Stochastic Dynamics and the Stationary Distribution of Wealth

Christopher Bliss

Received:

April, 2005

Abstract. *A model with wealth accumulation subject to i.i.d. random shocks is examined. The transfer function shows what* k_{t+1} *- wealth at* $t+1$ *- would be, given* k_t*, with no shock. It has a positive slope, but indeterminate concavity/convexity. The stationary distribution satisfies a Fredholm integral equation and can be examined by direct analysis of the stochastic process or via the Fredholm equation. The shape of the transfer function, particularly any non-linearities, and the distribution of shocks, affect the stationary distribution(s). Economic theory forces consideration of a broad range of cases, including some which violate* β*-convergence. "Twin peaks" in the stationary distribution cannot be excluded.*

1980 AMS Subject Classification: ***JEL Classification****: D3 E1*

Keywords and Phrases: Convergence, stochastic process, wealth distribution

0.1. Stochastic Dynamics and the Stationary Distribution of Wealth

Christopher Bliss[1]

Introduction. Consider a population of wealth holders of fixed size at time t. Next period $t + 1$ each member of the population will have another level of wealth, typically different, though conceivably the same. The process that maps wealth at t to wealth at $t + 1$ can be decomposed into the systematic and the stochastic. In other words wealth levels are generated by stochastic processes. For this reason the study of wealth dynamics should be a routine application of the mathematical theory of stochastic processes, a field that enjoys a large literature.

The justification for revisiting this field is that neither the pure mathematical theory, nor the numerous economic discussions of wealth dynamics that are available, provide a treatment that meets fully the needs of the economist. Economists are interested in cases thrown up by their theories, in particular a systematic relation between present and future wealth that is non-linear, and sometimes non-linear in a specific manner. The mathematical theory has focussed considerably, though not exclusively, on linear stochastic processes. These are appealing but may be of limited use to economists.

There is an old tradition in which wealth distribution, its change over time, and long-run equilibrium distributions, are modelled from some specification of the process which transfers individuals from one wealth-state to another. See Champernowne (1953), Steindl (1972) and Wold and Whittle (1957). This approach has been revived by Danny Quah. In Quah (1993) he considers income generation as a pure Markov process. See also Quah (1996a) and (1996b). He looks at observed transition patterns, without considering the stationary distribution. He does not derive his Markov transitions from economic theory. The aim is to confront theoretically derived convergence models with the hard facts shown by the data. Quah finds a tendency for convergence within two groups - high and low income. He names this finding "twin peaks". There are non-negligible probabilities that a country will shift from one group to the other, but these transition probabilities are too low to iron out the twin peaks in the distribution.

The economic theory foundations of the transition process is the focus of two papers Durlauf (1993 and 1996). This author uses non-convexity to produce agglomerations of agents that form local wealth distributions between which they cannot transit. In the 1996 paper agents locate in neighbourhoods within which they enjoy local externalities. The argument makes clear that the conclusions depend upon bounded shocks. In other words an agent cannot be thrown from one neighbourhood to another by a large shock. A similar

[1]Department of Economics, University of Oxford, and Nuffield College, Oxford OX1 1NF, England. christopher.bliss@nuf.ox.uk.ac

Helpful comments from David Cox, Bent Nielsen, Derek Ritzman and Neil Shephard are gratefully acknowledged.

point features in the analysis below, see Theorem 6.

The study of the time series generated by difference equations has become a central concern of modern econometric theory. Hamilton (1994) provides a wide-ranging account of this field, and other texts, for instance Øksendal. (1998), treat the theory of stochastic processes at a more advanced mathematical level. Despite the existence of a substantial literature, some questions that contemporary economic theory brings to the foreground are not answered by the econometrics literature. Much of the latter concentrates on linear difference equations, which are too limited for our needs here. Also while the stationary distribution is a natural long-run model for economic theory, econometricians have been more interested, understandably, in the estimation of equation parameters, or spectral analysis of time series generated by difference equations, and in models other than the i.i.d. case.

By far the deepest and most rigorous treatment of non-linear stochastic processes with an eye to economic applications known to the author is provided by Furia (1982). The Furia paper places stochastic processes in the mathematics of continuous mappings in separable topological spaces. This powerful abstract treatment of the subject offers only a limited elucidation of the results with specific economic models. Thus examples provided by Durlauf, and another example given below, show that Furia's uniqueness criterion - Furia (1982, p. 385) - is violated by fairly reasonable economic models. A reader of that paper, immersed in a formal and powerful mathematical argument, might fail to appreciate that great mathematical generality is not accompanied in this case by an all-embracing coverage of economic models. While the paper leaves out relevant points concerning uniqueness, its treatment of convergence to invariant distributions goes beyond anything to be found in the present paper. The shape of invariant distributions, a major concern here, is not considered by Furia.

Recently Quah (2005) examines the co-determination of growth and distribution in general stochastic models. This is ambitious work in progress, already far more extensive than a journal paper. It provides many results not in the present paper, while it omits the examination of how the structure of a particular economic model, together with the distribution of shocks, can shape the stationary distributions of wealth.

Binder and Pesaran (1999) parallels the approach of this paper in more than one respect. The focus on stationary solutions is similar and the idea of a growth process disturbed by shocks is common to both papers. An important difference is found in the way in which random shocks are incorporated. In Binder and Pesaran the model itself is stochastic, while the models below are determinate but random shocks are added. The second is more restrictive. However as a consequence it is possible to obtain many results concerning the shape of stationary distributions, while Binder and Pesaran are forced to work via general results concerning all the moments of a stationary distribution, reinforced by some excellent econometrics.

Convergence. As a case in point take the neoclassical convergence model. See Barro (1991), Barro and Sala-i-Martin (1992 and 1995). The model says that all units within the relevant population[2] - which might be countries, regions, even individual families, as desired - tend to converge to a common level of capital and output per head. The theory leads to a relationship similar to:

$$k_{t+1} = h\left[k_t\right] \tag{1}$$

where k_t is the logarithm of wealth (or income), and there is a unique stable value of $k = k*$, such that $k* = h\left[k*\right]$. The function $h\left[\cdot\right]$ will be called the *Transfer Function*, because it indicates how the underlying dynamics of the system take it from k_t to k_{t+1}. The reference to wealth acknowledges the fact that the economic theory to which the leading writers appeal applies to wealth accumulation. Empirical studies, however, typically use income rather than wealth, because income is far better measured. In what follows I shall always refer to wealth, even when discussing studies that use income. Nothing essential in what follows is affected by the income-wealth distinction.

Much of the convergence literature reads as if all units within the population will move closer and closer to $k*$. Or, in an approach, which Robert Barro has built and elaborated, individual countries have different $k*$ values, depending on numerous additional variables, such as democracy or the share of government expenditure in GDP. Barro (1997) is a convenient reference.

As an econometric model based on (1) has to include an error term, it is essential that random departures from the strict model be taken into account. Therefore (1) is modified to become:

$$k_{t+1} = h\left[k_t\right] + \epsilon_t \tag{2}$$

This is the equation of a non-linear stochastic process. Economic theory will impose various restrictions on its form, see below. Do the shocks ϵ_t show true exogenous random shocks, or might they represent the impact of missing variables? If shocks reflect missing variables the presence of unobserved effects for a particular unit can produce serial correlation. With unequal conditioning variables distributed arbitrarily across the population almost any stationary wealth distribution is possible.

Concentrating therefore on the case in which all units are identical except for initial conditions, it will not be the case that all will converge to a common $k*$ if random shocks constantly disturb the dynamic adjustment process (1). If random shocks are important in their magnitudes they affect the process of convergence. And if shocks are important, there

[2]For countries many recent studies claim convergence to be conditional on qualifying properties, such as economic openess, or the protection of property rights.

will never be strict convergence. Then the question is: what will the stationary distribution of k values be when a convergent process such as (1) is modified by random shocks?

The stationary distribution is that which replicates itself under the combined effects of the process (1) and the addition of random shocks. In particular one may ask, what are the respective contributions to the form of the stationary wealth distribution of:

- the shape of the function $h[\cdot]$
- the probability distribution of shocks

The problem addressed in this paper has been introduced in terms of the well-known convergence model. Formally, however, the same problem is encountered whenever an adjustment process of the form of (1) is shocked. Thus k_t might be the advertising budget of a firm with (1) representing slow adjustment of that budget to a long-run equilibrium level equal to $k*$. If this adjustment process is regularly shocked, a stationary frequency distribution of k values shows the long-run probability that k will be found to be in any particular interval at a randomly chosen time.

Note that treating k_t as being continuously distributed does not imply that k takes an uncountable infinity of values at any time. The integral over an interval in the density of the distribution may measure the probability of finding k within that interval of values.

Empirical Results. Returning to wealth accumulation, consider an empirical study based on a linearized and re-arranged version of the relation (2), viz:

$$k_{t+1} - k_t = f[k_t] - k_t = \alpha - \beta \cdot k_t + \epsilon_t \tag{3}$$

With the normal finding $0 < \beta < 1$, equation (3) says that on average poor units (countries) grow faster than rich units. This has been called $\beta-$*convergence*. This concept of convergence is not the same as $\sigma-$*convergence*, which means that the variance of the population of k values declines over time[3]. Friedman (1992) claims that interpreting a negative coefficient on k_t in a regression like (3) as convergence exhibits "Galton's Fallacy", on the ground that a negative coefficient is consistent with no tendency for the variance of k_t to decrease with time: it may even increase[4].

That point fits well with the argument of this paper. If k_t is distributed according to a stationary distribution, which replicates itself, there is plainly no σ-convergence. Take a value of k_t far from $k*$. The expected value of k_{t+1} conditional on such a value of k_t will be closer to $k*$. It is $\alpha - (\beta - 1) \cdot k_t$ when (3) applies. Here, while individual dispersed units

[3]For a clear exposition of the two concepts of convergence, and empirical discussion, see Sala-i-Martin (1996).

[4]See also Quah (1993) and Bliss (1999) and (2000).

tend to converge, their density is made good by units, including the less-dispersed, pushed outwards by random shocks. A stationary distribution of k values is invariant over time in the sense that it reproduces itself next period, although individual values will vary, partly systematically, showing β-convergence, and partly randomly, due to stochastic realisations of ϵ_t.

Asymptotic Properties and Stationary Distributions. This section is concerned with clarifying terminology and some concepts employed in the paper.

Definition 1. *A **stationary distribution** is one for which the probability of finding k within any interval is the same next period as in the current period.*

A stationary wealth distribution may be compared to a liquid in thermodynamic equilibrium. Its macro properties are invariant. Temperature and pressure in any part of the liquid do not alter with time. Even so, microscopic inspection will reveal local random changes (Brownian motion) involving a few molecules. In practice these movements always average out to no change at the macroscopic level. The implication that a thermodynamic equilibrium is only static in a statistical sense - the expected value of the system next observation period is the same as the current distribution - applies equally to a wealth distribution. Given the possibility of sufficiently large shocks, everyone may be twice as wealthy tomorrow as they are today. Unfortunately we may have to wait a very long time for this happy outcome to occur. The remaining age of the Earth may not be sufficient to make the probability as high as 0.01%.

A stationary distribution requires *stationarity* of the stochastic process. A stochastic process is stationary if the probability density which attaches to any sub-sequence of values is independent of the dates at which those events are observed. Then (2) defines a stationary process if the ϵ_t values are always drawn independently from the same distribution (are i.i.d. in the usual terminology). For in that case the sequence:

$$\{k_t, ..., k_{t+n}\} \tag{4}$$

requires the random variable ϵ to take specific values in the ordered sequence:

$$\{\epsilon_t, ..., \epsilon_{t+n-1}\} \tag{5}$$

Then those same ϵ values ordered as in (5) will generate the equally probable sequence:

$$\{k_T, ..., k_{T+n}\} \tag{6}$$

starting at a different date.

Stationarity is distinct from another property.

Definition 2. *A stochastic process which generates the values $\{x_1, ..., x_t, ...\}$ will be said to have* **long-memory** *if the probability that $x_T \in I$, where I is a closed interval of values of x, is not asymptotically independent of x_1 as $t \to \infty$.*

Long-memory corresponds to what is sometimes called *path-dependence*. This means that the economic system may be in a subset of its state space which it will never leave, although when it is not in that subset it never enters it. David (1993) discusses the implications of path-dependence in economic history. In another famous paper, David (1985), the same author argues that the QWERTY arrangement of keys on a standard typewriter is the result of historical accident, but is now locked in because the costs of change prohibit its substitution by a layout more efficient for modern machines. Further examples of this type are easily found.[5] Where a stochastic process describes the economy, the conditions for long-memory are more demanding than is the case with a deterministic economy, as will be made clear by the example that follows shortly. The opposite of long-memory is short-memory, taken here to mean simply that the condition specified by Definition 2 is not satisfied, so that we have asymptotic independence of x_1.

Bradford DeLong (1999) describes a short-memory system as *ergodic*. In this paper the term ergodic has been avoided. Too many meanings have been attached to it, ranging from the precise but highly abstract, to the vague and hand-waving (as Granger and Terasvirta (1993), p.10). For Binder and Peseran (1999) ergodic means asymptotic convergence of all moments of a distribution. Probably all would agree that a long-memory system is not ergodic. For that case we now have another precise term to hand.

Some of the distinctions made above are illustrated well by an example which falls outside Barro's model of optimal economic growth based on the Ramsey model. The objective function is the standard:

$$\int_0^\infty [c(t)]\, e^{-\delta t} dt \tag{7}$$

In place of the usual concave production function substitute a non-concave function with the property that the marginal product of capital goes almost to zero for medium levels of capital, but above that range the marginal product rises sharply, although it later falls away when capital is extremely high. This economy is like two concave economies stitched together, yet sufficiently separated to allow optimal growth paths to exist in two regions

[5] Consider the keywork on a standard modern flute. Theobald Boehm, who is credited with the design of the precursor of today's flutes, preferred an open G# design. This respects the simple principle that putting down fingers produces lower notes. However because flute players of that time were used to closed G# fingering, closed G# flutes were made by Boehm and his followers, and today this fingering is standard for mass-produced flutes and teaching manuals. This is so despite the fact that closed G# gives rise to formidable intonation difficulties for E in the third-octave, requiring more keywork to solve the problem. .

and to converge to different levels of k. Figure 1 illustrates the form of the $h[\cdot]$ function. The dynamics of the system when no random shocks intervene is plain. There are three values of k such that $k = h[k]$, labelled A, B and C. B is unstable, the other two points, A and C are stable.

Without random shocks the system has long memory. Ignoring the zero-measure point B, all initial values of k lie in one of two basins of attraction from which k asymptotes either to A or to C, according to which basin applies. Furthermore, as the curve through B is steep, k is thrown some distance to the left or to the right away from B by the transfer process by itself. Then if the distribution of ϵ values is uniform on $[-a, a]$ with a small, the random effects can never overcome the powerful centrifugal force of the transfer process, and the system can never move from one basin of attraction to the other. Even with random shocks in this case, there is long-memory.

The next theorem builds on the example illustrated in Figure 1 in the following sense. The properties on which it depends are precisely those which are violated by the example. A regularity condition on the distribution of ϵ values makes the proof of the theorem straightforward, and does not exclude any case likely to be of interest.

DEFINITION 3. *Let a probability density function for values of the random shocks ϵ be defined on $[-\infty, +\infty]$, and denoted $\pi(\epsilon)$. Then the distribution of shocks will be said to be* ***Regular*** *if:*
(i) $\pi(\epsilon)$ is continuous in ϵ; and
(ii) given any four values ϵ_1,ϵ_2,ϵ_3 and ϵ_4, with $\epsilon_1 < \epsilon_2 < \epsilon_3 < \epsilon_4$:

$$\int_{\epsilon_1}^{\epsilon_2} \pi(\epsilon)\, d\epsilon > 0 \text{ and } \int_{\epsilon_3}^{\epsilon_4} \pi(\epsilon)\, d\epsilon > 0 \tag{8}$$

implies:

$$\int_{\epsilon_2}^{\epsilon_3} \pi(\epsilon)\, d\epsilon > 0 \tag{9}$$

The definition excludes distributions with "holes" in them; that is ranges with zero probability density enclosed between ranges with positive probability density. Empty tails, such as are seen with the uniform distribution are not excluded.

THEOREM 1. *If*

$$k_{t+1} = h[k_t] \tag{10}$$

has a unique stable solution for a constant k $(k = h[k])$, and the values ϵ_t come from a regular i.i.d. generator, then

$$k_{t+1} = h[k_t] + \epsilon_t \tag{11}$$

defines a short-memory process.
<u>Proof</u>: *Is in two parts. Let the constant solution be $k*$. First it is shown that given any*

interval centered on $k*$, $I_\Delta = [k* - \Delta, k* + \Delta]$, *and any initial value* k_1, *the probability that* k_t *will never be in* I_Δ *tends to zero as* $t \to \infty$. *Secondly, given any closed interval* $I = [k', k'']$ *and an initial value* k_1, *the probability that* k_T *will be in* I *is a continuous function of* k_1.

These two results together imply the theorem. The first result entails that for any two starting values of k, k_1^1 *and* k_1^2, *both the paths leading from these values will enter* I_Δ *at some time (not necessarily the same time for each series) with limiting probability* 1. *Then we can reset the clocks for each realization of the process so that* k_1^1 *and* k_1^2 *will now both be in* I_Δ *at* $t = 1$. *By further choosing* Δ *suitably small, and now using the second part of the proof (continuity of the probability that* k_T *will be in* I *as a function of* k_1*), we can make the probabilities that the two paths will each be in* I *at* T *(on the respective reset clocks) as close together as desired. This contradicts long-memory.*

Part 1 *First suppose that* $k_1 > k* + \Delta$. *Then for the transfer process not to take* k *into* I_Δ, *either the average of* ϵ_t *realizations must be positive, however large* t, *or any negative realizations of* ϵ_t *must be* $\leq -2\Delta$. *Therefore* k *starting from a value* $> k* + \Delta$ *not to enter* I_Δ *eventually requires a hole in the distribution of* ϵ *at least over the interval* $[-\Delta, 0]$. *By symmetry, should* k *start from a value* $< k* + \Delta$, *a hole in the distribution of* ϵ *must be found at least over the interval* $[0, +\Delta]$. *In summary there must be a hole in the distribution of* ϵ *at least over the interval* $[-\Delta, +\Delta]$. *This contradicts the assumption that the distribution of* ϵ *is regular.*□

Part 2 *Take the interval* $I = [k', k'']$ *and an initial value* k_1. *Denote by* $Pr\langle k_t \in I \mid k_1\rangle$ *the probability that* k_t *will be in* I *conditional on* k *taking the value* k_1 *at* $t = 1$. *Then* $Pr\langle k_t \in I \mid k_1\rangle = \int_E \pi(\epsilon)$, *where* E *is the set of all realizations of the* ϵ *values such that* k_1 *is transferred to* k_t *by the process (11). Now consider* $Pr\langle k_t \in I \mid k_1 + \delta\rangle$, *where* δ *may be arbitrarily small, Then relative to any specific element of* E, *the realizations of* ϵ *may be modified so as to make* k *follow the sequence:*

$$\left\{k_1 + \delta, k_1 + \delta\frac{t-2}{t-1}, k_1 + \delta\frac{t-3}{t-1}, ..., k_t\right\} \tag{12}$$

Notice that (12) establishes a one-to-one correspondence between paths from k_1 *to* k_t *and paths from* $k_1 + \delta$ *to* k_t. *Plainly any path from* k_1 *to* k_t *has a unique partner defined by (12). Equally a path from* $k_1 + \delta$ *to* k_t *has a unique partner given by:*

$$\left\{k_1 - \delta, k_1 - \delta\frac{t-2}{t-1}, k_1 - \delta\frac{t-3}{t-1}, ..., k_t\right\} \tag{13}$$

The importance of this one-to-one correspondence lies in the fact that in integrating over all paths to k_t *to determine summed probabilities we will always be integrating over sets of equal measure regardless of starting point, so that only differences in probabilities will matter.*

For δ *sufficiently small alterations to the* ϵ *values can be made arbitrarily small, in which*

case the alteration to $\int_E \pi(\epsilon)$ may be made arbitrarily small. It follows that $Pr\langle k_t \in I \mid k_1\rangle$ is continuous in k_1 as required.$\square$
The proof is complete.$\square$

If capital accumulation is a short-memory process we reach the type of striking conclusion that if we only wait long enough, Bangladesh will be richer than the US in per capita terms at some t with probability 1. Barro's conditional convergence would apparently destroy that conclusion. However perhaps as $t \to \infty$ the US will certainly become undemocratic, closed and without secure property rights, etc.. These speculations underline the curious nature of asymptotic results.

Convergence and the Transfer Function. It is convenient to work with the logarithm of wealth because it is not bounded below by zero, which makes possible infinite-tail distributions, such as the normal. Obviously, were k to be normally distributed, wealth itself would be distributed as the log-normal distribution. Starting with the model (1), we add i.i.d. errors ϵ with mean zero to obtain:

$$k_{t+1} = h[k_t] + \epsilon_t \tag{14}$$

The equation (14) is a stochastic non-linear process in k_t. The transfer function shows how much capital would be held one period later, starting from a level k_t, were no random shock to arrive to throw the adjustment process off its intended path. The agent starts with k_t, turns that into $h[k_t]$ by saving, or dissaving as the case may be, and ends up with k_{t+1} after the shock has taken effect. One could assume:

$$k_{t+1} = h[k_t + \epsilon_t] \tag{15}$$

meaning that shocks affect wealth before the adjustment decision is made. However (14) fits most simply with existing econometric approaches.

The model is altered more radically if one assumes:

$$k_{t+1} = h[k_t, \epsilon_t] \tag{16}$$

which means that shocks affect the adjustment process in a non-linear manner. It is important in interpreting (14) to understand what is implied by the i.i.d. assumption. The additive i.i.d. shock entails that the value of $h[\cdot]$ is unaffected by the particular value taken by ϵ. That does not imply that $h[\cdot]$ is unaffected by the distributional properties of ϵ, in particular by the fact that ϵ does not always take the value zero. As $h[\cdot]$ shows an optimal saving (adjustment) rule, that rule may be influenced by the existence of uncertainty. Computing the properties of optimal saving rules under uncertainty is formidably difficult and will not be attempted. There is ambiguity for an important property of $h[\cdot]$ even when the saving decision takes no account of uncertainty. When $h[\cdot]$ shows an optimal saving rule which reflects the existence of uncertainty, it does not use information on the current value of ϵ.

For stability one must have:

$$\left[\frac{\partial h\left[k\right]}{\partial k}\right]_{k=k*} < 1 \tag{17}$$

The Transfer Function and β-Convergence. Recall that β-convergence requires that the growth rate of k should decline as k increases. So with k being the logarithm of capital, we have strict β-convergence if:

$$\frac{dh(k)}{dk} - 1 < 0 \tag{18}$$

From (17) plus continuity we must have β-convergence close to $k = k*$. Therefore if k tends asymptotically to $k*$ there will be β-convergence in the limit. With stochastic shocks there will not be strict asymptotic convergence to $k*$, so this last point is without force. In any case an asymptotic property cannot confine what may happen globally, see below.

Thinking on terms of the level of capital K, rather than its logarithm k, (18) can be written:

$$\frac{dK_{t+1}}{dK_t}\frac{1}{K_{t+1}} - \frac{1}{K_t} < 0 \tag{19}$$

which is the same as requiring that the elasticity of next period's capital with respect to current capital shall be less than unity. Can this property, which would both entail β-convergence, and also help to detail features of the stationary distribution, be assumed? An examination of this seemingly technical question throws up points which the current literature on convergence has pushed aside.

Suppose that (1) shows the outcome of the optimal wealth accumulation of a Ramsey-style agent, who maximizes:

$$\sum_{t=1}^{\infty} \delta^{t-1} U\left[f\left(K_t\right) + K_t - K_{t+1}\right] \tag{20}$$

where K is wealth itself, not its logarithm k. K_1 is given; $\delta < 1$; $U\left[\cdot\right]$ is a strictly concave utility function; $f\left(\cdot\right)$, which is the production function in per capita terms, is strictly concave, and the argument of $U\left[\cdot\right]$, $f\left(K_t\right) + K_t - K_{t+1}$ shows consumption at t.

The maximization of (20) with respect to K_t requires:

$$\delta^{t-1} U_1\left[f\left(K_t\right) + K_t - K_{t+1}\right]\left\{f_1\left(K_t\right) + 1\right\} - \delta^{t-2} U_1\left[f\left(K_{t-1}\right) + K_{t-1} - K_t\right] = 0 \tag{21}$$

where subscripts, apart from those indicating time, denote differentiation. Rearranging and showing consumption levels explicitly gives:

$$\frac{U_1[C_t]}{U_1[C_{t-1}]} = \frac{1}{\delta\{f_1(K_t)+1\}} \tag{22}$$

This is the equivalent to the optimality condition for the continuous case (see Barro and Sala-i-Martin (1995), p.63, equation (2.8)) and Blanchard and Fischer (1989), p.40 equation (7).

Taking logarithms of both sides of (22) gives:

$$\ln U_1[C_t] - \ln U_1[C_{t-1}] = -\ln\delta - \ln\{f_1(K_t)+1\} \tag{23}$$

Or,

$$\left[-C^m \frac{U_{11}[C^m]}{U_1[C^m]}\right]\left[-\frac{C_t - C_{t-1}}{C^m}\right] = -\ln\delta - \ln\{f_1(K_t)+1\} \tag{24}$$

where C^m is a value chosen to satisfy the requirement of the mean-value theorem. Denote $-C^m \frac{U_{11}[C^m]}{U_1[C^m]}$, the elasticity of marginal utility, by ζ. Equation (24) can be interpreted as follows:

$$\begin{aligned}&[\text{Elasticity of marginal utility} = \zeta]\,[\text{-Growth rate of consumption}]\\ &= -\ln\delta - \ln\{\text{Marginal product of capital}\}\end{aligned} \tag{25}$$

where the growth rate of consumption in (25) is measured on the base of a mean-value theorem level, and ζ is evaluated at the same point.

The larger is the value of K_t the smaller is $f_1(K_t)$, then the larger is the right-hand side of (25). For the left-hand side of (25) to be correspondingly larger, either the growth rate of consumption must be smaller, or ζ must be larger. If we follow Barro and Sala-i-Martin (1995, p.64) and assume ζ to be a *constant*, we arrive at a standard type of result. The growth rate of consumption must be smaller when K_t is larger. In that particular sense we have β-convergence.

To move from this β-convergence result to a conclusion concerning how the growth rate of K depends upon the level of K involves some complex calculation. See Barro and Sala-i-Martin (1995), Appendix 2C to Chapter 2, for analysis of the parallel. problem for the continuous case. Here there is no need to pursue the issue further. The supposition that the elasticity of marginal utility might be a constant is, despite its convenience and popularity, highly dubious. Therefore, if we cannot be sure that consumption will grow more rapidly when K is small, we surely cannot require that the growth rate of K should be smaller when K is large.

What the Poor Save and the Value of ζ. Empirical cross-section studies of economic growth show the hypothesis of β-convergence to be well supported for fairly homogenous cross-sections (OECD countries or US States from about 1960) and poorly supported for broader cross sections (all countries in the Summers-Heston data set). In the latter case in particular poor countries have not grown as fast as the β-convergence hypothesis would have one suppose. In a similar manner the wealth of poor families in national samples does not typically grow as rapidly as an optimistic convergence view would have one expect. Why is this?

To keep the discussion as simple as possible, the focus here will be on the saving rate of the poor, defined as the share of saving in total income. When the saving rate is independent of wealth, as in the Stigltiz (1969) model, we obtain β-convergence, and the wealth of the poor grows more rapidly than that of the rich. Under the approach of Barro and Sala-i-Martin, where growth is determined by consumer optimization, we again arrive at β-convergence, as long as ζ is constant. Then these authors, and Barro (1997), allow numerous extra variables, from the share of government expenditure in output to religion, to explain why poor countries do not grow as fast as a simple β-convergence model would predict.

Everything here hangs on the idea that the wealth of poor units will grow rapidly because:

- their saving causes their capital to grow faster, because they have less capital income relative to their labour income;
- they will enjoy a higher return to capital because they have little; and
- their utility discount rate will be the same as everyone else's

These points and equation (18) imply the consumption of the poor growing faster than that of the rich. Yet the pain of saving when one is poor is not only measured by the utility discount rate and the rate at which consumption grows over time. Inspection of equation (18) reveals that if capital holding is low, when the marginal product of capital will be high; and given the value of δ, the growth rate of consumption may yet be low if ζ should be large.

This is not an unreasonable case. The value of ζ could be said to measure the ease with which the intertemporal substitution of consumption takes place. A high value of ζ corresponds to intertemporal substitution being difficult. Imagine an agent poor and hungry. He can eat a little less and is guaranteed a high return on any such saving. In comparing his current marginal utility with its future value he is no more subject to myopic discounting than any one else (he shares the value of δ that applies to richer individuals). Yet the issue remains of how fast his consumption has to rise over time to equate the ratio of his discounted future marginal utility to his current marginal utility with the high net return on his saving. That growth rate of consumption may be low if ζ is low.

In this connection the argument of Barro and Sala-i-Martin (1995), which claims correctly that ζ must be constant asymptotically, is irrelevant. As all units in this type of model must converge in the limit to the same level of consumption, ζ must similarly converge. That proposition says nothing about the global constancy of ζ. Without assuming ζ to be invariant, we lose any guarantee of β-convergence, and we have to abandon the imposition of simple restrictions on the form of $h(k)$; beyond its root property of being an increasing function.

To conclude this part of the argument, the description of a stationary distribution for wealth needs to take into account wider possibilities for the transfer function than are found in a model which gives β-convergence. If the transfer function for the logarithm of wealth should be linear, β-convergence follows [see equation (18) above]. However non-linearities in the transfer function greatly enrich the range of possible stationary wealth distributions and they have to be taken into account.

Convergence and Scattering. The effect on the distribution of wealth in moving from one period to the next is the sum of two separate transformations. First each k value maps to $h\left[k\right]$. This is optimal adjustment without any shock; called *h-transfer*. Next all values are scattered by the addition of random shocks ϵ_t. We call this *scattering*.

Applying the mathematical expectation operator E to (3) gives:

$$Ek = Eh\left[k\right] \tag{26}$$

The relation of $Eh\left[k\right]$ to $h\left[Ek\right]$ depends upon the concavity/convexity of $h\left[\cdot\right]$, which is ambiguous. Subtracting (3) from (25) and rearranging gives:

$$\begin{gathered} E\left[k_t - Ek\right]^2 = E\left\{\left[h\left[k_t\right] - Eh\left[k\right] + \epsilon_t\right]^2\right\} \\ = E\left\{h\left[k\right]^2\right\} + E\left\langle\left\{Eh\left[k\right]\right\}^2\right\rangle + E\left\{\epsilon_t^2\right\} - 2E\left\{h\left[k\right]Eh\left[k\right]\right\} \\ = E\left\{\epsilon^2\right\} + E\left\{h\left[k\right]^2\right\} - \left\{Eh\left[k\right]\right\}^2 \end{gathered} \tag{27}$$

where time subscripts have been dropped, because they are irrelevant when a stationary distribution is under consideration. On account of ϵ_t being i.i.d., expectations of products involving ϵ_t have been equated to zero.

Notice that the variance of $h(k)$ is given by:

$$\begin{aligned} E\left\{h(k) - Eh(k)\right\}^2 &= E\left\{h(k)^2 + \left[Eh(k)\right]^2 - 2h(k)Eh(k)\right\} \\ &= E\left\{h(k)^2\right\} - \left\{Eh(k)\right\}^2 \end{aligned} \tag{28}$$

Now equations (27) and (28) together can be interpreted in a very natural result.

THEOREM 2. *An $h-$transformation always subtracts variance from the distribution of k. For a stationary distribution it subtracts precisely the amount of variance that is added by scattering.*
Proof: *Notice that the result is not trivial. While $h-$transformation obviously moves every k closer to $k*$, there is no immediate guarantee that it moves every k closer to the mean of the k values. However from (27) the second moment of the distribution of k in general, and hence the same moment in a stationary distribution, is the sum of the variances of $h(k)$ and of ϵ. In that case, k itself must have a larger variance than $h(k)$. Evidently scattering restores equality, as required.*□

Similar calculations for the third moment of the distribution of k, assuming $E\left\{\epsilon^3\right\}=0$, produce:

$$E\left\{k-Ek\right\}^3 = E\left\{h(k)^3\right\} - 3E\left\{h(k)\right\}E\left\{h(k)^2\right\} + 2\left[E\left\{h(k)\right\}\right]^3 \tag{29}$$

THEOREM 3. *If the distribution of shocks ϵ_t has a third moment about its mean equal to zero; hence in particular if it is symmetrical about zero; an $h-$transformation applied to a stationary distribution does not affect the third moment of k about its mean.*
Proof: *The right-hand side of (29) is the third moment $h(k)$ about its mean. Therefore the result follows immediately.*□

The Shape of the Stationary Distribution: Transfer Plus Scattering. In this section, like the last, the joint effects of h-transfer and scattering are taken into account without making simplifying assumptions on either side. Thus confers the benefit of great generality. The cost is that one is then confronted with the complex product of the two effects, in manner which will be made precise. In that case it is not always easy to see how the separate influences of convergence and scattering affect the shape of a stationary wealth distribution. In later sections simpler cases will be displayed which make the separate effect of one or other of the two influences more transparent.

In analysing the distribution of k values, it is sometimes convenient to work in terms of the cumulative distribution. Hence $\Delta(k)$ is the proportion of the population with wealth not greater than k. Clearly $\Delta(-\infty)=0$ and $\Delta(\infty)=1$.

Recall that the effect on the distribution of wealth in moving from one period to the next is the sum of two separate transformations. First each k value maps to $h\left[k\right]$; that is *h-transformation*. Next all values are *scattered* by the addition of random shocks ϵ_t. Consider the first step. Before h-transformation the probability density of k, $\Lambda\left[k\right]$, is given by:

$$\Lambda\left[k\right] = \frac{d\Delta\left[k\right]}{dk} \tag{30}$$

Whereas the cumulative distribution of k after h-transformation, $\Gamma\left[k\right]$, satisfies:

$$\Gamma[k] = \Delta\left[h^{-1}[k]\right] \tag{31}$$

Then:

$$\frac{d\Gamma[k]}{dk} = \frac{d\Delta[h^{-1}[k]]}{d[h^{-1}[k]]}\frac{d[h^{-1}[k]]}{dk} = \frac{\Lambda[h^{-1}[k]]}{\frac{dh[k]}{dk}} \tag{32}$$

is the density of wealth distribution after $h-$transformation. Equation (32) defines how the adjustment function affects the distribution of wealth in the absence of random effects.

Denote the transformed distribution by $\Phi(k)$. So:

$$\Phi(k) = \frac{\Lambda[h^{-1}[k]]}{\frac{dh[k]}{dk}} \tag{33}$$

Consider a maximum of $\Phi(k)$ at $k = k^0$. Then:

$$\left\{\frac{d\Lambda[h^{-1}[k]]}{dk} - \Lambda\left[h^{-1}[k]\right]\frac{d^2h[k]}{dk^2}\right\}_{k=k^0} = 0 \tag{34}$$

Equation (34) is useful when locating a maximum, including a mode, of a wealth distribution after $h-$transformation when the location of a maximum of $\Lambda[k]$ is known. Suppose, for instance, that $h[k]$ is so nearly linear in the relevant range that $\frac{d^2h[k]}{dk^2}$ may be replaced by zero. Then (34) says that one should look for a maximum of $\Phi(k)$ to the left (right) of a maximum of $\Lambda[k]$ according as k is less than (greater than) $k*$.

The sequential effects of $h-$transfer and scattering in a stationary case can be exhibited mathematically as follows. Take any value of k. Suppose $k < k*$. A symmetrical argument works for the other side. For any level of wealth between $h^{-1}[k]$ and k, $h-$transfer will carry wealth across the border marked by k from lower to higher values of wealth. Next, after $h-$transformation, scattering will carry a certain mass of wealth across the same border, travelling in the same direction, while scattering will also carry another mass of wealth across the border in the opposite direction. It is an evident equilibrium condition for a stationary distribution that the net movement of wealth across the border shall be zero. That condition is expressed in the following equation.

$$\int_{h^{-1}[k]}^{k}\Lambda[\kappa]\,d\kappa + \int_{-\infty}^{k}\{1-\Pi(k-\kappa)\}\,\Phi(\kappa)d\kappa = \int_{k}^{+\infty}\Pi(k-\kappa)\,\Phi(\kappa)d\kappa \tag{35}$$

where $\Pi(\cdot)$ is the cumulative distribution of i.i.d. shocks; that is the probability that ϵ_t will be $\leq$ the argument of $\Pi(\cdot)$.

As (35) holds as an identity in k, we may differentiate it with respect to k to obtain:

$$\Lambda[k] - \frac{\Lambda[h^{-1}[k]]}{\frac{dh[k]}{dk}} + \{1-\Pi(0)\}\,\Phi(k) - \int_{-\infty}^{k}\pi(k-\kappa)\,\Phi(\kappa)d\kappa$$

$$= -\Pi(0)\,\Phi(k) + \int_{k}^{+\infty} \pi(k-\kappa)\,\Phi(\kappa)d\kappa \tag{36}$$

Notice that if the distribution of shocks is symmetrical about 0, then $\Pi(0) = \frac{1}{2}$. However the argument does not use that property. Simplifying (36) gives:

$$\Lambda[k] = \int_{-\infty}^{+\infty} \pi(k-\kappa)\,\frac{\Lambda\left[h^{-1}[\kappa]\right]}{\frac{dh[\kappa]}{d\kappa}}d\kappa \tag{37}$$

Changing variable in (37) - $\lambda = h^{-1}(\kappa)$ - provides another integral equation description of a stationary equilibrium.

$$\Lambda[k] = \int_{-\infty}^{+\infty} \pi[k - h[\lambda]] \cdot \Lambda[\lambda]\,d\lambda \tag{38}$$

Or, denoting the variable of integration again by κ:

$$\Lambda[k] = \int_{-\infty}^{+\infty} \pi[k - h[\kappa]] \cdot \Lambda[\kappa]\,d\kappa \tag{39}$$

The integral on the right-hand side of (39) is the sum of all transitions from κ to k weighted by the probability that the initial value is κ, which is $\Lambda[\kappa]$, and the probability of a transition to k, which is the probability that ϵ_t takes the value $k - h[\kappa]$. Placing the same function $\Lambda[\cdot]$ on both sides of (39) identifies a stationary fixed point outcome.

Equation (39) is a *Fredholm Equation* of the second kind[6]. This derivation is somewhat similar to the so-called *Theory of Breakage* which leads to the equation:

$$F_j(x) = \int_{u} H_j\left[\frac{x}{u}\right] dF_{j-1}[u] \tag{40}$$

for which see Aitchison and Brown (1957), pp.26-7. The observation that the stationary states of a Markov stochastic process can be expressed by a Fredholm equation is not new. Furia (1982) uses the topological generalization of a Fredholm equation for a similar purpose, although he does not employ the same terminology.

The process:

$$k_{t+1} = h[k_t + \epsilon_t] \tag{41}$$

generates another *Fredholm Equation, viz:*

$$\Lambda[k] = \int_{-\infty}^{+\infty} \pi\left[h^{-1}[k] - \kappa\right] \cdot \Lambda[\kappa]\,d\kappa \tag{42}$$

which is quite similar.

To keep things simple, we concentrate below on the Fredholm Equation (39).

[6]See Hildebrand (1961) p. 381-2. In section 4.5 of the same chapter the author explains the connection between this type of equation and the joint effect of many causes.

Results from the Fredholm Equation. To be able to write down an equation showing a stationary solution as (39) is encouraging. Unfortunately this equation cannot be solved for $\Lambda[k]$. However it yields two useful results.

Theorem 4. *The set of functions satisfying (39) is convex*[7].
Proof: Is immediate. If $\Lambda^1[k]$ and $\Lambda^2[k]$ both satisfy (39), then:

$$\lambda \cdot \Lambda^i[k] = \int_{-\infty}^{+\infty} \pi[k - h[\kappa]] \cdot \lambda \cdot \Lambda^i[\kappa]\, d\kappa \tag{43}$$

for $i = 1$ or 2, and for any value of λ. Hence:

$$\lambda \cdot \Lambda^1[k] + (1-\lambda) \cdot \Lambda^2[k]$$
$$= \int_{-\infty}^{+\infty} \pi[k - h[\kappa]] \cdot \left\{\lambda \cdot \Lambda^1[\kappa] + (1-\lambda) \cdot \Lambda^1[\kappa]\right\} d\kappa \tag{44}$$

□

The next theorem uses the Fredholm equation to establish continuity of $\Lambda[k]$ with respect to k. It assumes that $\pi[\cdot]$ is uniformly continuous. For a probability density function this is a mild condition. For note that, because any continuous function is uniformly continuous on a compact support; if there is any problem with uniform continuity of $\pi[\cdot]$, it can only arise from extraordinary behaviour of the function in its tails.

Theorem 5. *If $\pi[\cdot]$ is uniformly continuous, a stationary distribution value for $\Lambda[k]$ is continuous in k.*
Proof: From (33) it will be seen that for a stationary distribution of k, the continuity of $\Lambda[k]$ is implied by the continuity of:

$$\int_{-\infty}^{+\infty} \pi[k - h[\kappa]] \cdot \Lambda[\kappa]\, d\kappa \tag{45}$$

Take a sequence of values $k_1, k_2, ..., k_n, ..$ with limit k. Then the sequence of values

$$\Lambda[k] - \Lambda[k_1], \Lambda[k] - \Lambda[k_2], ..., \Lambda[k] - \Lambda[k_n], .. \tag{46}$$

are given by terms of the form:

$$\int_{-\infty}^{+\infty} \{\pi[k - h[\kappa]] - \pi[k_n - h[\kappa]]\} \cdot \Lambda[\kappa]\, d\kappa \tag{47}$$

which are less than or equal to:

$$\int_{-\infty}^{+\infty} |\pi[k - h[\kappa]] - \pi[k_n - h[\kappa]]| \cdot \Lambda[\kappa]\, d\kappa \tag{48}$$

[7] To say that the set of functions is convex is not, of course, to say that the functions are convex functions.

From uniform continuity it follows that for any $\theta > 0$, $|k - k_n|$ sufficiently small implies that (48) is less than:

$$\theta \int_{-\infty}^{+\infty} \Lambda[\kappa]\, d\kappa = \theta \tag{49}$$

From which the continuity of $\Lambda[k]$ follows.□

Continuity of $\Lambda[k]$ does not rule out the possibility that an stationary density might split into two or more disjoint segments: say a high wealth segment with positive density; a low wealth segment with positive density; and a region between these two where $\Lambda[k] = 0$. We have encountered such a possibility with the long-memory example of page ?? above. There in the context of long/short memory we saw that in the limit k will be found in one of two (but in general it could be three or more) zones of attraction, and once inside such a zone k will always remain in that zone. It is immediate that this situation can only arise if the distribution of shocks ϵ has one or more bounded tails. Otherwise a shock greater than or equal in absolute value to that required to throw the system form one zone to another will occur with probability 1 in the limit, and the system will have short-memory.

Definition 4. *The distribution of shocks ϵ will be said to be* **regular** *if the set of closed intervals over which the probability measure for ϵ is positive is a convex set.*

The definition will surely be satisfied by any distribution likely to be of interest. Here specifically the definition excludes the possibility that zero probability will attach to a closed interval of values of ϵ, while positive probability attaches to closed intervals for values of ϵ both above and below that interval.

Theorem 6. *If the distribution of shocks ϵ is regular, and*

$$k_{t+1} = h[k_t] \tag{50}$$

is a short-memory process, a stationary density cannot be disjoint.

Proof: *If the distribution is disjoint there will exist at least two open intervals of values of k such that $\Lambda[k] > 0$ for values of k in those intervals. Call the said intervals $I^{''}$ and $I^{\cdot''}$. Between $I^{'}$ and $I^{''}$ will be found an interval such that $\Lambda[k] = 0$ for all values of k in that interval. Call this last interval $^{I'''}$. Because the process (50) is short-memory, k will transit between $I^{'}$ and $I^{''}$ with probability 1 during any infinite history. Consider such a transit which will satisfy:*

$$k_2 = h[k_1] + \epsilon_1 \tag{51}$$

for $k_1 \in I^{'}$ and $k_2 \in I^{''}$. Then a transit from $k_1 \in I^{'}$ to $k_3 \in I^{'''}$ could occur, with k_3 closer to k_1 than is k_2, if ϵ were to take a value closer in absolute value to zero than is ϵ_1. As the mean of ϵ is zero, the support of ϵ must include both positive and negative intervals. Hence *because the distribution of shocks is regular, positive probability must attach to transits to I'''. Then there cannot be zero probability density on I''', not in the limit of any history, hence not in particular for a stationary distribution.*$\square$

If the transfer function plus the distribution of shocks amount to a long memory process, then of course the distribution may be disjoint and non-uniqueness of the stationary distribution will be found. This is clear from the disjoint Ramsey economies discussed above.

The Shape of the Stationary Distribution: Special Cases. The general analysis of Section 7 above denies us simple insights into the way in which the stationary distribution is shaped by the separate forces operating on it. For that understanding it is helpful to look at models which are designed to isolate particular effects, while other influences which combine to make up the Fredholm integral are muted by simplifying special assumptions.

In the following section the focus is on the form of the transfer function, in particular the effect of non-linearities in the transfer function; so only the most simple specifications for the distribution of shocks are admitted. In section after the next the focus is on the influence of the shocks themselves. In that case a linear transfer function is ideal, and the outcome with that assumption will be examined.

The Shape of the Stationary Distribution: The form of the Transfer Function. To elucidate the stationary distributional properties of a variable generated by the stochastic process (2), the following strange, yet understandable, hydraulic model may be helpful.

A Hydraulic Model. In the centre is a rift valley, running due North-South, and viewed in cross section from the South. Rivers flow down from highlands on the east side and from the west. Position is measured by a variable k which runs from $-\infty$ (indefinitely far west) to $+\infty$ (indefinitely far east).

These are not normal rivers, fed by springs, and rainfall originating outside the river system. The system is completely closed. All rainfall originates from water in the rivers themselves. Evaporation constantly redistributes water within the system. The amount of water evaporated depends on the volume at a point. One molecule of water may travel any distance, east or west. The probability of any such journey depends upon the absolute distance travelled, and it decreases monotonically with absolute distance. Elevation as such has no effect on precipitation. Indeed the high highlands are dry, because they are far from the great mass of water. Finally water runs down hill and it runs faster the steeper the absolute gradient.

The bottom of the rift valley is at $k*$. The flow of river water towards the valley represents non-stochastic transformation of values of k through the function $h[k]$. Evaporation and the random redistribution of water represent the effect of i.i.d. shocks, which are called *scattering* as above. The depth of water at any point k represents the density of wealth at that point. When this hydraulic system is in a stationary state, depth is constant at any point. The rivers flow always towards $k*$. However evaporation and the random redistribution of water frustrate that process. A deep lake may build up around $k*$. Yet if redistribution is significant, the lake can never contain all the water in the system, because redistribution will always throw some water back into the highlands.

The hydraulic model offers a helpful mental picture of how the type of non-linear stochastic process under consideration in this paper might appear in a stationary equilibrium. It is plain that non-linearities in the slopes of the valley walls will shape the stationary distribution, almost as a potter's hands shape the final pot. More detail on how non-linearities have their effects will follow below.

The Shape of the Stationary Distribution: The Influence of Shocks. Take an arbitrary value of k, k_0, and follow its random path as it is repeatedly transformed by the process (3). This can be written as:

$$k_{t+1} - k* = h\left[k_t\right] - h\left[k*\right] + \epsilon_t \tag{52}$$

Or,

$$\widetilde{k}_{t+1} = \widetilde{k}_t h'\left[k_t^M\right] + \epsilon_t \tag{53}$$

where $\widetilde{k}_t = k_t - k*$, the prime $'$ denotes differentiation, and k_t^M is the value of k which makes (53) correct. The mean-value theorem says that such a value lying between k and $k*$ always exists. From (53):

$$\widetilde{k}_{t+1} = \left\{\widetilde{k}_{t-1} h'\left[k_{t-1}^M\right] + \epsilon_{t-1}\right\} h'\left[k_t^M\right] + \epsilon_t \tag{54}$$

Which simplifies to:

$$\widetilde{k}_{t+1} = \widetilde{k}_{t-1} \Pi_{s=t-1}^{t} h'\left[k_s^M\right] + \epsilon_{t-1} h'\left[k_t^M\right] + \epsilon_t \tag{55}$$

Similarly, repeated substitutions give:

$$\widetilde{k}_{t+1} = \widetilde{k}_0 \Pi_{s=1}^{t} h'\left[k_s^M\right] + S \tag{56}$$

where S is equal to:

$$\epsilon_t + \epsilon_{t-1} h'\left[k_t^M\right] + \epsilon_{t-2} h'\left[k_t^M\right] h'\left[k_{t-1}^M\right] + .. + \epsilon_1 \Pi_{q=2}^{t} h'\left[k_q^M\right] \tag{57}$$

If the unshocked system is globally stable, the first term on the right-hand side of (56) will go to zero as $t \rightarrow \infty$. A sufficient condition for that property is:

$$h'\left[k\right] \leq \zeta < 1 \text{ all } k \tag{58}$$

While (5) ensures that (58) is satisfied at $k = k*$, the fact that $h[\cdot]$ may have any concavity/convexity, shown above, implies that is not necessarily satisfied everywhere.

Given global stability and that the first term of the right-hand side of (56) goes to zero, the limit of (56) as $t \to \infty$ gives the frequency distribution of k. It would be incorrect to conclude that because this distribution is independent of k_0, the process must be short-memory. So long as $h[\cdot]$ is non-linear, the influence of k_0 may make itself felt via the particular values of $h'\left[k_s^M\right]$ that appear in (57). Should $h[\cdot]$ be linear, which possibility cannot be excluded, the limiting distribution of k depends only upon the distribution of the shock values ϵ and upon powers of the constant slope coefficient.

Equation (57) can be employed to give insight into how non-linearity in the adjustment function $h[k]$ translates into asymmetry in the stationary distribution. Suppose that $h[k]$ is linear to the left of $k*$, and also to the right of $k*$, but the two slopes differ. Let the slope to the left, h^l, be larger than the slope to the right, h^r. As negative realizations of the random variable ϵ will on average be associated with $k < k*$, and positive realizations of the random variable ϵ will on average be associated with $k > k*$, (57) can be read to say that negative values of ϵ will be more heavily weighted, and the stationary distribution will have greater density to the left of $k*$ than to the right. The fact that convergence to $k*$ is more rapid from the right than from the left accounts for this asymmetry.

Figure 2 illustrates this case in which non-linearity of $h(k)$ combines with symmetrical shocks to produce an asymmetrical distribution. If capital is at the point A, and ignoring shocks at this point, it is seen that increase in k is relatively rapid. Starting from point B, however, the decline in capital is slower. These effects will cause the stationary distribution to bunch to the right of $k*$ and to spread to the left of $k*$, as required.

Equation (57) provides further insight if we transfer attention to another simple case. Now $h[k]$ is linear, but the distribution of ϵ values is asymmetric about zero. It is clear that the asymmetry of the distribution $\pi[\epsilon]$ is reflected in a similar asymmetry in the stationary distribution of k values. In the context of the theory of wealth accumulation this possibility is intriguing. We can always make the expected value of ϵ equal zero, as adding or subtracting a constant to the ϵ values, and subtracting or adding the same value to $h[k]$ makes no difference. That point does not dispose of the possibility that higher probability density may attach to large negative shocks (bad growth set-backs) than to large positive shocks of similar size. In the type of case just described the stationary distribution will be fat to the left of $k*$ relative to its density at a similar distance from $k*$ to the right.

Theorem 7. *If the stochastic process is standard and $h[k]$ is linear, a stationary distribution is symmetric with its centre at $k*$.*
Proof: If $h(k)$ is linear, it follows that $h'(k)$ in (57) will be a constant, denoted H $(0 < H < 1)$, and that this expression will become:

$$\epsilon_t + \epsilon_{t-1}H + \epsilon_{t-2}H_t^2 + .. + \epsilon_1 H^{t-1} \tag{59}$$

In this case the process is short-memory, in the sense that the limiting probability that k lies in any interval is independent of its initial value. Suppose that the limiting probability that k lies in the closed interval of positive values $[k^-, k^+]$ is p_0. This is equivalent to:

$$Lim_{t\to\infty} \int_P \left\{\epsilon_t + \epsilon_{t-1}H + .. + \epsilon_1 H^{t-1}\right\} \Pi_{\theta=1}^t \pi\left[\epsilon_\theta\right] = p_0 \tag{60}$$

where the integration over P in (60) is over all values of ϵ_θ such that the integral, and hence its limit, is equal to p_0 . In that case it seems that (60) says nothing. It being the case that the integration on the left-hand side is over all values of ϵ such that the integral takes the value p_0, it conveys no information to state that all the integrals, and hence their limit as $t \to \infty$, take the value p_0. For our present purposes however, what matters is not that (60) is satisfied, but rather the manner, exhibited in (60), in which the various values of ϵ and their probability densities combine to produce a value p_0. Consider the closed interval of negative values $[-k^+, -k^-]$, and the limit of a sequence of integrals:

$$Lim_{t\to\infty} \int_{P-} \left\{\epsilon_t + \epsilon_{t-1}H + .. + \epsilon_1 H^{t-1}\right\} \Pi_{\theta=1}^t \pi\left[\epsilon_\theta\right] \tag{61}$$

where integration over $P-$ is over all values of ϵ_θ such that k lies in the closed interval of negative values $[k^-, k^+]$. By symmetry of the density function $\pi[\epsilon]$, (61) takes the value p_0, which is the result required.□

The Shape of the Stationary Distribution:
Asymmetry and Single-Peakedness.

Definition 5. *The wealth distribution will be said to be single-peaked if all its local maxima are attained on one convex set of values of k.*

The definition allows a "table mountain" case in which the maximum value is attained over a connected range of values of k. That case apart, the definition rules out multiple local maxima as distinct peaks. In the standard case, a linear $h[k]$ function produces a symmetric stationary distribution. That does not by itself imply single-peakedness, as a symmetric distribution might have many isolated local maxima. However an argument similar to the proof of Theorem 7 shows that if the distribution of ϵ values is symmetric around zero with probability density a declining function of the absolute distance from zero, a stationary distribution of k values has the same qualitative form.

Without excluding any case likely to be of interest, we may confine attention to distributions of ϵ values which are centred on zero and with density monotonically decreasing in the absolute distance from zero, but not necessarily at the same rate for positive and negative values of ϵ. Within that class of cases, we may, by collecting implications of results derived above, throw considerable light on the question of asymmetry in a stationary distribution of k values. Asymmetry can come about only from at least one of the following features:

- non-linearity of the adjustment function $h[k]$;
- asymmetry of the density $\pi[\epsilon]$.

The hydraulic system described in above can also throw light on single-peakedness. Suppose that the steepness of the rift valley walls on both sides increases monotonically with absolute distance from $k*$, and is symmetrical on the two sides. Far from $k*$ water is moved quickly towards $k*$. All the water that far out has been transported a long distance. There cannot be much of it, and swift running rivers must be shallow. As one moves closer to the floor of the valley, the absolute gradient becomes lower and rainfall rises, because the total water not too far away increases. Now rivers flow slowly and are deep. Therefore the depth of water rises until it reaches its maximum at $k*$. The density of water is symmetric around $k*$.

Now modify the model just described. On the west side insert a range of values of k along which the gradient is quite flat. Follow it by a very steep interval closer to $k*$, and then return to a similar gradient to that prevailing on the opposite valley wall. The amount of water above these ranges will hardly be affected if redistribution is strong and the ranges described cover short intervals. Therefore water will move through the intervals first slowly, next rapidly, then it will slow down. Depth will be high, then lower, then high again. The depth of water, which is to say the density of wealth, will exhibit twin peaks.

For the accumulation of wealth the model just described corresponds to the following state of affairs. For a range of low levels of wealth, is accumulated towards $k*$, but at a slow rate. Then, when wealth gets a bit higher, the pace of accumulation picks up sharply. Later it moderates. If we allow the elasticity of marginal utility to vary with wealth, economic theory cannot exclude such a case. The only way to avoid a twin-peak outcome in such a case is to have a high density in the steep (fast-flowing) section. That will never be a stationary equilibrium because that high density would be rapidly dissipated by flow towards $k*$ which rainfall could not replace.

An informal mathematical version of this pictorial argument runs as follows. Take a regular model with no twin-peaks in a stationary wealth distribution. Over a range of values of $k < k*$ which is small relative to a range which contains much of the density of $\pi[\cdot]$, distort the $h[\cdot]$ function so as to make its derivative large. Figure 3 sketches this case. Recall that with k the logarithm of wealth, the value $h(k) - k$ measures the rate of growth of capital. In the figure this growth rate is the height of the thick non-linear curve above the linear 45^0 line through the origin. It will be seen that starting from the lowest levels of capital the growth rate of capital is successively rapid, slow, rapid slow. Ranges with these respective qualities are shown on the figure by the letters R and S. Given a suitable distribution of shocks the case illustrated points to the possibility of a twin-peak stationary distribution.

An example of this type depends upon the magnitude of $\frac{dh[k]}{dk}$ varying considerably over a narrow range: first rising then falling. It has been argued above that such severe non-linearities cannot be assumed away. Within the family of standard Ramsey growth models are to found examples in which growth proceeds slowly for very poor units, then rapidly for medium-income units, and the again slowly closer to $k*$. To say that such an example contradicts the constancy of the elasticity of marginal utility is like saying that total sales revenue for a market falling as total sales increase contradicts the assumption that the elasticity of demand for that market is unity. The statement is correct but uninformative.

When Quah[8] published his empirical evidence showing the twin-peak pattern in international cross section per capita income data, I read it as evidence against the simple convergence model, as no doubt did many other readers. It is interesting to note that Quah himself advances no such claim. First he is very clear that he is describing the development of income distribution over a short period of time. Secondly, Quah is aware of the possibility that even the apparently disconnected distributions he observes may be generated by a process which in the long run is ergodic. Now the theoretical investigation of a stationary wealth distribution has shown that it may have twin (indeed multiple) local peaks. So it seems that even such a surprising feature may be completely consistent with a standard convergence model.

That is not a good way of looking at matters. To be worthy of study, an economic model has not only to be true in some high abstract sense; it has to be useful. A twin peak case can only arise when $h[k]$ is severely non-linear. The estimation of such a model presents many difficulties. In a way the strength of the Baumol-Barro convergence model is its crude simplicity. If it has to be rescued by refined mathematical argument, it loses its appeal. Also, in the example, and more generally, twin peaks in a stationary distribution can only happen over a range within the reach of a single-period realization of the random shock. Therefore if twin peaks are an important feature of the distribution, it must be the case that shocks are large in absolute value. This is another way of saying that the explanatory power of the model is weak.

Concluding Remarks. The long history of the analysis of income or wealth distributions, going back to Pareto, includes different approaches. One is purely empirical. The shape of the distribution is examined and the fitness of simple mathematical specifications is investigated. Another approach is to start with postulates concerning the process which generates the distribution and then to investigate mathematically what is the limiting distribution which results. Yet the limiting distribution does not have to be the object of concern. The shorter term conditional transfer process can itself be the focus of investigation. Indeed for some neoclassical convergence theorists that is all that can be done, because for them the limiting distribution is trivial, being a state in which all countries - or

[8] See Quah (1996a) and (1996b).

individuals in the case of a personal distribution - are at the common limit point $k*$. When the adjustment process is taken to include random effects there are wider possibilities than when it is modelled using non-stochastic economic theory.

The present paper marries two different traditions. They are the pure neoclassical approach, according to which wealth accumulation is systematic and deliberate; and the random shocks approach, according to which wealth accumulation is purely haphazard. As would be expected, such a model is complicated, and direct mathematical solution is hardly possible. Even so, we have been able to obtain a series of results which together reveal many features of a stationary distribution of wealth levels.

Bibliography

[1] Aitchison, J. and J.A.C. Brown (1957), *The Lognormal Distribution*, Cambridge University Press

[2] Barro, Robert J. (1991), "Economic growth in a cross section of countries", *Quarterly Journal of Economics*, CVI, 407-444

[3] Barro, Robert J. (1997), *Determinants of Economic Growth: A Cross-Country Comparative Study,* MIT Press

[4] Barro, Robert J. and X. Sala-i-Martin (1992), "Convergence", *Journal of Political Economy*, 100, 223-251

[5] Barro, Robert J. and X. Sala-i-Martin (1995), *Economic Growth*, McGraw-Hill

[6] Binder, Michael and Pesaran, M. Hashem (1999), "Stochastic Growth Models and their Econometric Implications", *Journal of Economic Growth*, Vol. 4, 139-183

[7] Blanchard, Olivier Jean and Fischer, Stanley (1989) *Lectures in Macroeconomics*, MIT Press

[8] Bliss C. (1999), "Galton's Fallacy and Economic Convergence", *Oxford Economic Papers*, 51, January, pp 4-14

[9] Bliss C. (2000), "Galton's Fallacy and Economic Convergence: A Reply ", *Oxford Economic Papers*, 52, April, pp. 420-422

[10] Champernowne, D.G. (1953), "A Model of Income Distribution", *Economic Journal*, **63**, 318

[11] David, Paul (1985), "Clio and the Economics of QWERTY", American Economic Review, 75:2(May), pp. 332-337

[12] David, Paul (1993), "Historical Economics in the Long-Run: Some implications of Path-Dependence", in G.D. Snooks, ed., *Historical Analysis in Economics*, London, Routledge, pp. 29-40

[13] DeLong, Bradford (1999), Brad DeLong's Website, http://www.j-bradford-delong.net/Teaching, "Reading Notes for August 25, Introduction".

[14] Steven N. Durlauf (1993), "Nonergodic economic growth", *Review of Economic Studies*, 60(2):349-366

[15] Steven N. Durlauf (1996), "A theory of persistent income inequality", *Journal of Economic Growth*, 1(1):75-93

[16] Friedman, M. (1992), "Do Old Fallacies Ever Die?", *Journal of Economic Literature*, XXX, December, 2129-32

[17] Carl Futia (1982), "Invariant distributions and the limiting behaviour of Markovian economic models", *Econometrica*, 50(1):377-408

[18] Granger, Clive W.J. and Terasvirta, Timo (1993), *Modelling Nonlinear Economic Relationships*, New York; Toronto and Melbourne: Oxford University Press

[19] Hamilton, James D. (1994), *Time Series Analysis*, Princeton University Press

[20] Hildebrand, F.B. (1961), *Methods of Applied Mathematics*, Englewood Cliffs, N.J., Prentice-Hall.

[21] Øksendal, B. (1998), *Stochastic differential equations*, 5th Edition, Springer-Verlag, New York

[22] Quah, Danny (1993), "Galton's Fallacy and Tests of the Convergence Hypothesis", *Scandinavian Journal of Economics*, 95(4): 427-433

[23] Quah, Danny (1996a), "Twin Peaks: Growth and convergence in models of distribution dynamics", *Economic Journal*, 106(437):1045-55.

[24] Quah, Danny (1996b), "Empirics for Economic Growth and Convergence", *European Economic Review*, 40(6):1353-1376.

[25] Quah, Danny (2005), *Growth and Distribution*, incomplete manuscript available from Danny Quah's homepage.

[26] Sala-i-Martin, X. (1996), "The Classical Approach to Convergence Analysis", *Economic Journal*, 106, July, 1019-36.

[27] Steindl, J. (1972), "The Distribution of Wealth after a Model of Wold and Whittle", *Review of Economic Studies*, 39(3) July, 263-79

[28] Wold, H.O.A. and P. Whittle (1957), "A Model Explaining the Pareto Distribution of Wealth", *Econometrica*, October.

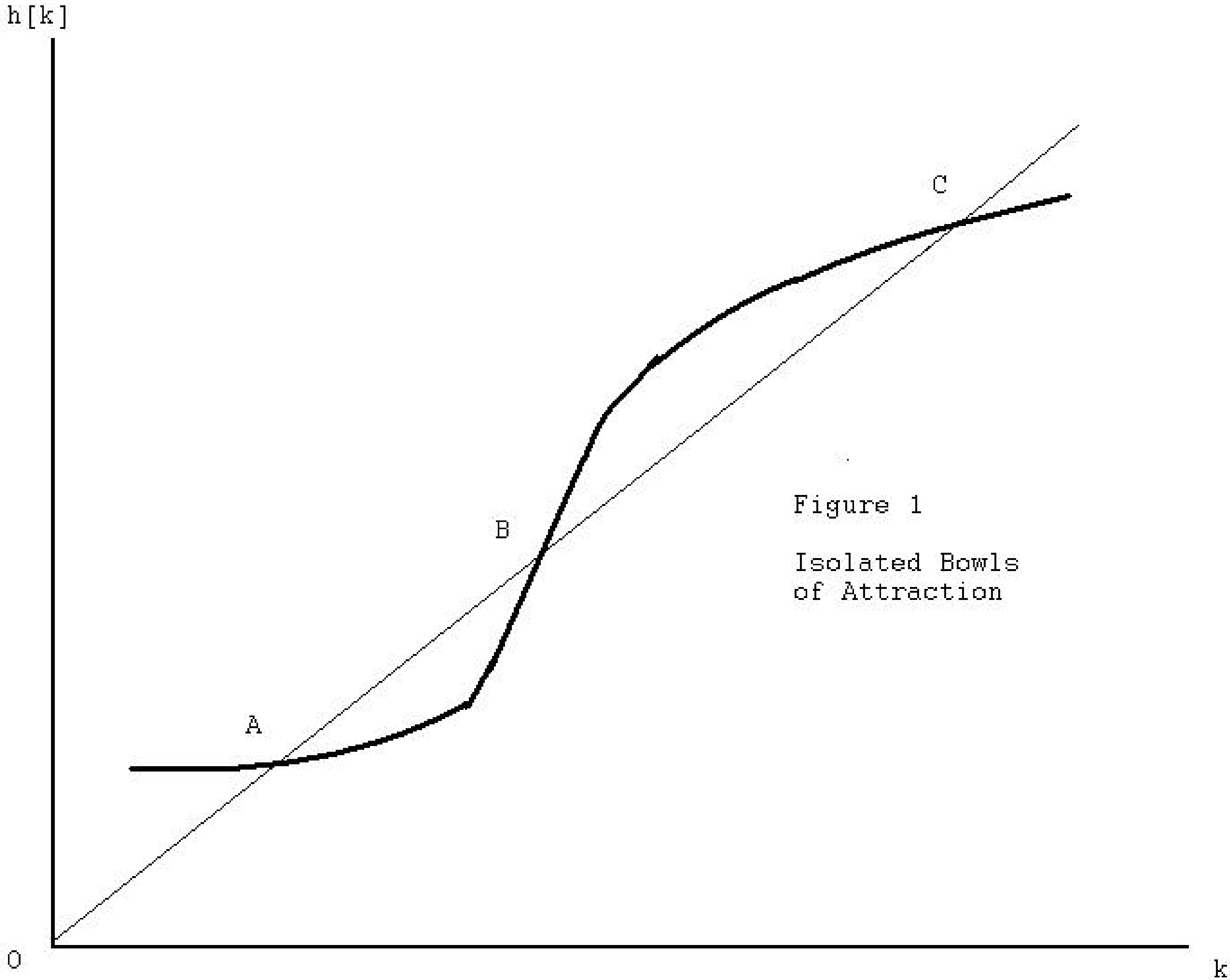


Figure 1

Isolated Bowls of Attraction

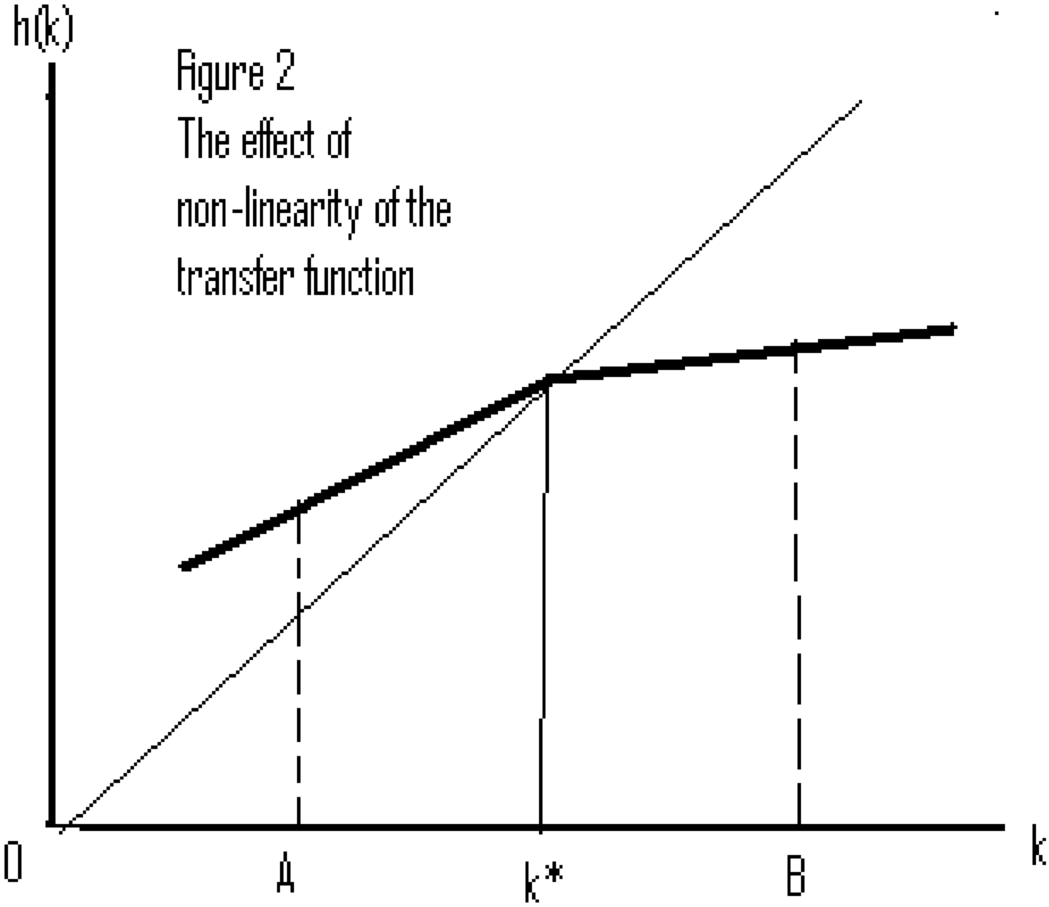
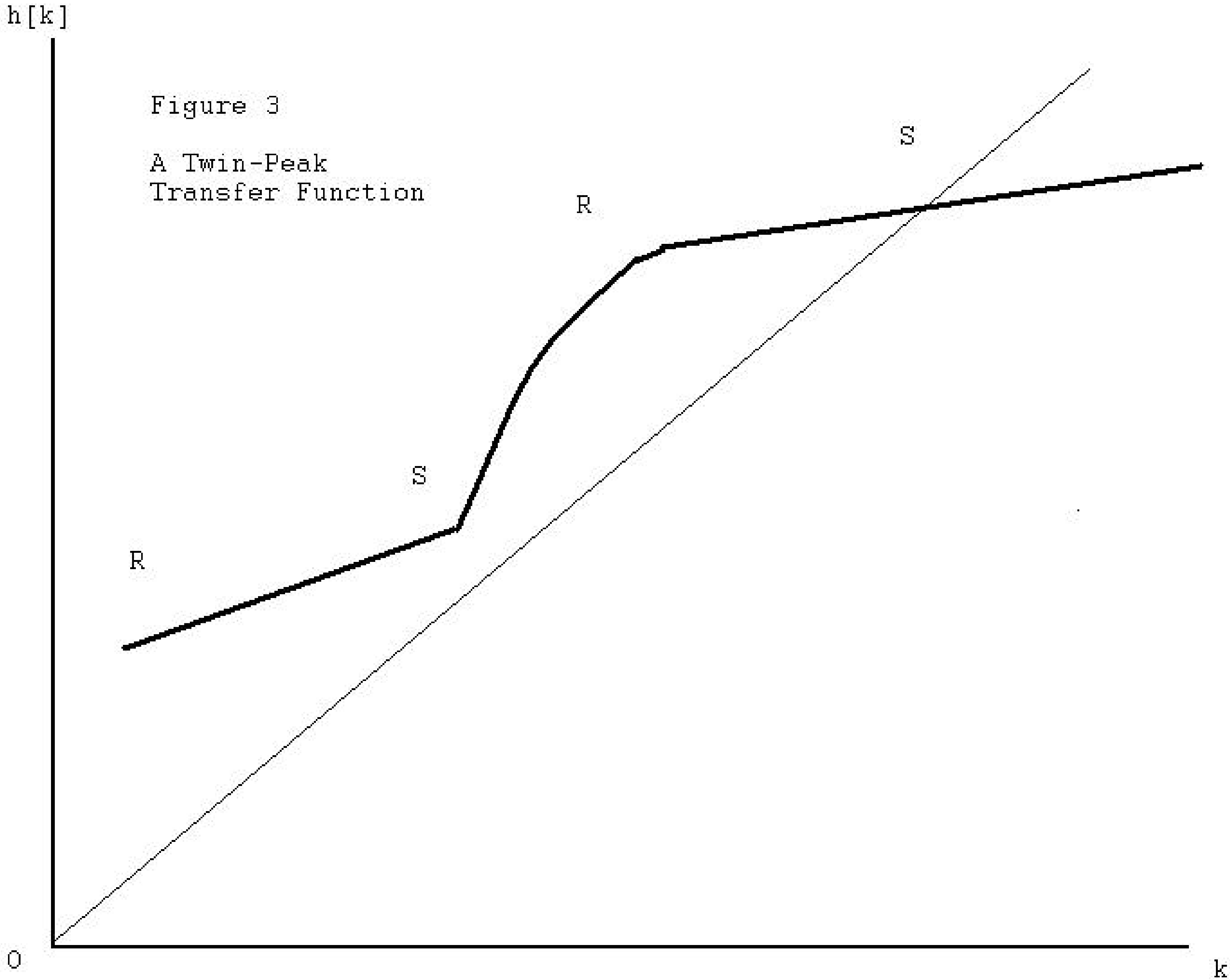


Figure 2
The effect of non-linearity of the transfer function


h[k]
Figure 3
A Twin-Peak
Transfer Function
S
R
S
R
O
k