *becomes*

$$\mathbf{G}\mathbf{R}_\lambda \phi = \lambda \mathbf{R}_\lambda \phi - \phi \implies \mathbf{R}_\lambda = (\lambda - \mathbf{G})^{-1}.$$

Q.E.D.

Proof of Theorem 4.1 *Consider* $y > 0$ *a fixed but otherwise arbitrary value for income. Denote by* t_0 *that epoch at which* y *last realized as the income level at the frontier, i.e.,*

$$t_0 = \sup \{ \, t : y_{max}(t) \leq y \, \} \, ,$$

or, as $y_{max}(t)$ *has its timepath continuous and increasing,*

$$y = y_{max}(t_0).$$

Let $\phi : [t_0, \infty] \to [0, 1]$ *have* $\phi(t)$ *denote the fraction of economies at time* $t \geq t_0$ *having per capita incomes no greater than* y*, so that*

$$\dot{\phi}(t) = -\phi(t) \times \mathsf{h}(t) \implies \phi(t) = e^{-\int_{t_0}^{t} \mathsf{h}(s) \, ds}, \quad t \geq t_0.$$

Moreover,

$$y_{max}(t) = y_{max}(t_0) e^{\int_{t_0}^{t} \zeta \mathsf{h}(s) \, ds} = y \times e^{\int_{t_0}^{t} \zeta \cdot \mathsf{h}(s) \, ds}$$

so that

$$\phi(t) = [\, y / y_{max}(t) \,]^{1/\zeta} \, . \tag{9.22}$$

Equation (9.22) *confirms that the fraction of economies with incomes no greater than the fixed level y is constantly declining. But* (9.22) *also says that the fraction of economies with relative incomes no greater than $x = y/y_{\max}$ is time-invariant at value $x^{1/\zeta}$, thus confirming the shape of the density given in the Theorem.* *Q.E.D.*

Proof of Theorem 6.1 *Call* f *the density for* F. *Since, by hypothesis,* F *is continuous and strictly increasing, for any $x \in (0,1)$, the inverse* $\mathsf{F}^{-1}(x)$ *exists and is unique. Write the Lorenz curve*

$$\mathcal{Z}(h) = \mathcal{E}^{-1}\int_0^x x'\,\mathsf{f}(x')\,dx', \quad \textit{with } x = \mathsf{F}^{-1}(h).$$

Therefore,

$$\frac{d\mathcal{Z}}{d\mathsf{F}} = \frac{d\mathcal{Z}}{dx} \Big/ \frac{d\mathsf{F}}{dx} = \mathcal{E}^{-1}x\,\mathsf{f}(x)/\,\mathsf{f}(x)$$
$$= \mathcal{E}^{-1}x = \mathcal{E}^{-1}\mathsf{F}^{-1}(h) > 0$$

and

$$\frac{d^2\mathcal{Z}}{d\mathsf{F}^2} = \frac{d}{dx}\left(\frac{d\mathcal{Z}}{d\mathsf{F}}\right) \Big/ \frac{d\mathsf{F}}{dx} = \mathcal{E}^{-1} \Big/ \mathsf{f}(x) > 0,$$

so that the Lorenz curve has tangent whose slope is always positive and increasing, i.e., $\mathcal{Z}$ is convex from below. Thus $\mathcal{Z}(\mathsf{F}) - \mathsf{F}$, the Lorenz curve's horizontal separation from the 45-degree line, is maximized at:

$$0 = \frac{d}{dx}\left[\,\mathcal{Z}(\mathsf{F}) - \mathsf{F}\,\right] = \mathcal{E}^{-1}x\,\mathsf{f}(x) - \mathsf{f}(x),$$

i.e., at $x = \mathcal{E}$ or $h = \mathsf{F}(x) = \mathsf{F}(\mathcal{E})$. Therefore, since $h = \mathsf{F}(\mathcal{E})$ maximizes $\mathcal{Z}(h) - h$, that same value for h also gives $\mathcal{Z}'(h) = 1$.

Q.E.D.

Proof of Theorem 6.2 *The mean income gap in the definition of $\mathcal{I}_G$ in* *equation* (6.1) *can be rewritten:*

$$\int_0^\infty \int_0^\infty |x - x'|\, dF(x') dF(x) = \int_0^\infty \left[\int_0^x (x - x')\, dF(x') + \int_x^\infty (x' - x)\, dF(x') \right] dF(x).$$

But since

$$\int_0^\infty \int_0^\infty (x - x')\, dF(x')\, dF(x) = 0$$
$$\Longrightarrow \int_0^\infty \left[\int_0^x (x - x')\, dF(x') - \int_x^\infty (x' - x)\, dF(x') \right] dF(x) = 0$$
$$\Longrightarrow \int_0^\infty \int_0^x (x - x')\, dF(x')\, dF(x) = \int_0^\infty \int_x^\infty (x' - x)\, dF(x')\, dF(x),$$

the mean income gap is therefore also

$$\begin{aligned}\int_0^\infty \int_0^\infty |x - x'|\, dF(x')\, dF(x) &= 2 \times \int_0^\infty \int_0^x (x - x')\, dF(x')\, dF(x) \\ &= 2 \times \left[\int_0^\infty x\, F(x)\, dF(x) - \int_0^\infty \int_0^x x'\, dF(x')\, dF(x) \right].\end{aligned}$$

Definitions (6.1) *and* (6.2) *for* $\mathcal{I}_G$ *and* $\mathcal{Z}$ *then give:*

$$\begin{aligned}\mathcal{I}_G &= \mathcal{E}^{-1} \int_0^\infty x\, F(x)\, dF(x) - \int_0^\infty \mathcal{Z}(F(x))\, dF(x) \\ &= \int_0^1 h\, d\mathcal{Z}(h) - \int_0^1 \mathcal{Z}(h)\, dh \\ &= \left[\left(\frac{1}{2} + \int_0^1 [h - \mathcal{Z}(h)]\, dh \right) - \left(\frac{1}{2} - \int_0^1 [h - \mathcal{Z}(h)]\, dh \right) \right] \\ &= 2 \int_0^1 [\, h - \mathcal{Z}(h)\,]\, dh.\end{aligned}$$

Q.E.D.

Proof of Theorem 6.8 *From* (6.9) *and* (6.11),

$$U(C(t)) = U(e^{Z(t)}e^{\xi t}) = \frac{e^{(1-R)Z(t)}e^{(1-R)\xi t} - 1}{1-R}$$
$$\Longrightarrow\ e^{-\rho t}U(C(t)) = \frac{e^{(1-R)Z(t)}e^{-[\rho-(1-R)\xi]t} - e^{-\rho t}}{1-R}$$

Using the definitions for $\widetilde{\rho}$ *and* $\widetilde{U}$, *this gives*

$$e^{-\rho t}U(C(t)) = e^{-\widetilde{\rho}t}\widetilde{U}(Z(t)) + \begin{cases} -(1-R)^{-1}e^{-\rho t} & \text{for } R \neq 1; \\ \xi t e^{-\rho t} & \text{otherwise.} \end{cases}$$

Then, conditional on $C(0) = e^z$,

$$W(0) = E\left[\int_0^\infty e^{-\rho t}U(C(t))\,dt \mid C(0) = e^z\right]$$
$$= (\mathbf{R}_{\widetilde{\rho}}\widetilde{U})(z) + \begin{cases} -\int_0^\infty (1-R)^{-1}e^{-\rho t}\,dt & \text{for } R \neq 1; \\ \int_0^\infty \xi t e^{-\rho t}\,dt & \text{otherwise.} \end{cases}$$

The conclusion of the Theorem then follows from calculating, for $R \neq 1$,

$$\int_0^\infty -(1-R)^{-1}e^{-\rho t}\,dt = (1-R)^{-1}\rho^{-1} \times \left(e^{-\rho t}\right)\Big|_0^\infty$$
$$= -(1-R)^{-1}\rho^{-1};$$

and, for $R = 1$,

$$\int_0^\infty \xi e^{-\rho t}t\,dt = \xi \times (-\rho^{-2}) \times \left[(1+\rho t)e^{-\rho t}\right]\Big|_0^\infty = \xi\rho^{-2}.$$

Q.E.D.

Proof of Theorem 6.9 *From the resolvent kernel* (3.32) *and equation* (3.35) *of Theorem 3.14,*

$$(\mathbf{R}_{\widetilde{\rho}}\widetilde{U})(z^\dagger) = \int_{-\infty}^\infty \widetilde{U}(z) \times \sigma^{-1}\frac{1}{\sqrt{2\widetilde{\rho}}} \times e^{-|z-z^\dagger|\sqrt{2\widetilde{\rho}}/\sigma}\,dz.$$

For $R \neq 1$ this becomes.

$$(\mathbf{R}_{\widetilde{\rho}}\widetilde{U})(z^\dagger) = \frac{(2\widetilde{\rho})^{-1/2}\sigma^{-1}}{1-R}\int_{-\infty}^{\infty} e^{-(2\widetilde{\rho}/\sigma^2)^{1/2}\times|z-z^\dagger|} \times e^{(1-R)z}\,dz.$$

The integral on the right can be rewritten:

$$\int_{-\infty}^{z^\dagger} e^{-(2\widetilde{\rho}/\sigma^2)^{1/2}(z^\dagger - z)} \times e^{(1-R)z}\,dz + \int_{z^\dagger}^{\infty} e^{-(2\widetilde{\rho}/\sigma^2)^{1/2}(z-z^\dagger)} \times e^{(1-R)z}\,dz$$

$$= \frac{e^{-(2\widetilde{\rho}/\sigma^2)^{1/2}z^\dagger}}{(2\widetilde{\rho}/\sigma^2)^{1/2} + (1-R)} \times e^{[(2\widetilde{\rho}/\sigma^2)^{1/2}+(1-R)]z^\dagger} + \frac{e^{(2\widetilde{\rho}/\sigma^2)^{1/2}z^\dagger}}{(2\widetilde{\rho}/\sigma^2)^{1/2} - (1-R)} \times e^{-[(2\widetilde{\rho}/\sigma^2)^{1/2}-(1-R)]z^\dagger}$$
$$= e^{(1-R)z^\dagger} \times \frac{(2\widetilde{\rho}/\sigma^2)^{1/2} \times 2}{(2\widetilde{\rho}/\sigma^2) - (1-R)^2}$$

where convergence of the integral follows from $\widetilde{\rho} > \frac{1}{2}(1-R)^2\sigma^2$. Then

$$\begin{aligned}(\mathbf{R}_{\widetilde{\rho}}\widetilde{U})(z^\dagger) &= \frac{e^{(1-R)z^\dagger}}{1-R} \times \frac{2\sigma^{-2}}{(2\widetilde{\rho}/\sigma^2) - (1-R)^2} \\ &= \frac{e^{(1-R)z^\dagger}}{1-R} \times \frac{1}{\rho - (1-R)\xi - \frac{1}{2}(1-R)^2\sigma^2} \\ &= r(R,\rho,\xi,\sigma)^{-1} \times e^{(1-R)z^\dagger}.\end{aligned}$$

For $R = 1$ analogous steps apply to

$$(\mathbf{R}_{\widetilde{\rho}}\widetilde{U})(z^\dagger) = (2\widetilde{\rho})^{-1/2}\sigma^{-1}\int_{0}^{\infty} e^{-(2\widetilde{\rho}/\sigma^2)^{1/2}\times|z-z^\dagger|} \times z\,dz.$$

Again rewrite the integral on the right

$$\begin{aligned}
&\int_{-\infty}^{z^\dagger} e^{-(2\widetilde{\rho}/\sigma^2)^{1/2}(z^\dagger - z)} \times z\,dz + \int_{z^\dagger}^{\infty} e^{-(2\widetilde{\rho}/\sigma^2)^{1/2}(z - z^\dagger)} \times z\,dz \\
&= \frac{e^{-(2\widetilde{\rho}/\sigma^2)^{1/2} z^\dagger}}{-(2\widetilde{\rho}/\sigma^2)} \left[(1 - (2\widetilde{\rho}/\sigma^2)^{1/2} z) e^{(2\widetilde{\rho}/\sigma^2)^{1/2} z} \right]\Big|_{-\infty}^{z^\dagger} \\
&\qquad + \frac{e^{(2\widetilde{\rho}/\sigma^2)^{1/2} z^\dagger}}{-(2\widetilde{\rho}/\sigma^2)} \left[(1 + (2\widetilde{\rho}/\sigma^2)^{1/2} z) e^{-(2\widetilde{\rho}/\sigma^2)^{1/2} z} \right]\Big|_{z^\dagger}^{\infty} \\
&= \sqrt{2} \left[\sigma^2/\widetilde{\rho} \right]^{1/2} z^\dagger
\end{aligned}$$

so that

$$(\mathbf{R}_{\widetilde{\rho}}\widetilde{U})(z^\dagger) = (2\widetilde{\rho})^{-1/2} \sigma^{-1} \sqrt{2} \left[\sigma^2/\widetilde{\rho} \right]^{1/2} z^\dagger = \rho^{-1} z^\dagger .$$

Q.E.D.

Of course equation (6.13) combining Theorems 6.8–6.9 can also be obtained directly using

$$E\left[e^{Z(t)} \mid Z(0) = z \right] = \exp\left(z + \sigma^2 t/2 \right),$$

but such a direct approach would not generalize the way that using the resolvent operator allows.

Proof of Theorem 6.10 *From the definition*

$$\begin{aligned}
(\mathbf{R}_{\widetilde{\rho}}\widetilde{U})(\overline{z}_m) &= \int_0^\infty e^{-\widetilde{\rho}t} E\left[\widetilde{U}(Z(t)) | Z(0) = \overline{z}_m \right] dt \\
&= \int_0^\infty e^{-\widetilde{\rho}t} \left[\sum_{m=0}^{M-1} P_t(m, m') \widetilde{U}(\overline{z}_{m'}) \right] dt
\end{aligned}$$

Collecting the right side across m gives the vector

$$(\mathbf{R}_{\widetilde{\rho}}\widetilde{U})(\overline{z}_m), \quad m = 0, 1, \ldots, M-1$$

as

$$\int_0^\infty e^{-\widetilde{\rho}t} \left[P_t \overline{U} \right] dt = (\,\widetilde{\rho} - \mathbf{G}\,)^{-1}\, \overline{U},$$

thus establishing the Theorem. *Q.E.D.*

Proof of Theorem 8.1 *Performing the calculations directly for this case allows checking the results obtained more generally by resolvent operators. To that end use the conditional distributions in equation* (8.5) *to write:*

$$E_0 \log \epsilon_j(t) = \alpha^t \log \epsilon_j(0) - \frac{1-\alpha^t}{1-\alpha^2}\frac{\sigma_j^2}{2}.$$

Now take each of the cases in turn.

(i) Suppose $R = 1$. *When* $t > 0$,

$$\begin{aligned} E_0 U(C_j(t)) &= E_0 \log C_j(t) \\ &= E_0 \left[\log \overline{z}_j + t \log \xi + \log \epsilon_j(t)\right] \\ &= \left[\log \overline{z}_j - (1-\alpha^2)^{-1}\sigma_j^2/2\right] + (\log \xi) \times t \\ &\quad + \left[\log \epsilon_j(0) + (1-\alpha^2)^{-1}\sigma_j^2/2\right] \alpha^t, \end{aligned}$$

so that

$$\begin{aligned} &E_0 \sum_{t=0}^{\infty} \delta^t U(C_j(t)) \\ &= U(C_j(0)) + E_0 \sum_{t=1}^{\infty} \delta^t U(C_j(t)) \\ &= \log \overline{z}_j + \log \epsilon_j(0) + \left[\log \overline{z}_j - (1-\alpha^2)^{-1}\sigma_j^2/2\right] \frac{\delta}{1-\delta} \\ &\quad + \frac{\delta}{(1-\delta)^2} \log \xi + \left[\log \epsilon_j(0) + (1-\alpha^2)^{-1}\sigma_j^2/2\right] \frac{\delta\alpha}{1-\delta\alpha}, \end{aligned}$$

using $\sum_{t=0}^{\infty} t\delta^t = \delta \sum_{t=1}^{\infty} t\delta^{t-1} = \delta \frac{d}{d\delta} \sum_{t=1}^{\infty} \delta^t = \delta \frac{d}{d\delta} \frac{\delta}{1-\delta} = (1-\delta)^{-2}\delta$. *Collecting terms then gives*

$$\begin{aligned} &E_0 \sum_{t=0}^{\infty} \delta^t U(C_j(t)) = (1-\delta\alpha)^{-1} \log \epsilon_j(0) \\ &+ \frac{\delta}{(1-\delta)^2} \log \xi + \frac{1}{1-\delta} \log \overline{z}_j - \frac{\delta}{(1-\delta)(1-\delta\alpha)(1+\alpha)} \frac{\sigma_j^2}{2}. \end{aligned}$$

(ii) Suppose $R \neq 1$. When $t > 0$,

$$
\begin{aligned}
E_0 U(C_j(t)) &= (1-R)^{-1} E_0 \left[(\overline{z}_j)^{1-R} \xi^{-(R-1)t} \epsilon_j(t)^{1-R} - 1 \right] \\
&= (1-R)^{-1} \Big[(\overline{z}_j)^{1-R} \xi^{-(R-1)t} \times \\
&\quad \exp\Bigg([1-R] \left\{ \alpha^t \log \epsilon_j(0) - \frac{1-\alpha^t}{1-\alpha^2} \frac{\sigma_j^2}{2} \right\} \\
&\quad + \frac{1}{2}[1-R]^2 \frac{1-\alpha^{2t}}{1-\alpha^2} \sigma_j^2 \Bigg) \\
&\quad -1 \Big].
\end{aligned}
$$

Collecting time-dependent terms in the exponent gives

$$
\begin{aligned}
&[1-R] \left\{ \log \epsilon_j(0) + (1-\alpha^2)^{-1} \sigma_j^2/2 \right\} \alpha^t \\
&\qquad - [1-R]^2 (1-\alpha^2)^{-1} \sigma_j^2 / 2 \alpha^{2t},
\end{aligned}
$$

while the time-invariant terms there become

$$
\begin{aligned}
&- [1-R](1-\alpha^2)^{-1} \sigma_j^2/2 + [1-R]^2 (1-\alpha^2)^{-1} \sigma_j^2/2 \\
&\qquad = -[1-R] R (1-\alpha^2)^{-1} \sigma_j^2/2.
\end{aligned}
$$

For convenience, let

$$
\begin{aligned}
D_1 &= [1-R] \left\{ \log \epsilon_j(0) + (1-\alpha^2)^{-1} \sigma_j^2/2 \right\} \\
D_2 &= [1-R]^2 (1-\alpha^2)^{-1} \sigma_j^2/2 > 0.
\end{aligned}
$$

Then

$$
\begin{aligned}
E_0 U(C_j(t)) = (1-R)^{-1} \Big[&(\overline{z}_j)^{1-R} e^{-[1-R]R(1-\alpha^2)^{-1}\sigma_j^2/2} \times \\
&\xi^{-(R-1)t} e^{D_1 \alpha^t} e^{-D_2 \alpha^{2t}} - 1 \Big]
\end{aligned}
$$

so that

$$
\begin{aligned}
&E_0 \sum_{t=0}^{\infty} \delta^t U(C_j(t)) \\
&\qquad = \frac{(\overline{z}_j)^{1-R} \epsilon_j(0)^{1-R} - 1}{1-R} + (1-R)^{-1} \times \\
&\left[(\overline{z}_j)^{1-R} e^{-[1-R]R(1-\alpha^2)^{-1}\sigma_j^2/2} \sum_{t=1}^{\infty} \xi^{-(R-1)t} e^{D_1 \alpha^t} e^{-D_2 \alpha^{2t}} \right. \\
&\qquad\qquad \left. - (1-\delta)^{-1}\delta \right] \\
&\qquad = (1-R)^{-1} \left[-(1-\delta)^{-1} + (\overline{z}_j)^{1-R} \times \right. \\
&\left\{ e^{-[1-R]R(1-\alpha^2)^{-1}\sigma_j^2/2} \sum_{t=1}^{\infty} \xi^{-(R-1)t} e^{D_1 \alpha^t} e^{-D_2 \alpha^{2t}} \right. \\
&\qquad\qquad\qquad \left. \left. + \epsilon_j(0)^{1-R} \right\} \right].
\end{aligned}
$$

Q.E.D.

References

Moses Abramowitz. Catching up, forging ahead, and falling behind. *Journal of Economic History*, 46:385–406, June 1986.

Philippe Aghion and Peter Howitt. A model of growth through creative destruction. *Econometrica*, 60(2):323–351, March 1992.

Philippe Aghion and Peter Howitt. *Endogenous Growth Theory*. MIT Press, Cambridge, 1998.

Philippe Aghion, Eve Caroli, and Cecilia García-Peñalosa. Inequality and economic growth: The perspective of the new growth theories. *Journal of Economic Literature*, 37(4):1615–1660, December 1999.

Alberto Alesina and Dani Rodrik. Distributive politics and economic growth. *Quarterly Journal of Economics*, 109(2):465–490, May 1994.

Gordon Anderson. Toward an empirical analysis of polarization. *Journal of Econometrics*, 122(1):1–26, September 2004.

Robert B. Ash. *Real Analysis and Probability*. Academic Press, New York, 1972.

Anthony B. Atkinson. On the measurement of poverty. *Econometrica*, 55(4):749–764, July 1987.

Anthony B. Atkinson. Bringing income distribution in from the cold. *Economic Journal*, 107(441):297–321, March 1997.

Costas Azariadis and Allan Drazen. Threshold externalities in economic development. *Quarterly Journal of Economics*, 105(2): 501–526, May 1990.

Costas Azariadis and John Stachurski. A forward projection of the cross-country income distribution. Discussion Paper 570, Kyoto University, May 2003.

Abhijit V. Banerjee and Esther Duflo. Inequality and growth: What can the data say? *Journal of Economic Growth*, 8(3):267–299, September 2003.

Robert J. Barro. Inequality and growth in a panel of countries. *Journal of Economic Growth*, 5(1):5–32, March 2000.

Robert J. Barro and Xavier Sala-i-Martin. Convergence. *Journal of Political Economy*, 100(2):223–251, April 1992.

William J. Baumol. Productivity growth, convergence, and welfare. *American Economic Review*, 76(5):1072–85, December 1986.

Roland Bénabou. Inequality and growth. In Ben Bernanke and Julio Rotemberg, editors, *Macroeconomics Annual*, volume 11, pages 11–74. NBER and MIT Press, Cambridge, 1996.

Andrew B. Bernard and Charles I. Jones. Productivity across industries and countries: Time series theory and evidence. *Review of Economics and Statistics*, 78(1):135–146, February 1996.

Marco Bianchi. Testing for convergence: Evidence from nonparametric multimodality tests. *Journal of Applied Econometrics*, 12(4):393–409, July 1997.

Peter J. Bickel and Murray Rosenblatt. On some global measures of the deviations of density function estimates. *Annals of Statistics*, 1(6):1071–1095, November 1973. Correction: 3(6):1370, November 1975.

Patrick Billingsley. *Convergence of Probability Measures*. John Wiley, New York, 1968.

David E. Bloom, David Canning, and Jaypee Sevilla. Geography and poverty traps. *Journal of Economic Growth*, 8(4):355–378, December 2003.

I. Blumen, M. Kogan, and P. J. McCarthy. *The Industrial Mobility of Labor as a Probability Process*, volume 6 of *Cornell Studies of Industrial and Labor Relations*. Cornell University Press, Ithaca, 1955.

Richard Blundell and Ian Preston. Consumption inequality and income uncertainty. *Quarterly Journal of Economics*, CXIII(2): 603–640, May 1998.

François Bourguignon. The growth elasticity of poverty reduction: Explaining heterogeneity across countries and time periods. In Theo Eicher and Steve Turnovsky, editors, *Growth and Inequality: Issues and Policy Implications*, chapter 1, pages 3–26. MIT Press, Cambridge, 2003.

François Bourguignon. Decomposable income inequality measures. *Econometrica*, 47(4):901–920, July 1979.

François Bourguignon and Christian Morrisson. Inequality among world citizens: 1820–1990. *American Economic Review*, 92(4): 727–744, September 2002.

Sandra Bulli. Distribution dynamics and cross-country convergence: A new approach. *Scottish Journal of Political Economy*, 48(2): 226–243, May 2001.

Theophilos Cacoullos. Estimation of a multivariate density. *Annals of the Institute of Statistical Mathematics*, 18(2):179–189, 1966.

Gavin Cameron, James Proudman, and Stephen Redding. Productivity convergence and international openness. In James Proudman and Stephen Redding, editors, *Openness and Growth*, chapter 6, pages 221–260. Bank of England, London, 1998.

Shaohua Chen and Martin Ravallion. How did the world's poorest fare in the 1990s? *Review of Income and Wealth*, 47(3):283–300, September 2001.

Shaohua Chen and Martin Ravallion. How did the world's poorest fare since the early 1980s? *The World Bank Research Observer*, 19(2):141–169, August 2004a.

Shaohua Chen and Martin Ravallion. How have the world's poorest fared since the 1980s? Working paper, DRG World Bank, 2004b.

Xiahong Chen, Oliver Linton, and Peter M. Robinson. The estimation of conditional densities. Working Paper EM/01/415, STICERD, LSE, May 2001.

Kai Lai Chung. *Markov Chains with Stationary Transition Probabilities*. Springer-Verlag, New York, second edition edition, 1967.

Kai Lai Chung. *A Course in Probability Theory*. Academic Press, New York, second edition, 1974.

David T. Coe and Elhanan Helpman. International R&D spillovers. *European Economic Review*, 39(5):859–887, May 1995.

Frank Cowell. Measurement of inequality. In Anthony B. Atkinson and Francois Bourguignon, editors, *Handbook of Income Distribution*, volume 1, chapter 2, pages 87–166. North Holland Elsevier Science, Amsterdam, 2000.

Frank Cowell. *Measuring Inequality*. Prentice Hall/Harvester Wheatsheaf, Hemel Hempstead, third edition, 1995.

D. R. Cox and H. D. Miller. *The Theory of Stochastic Processes*. Chapman and Hall, New York NY, 1965.

Angus Deaton. Measuring poverty in a growing world (or measuring growth in a poor world). *Review of Economics and Statistics*, 87(1):1–19, February 2005.

Klaus Deininger and Lyn Squire. New ways of looking at old issues: Inequality and growth. *Journal of Development Economics*, 57: 259–287, 1998.

Luc Devroye. *A Course in Density Estimation*. Birkhauser, Boston, 1987.

Luc Devroye and László Györfi. *Nonparametric Density Estimation: The L_1 View*. Probability and Mathematical Statistics. John Wiley, New York, 1985.

Jared Diamond. *Guns, Germs, and Steel*. W. W. Norton, New York, 1997.

David Dollar and Art Kraay. Growth is good for the poor. *Journal of Economic Growth*, 7(3):195–225, September 2002.

Joseph L. Doob. *Stochastic Processes*. John Wiley, New York NY, 1953.

Darrell Duffie and Peter Glynn. Estimation of continuous-time markov processes sampled at random time intervals. Working paper, Stanford University, Stanford CA, June 2004.

Gilles Duranton and Henry G. Overman. Testing for localisation using micro-geographic data. Working paper, LSE, London, December 2002.

Steven N. Durlauf. Nonergodic economic growth. *Review of Economic Studies*, 60(2):349–366, April 1993.

Steven N. Durlauf. A theory of persistent income inequality. *Journal of Economic Growth*, 1(1):75–93, March 1996.

Steven N. Durlauf and Paul A. Johnson. Multiple regimes and cross-country growth behavior. *Journal of Applied Econometrics*, 10 (4):365–384, October 1995.

Steven N. Durlauf and Danny Quah. The new empirics of economic growth. In John B. Taylor and Michael Woodford, editors, *Handbook of Macroeconomics*, volume 1A, chapter 4, pages 231–304. North Holland Elsevier Science, Amsterdam, 1999.

William Easterly and Ross Levine. It's not factor accumulation: Stylized facts and economic growth. *World Bank Economic Review*, 15:177–219, 2001.

Jonathan Eaton and Samuel Kortum. International technology diffusion: Theory and measurement. *International Economic Review*, 40(3):537–570, August 1999.

Philip Epstein, Peter Howlett, and Max-Stephan Schulze. Distribution dynamics: Stratification, polarization, and convergence among OECD economies, 1870–1992. *Explorations in Economic History*, 40(1):78–97, 2003.

Joan-María Esteban and Debraj Ray. On the measurement of polarization. *Econometrica*, 62(4):819–851, July 1994.

William Feller. *An Introduction to Probability Theory and Its Applications*, volume 1. John Wiley, New York, 3rd edition, 1968.

James Feyrer. Convergence by parts. Working paper, Dartmouth College, Hanover, December 2001.

Davide Fiaschi and Andrea Mario Lavezzi. Distribution dynamics and nonlinear growth. *Journal of Economic Growth*, 8(4):379–401, December 2003.

Glenn Firebaugh. *The New Geography of Global Income Inequality*. Harvard University Press, Cambridge, 2003.

Kristin J. Forbes. A reassessment of the relationship between inequality and growth. *American Economic Review*, 90(5):869–887, December 2000.

Carl Futia. Invariant distributions and the limiting behavior of Markovian economic models. *Econometrica*, 50(1):377–408, January 1982.

Oded Galor and Joseph Zeira. Income distribution and macroeconomics. *Review of Economic Studies*, 60(1):35–52, January 1993.

Iosif I. Gihman and Anatoliĭ V. Skorohod. *The Theory of Stochastic Processes II.* Springer-Verlag, Berlin, 1975.

Wolfgang Hardle and Oliver Linton. Applied nonparametric methods. In Robert F. Engle and Daniel McFadden, editors, *Handbook of Econometrics*, volume 4, chapter 38, pages 2295–2339. North Holland Elsevier Science, Amsterdam, 1994.

Elhanan Helpman. Innovation, imitation, and intellectual property rights. *Econometrica*, 61(6):1247–1280, November 1993.

Alan Heston, Robert Summers, and Bettina Aten. Penn World Table version 6.1. Dataset, Center for International Comparisons at the University of Pennsylvania (CICUP), October 2002.

Peter Howitt. Endogenous growth and cross-country income differences. *American Economic Review*, 90(4):829–846, September 2000.

Charles I. Jones. On the evolution of the world income distribution. *Journal of Economic Perspectives*, 11(3):19–36, Summer 1997.

Charles I. Jones. *Introduction to Economic Growth.* W. W. Norton, New York, 1998.

Eric L. Jones. *Growth Recurring: Economic Change in World History.* Oxford University Press, Oxford, 1988.

Samuel Karlin and Howard M. Taylor. *A Second Course in Stochastic Processes.* Academic Press, New York, 1981.

Louise C. Keely. *Ideas and Incentives in the Innovation Process: Implications for Economic Growth.* PhD thesis, LSE, London, June 1999.

Wolfgang Keller. International technology diffusion. *Journal of Economic Literature*, September 2004. Forthcoming.

Wolfgang Keller. Are international R&D spillovers trade-related? *European Economic Review*, 42(8):1469–1481, September 1998.

Michael Kremer, Alexei Onatski, and James Stock. Searching for prosperity. *Carnegie-Rochester Conference Series on Public Policy*, 55(1):275–303, December 2001.

Simon Kuznets. Economic growth and income inequality. *American Economic Review*, 45(1):1–28, March 1955.

David S. Landes. *The Wealth and Poverty of Nations*. Little, Brown and Company, London, 1998.

Hongyi Li, Lyn Squire, and Hengfu Zou. Explaining international and intertemporal variations in income inequality. *Economic Journal*, 108:1–18, January 1998.

Michael Loeve. *Probability Theory I*. Springer-Verlag, New York, 4th edition, 1977.

Robert E. Lucas. Some macroeconomics for the 21st century. *Journal of Economic Perspectives*, 14(1):159–168, Winter 2000.

Robert E. Lucas. *Studies in Business Cycle Theory*. MIT Press, Cambridge, 1981.

Robert E. Lucas. Making a miracle. *Econometrica*, 61(2):251–271, March 1993.

Angus Maddison. *The World Economy in the 20th Century*. Development Centre of the OECD, Paris, 1989.

N. Gregory Mankiw, David Romer, and David N. Weil. A contribution to the empirics of economic growth. *Quarterly Journal of Economics*, 107(2):407–437, May 1992.

Kiminori Matsuyama. Financial market globalization and endogenous inequality of nations. Working paper, Economics Department, Northwestern University, Evanston, June 2000.

Robert C. Merton. *Continuous-time Finance*. Basil Blackwell, Cambridge, Massachusetts, 1990.

Branko Milanovic. True world income distribution, 1988 and 1993: First calculation based on household surveys alone. *Economic Journal*, 112(476):51–92, January 2002.

Joel Mokyr. *The Lever of Riches: Technological Creativity and Economic Progress*. Oxford University Press, Oxford, 1990.

Benny Moldovanu and Eyal Winter. Order independent equilibria. *Games and Economic Behavior*, 9(1):21–34, 1995.

Arthur M. Okun. *Equality and Efficiency: The Big Tradeoff*. Brookings Institution, Washington DC, 1975.

Richard Paap and Herman K. van Dijk. Distribution and mobility of wealth of nations. *European Economic Review*, 42(7):1269–1293, July 1998.

Adrian Pagan and Amman Ullah. *Nonparametric Econometrics*. Cambridge University Press, New York, 1999.

Stephen L. Parente and Edward C. Prescott. *Barriers to Riches*. Walras-Pareto Lectures. MIT Press, Cambridge, 2000.

Stephen L. Parente and Edward C. Prescott. Changes in the wealth of nations. *Federal Reserve Bank of Minneapolis Quarterly Review*, 17(2):3–16, Spring 1993.

Stephen L. Parente and Edward C. Prescott. Barriers to technology adoption and development. *Journal of Political Economy*, 102 (2):298–321, April 1994.

Emmanuel Parzen. On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33:1065–1076, 1962.

Roberto Perotti. Growth, income distribution, and democracy: What the data say. *Journal of Economic Growth*, 1(2):149–187, June 1996.

Torsten Persson and Guido Tabellini. Is inequality harmful for growth? *American Economic Review*, 84(3):600–621, June 1994.

Kenneth Pomeranz. *The Great Divergence: China, Europe, and the Making of the Modern World Economy*. Princeton University Press, Princeton, 2000.

B. L. S. Prakasa Rao. *Nonparametric Functional Estimation*. Academic Press, New York, 1983.

Lant Pritchett. Divergence, big time. *Journal of Economic Perspectives*, 11(3):3–17, Summer 1997.

Danny Quah. The weightless economy in economic development. In Matti Pohjola, editor, *Information Technology, Productivity, and Economic Growth*, UNU/WIDER and Sitra, chapter 4, pages 72–96. Oxford University Press, Oxford, 2001a.

Danny Quah. Cross-country growth comparison: Theory to empirics. In Jacques Dreze, editor, *Advances in Macroeconomic Theory*, volume 133 of *Proceedings of the Twelfth World Congress of the International Economic Association, Buenos Aires*, chapter 16, pages 332–351. Palgrave, London, 2001b.

Danny Quah. Spatial agglomeration dynamics. *American Economic Review (Papers and Proceedings)*, 92(2):247–252, May 2002.

Danny Quah. One third of the world's growth and inequality. In Theo Eicher and Steve Turnovsky, editors, *Growth and Inequal-*

ity: Issues and Policy Implications, chapter 2, pages 27–58. MIT Press, Cambridge, 2003.

Danny Quah. Empirical cross-section dynamics in economic growth. *European Economic Review*, 37(2/3):426–434, April 1993a.

Danny Quah. Galton's fallacy and tests of the convergence hypothesis. *The Scandinavian Journal of Economics*, 95(4):427–443, December 1993b.

Danny Quah. Convergence empirics across economies with (some) capital mobility. *Journal of Economic Growth*, 1(1):95–124, March 1996a.

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

Danny Quah. Empirics for growth and distribution: Polarization, stratification, and convergence clubs. *Journal of Economic Growth*, 2(1):27–59, March 1997.

Danny Quah. Life in unequal growing economies. Working paper, Economics Dept., LSE, April 2007.

Danny Quah and Helen Simpson. Spatial cluster empirics. Working paper, LSE, June 2003.

C. R. Rao. *Linear Statistical Inference and Its Applications*. John Wiley and Sons, New York, second edition, 1973.

Debraj Ray and Rajiv Vohra. A theory of endogenous coalition structures. *Games and Economic Behavior*, 26(2):286–336, January 1999.

L. C. G. Rogers. The potential approach to the term structure of interest rates and foreign interest rates. *Mathematical Finance*, 7 (2):157–176, April 1997.

L. C. G. Rogers and David Williams. *Diffusions, Markov Processes, and Martingales*, volume 1. John Wiley, 1994.

Paul M. Romer. Endogenous technological change. *Journal of Political Economy*, 98(5, part 2):S71–S102, October 1990.

Murray Rosenblatt. Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, 27:832–837, 1956.

George G. Roussas. Nonparametric estimation in Markov processes. *Annals of the Institute of Statistical Mathematics*, 21(1):73–87, 1969.

Xavier Sala-i-Martin. The disturbing 'rise' of global income inequality. Working Paper 8904, NBER, Cambridge, April 2002a.

Xavier Sala-i-Martin. The world distribution of income (estimated from individual country distributions). Working paper, Columbia University, New York, May 2002b.

T. Paul Schultz. Inequality in the distribution of personal income in the world: How it is changing and why. *Journal of Population Economics*, 11:307–344, 1998.

Anthony F. Shorrocks. The class of additively decomposable inequality measures. *Econometrica*, 48(3):613–625, April 1980.

Bernard W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, New York NY 10001, 1986.

Burton Singer and Seymour Spilerman. The representation of social processes by Markov models. *American Journal of Sociology*, 82 (1):1–54, January 1976.

Robert M. Solow. A contribution to the theory of economic growth. *Quarterly Journal of Economics*, 70(1):65–94, February 1956.

Robert M. Solow. *Learning from 'Learning by Doing': Lessons for Economic Growth.* Stanford University Press, 1997. The Kenneth J. Arrow Lectures.

Hyun-Hwa Son. A note on pro-poor growth. *Economics Letters*, 82 (3):307–314, March 2004.

Nancy L. Stokey and Robert E. Lucas. *Recursive Methods in Economic Dynamics.* Harvard University Press, Cambridge, 1989.

Alan Stuart and J. Keith Ord. *Kendall's Advanced Theory of Statistics: Vol. 1 Distribution Theory.* Halsted Press, John Wiley and Sons, 1993.

Robert Tamura. From decay to growth: A demographic transition to economic growth. *Journal of Economic Dynamics and Control*, 20:1237–1261, 1996.

UNU. *UNU/WIDER-UNDP World Income Inequality Database version 2.0 beta.* United Nations University/WIDER, 2004.

Halbert White. *Asymptotic Theory for Econometricians.* Academic Press, New York NY, 1984.

Michael Wolfson. Diverging inequalities. *American Economic Review*, 84(2):353–358, May 1994.

World Bank. *World Development Report: Attacking Poverty.* Oxford University Press, New York, 2000.

World Bank. *Globalization, Growth, and Poverty.* Oxford University Press, New York, 2002.

Sang-Seung Yi. Stable coalition structures with externalities. *Games and Economic Behavior*, 20(2):201–237, August 1997.

Feng Zhu. A nonparametric analysis of the shape dynamics of the US personal income distribution, 1962–2000. Working paper, Economics Department, Yale University, January 2003.