NBER WORKING PAPERS SERIES

THE DIFFUSION OF TECHNOLOGY AND INEQUALITY AMONG NATIONS

Boyan Jovanovic

Saul Lach

Working Paper No. 3732

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 1991

We thank the C.V. Starr Center for Applied Economics at NYU for Technical and financial help, Michael Gort for providing the data, Ray Atji and Ken Rogoza for capable research assistance, and a referee and many other colleagues for useful comments. This paper is part of NBER's research program in Growth. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Working Paper #3732
June 1991

THE DIFFUSION OF TECHNOLOGY AND INEQUALITY AMONG NATIONS

ABSTRACT

One usually accounts for output growth in terms of the growth of the primary inputs: labor, physical capital, and possibly human capital. In this paper we account for growth with labor and with intermediate goods. Because we have no measures of the extent of adoption of most intermediate goods in most countries, we have to assume something about how they spread, based on what we see in U.S. data. We find that if all countries have (a) the same production function, (b) the same speed of adoption technology, and (c) imperfectly correlated technology shocks, then we can easily account for the extent and persistence of inequality among nations. Unfortunately, while it easily generates the sorts of low frequency movements that we observe, our technology shock seems to have little to do with high frequency movements in GNP so that if our definition of this shock is correct, real business cycle models are way off the mark.

Boyan Jovanovic
Department of Economics
New York University
New York, NY 10003
and
NBER

Saul Lach
Department of Economics
Hebrew University
Mount Scopus, Jerusalem
91905 Israel
and
NBER

1. Introduction.

Societies differ in their economic well-being, and their differences persist for long periods of time. Societies also differ in the technologies they use. These two facts surely are linked, but how? This paper accounts for the variation in per capita GNP among countries by positing that continuously arriving technologies differ in how well they suit each country. We show that this mechanism can not affect long-run growth, but that it can create huge level effects that persist for long periods of time. Because diffusion takes a long time, technological shocks are highly persistent, which is why there is so little turnover in the world distribution of per-capita GNP. Since we lack the comparable micro data on diffusion that we would need in order to make an extensive cross-country comparison of diffusion speeds with macro performance, our estimates are based on micro data from the U.S. We use these estimates to project what the rest of the world would do if it had the same speed of diffusion and the same production function as the U.S., but different, imperfectly correlated technology shocks. We show that there is no need for country specific permanent differences in diffusion speeds or in anything else, and that the bulk of the rigidity in the distribution of the world's income stems from the slowness of diffusion.

Our model says that countries that find recent technologies unsuitable will have lower per-capita GNP. Qualitatively, this statement must be valid, since poorer societies rely more on older technologies.[1] But how much poorer will a slow-adopting society become? This paper will provide quantitative estimates of the link between diffusion lags and development.

This is a growth-accounting exercise, although the framework differs from that of Robert Solow (1957) who accounted for growth in terms of the primary inputs, namely labor and capital. Dennison (1962) added education to the list of inputs. We account for the growth of countries in terms of their labor input and their intermediate goods inputs. We conclude that cross-country inequality and its persistence can be supported by positing that each country adopts intermediate goods with, on average, the same delays.

We do not explain why adoptions lags are as long as they are. We refrain from this for two reasons. First, it helps to know how strong the link between diffusion and development is quantitatively before trying too hard to tie them

---

[1] For instance, Lars Nabseth and George Ray (1974, p. 40) present evidence that higher wage societies are quicker to adopt labor saving technology. Moreover, information from their Table 2 on p. 17 shows that faster diffusion in countries relates strongly and positively to their per capita incomes. The product cycle literature has shown that advanced economies export (and therefore persumably produce and consume) more advanced goods (Donald Keesing, 1967). Bradford DeLong and Lawrence Summers (1990) find that more advanced economies enjoy lower prices and higher quantities of producers' durable equipment than less advanced economies. In the U.S., regions with higher per capita incomes seem to use more advanced technologies. Hybrid corn, for example, was first introduced on a wide scale in the north, and only later in the south (Griliches 1957, figure 2). And data from John Putnam (1991) shows that inventors tend to patent their inventions in richer countries, persumably because such countries are more likely to have the technical ability to copy the inventions.

A different strand of the literature tries to link economic development to technological <u>inventiveness</u>; e.g. Jan Fagerberg (1987). Our emphasis here is not on inventiveness but on implementation or adoption. An early paper by Richard Nelson (1968) takes a view similar to ours.

together theoretically. And second, the speed of diffusion should hinge on a diverse set of circumstances such as a society's patent laws and its endowments of human and physical capital, and to theorize about an analytically manageable subset of determinants seems to us too speculative a task at this point.[2]

We start with Paul Romer's (1990) formulation of a production function that relates a society's output of final goods to its use of intermediate-good inputs.[3] Each society can use each new technology right away, and each technology is on average of equal value to each society.[4] But there are random

---

[2] Existing models identify at least four reasons why the diffusion of new products and processes takes time: (a) Sunk costs in older technologies allow them to survive in spite of their higher variable costs (Chari and Hopenhayn 1991, Jovanovic and Lach, 1989); (b) Learning by doing in a new product or technology takes time (Stokey 1988, Jovanovic 1982); (c) The extent of patent protection and the ease of licensing the use of a technology or the freedom with which new technologies can be imported -- all institutional restrictions of the kind that Parente and Prescott (1991) stress, and (d) Frictions and lags in communication and in the transfer of information (Jovanovic and Rob (1989), Jovanovic and MacDonald (1988)).

[3] An alternative production function, emphasizing variety in process rather than product innovation is Jovanovic and Rob (1990, eq. 2.1). We use Romer's production function because our micro data are on new intermediate products rather than new production processes.

[4] Models in which growth is tied to the growth of a nonrivalrous good (such as knowledge) will have tremendous scale effects, "scale" being the extent of the market, and "market" being the area over which knowledge is nonrivalrous. The size of these markets is hard to measure (although Sokoloff (1988) does so with some success), and cheaper communication and transportation have led to a steady rise in the size of markets. Our assumption that the market is the whole world pays off in two ways. First, it saves us from having to measure market size or assume (incredibly) that it coincides with countries' borders. And second, it makes the theoretical prediction that larger markets will have higher growth rates irrelevant for cross-country comparisons -- a virtue because large countries do not seem to grow any faster than small ones.

deviations from that average, both over countries, and over time. We measure the value of a new product by its eventual penetration. It is this penetration that differs over products in a country, and over countries for a given product. But the speed with which that penetration is reached is assumed to be the same for all countries and all products. There are three parameters governing their speed of diffusion, one parameter describing the nature of the selection bias governing our sample of new products, and finally, one parameter describing the share of labor. We thus end up with a six parameter model.

We estimate three of the six parameters by using Michael Gort's and Steven Klepper's (1982) data on the diffusion of new products in the U.S. Labor's share is set at two thirds. The remaining two parameters represent the world's long-run growth-rate, and the fraction of newly arriving technologies that are unsuitable. Information about this last parameter can be found in recent work by Mark Schankerman and Ariel Pakes (1986) and Schankerman (1990). We then use these parameter estimates to project the extent of world inequality and its persistence. The projections match the Heston-Summers data well. There is no need for country-specific permanent differences.[5] Although the context is quite different, the methodology and the conclusions, positive ones we think, parallel those of Jess Benhabib and Jovanovic (1991). The basic conclusion is that each country's per capita GNP obeys the same stochastic process with partially

[5] The Solow model with permanent differences in savings and population growth was recently estimated by Gregg Mankiw, David Romer, and David Weil (1990). Permanently different institutional arrangements that lead to different effective tax rates on the returns to adopting technologies were the focus in Parente and Prescott (1991).

independent realizations of the shocks, and that the properties of that process are consistent with micro data from the U.S.

Surprisingly, our technology shocks have little success in generating movements at higher frequencies. The reason for this is simple: Technology shocks are embodied in new products. But new products take a long time to spread. By the time they have spread, newer products have appeared on the scene, and these independent technology shocks are subjected to too much averaging to have an impact on aggregates at frequencies associated with the business cycle. In sum, the technology shock is a useful vehicle for generating long waves, not short ones.

2. Using Diffusion Lags to Account for Inequality and its Persistence.

The aim in this section is to find out how much inequality one would see in the world if every country had the same production function as the US, the same diffusion lags as the US, the same probability distribution for the technology-specific shocks as the US, but different and imperfectly correlated realizations of those shocks. US data on diffusion of new products generate estimates for the stochastic process that US GNP per capita should follow. We shall compare the steady state variance of this process to the empirical cross sectional variance of GNP per capita among countries. We shall also look at how much turnover within the world distribution the US-estimated model implies. Throughout this section, the only reason why countries' incomes differ is that technologies suit different countries differently.

The story begins with a version of Romer's (1990) relation between an economy's output of final goods, Y_t, and its use of intermediate inputs:

$$Y_t = L_t^{1-\alpha}\int_0^{A_t} q_{it}^{\alpha}\,di\ . \qquad (1)$$

Here L_t is the labor input, $1-\alpha$ is labor's share, A_t is the list of intermediate inputs available at t, q_{it} is the quantity at date t of the i^{th} intermediate input, and α is, under further assumptions, the share of physical capital.[6]

Romer looks at constant growth-rate paths along which the output of an intermediate good invented at date s jumps from zero to, say, θ, where it remains thereafter. This is shown in Figure 1 as time-path B. The data from Michael Gort and Steven Klepper (1982) will show, however, that the output of a new product resembles a path better described by line C.[7]

---

[6] Romer assumes that physical capital alone is specific to the intermediate goods. Specificity of physical capital to new inventions is supported by casual observation and by a host of studies, eg. Roger Gordon, Mark Schankerman and Richard Spady (1986). But it is just as apparent that much human capital is also specific to new inventions: Jacob Mincer (1989) shows that workers in sectors that enjoy higher TFP growth seem to also get more on-the-job training, training that is presumably at least partly specific to the technologies that fuel TFP growth in their sectors, Gregory Clark (1987) and Howard Pack (1984) show that societies with less human capital will use a given technology less efficiently, and David Teece (1977) shows that the presence of skilled labor makes it easier for multinational firms to transfer technology across borders. Chari and Hopenhayn (1991) build a model in which diffusion lags stem entirely from the specificity of human capital.

Although we shall assume the truth of eq. (1) throughout, it will be clear that our analysis easily extends to accommodate human capital and possibly raw physical capital in the production function for final goods. We lack good measures of these two types of capital for most countries, however, so that empirical work with a more complicated production function would be severely limited in scope. Moreover, the paper aims to give a general flavor of the sorts of results that one can expect with this approach, as well as the kinds of questions that one can ask in this, as well as more detailed formulations. Finally, since little agreement exists on how to specify production functions that use intermediate goods (for instance, compare (1) to John Long and Charles Plosser's (1983) Cobb-Douglas production function or to Dale Jorgenson, Frank Gollop, and Barbara Fraumeni's (1987) translog specification), we shall stick to the simplest case.

[7] In Romer (1990), θ is written as $\bar{x}$ and it equals $\eta K/A$, where K is the capital stock and η the number of units of capital needed to make a unit of the intermediate good.

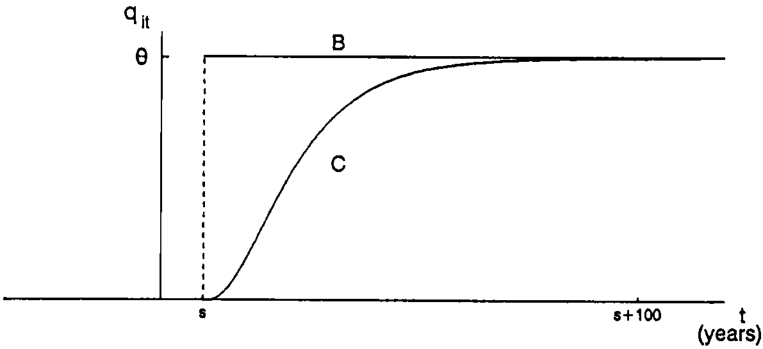

Figure 1: The Growth of the Average Gort-Klepper New Product

We need a scale-invariant representation for a society's tendency to adopt new technology, and the natural scaling factor is population. So if s_i is the vintage of product i, let

$$q_{it} = L_t h_i(t - s_i),$$

where h(x) is positive, and vanishes for x < 0. Now let

$$h_i(t - s_i)^{\alpha} = \hat{f}(t - s_i, \theta_{s_i}, \epsilon_{it}) . \tag{2}$$

Here θ_s is a shock specific to all products of vintage s,[8] while ϵ_{it} is

[8] Vintage-specific shocks also go by the name "cohort effects". Later on we shall compare our cohort effects to those that Schankerman and Pakes (1986) estimate for values of patent rights.

independent over i and t, and has CDF $G(\epsilon)$. Because a continuum of vintages arrives at each date, the ϵ's will wash out in the aggregate, so that the collective contribution at t of the products of vintage s will be

$$f(t-s,\theta) = \int \hat{f}(t-s,\ \theta,\ \epsilon)dG(\epsilon).$$

With (1), this means that per capita income is

$$y_t = \frac{Y_t}{L_t} = \int_0^{A_t} f(t - s_i,\ \theta_{s_i})di\ . \tag{3}$$

Now suppose that $\dot{A}/A = \lambda$ so that with an initial condition $A_0 = 1$,[9] $A_t = e^{\lambda t}$. Then the number of technologies of vintage s is $\lambda e^{\lambda s}$, and since they contribute

[9] This implies a mass point, a pulse, at date zero that we shall ignore in (4) and beyond. The size of this pulse does not matter for the steady-state implications because it fades relative to y_t as the latter grows without bound, and we omit it to avoid clutter.

That λ is constant is questionable since incentives for basic inventions should fluctuate over time: James Adams (1990), Zvi Griliches (1990) and Robert Evenson (1984) detect fluctuations with their proxies for A. For a summary of earlier findings on bunching of innovations see Christopher Freeman, John Clark and Luc Soete (1982). These studies suggest that a reliable measure of A_t would improve on our treatment of A_t as a deterministic, trend-stationary unobervable. But the obvious candidate -- number of patents -- is unsatisfactory for many reasons, some of which are summarized by John Jewkes, David Sawers and Richard Stilleman (1968, pp. 88-90). The main problem is that changes in the number of patents are usually accompanied by largely offsetting changes in the quality of patents (Schankerman and Pakes 1986, section 5). Pakes and Simpson (1989, pp. 398-9) list several possible reasons for this phenomenon.

$f(t-s,\ \theta_s)$ to output at date t,[10]

$$y_t = \lambda \int_0^t e^{\lambda s} f(t-s,\ \theta_s)ds = \lambda e^{\lambda t} \int_0^t e^{-\lambda \tau} f(\tau,\ \theta_{t-\tau})d\tau . \tag{4}$$

A high θ_s means that there was a basic advance in technology that was a common input into many products of vintage s, and that vintage <u>only</u>. At the aggregate level this is indistinguishable from assuming that $\lambda e^{\lambda s}\theta_s$ is the <u>number</u> of new products invented at s, but at the micro level the implications differ. We shall stick to the view that it is the quality of inventions that fluctuates, not their number. Being this specific about the shocks pays off in that we can estimate them from micro data rather than as an aggregate or sectoral residual.

If (θ_t) is stationary, the distribution of $(\theta_{t-\tau})_{\tau\in[0,\infty)}$ is the same as that of $(\theta_{-\tau})_{\tau\in[0,\infty]}$, so that as $t \to \infty$[11],

---

[10] The fraction, at date t, of the output of all intermediate goods that are of vintage s is $\hat{\pi}_s^t \equiv e^{\lambda s} f(t-s,\theta_s) / \int_0^t e^{\lambda \tau} f(t-\tau,\theta_\tau)d\tau$. If $\pi_\tau = \lim_{t\to\infty} \hat{\pi}_{t-\tau}^t$, then π_τ is our counterpart of the μ_τ that Chari and Hopenhayn (1991) plot in their figure 1. If f is bounded, $\pi_\tau \to 0$ as $\tau \to \infty$. To obtain $\pi_\tau = 0$ for τ large enough (as Chari and Hopenhayn do) we need $f(\tau,\ \cdot) = 0$ for large enough τ. The specific example we use in this section does not have the latter property.

[11] The expression in (5) is finite only if asymptotically f grows at a rate less than λ. Since X is defined for a fixed realization of $(\theta_t)_0^\infty$, it will therefore have a nondefective distribution only if $e^{-\lambda\tau}f$ is integrable for almost all $(\theta_t)_0^\infty$. This condition asks that each intermediate product's share in GNP go to zero as $t \to \infty$. The condition will hold if f is bounded uniformly in θ, or if products are transitory as in Stokey's (1988) model so that $f \to 0$ as $\tau \to \infty$.

$$\frac{y_t}{e^{\lambda t}} \rightarrow X = \lambda \int_0^\infty e^{-\lambda\tau} f(\tau, \theta_{-\tau}) d\tau \tag{5}$$

A stationary θ implies a stationary X, and so we arrive at

<u>Proposition 1</u>: log y_t is stationary around the trend λ.

This means that detrended log y_t has a stationary distribution, and that it stochastically regresses towards its mean. So, if countries are driven by independent or positively correlated θ's, their y's should converge.[12]

Proposition 1 also implies that all countries have the same long-run growth rate λ <u>regardless of their f</u>. That is, if a country takes a long time to adopt technology, this will lower the level of its GNP, but not its long-run growth rate.

Characterizing inequality in levels of y and its persistence requires

---

[12] Not to a constant, of course, but to the same long-run average level. Baumol and Wolff (1988, p. 1157) show that among countries that in 1950 had per capita incomes above $1,300 (in 1975 dollars) there was significant convergence in log-levels by 1980. But this is not true of the entire sample of countries, according to these authors and others -- most recently Quah (1990). At any rate, while ln y has no permanent components, the long diffusion lags render it so highly persistent, as we shall show below, that it is probably indistinguishable from a random walk with drift in currently available time series.

further assumptions about f. Let[13]

$$\hat{f}(x, \theta, \epsilon) = \begin{cases} (1 - e^{-\rho x + \epsilon})\theta & \text{for } x > 0 \\ 0 & \text{for } x \le 0 . \end{cases}$$

Assuming that $E(e^{\epsilon}) = 1$ results in $f(x, \theta) = (1 - e^{-\rho x})\theta$.

The parameter ρ measures the speed of diffusion. If $\theta_t = \theta$ (all t), the function f attains path B in Figure 1 as $\rho \to \infty$. To keep things simple, ρ is the same for all products; all grow at the same speed.[14]

Since it is vintage-specific rather than time-specific, and since it enters multiplicatively, the parameter θ does not affect a product's growth rate, only

---

[13] This functional form makes product quantities converge to θ, so that they last for ever. How good this assumption is will depend on the level of aggregation. For instance, out of 70 US manufacturing subsectors, only three (Iron and Steel Foundries, Railroad Equipment, and Leather Goods) had a decrease in real output between January of 1954 and December of 1988. Over shorter periods, however, cyclical forces are more important and transitory declines in sectoral output will be more frequent (Pankaj Ghemawat and Barry Nalebuff 1990). The more disaggregated Gort and Klepper data also do not show much evidence of downturns in output as products age. But at still higher levels of disaggregation, there are bound to be many products that have disappeared, or that will eventually do so.

[14] If high θ inventions are more profitable, then Zvi Griliches's (1957, Tables 2.6 and 2.7) and Edwin Mansfield's (1963) evidence that more profitable inventions spread faster suggests that if allowed for heterogeneity in ρ over products, we would have a positive correlation of θ and ρ over products. In the next section, we shall estimate product specific ρ's, correlate them with the product specific θ's, and find a weak but positive relation between them, thereby confirming Mansfield's findings. Chari and Hopenhayn (1991, Proposition 5) supply the logic for this finding: the greater the productivity of a newly arriving technology relative to old technologies, the faster is the diffusion of new technologies in the steady state.

the level of its eventual penetration, or "ceiling" in Griliches's (1957) terminology. If one views $\dot{f}$ as the reduced form for the time path of the product's equilibrium quantity, then a large θ denotes products for which demand is high, or products for which production costs are low, or both. Strictly speaking, however, the production function (1) does not allow heterogeneity on the demand side because each intermediate good enters in the same way.[15] Substitution into eq. (5) yields

$$X_t = \lambda\int_0^\infty e^{-\lambda\tau}(1 - e^{-\rho\tau})\theta_{t-\tau}d\tau \qquad (6)$$

if the process has gone on for long enough. The process (θ_t) is assumed to be stationary and serially uncorrelated:[16]

[15] This is easily changed, of course. A demand-based rationale for θ would replace (1) by something like $Y = L^{1-\alpha}\int_0^A \theta_i q_i^\alpha di$. A supply-based story could have $\eta = \theta_s^{-1}$ for products of vintage s (see note 7). Bertola (1991) introduces such differing efficiencies among intermediate goods producers.

[16] Since θ_t is the limiting size of the average intermediate good of vintage t, stationarity of θ_t means that new products are no more and no less important than the old. Surprisingly, perhaps, the micro data indicate that θ has no trend, and no autocorrelation, in θ, so that all persistence will originate in the model, and not outside it.

$$E(\theta_t) = \bar{\theta} \quad \text{and} \quad Cov(\theta_t, \theta_s) = \begin{cases} \sigma^2 & \text{if } t = s \\ 0 & \text{if } t \neq s \end{cases} \tag{6'}$$

The expression in (6) is random because it is a <u>sum</u>, not an <u>average</u>. To explain this better, we give two examples:

<u>Example 1</u>. Let $\theta_{t-\tau}d\tau = \bar{\theta}d\tau + \sigma dW_\tau$, where W_τ is Brownian motion.[17] Then $E\theta_{t-\tau}d\tau = \bar{\theta}d\tau$ and $Var\ \theta_{t-\tau}d\tau = \sigma^2 d\tau$, which is consistent with (6').

<u>Example 2</u>. Let $\theta_{t-\tau}d\tau = \delta_\tau(md\tau + sdW_\tau)$, where $\delta_\tau \in \{0, 1\}$ is iid and binomialy distributed with $P(\delta_\tau = 1) = \mu$. The distribution of $\theta_{t-\tau}d\tau$ is now a weighted average of a distribution that has mean $md\tau$ and variance $s^2d\tau$, and a distribution that is concentrated at zero. Then, $E(\theta_\tau d\tau) = \bar{\theta}d\tau = \mu m$, and $Var(\theta_\tau d\tau) = \sigma^2 d\tau = \mu[s^2 + (1-\mu)m^2]d\tau$, using the formulas for the mean and variance of a mixture.[18] This example builds on the first one. It assumes that the τ-cohort of inventions is either a success ($\delta = 1$) or a failure ($\delta = 0$). If it is a success its contribution to output is proportional to $\bar{\theta}d\tau + \sigma dW_\tau$. If it is a failure, it contributes nothing. Thus μ is the fraction of successful cohorts.

---

[17] Brownian motion has independent, normally distributed increments with $E(W_{t+k} - W_t) = 0$ and $Var(W_{t+k} - W_t) = k$.

[18] See William Feller (1971, p. 167).

Our empirical work will build on the second example.[19]

Equations (6) and (6') imply that

$$EX_t = \frac{\rho\bar{\theta}}{\rho + \lambda},$$

and the appendix shows that for each $k \geq 0$,

$$\mathrm{Cov}(X_t, X_{t-k}) = \frac{\lambda\sigma^2\rho^2 e^{-\lambda k}}{2(\lambda+\rho)(2\lambda+\rho)} [1 + \frac{\lambda}{\rho}(1-e^{-\rho k})] .$$

For given k, the covariance between ln y_t and ln y_{t-k} is the same as that between ln X_t and ln X_{t-k}:[20]

$$\mathrm{Cov}(\ln y_t, \ln y_{t-k}) \approx \frac{\lambda\sigma^2(\lambda+\rho)}{2\bar{\theta}^2(2\lambda+\rho)} e^{-\lambda k} \left[1 + \frac{\lambda}{\rho}(1-e^{-\rho k})\right] \tag{7}$$

so that its autocorrelation coefficient is

$$r_k = e^{-\lambda k} \left[1 + \frac{\lambda}{\rho}(1-e^{-\rho k})\right]. \tag{7'}$$

While r_k decreases with ρ for all k, it decreases with λ only for large

---

[19] All this can be done in discrete time and with discrete numbers of innovations; the results would be about the same. Indeed, with continua of independent random variables, some readers may find it easier to think of our integrals as limits of discrete sums. Nothing changes. But our setup leads to shorter, simpler expressions.

[20] A linear expansion of the logarithmic function leads to the approximation $\mathrm{Cov}(\ln u, \ln v) \approx \mathrm{Cov}(u, v)/E(u)E(v)$. This approximation underlies the expression in eq. (7).

k, and it increases with λ for small k. As $\rho \to \infty$, $r_k \to e^{-\lambda k}$, and the variance of ln y tends $\lambda\sigma^2/2\bar{\theta}^2$. Thus faster diffusion reduces the persistence, but it increases the larger steady-state variance around trend.

In equation (6), the weight on $\theta_{t-\tau}$ is $\lambda e^{-\lambda\tau}(1 - e^{-\rho\tau})$. At $\tau = 0$, this weight is zero. It then rises monotonically until vintage $\hat{\tau} = \rho^{-1}\ln[(\lambda+\rho)/\lambda]$ after which it declines and converges to zero. If $\hat{\tau}$ is large, the effect of a shock will take a long time to build up. In fact the parameter estimates yield a $\hat{\tau}$ of about twenty seven years, a surprisingly large number that underlies the high persistence of ln y_t in this trend-stationary model. Technological shocks do not have permanent effects on y_t because while a particular cohort of technologies is affected permanently by a shock, this permanent <u>absolute</u> effect becomes insignificant relative to y_t as the latter grows without bound.[21]

In spite of being trend-stationary, ln y_t is highly persistent. Another way to describe this persistence is with John Cochrane's (1988) statistic involving the variance of the k-differences in y_t which we shall rewrite in the form that John Campbell and Gregory Mankiw (1989) use:

$$V^k \equiv \frac{1}{k}\frac{\mathrm{var}(\ln y_{t+k} - \ln y_t)}{\mathrm{var}(\ln y_{t+1} - \ln y_t)}$$

Appendix B shows that the following approximation holds in our model:

[21] Footnote 10 defined π_τ to be the limiting fraction of technologies in use that are τ years old. In this sample π_τ is proportional to $e^{-\lambda\tau}(1-e^{-\rho\tau})$. That is, π_τ is proportional to the impulse response function.

$$v^k \approx \frac{1}{k}\frac{1 - r_k}{1 - r_1}$$

The model's predictions about v^k can be summarized as follows:

k	1	5	10	20	50	100	200	500	1000
v^k	1	4.5	7.9	12.4	16.2	13.7	8.3	3.4	1.7

The maximal value of v^k takes 50 years to reach and thereafter the decay is extremely slow. The maximal v^k that Campbell and Mankiw report is for Japan -- 13.7 at k = 60. For other countries the maximal v^k is far smaller, reflecting the greater importance of transitory shocks whose effects die off relatively rapidly. Our model omits all aggregate shocks other than the technology shock θ, and since the omitted shocks are probably more transitory, it is not surprising that our model overpredicts persistence.[22] This all means that our technology shocks do not generate enough movement at high frequencies. Indeed, if our conception of a technology shock is correct, one must look elsewhere for an explanation of business cycles. And, by implication, the Solow residual, as usually measured, must have a sizeable non-technological component.

Equation (7) and proposition 1 describe the properties of the time series $\ln y_t$ in one country. To direct these two propositions at the Heston Summers (1984) panel, we shall make the following four assumptions: First, all countries have the same production function (1). Second, the parameters λ, ρ, $\bar{\theta}$ and σ^2 are

[22] It is the slowness of diffusion that causes the high predicted values of v^k. As ρ gets large v^k approaches $(1-e^{-\lambda k})/k(1-e^{-\lambda})$ which declines monotonically from unity when k = 1, to zero as k gets large.

the same in each country. Third, A is the same in each country -- each has immediate access to every new intermediate good as soon as it appears on the scene.[23] And fourth, we shall start with the benchmark case of θ_t independent over countries.[24] What we shall do here is much like what Benhabib and Jovanovic (1991) did when they looked at how much cross country variability one

---

[23] That all countries can freely access A is not that unrealistic, at least not in the post war period during which the world economy has become highly integrated through increased trade, a greater presence of multinationals, and a growing number of cooperative agreements among firms of different nationalities. For instance, John Dunning, Bruce Kogut and Magnus Blomström (1990, pp. 74-75) show that during extended post-war periods, automobile and semiconductor firms in the U.S., Europe and Japan made as many cooperative agreements with firms of nationalities other than their own, as they did with firms of their own nationalities. International patenting is pervasive (Evenson, 1984) but even less effective than domestic patent protection (Robinson 1988, ch. 9) ,and moreover intermediate goods can be imported rather than domestically produced. Mansfield (1984, p. 136) notes that about 75 percent of the technologies transferred by U.S. firms to their subsidiaries were less than five years old. Clark (1987) for instance documents the quickness with which mechanized weaving methods spread from England to the rest of the world in spite of England's utmost efforts to contain them domestically. And Henderson, Jaffee and Trajtenberg (1990) show that in patent citations, a domestic resident is only slightly more likely to cite another domestic resident than a foreigner. "Basic" knowledge, (which is what A _is_) thus arguably bestows what Robert Lucas (1988, p. 37) has called global external effects. We treat A as a global public good.

The assumptions on f mean that all countries begin adopting each invention immediately. This is counterfactual, of course: Lumpiness in the invention process means that each invention will be implemented first _somewhere_. But in their small sample Nabseth and Ray (1974) show that in spite of this, no one country led others in first adoptions.

[24] Then we shall make some allowance for cross country correlation of the θ's. On the one hand, one expects a substantial country-specific component to θ_t. The railroad, for instance, was more important for the U.S. than for a mountainous country like Bolivia. On the other hand, the railroad must have mattered more to both than the toothpick has, so one also expects a good deal of covariation in θ_t over countries.

would get if one gave each the same production function, but independent Solow residuals. But we shall go farther because unlike them, we shall use micro data to estimate the parameters of the process governing the shocks, and we shall allow shocks to be dependent over countries.

The k^{th} order autocorrelation r_k depends on two parameters only: λ and ρ. From proposition 1, λ is set equal to the world's growth-rate which, between 1960 and 1985, was .02. The Gort-Klepper data on twenty-one U.S. products yield an estimate of ρ of .04. Figure 2 reports the correlation between countries' GNP's in 1960 and their GNP's t periods hence. It does the same thing with 1950 as the base year. It plots the empirical autocorrelations: $r_k(60) = \text{Corr}(\log y_{i,1960}, \log y_{i,1960+k})$[25] for the full sample, and $r_k(50) = \text{Corr}(\log y_{i,1950}, \log y_{i,1950+k})$ for the sample of 59 countries for which coverage begins in 1950, as well as the model's predicted r_k.

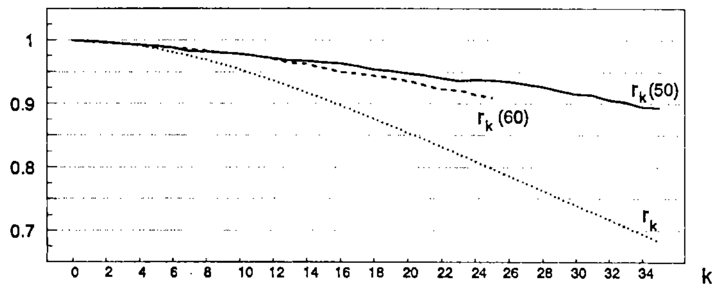

Figure 2: Actual and Predicted Autocorrelations.

[25] The calculated r_k would have been the same if we had taken out a common trend from all the $\ln y_{it}$ and then correlated the residuals. To see why, let $x_{it} = a + bt + u_{it}$. For given k and t, $\text{Cov}(x_{it}, x_{it+k}) = \text{Cov}(u_{it}, u_{it+k})$. Figure 2 fixes t at either 1950 or 1960 and then computes the correlation coefficients for each fixed k. The point is that the calculation of each r_k holds t and k fixed.

The predicted r_k is in equation (7′). But for it to be the theoretical counterpart of $r_k(50)$ and $r_k(60)$, the y_{it} processes must be independent over i. If they were, the time series (y_{it}) could, for each i, be viewed as a randomly selected sample path of the stochastic process whose autocorrelations appear in equation (7′). But the time series are not independent. Letting $z_{it} = \ln y_{it} - \lambda t$ and letting $\bar{z}_t$ be the year-t mean of z_{it}, the ratio

$$w_k = \frac{Var_t(\bar{z}_t)}{Var_t(\bar{z}_t) + Var_i(z_{ik})}$$

measures the fraction of the year k variance of $\ln y_{it} - \lambda t$ accounted for by shocks common to all countries. This ratio ranged from a low of 0.23 in 1970, to a high of 0.62 in 1960; on average it hovered around three-tenths. Now if thirty percent of fluctuations are common to countries, the correlation between countries' initial and their period t positions will be higher than r_k. Indeed if w_k were equal to one so that all movement in y was common to countries, r_k would also be equal to one for all k. So cross-country correlation will at least partly close the gap between r_k and r(50) and r(60), but exactly by how much will depend on the distribution of the common and idiosyncratic components.

The economic reason why the model generates so much persistence is that although the θ_s themselves are serially uncorrelated, their impulse responses peak at only twenty seven years after a vintage of products is introduced. This, of course, is because the diffusion of these products is so slow. The predicted r_k is certainly higher than table 4 of Charles Nelson and Charles Plosser (1982) indicates -- their autocorrelation coefficient for the US falls to zero within

6 years.

So much for the persistence of inequality. Now what about its extent? Equation (7) shows that this will hinge on the ratio $\sigma^2/\bar{\theta}$, the squared coefficient of variation of θ. The Gort-Klepper sample contains only partial and biased information about this ratio. The sample includes only successful inventions, whereas we know from the work of Mark Schankerman and Ariel Pakes (1986) that most inventions are relatively worthless, and that the fraction of such worthless inventions is close to one.[26] If θ_s is as described in our second example, the squared coefficient of variation of θ is:

$$\frac{\sigma^2}{\bar{\theta}^2} = \frac{s^2}{\mu m^2} + \frac{(1-\mu)}{\mu} .$$

Assuming that the Gort-Klepper sample is an unbiased selection of <u>successful</u> inventions, it tells us that $m = 1.26$ and that $s^2 = .32$. And Schankerman's and Pakes's (1986) results and other micro evidence suggest that μ should be close to zero. Along with $\lambda = .02$ and $\rho = .04$, such parameter values do an excellent job for accounting for the extent of inequality. Figure 3 plots the variances $\sigma_t^2(1950)$ and $\sigma_t^2(1960)$ of $\ln y_{it}$ in the small sample and in the large sample respectively. These variances increased during the sample period, and ranged

---

[26] Their third figure shows that in spite of negligible renewal fees in the U.K., France, and Germany, only about half of all patents were renewed through age 10, and less than 10 percent were renewed through the age of 18. The evidence from U.S. data also points to distributions of values of patent-rights with medians very close to zero. Griliches (1990, pp. 1679-80) and Pakes (1986, p. 779) survey this evidence.

from .778 to 1.324. There was, in other words, substantial divergence over this period. That $\sigma_t^2(1960) = \sigma_t^2(1950)$ for t = 1960 is a coincidence.

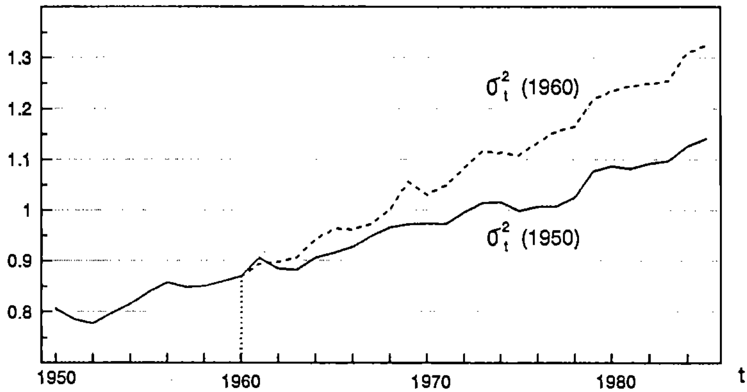

Figure 3: The Time Path of Relative Inequality

Because θ is correlated over countries, and because w_t is on average around three tenths, equation (7) (evaluated at k = 0) predicts a steady-state variance of ln y among countries of about

$$\left(\frac{7}{10}\right) \frac{\lambda\sigma^2(\lambda+\rho)}{2\bar{\theta}^2(2\lambda+\rho)} .$$

According to figure 3, this quantity must be between .78 and 1.32. For this to be true, we need μ to be between .005 and .0084, that is, a little less than one percent. While this is a bit on the low side, it is not far from what the micro evidence would lead one to expect. We shall return to this point.

Superficially it might seem that unlike the models of Prescott (1986) and Benhabib and Jovanovic (1991), the shocks are serially uncorrelated. But the state of knowledge or the state of technology is the accumulation of the θ's, and this quantity is serially correlated. The big difference between this model and conventional Solow residual models is in the shape of the impulse responses which here are hump-shaped with a peak at about twenty seven years in contrast to the conventional model where they decay geometrically and much more rapidly.[27] In contrast to the conventional real business cycle model, a series of favorable technological shocks will produce a boom about three decades hence. As it stands, then, this model is better suited for understanding low frequency movements; technological shocks spread too slowly to have much to do with the business cycle.

[27] Marco Lippi and Lucrezia Reichlin (1990) also discuss the effect that diffusion lags have on the time series properties of GNP. Unlike ours, their model has a permanent component and diffusion lags affect the way that underlying technological shocks affect that component.

3. Parameter Estimates from the Micro Data

The parameters are estimated using the data set assembled by Gort and Klepper (1982). These authors document the historical development of 46 new products in terms of their sales, price, quantity of output, and numbers of firms selling the new product over (part of) the life-cycle of each product. Table 1 lists the 21 products for which we have sales data, as well as the year in which each product was introduced into the market. There are old products such as records, dating from 1887, as well as relatively new ones such as lasers, which became available in 1960. The last year for which data were collected was 1972 and, in general, sales and quantity of output figures were not available for the whole life of the product. The age range for which there are data appears in the second column. The third and fourth columns give the products' average age in the sample, and the average level of sales in 1967 dollars (using the Wholesale Price Index). Most, but not all of the products seem to qualify as intermediate inputs in the sense of eq. (1). Since the number of products is not that large, we analyze them all.

The production function in (1) treats intermediate products as exchangeable inputs: One unit of product i and two units of product j can produce as much final output as two units of product i and one unit of product j. Now computers and ballpoint pens are surely not exchangeable in this sense, and something must be done to bring them into common units. We shall do this by expressing everything in units of the 1967 "consumption good", so that for q_{it} we shall use product i's sales at t, deflated to 1967 dollars by the Wholesale Price Index.

Product	Initial Year	Age Range	Mean Age	Sales*	Sample Size
1. Computers	1935	20-36	28	8567.6	17
2. Crystals, Piezo	1936	25-36	30.5	45.6	12
3. DDT	1943	1-27	14	23.3	27
4. Electrocardiographs	1914	47-58	52.5	8.8	12
5. Electric Blankets	1911	35-61	48	61.5	27
6. Electric Shavers	1930	1-42	22	83	39
7. Fluorescent Lamps	1938	0-34	17	94.6	35
8. Freezers, Home and Farm	1929	18-43	30.5	335	26
9. Gyroscopes	1911	52-61	56.5	15.1	10
10. Lasers	1960	3-11	7	41.9	9
11. Missiles, Guided	1942	9-30	19.5	2150.8	22
12. Motors, Outboard	1908	42-64	53	138.6	23
13. Penicillin	1943	2-28	15	72.7	27
14. Pens, Ballpoint	1945	6-27	16	92.1	22
15. Records, Phonograph	1887	34-85	59.5	372.7	52
16. Streptomycin	1945	1-27	14	9.5	27
17. Styrene	1935	8-36	22	85.8	29
18. Tapes, Recording	1947	14-25	19.5	159.7	12
19. Television, Apparatus, Parts	1929	17-43	30	1355.1	27
20. Transistors	1948	6-24	15	266.5	19
21. Tubes, Cathode Ray	1922	26-50	38	157.5	25
Average			29.7	578.2	499

Notes: * Deflated to 1967 dollars by the Wholesale Price Index, in millions.

Table 1: Variable Means

Gort and Klepper discuss their data at length. One property that they point to is that on average there is a rapid decline in the rate at which sales and quantity of output grow with the age of the product, and that their growth rates asymptote to zero. A model with this property is the one presented in section 2:

$$(q_{it}/L_t)^{\alpha} = \left[1 - \exp(-\rho(t-s) + \epsilon_{it})\right]\theta_s \qquad \text{for } t \geq s.$$

where the q_i are sales in 1967 dollars, L is the population of the US, and α = 1/3. As Table 1 shows, the mean value of sales differs quite a bit over products. Since the time series for each product is not too long, and since we assume that the parameter ρ is the same over products, we shall estimate ρ from the pooled data. To eliminate the heterogeneity, and to transform the nonlinear estimation problem into a linear one, we substitute the maximum value over t of $(q_{it}/L_t)^{\alpha}$ in each product i for θ_i.[28] Denote this estimator by $\hat{\theta}_i^*$. Note that since for all practical purposes there is only one product per vintage, we identify each product i with a different vintage s, i.e., s = i. Let p_{it} = log $[1 - (q_{it}/L_t)^{\alpha}/\hat{\theta}_i^*]$. Then ρ can be estimated from the regression $p_{it} = \gamma - \rho\, AGE_{it} + u_{it}$, where u_{it} equals ϵ_{it} plus an additional error term

[28] This is a crude estimate of θ, since products exist forever and never actually reach θ. This, however, is a limitation of the model since in reality some products become obsolete and physically disappear from the economy, implying that they reach their peak in finite time. A model in which products are transitory is Stokey (1988). We reemphasize, however, that Proposition 1 is valid in that context as well as long as λ is indeed fixed, and so we believe that most of our results are robust with respect to this assumption.

generated by estimating θ_i, where AGE = t - s, and where γ captures the nonzero mean of ϵ. This equation has been estimated by OLS for the pooled sample, as well as for each product individually.

The results are in Table 2. The last row presents the estimates for the pooled sample (499 observations) and gives an estimated ρ of 4.3 percent. The individual estimates, however, reveal that this estimate is not that representative -- it is about half the average of the individual estimates of ρ. In fact, an F-test for the equality of the ρ's rejects the null hypothesis of equal ρ's over products. Hence, for the purposes of estimating the individual θ's we use the product-specific estimates of ρ in the following way: First we solve for $\hat{\theta}_{it} = (q_{it}/L_t)^{\alpha}[1 - \exp(-\hat{p}_{it})]$, where $\hat{p}_{it}$ is the predicted p_{it} from the product-specific regression. We then average over all ages t to obtain an estimate of θ_i.[29] This estimate and its standard deviation within each product are in columns 4 and 5 of the table. Finally, in column 5 we present the preliminary estimate of θ_i, namely $\hat{\theta}_i^*$. The two sets of estimates do not differ much.

These results underlie the exercise that we went through in the previous section.[30] For m we used the average of the individual $\hat{\theta}_i$'s, namely 1.26, although this does not differ much from the pooled estimate. For s^2, we used the

---

[29] In computers, lasers, and records, some of the θ_{it} were highly negative. Since this does not make any sense these outliers were removed from the estimation procedure.

[30] We also used them to plot the curve C in figure 1. The plotted function is $[1 - \exp(-.04(t-s))]^3$, and it is based on $\alpha = 1/3$ and the pooled estimate of ρ of .04. The figure also assumes that $\theta = 1$.

Product	$\hat{\rho}_i$	SD of $\hat{\rho}_i$	$\hat{\theta}_i$	SD of $\hat{\theta}_i$	$\hat{\theta}_i^* = \max_t \left(\frac{q_{it}}{L_t}\right)^\alpha$	R^2	N
Computers	.44	.19	4.33[a]	.76	4.81	.42	17
Crystals	-.02	.08	.65	.05	.69	.00	12
DDT	-.08	.03	.59	.06	.63	.11	27
Electrocardiographs	.14	.11	.36	.04	.38	.10	12
Electric Blankets	.06	.02	.77	.06	.80	.12	27
Electric Shavers	.05	.01	.89	.20	.97	.32	39
Lamps	.12	.02	1.12	.42	.98	.65	35
Freezers	-.09	.03	1.42	.12	1.47	.22	26
Gyroscopes	-.38	.19	.56	.16	.52	.44	10
Lasers	.74	.29	.68	.12	.80	.58	9
Missiles	.07	.04	2.38	.59	2.77	.04	22
Motors	.12	.06	1.04	.08	1.06	.29	23
Pencillin	-.11	.03	1.19	.69	1.03	.47	27
Pens	.19	.03	.89	.05	.89	.70	22
Records	.08	.02	2.03	1.99	1.77	.47	52
Streptomycin	-.08	.04	1.07	1.77	.71	.18	27
Styrene	.07	.02	.85	.07	.88	.17	29
Tapes	.23	.08	1.01	.04	1.03	.32	12
Television	.12	.04	2.19	.50	2.37	.21	27
Transistors	.14	.06	1.20	.25	1.35	.16	19
Tubes	-.13	.06	1.18	.14	1.21	.25	25
Average[b]	.08		1.26				
Pool[c]	.043	.013	1.19	.80		.14	499

[a] Units of $\hat{\theta}_i$ and $\hat{\theta}_i^*$ are dollars per capita.
[b] Average over 21 products.
[c] Includes Dummies for the war, the post-war period and for product-specific intercepts.

<u>Table 2: Sales Regressions</u> $p_{it} = \gamma - \rho AGE + u_{it}$

variance of the $\hat{\theta}_i$ values, but adjusted for "measurement error" whose variance was estimated as the average of the sum of squares of the "S.D. of $\hat{\theta}_i$" column.[31] In other words, we treated the square of the standard error as the variance of the measurement error on θ_i. This procedure reduced the estimate of s^2 from its starting value of .77 down to .32. While somewhat arbitrary, any adjustment of a similar magnitude would have little bearing on the outcome of the exercise in the previous section.[32] For ρ we used the estimate from the pooled data.

The frequency distribution of $\hat{\theta}$ is in Figure 4. In their study of patent-

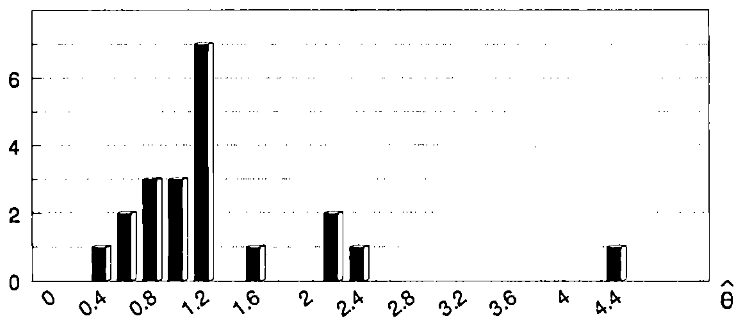

Figure 4: Frequency Distribution of $\hat{\theta}$ (in 1967 $'s per capita)

[31] This adjustment is necessary because the $\hat{\theta}$ are estimates so that their variance partly reflects sampling variability.

[32] Since they are based on BLS figures, the Gort-Klepper sales data probably understate s^2 because they do not satisfactorily control for quality change, quality change that gets passed on to the consumer and therefore does not affect sales. Adjusting for quality changes can make a huge difference, possibly doubling our estimates of some of the θ's (see Chapter 12 of Gordon, 1990). Looking at the products in table 1 it is clear that this underestimate of quality change was the largest for those products for which our $\hat{\theta}_i$ is the largest, such as computers, television, and transistors.

renewal data from Germany, France, and the United Kingdom, Schankerman and Pakes (1986) and Schankerman (1990) find that the distribution of private values of patent-rights is highly skewed: most are worth little, only a few are worth a lot. Now a Gort-Klepper product is an outgrowth of a highly successful idea, and the latter is presumably a draw from the extreme right tail of the distribution that Schankerman and Pakes talk about. Since this tail is very thin, it can have just about any shape and still be part of a highly skewed distribution. Still, the skewness shows up in this distribution as well. The calculations of the previous section assume that the Gort-Klepper data contain a representative sample of successful inventions, but that these successful inventions comprise an unknown fraction μ of all inventions, the remainder being worthless.

Figure 5 provides information that bears on our assumption that θ is stationary and, indeed, serially uncorrelated. The figure plots the size of $\hat{\theta}$ on its vintage, and it reveals neither trend nor autocorrelation, which is consistent with the assumptions on θ_s.

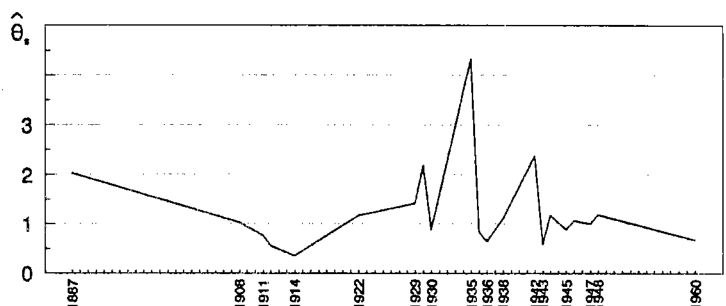

Figure 5: The estimated time series $\hat{\theta}_s$

We have assumed that ρ is a constant, independent of time. Conceivably, the speed of technological diffusion may have been changing.[33] We now examine this by looking at the time series of $\hat{\rho}_i$ plotted against the vintage of product i in Figure 6. There is a slight upward trend in these data,[34] and this is in spite of there being a slight <u>downward</u> trend in θ_i. One is then led to ask if this trend in ρ_i becomes significant once θ_i is held constant. The following

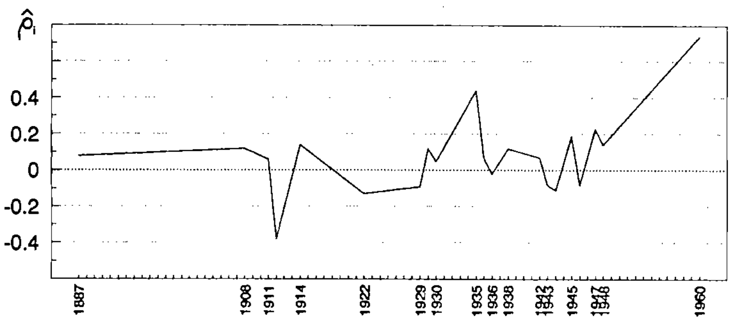

Figure 6: The estimated time series $\hat{\rho}_i$

---

[33] Proposition 1 implies that such changes in ρ can not affect the rate of growth. But they would have huge level effects. To get a feel for how large such effects would be, set $\theta_t = \theta$ for all t. If countries "one" and "two" differed only with respect to their ρ's and nothing else, their steady-state positions would satisfy

$$\lim_{t\to\infty} \frac{y_{1t}}{y_{2t}} = \frac{\rho_1}{\lambda+\rho_1} \Big/ \frac{\rho_2}{\lambda+\rho_2}$$

which increases from unity when $\lambda=0$, to ρ_1/ρ_2 as λ gets large.

[34] Since intermediate goods can be imported, the international migration of these goods will generally be faster than the migration of production methods used to make such goods. The latter kind of migration is the object of attention in the product cycle literature, and Mansfield (1984) argues that it has been accelerating. If so, one would expect that domestic diffusion would have been picking up speed.

least squares regression still shows no significant trend:

$$\hat{\rho}_i = \underset{(.26)}{.05} + \underset{(1.17)}{.07}\,\hat{\theta}_i - \underset{(.4)}{.001}s_i .$$

The absolute values of the t statistics are in parentheses, and s_i is the vintage of product i.

Do more valuable inventions spread faster? Griliches (1957) and Mansfield (1963) found evidence that they do. Our data, plotted in Figure 7, provide only weak support for this hypothesis. The correlation coefficient between $\hat{\rho}_i$ and $\hat{\theta}_i$

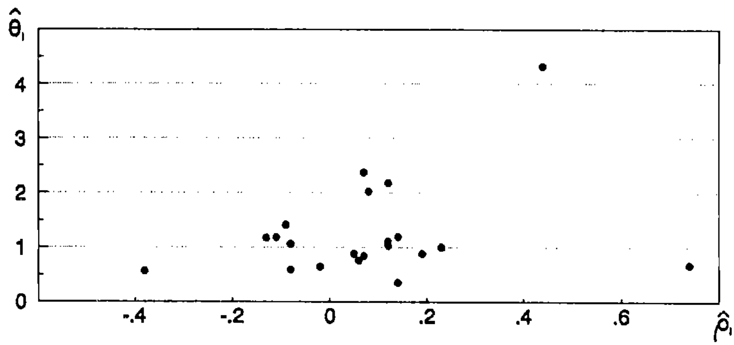

Figure 7: The size of inventions and the speed of their diffusion.

is .29, and it differs significantly from zero only at the 20% level of confidence.

Consistency check with the Schankerman-Pakes study.

The vintage effects θ are related to cohort effects on the value of patent holdings that the estimates of Schankerman and Pakes (1986) and Schankerman (1990) imply. Suppose that a patent is there to protect the invention of an intermediate good, and that it will yield its owner a fraction of the expected sales of the good.[35] If this fraction is about the same over products regardless of their expected sales, Schankerman's and Pakes's cohort effects will have a coefficient of variation that we can predict in terms of the parameters of our model. Their cohort effects are yearly vintage effects. Since δ_s averages out to μ over the year, the total value of the eventual sales of all inventions made in a particular year is

$$\int_0^1 \theta_s ds = m\mu + s\int_0^\mu dW_s = m\mu + sW(\mu),$$

which yields a coefficient of variation of $s/m\sqrt{\mu}$. Since $\mu < 1$ implies that

$$\frac{s^2}{\mu m^2} < \frac{\sigma^2}{\bar{\sigma}^2} \qquad \left(= \frac{s^2}{\mu m^2} + \frac{1-\mu}{\mu} \right),$$

the coefficient of variation of θ will exceed that of $\int_0^1 \theta_s ds$ because of averaging over independent cohorts within the year. Now in the previous section we found that to support the world inequality depicted in figure 3 we needed μ to be somewhere between .005 and .0084. Taking the midpoint in this range, $\mu = .0067$, we find that $s/m\sqrt{\mu} = 5.5$. This is what Schankerman and Pakes should have found

[35] For a process invention that reduces average costs, expected returns to the patent holder would also increase with market size.

given the truth of our model and the maintained assumption about what the patent holder can recover.

In fact, their estimates imply a coefficient of variation that is at most about one-tenth of this value, and it does not matter which of several possible ways one uses to get the estimate. First one can use the estimated value of all patents in various cohorts (V_s) that Schankerman and Pakes report in their table 5. Our model implies that on average V should grow at the rate λ, which is two percent per year. If one transforms their V_s estimates into discounted estimates using a rate of two percent per year, then for these transformed estimates, depending on the country, one gets a coefficient of variation of at most .4 (for France). Second, one can use their estimates of the value of the average patent in a cohort ($\bar{V}$) or the value of all patents per scientist and engineer in their table 6 (V/SE). And third, one can use information implicit in their table 3. Each of these methods leads to the same conclusion, namely a coefficient of variation that is at best an order of magnitude below our estimate. This discrepancy seems too large to attribute to sampling noise, and we devote what is left of this section to listing some possible reasons for it.

First, it is likely that the inventor can expect to capture a smaller fraction of revenues from large inventions. Large, profitable inventions invite faster imitation. Griliches (1958) and Mansfield et al (1977) argue that successful inventions give rise to social returns much bigger than private returns; the two can differ by hundreds of percentage points. This tends to reduce the standard deviation of patent values by more than it reduces their mean, thereby lowering the estimate of the coefficient of variation of patent values.

Second, the number of patents that an inventor will apply for in connection with a particular invention is probably larger for a big invention; in that way he can better protect it from imitators. Moreover, bigger inventions are probably more complex and may on those grounds require multiple patents. This tends to compress the distribution of patent values relative to the distribution of values of inventions. But if this really was the source of the problem, it could be solved by using the value of all patents in a cohort instead of the value of the average patent, and yet, as we mentioned above, the two procedures yield results that do not differ much.

Third, Schankerman and Pakes estimate cohort effects based on renewal behavior over the initial 15 to 18 years of patents' lives, whereas our θ measures the final contribution of inventions. But as Pakes (1986, p. 780) points out, as each patent ages and as uncertainty about it is resolved it is likely that the distribution of revenues across patents should become more dispersed, and that their coefficient of variation should go up.[36]

Fourth, Schankerman and Pakes estimate the coefficient of variation of patent values (including variations due to things other than cohort effects) at

---

[36] To drive this point home, assume that patents are initially all alike, and that the passage of time gradually informs patent holders about the patents' qualities. Initially the expected revenues on each patent would be either the same for all patents, and they would either all be renewed, or none would be. The coefficient of variation of expected revenues (and it is expected, and not actual revenues that determine whether a patent is renewed) would be initially zero, and would, in an untruncated sample, be monotonically increasing with the age of the sample. Now the truncation of the sample due to non-renewal will mitigate this effect, but will not reverse it.

.26 for the UK, .18 for Germany, and .45 for France. Now these findings are themselves an order of magnitude below comparable estimates for U.S. data that Griliches (1990, p. 1679) summarizes. So, if they have underestimated the variability of patent values due to all sources, it is likely that they have also underestimated the variability that stems from the cohort effects.

4. Conclusion

If societies take a long time to adopt technology, and if technologies suit them differently, this will create a good deal of persisting inequality. We analyze a mechanism through which this may occur, and quantify the effect that diffusion lags have on the aggregates. The resulting inequality in the levels of countries per capita GNP's and the persistence of that inequality is surprisingly large, large enough to match what the Heston-Summers data show.

Just as important, this exercise moves us closer to capture of the elusive "technological shock", closer to understanding how it affects aggregates. We specify what the shock is. It is the importance of a newly-arrived product, a new technology. Estimating twenty-one such shocks from micro data takes us beyond the aggregate and the sectoral Solow residual and shows that these shocks have little to do with the business cycle. The intuition is simple: they take so long to spread that averaging prevents them from having an impact at high frequencies. This negative conclusion is offset by the surprising ability this model has to explain movements at lower frequencies. This is a model of long waves in economic activity.

References:

Adams, James, "Fundamental Stocks of Knowledge and Productivity Growth," Journal of Political Economy 98, no. 4 (August 1990): 675-702.

Baumol, William, and Edward Wolff, "Productivity Growth, Convergence, and Welfare: Reply," American Economic Review 78, no. 5 (December 1988): 1155-9.

Benhabib, Jess, and Boyan Jovanovic, " Externalities and Growth Accounting," American Economic Review, 81, no. 1 (March 1991): 82-113.

Bertola, Giuseppe, "Flexibility, Investment, and Growth," unpublished paper, Princeton University, January 1991.

Campbell, John, and Gregory Mankiw, "International Evidence on the Persistence of Economic Fluctuations," Journal of Monetary Economics 23, no. 2 (March 1989): 319-34.

Chari, V.V., and Hugo Hopenhayn, "Vintage Human Capital, Growth, and the Diffusion of Technology," Journal of Political Economy 99, no. 6 (December 1991).

Clark, Gregory, "Why Isn't the Whole World Developed? Lessons from the Cotton Mills," Journal of Economic History (March 1987): 141-73.

Cochrane, John, "How Big is the Random Walk in GNP," Journal of Political Economy 96, no. 5 (October 1988): 893-20.

DeLong, Bradford, and Lawrence Summers, "Equipment Investment, Relative Prices, and Economic Growth," unpublished paper, Harvard University, April 1990.

Dennison, Edward, The Sources of Economic Growth in the United States, New York: Committee for Economic Development, 1962.

Dunning, John, Bruce Kogut and Magnus Blomström, Globalization of Firms and the Competitiveness of Nations, Lund, Sweden: Lund University Press, 1990.

Evenson, Robert, "International Invention: Implications for Technology Market Analysis," in Z. Griliches, ed., R&D, Patents and Productivity, University of Chicago Press (for NBER), Chicago, 1984.

Fagerberg, Jan, "A Technology Gap Approach to Why Growth Rates Differ," Research Policy 16, no. 2 (August 1987): 87-99.

Feller, William Probability Theory and its Applications II, New York: Wiley, 1971.

Freeman, Christopher, John Clark, and Luc Soete, Unemployment and Technical Innovation, Westport, Connecticut: Greenwood Press, 1982.

Ghemawat, Pankaj, and Barry Nalebuff, "The Devolution of Declining Industries," Quarterly Journal of Economics, 105, no. 1 (February 1990): 167-86.

Gordon, Robert, The Measurement of Durable Goods Prices, NBER, Chicago: University Press, 1990.

Gordon, Roger, Mark Schankerman, and Richard Spady, "Estimating the Effects of R&D on Bell System Productivity: A Model of Embodied Technical Change," in Prices, Competition and Equilibrium, edited by Maurice Peston and Richard Quandt, New York: Barnes and Noble, 1986, 164-188.

Gort, Michael and Steven Klepper, " Time Paths in the Diffusion of Product Innovations," Economic Journal, 92 (September 1982): 630-653.

Griliches, Zvi, Hybrid Corn: An Exploration in the Economics of Technological Change, Ph.D. Dissertation, Department of Economics, University of Chicago, August 1957.

Griliches, Zvi, "Hybrid Corn: An Exploration in the Economics of Technical Change," Econometrica 25, no. 4 (October 1957):

Griliches, Zvi, "Research Costs and Social Returns: Hybrid Corn and Related Innovations," Journal of Political Economy 66, no. 5 (October 1958):

Griliches, Zvi, "Patent Statistics as Economic Indicators: A Survey," Journal of Economic Literature 28 (December 1990): 1661-1707.

Henderson, Rebecca, Adam Jaffe, and Manuel Trajtenberg, "Telling Trails out of School: University and Corporate Patents and their Citations," unpublished, Harvard University, December 1990.

Heston, Alan and Robert Summers " Improved International Comparisons of Real Product and its Composition," Review of Income and Wealth, (June 1984): 207-26.

Jewkes, John, David Sawers, and Richard Stillerman, The Sources of Invention, Second Edition, New York: Norton, 1969.

Jorgenson, Dale, Frank Gallop, and Barbara Fraumeni, Productivity and U.S. Economic Growth, Cambridge, Massachusetts: Harvard University Press, 1987.

Jovanovic, Boyan, "Selection and the Evolution of Industry," Econometrica, (May 1982): 649-70.

Jovanovic, Boyan, and Saul Lach, "Entry, Exit, and Diffusion with Learning by Doing," American Economic Review, (September 1989): 690-99.

Jovanovic, Boyan, and Glenn MacDonald, "Competitive Diffusion," unpublished paper, University of Western Ontario, October 1988.

Jovanovic, Boyan, and Rafael Rob, "The Diffusion and Growth of Knowledge," Review of Economic Studies, (October 1989): 659-82.

Jovanovic, Boyan, and Rafael Rob, "Long Waves and Short Waves: Growth through Intensive and Extensive Search," Econometrica, (November 1990), 1391-1409.

Keesing, Donald, "The Impact of Research and Development on U.S. Trade," Journal of Political Economy 67, no. 1 (February 1967): 38-48.

Lippi, Marco, and Lucrezia Reichlin, "Diffusion of Technical Change and the Identification of the Trend Component in Real GNP," unpublished paper, O.F.C.E., Paris, January 1990.

Long, John, and Charles Plosser, " Real Business Cycles," Journal of Political Economy, 91, no. 1 (February 1983): 33-69.

Lucas, Robert E., Jr. "On the Mechanics of Economic Development," Journal of Monetary Economics 22 (January 1988): 3-32.

Mankiw, Gregg, Romer, David, and Weil, David, "A Contribution to the Empirics of Economic Growth," unpublished paper, Harvard University, 1990.

Mansfield, Edwin, "The Speed of Response of Firms to New Technologies," Quarterly Journal of Economics 77, no. 2 (May 1963): 290-309.

Mansfield, Edwin, "R&D and Innovation: Some Empirical Findings," in R&D, Patents and Productivity, Zvi Griliches (editor), NBER, Chicago: University of Chicago Press (1984): 126-54.

Mansfield, Edwin, Rapoport, John, Romeo, Anthony, and Beardsley, George, "Social and Private Rates of Return from Industrial Innovations," Quarterly Journal of Economics 91, no. 2 (May 1977): 221-40.

Mincer, Jacob, "Human Capital Responses to Technical Change in the Labor Market," NBER working paper no. 3207, December 1989.

Nabseth, Lars, and George Ray, The Diffusion of New Industrial Processes, London: Cambridge University Press, 1974.

Nelson, Charles, and Charles Plosser, "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications," Journal of Monetary Economics 10, no. 2 (September 1982): 139-62.

Nelson, Richard, "A Diffusion Model of International Productivity Differences in Manufacturing Industry," American Economic Review (1968?): 1219-48.

Pack, Howard, "Productivity and Technical Choice," Journal of Development Economics 16 (1984): 153-176.

Pakes, Ariel, "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks," Econometrica 54, no. 4 (July 1986): 755-84.

Pakes, Ariel, and Simpson, Margaret, "Patent Renewal Data," Brookings Papers on Economic Activity, Microeconomics, 1989: 331-401.

Parente, Stephen and Prescott, Edward, "Technology Doption and Growth," Federal Reserve Bank of Minneapolis, Report 136, February 1991.

Prescott, Edward, "Theory Ahead of Measurement in Business Cycle Research," Carnegie Rochester Conference on Public Policy, Autumn 1986, 25, 11-44.

Putnam, John, "Invention and Information Disclosure with International Patent Rights," Ph.D. dissertation, Department of Economics, Yale University, 1991.

Quah, Danny, " International Patterns of Growth: Persistence in Cross-Country Disparities," unpublished paper, Massachusetts Institute of Technology, January 1990.

Robinson, Richard, The International Transfer of Technology, Cambridge Massachusetts: Balinger 1988.

Romer, Paul, "Endogenous Technical Change," Journal of Political Economy 98, no 5, pt. 2 (October 1990): S71-S102.

Rosenberg, Nathan, Inside the Black Box: Technology and Economics, New York: Cambridge University Press, 1982.

Schankerman, Mark, and Ariel Pakes, "Estimates of the Value of Patent Rights in European Countries During the Post-1950 Period," Economic Journal, 96, (December 1986): 1052-76.

Schankerman, Mark, "The Private Value of Patent Rights in France, 1969-1987: An Empirical Study of Patent Renewal Data," report to the National Science Foundation, January 1990.

Sokoloff, Kenneth, "Inventive Activity in Early Industrial America: Evidence from Patent Records 1790-1846," Journal of Economic History 48 (December 1988): 813-50.

Solow, Robert, "Technical Change and the Aggregate Production Function," Review of Economics and Statistics, August 1957, 39, 312-20.

Stokey, Nancy, "Learning By Doing and the Introduction of New Goods," Journal of Political Economy, 96, no. 4 (August 1988): 701-18.

Teece, David "Technology Transfer by Multinational Firms: The Resource Cost of Transfering Technological Knowhow," Economic Journal 87 (June 1977): 242-61.

Appendix A: Calculating $Cov(X_t, X_{t+k})$.

Let $a_\tau = e^{-\lambda\tau}(1-e^{-\rho\tau})$. Then

$$Cov(X_{t+k}, X_t) = \lambda^2 Cov\left[\int_0^\infty a_\tau \theta_{t+k+\tau} d\tau, \int_0^\infty a_\tau \theta_{t-\tau} d\tau\right].$$

But if we change variables from τ to $s = \tau - k$, then $\tau = 0 \Rightarrow s = -k$, $\tau = \infty \Rightarrow s = \infty$, $t+k-\tau = t-s$, and $\tau = k+s$. Therefore, since $ds = d\tau$,

$$Cov(X_{t+k}, X_t) = \lambda^2 Cov\left[\int_{-k}^\infty a_{s+k}\theta_{t-s} ds, \int_0^\infty a_s \theta_{t-s} ds\right]$$

$$= \lambda^2 Cov\left[\int_0^\infty a_{s+k}\theta_{t-s} ds, \int_0^\infty a_s \theta_{t-s} ds\right]$$

$$= \lambda^2\sigma^2 \int_0^\infty a_{s+k} a_s ds \qquad = \lambda^2\sigma^2\int_0^\infty e^{-\lambda k} e^{-2\lambda s}(1-e^{-\rho(k+s)})(1-e^{-\rho s}) ds$$

$$= \lambda^2\sigma^2 e^{-\lambda k}\int_0^\infty \left[e^{-2\lambda s}(1-e^{-\rho s}) - e^{-\rho k} e^{-2\lambda s} e^{-\rho s}(1-e^{-\rho s})\right] ds$$

$$= \lambda^2\sigma^2 e^{-\lambda k}\left[\int_0^\infty a^{-2\lambda s} ds - \int_0^\infty e^{-(2\lambda+\rho)s} ds - e^{-\rho k}\int_0^\infty e^{-(2\lambda+\rho)s} ds + e^{-\rho k}\int_0^\infty e^{-2(\lambda+\rho)s}\right] ds$$

$$= \lambda^2\sigma^2 e^{-\lambda k}\left[\frac{1}{2\lambda} - \frac{1}{2\lambda+\rho} - e^{-\rho k}\left(\frac{1}{2\lambda+\rho} - \frac{1}{2(\lambda+\rho)}\right)\right]$$

$$= \lambda^2\sigma^2 e^{-\lambda k}\left[\frac{\rho}{2\lambda(2\lambda+\rho)} - \frac{e^{-\rho k}\rho}{(2\lambda+\rho)2(\lambda+\rho)}\right] = \frac{\rho\lambda^2\sigma^2 e^{-\lambda k}}{2(2\lambda+\rho)}\left[\frac{1}{\lambda} - \frac{e^{-\rho k}}{\lambda+\rho}\right].$$

To obtain the correlation coefficient between X_t and X_{t+k}, note that

$$Cov(X_t, X_{t+k}) = \frac{\rho\lambda\sigma^2 e^{-\lambda k}}{2(2\lambda+\rho)(\lambda+\rho)}\left(\lambda + \rho - \lambda e^{-\rho k}\right), \qquad (A.1)$$

which leads to the expression in the text. Moreover, dividing this expression by the variance of X we get

$$r^2(X_t, X_{t+k}) = e^{-\lambda k}\left[1 + \frac{\lambda}{\rho}(1 - e^{-\rho k})\right].$$

Appendix B: Deriving the v^k statistic.

Since $\ln y_{t+k} - \ln y_t = \lambda k + \ln X_{t+k} - \ln X_t$, and since, to a first approximation,

$$\mathrm{Var}(\ln X_{t+k} - \ln X_t) \approx [E(X)]^{-2}\,\mathrm{Var}(X_{t+k} - X_t),$$

we find that

$$\mathrm{Var}(\ln y_{t+k} - \ln y_t) \approx [E(X)]^{-2}\,\mathrm{Var}(X_{t+k} - X_t),$$

so that

$$v^{k-1} \approx \frac{1}{k}\,\frac{\mathrm{Var}(X_{t+k} - X_t)}{\mathrm{Var}(X_{t+1} - X_t)} \qquad (A.2)$$

Since X_t is stationary,

$$\begin{aligned}\mathrm{Var}(X_{t+k} - X_t) &= E(X_{t+k} - X_t)^2 \qquad (A.3)\\ &= 2\mathrm{Var}(X_t) - 2\mathrm{Cov}(X_{t+k}, X_t)\,.\end{aligned}$$

Now (A.1), (a.2) and (A.3) imply the expression for v^{k-1} given in the text.