working paper
department
of economics

GALTON'S FALLACY
AND
TESTS OF THE CONVERGENCE HYPOTHESIS

Danny Quah

No. 552 May 1990

massachusetts
institute of
technology

50 memorial drive
cambridge, mass. 02139

GALTON'S FALLACY
AND
TESTS OF THE CONVERGENCE HYPOTHESIS

Danny Quah

No. 552 May 1990

Galton's Fallacy and Tests of the Convergence Hypothesis
by
Danny Quah
Economics Department, MIT
10 May 1990

Abstract

Recent tests for the convergence hypothesis derive from regressing average growth rates on initial levels: a negative initial level coefficient is interpreted as convergence. These tests turn out to be plagued by Francis Galton's classical fallacy of regression towards the mean. Using a dynamic version of Galton's fallacy, we establish that, in fact, coefficients of arbitrary signs in such regressions are consistent with an unchanging cross-section distribution of incomes.

Feathers hit the ground before their weight can leave the air.

R.E.M.

1. Introduction

Do the incomes or productivity levels of different economies have a tendency to converge? Numerous researchers have recently examined this issue by calculating the cross-section regression of measured growth rates on initial levels. See for instance Barro (1989), Baumol (1986), DeLong (1988), Dowrick and Nguyen (1989), Murphy, Shleifer, and Vishny (1990), and many others. Murphy, Shleifer, and Vishny have called this the "Barro regression." Evidently, in the Barro regression, a negative coefficient on initial levels is taken to indicate convergence.

This paper clarifies what such *initial level regressions* are able to uncover. As used in this literature, the term "convergence" can mean a number of different things:

(a) Countries originally richer than average are more likely to turn below average eventually, and vice versa; the cycle repeats;

(b) Whether a country income is eventually above or below average is independent of that economy's original position;

(c) Income disparities between countries have neither unit roots nor deterministic time trends; and

Galton's Fallacy and Tests of the Convergence Hypothesis

by

Danny Quah *

10 May 1990

* Department of Economics, MIT and NBER (*dquah@athena.mit.edu*). I thank Roland Benabou, Olivier Blanchard, John Heaton, Marc Nerlove, and Julio Rotemberg for conversations. Errors and misinterpretations are mine alone.

(d) Each country eventually becomes as rich as all the others; the cross-section dispersion diminishes over time.

Cases (a) and (b) vaguely correspond to the notion of *mixing* in econometrics (see e.g. White (1984)). Case (c) is one formulation of persistence in income disparities: from a time-series perspective, it is the natural way to examine dependence on initial conditions. This particular probability model raises interesting econometric issues in the context of unit root random fields (see Quah (1990a)); it is, however, quite different in spirit and in substance from initial level regressions. Case (d) is closest to the notion of poorer countries eventually catching up with richer countries.

If (a) and (b) are the cases of interest, models for studying transitional characteristics—for example, that used in the income distribution and earnings mobility literature—would seem appropriate. Thus, Quah (1990b) attempts to uncover such effects in the context of heterogeneous Markov chains. Overall, however, the work using initial levels regressions strongly suggests case (d) as being of interest.

This paper shows that the widely-used initial level regressions, in fact, shed no light on convergence in the sense of (d). I develop an analogy between those regressions and Galton's classical fallacy of regression towards the mean. Recall that Galton, in his aristocratic manner, was concerned about the sons of tall fathers regressing into a pool of mediocrity along with the sons of everyone else. Galton inferred this from observing that taller-than-average fathers had sons who turned out to be not as much above average as the fathers themselves. However, he could not reconcile this with the fact that the observed population of male heights continued to display significant cross-section dispersion. I show—using exactly the same reasoning that reveals Galton's error—that a negative cross-section regression coefficient on initial levels is, in fact, perfectly consistent with absence of convergence in the sense of (d).

While Galton's formulation is convenient for analyzing observations at two points in time, it offers little by way of interesting dynamics. Extending the analysis to permit such dynamics, I show in Section 3 below that a given cross-section distribution—replicating itself over time—is consistent with *arbitrary* signs on the cross-section initial levels regression coefficient. In other words, the sign of the initial levels regression coefficient says nothing about whether there is convergence or divergence.

The final two sections below consider alternative probability models that might justify these initial levels regressions. We will see, however, that there are significant econometric difficulties in interpreting the estimation results from such models.

2. Galton's Fallacy for the Convergence Hypothesis

To make the point clearly, consider the simplest case. Let $Y_j(t)$ denote (the logarithm of) measured per capita income or productivity in country j and period t. For t_2 different from t_1, the cross-section regression of $Y(t_2)$ on a constant and $Y(t_1)$ is:

$$P[Y(t_2) \mid 1,\ Y(t_1)] = E_C Y(t_2) + \lambda \cdot (Y(t_1) - E_C Y(t_1)), \tag{1}$$

where

$$\lambda = \mathrm{Var}_C(Y(t_1))^{-1} \mathrm{Cov}_C(Y(t_2), Y(t_1)).$$

In (1), $P[\ \mid\ ]$ and E_C, Var_C, and Cov_C indicate projection and cross-section expectation, variance, and covariance respectively. Suppose that there is no convergence, i.e.,

$$\mathrm{Var}_C(Y(t_1)) = \mathrm{Var}_C(Y(t_2)).$$

The Cauchy-Schwarz inequality immediately implies that the regression coefficient λ is less than 1 in absolute value. This of course is simply the Galton fallacy: economies with higher than average incomes at t_1 (tall parents) have incomes that are not as high above average at t_2 (offspring regressing towards mediocrity). Note that this happens *exactly* when the cross-section variances at t_1 and t_2 are equal, i.e., when there is no convergence of cross-section incomes.

Equation (1) then implies:

$$P[Y(t_2) - Y(t_1) \mid 1,\ Y(t_1)] = \mu - (1 - \lambda) \cdot Y(t_1) \tag{2}$$

for some μ and $\lambda \leq 1$. By the last inequality, the cross-section regression coefficient on $Y(t_1)$ in (2) is non-positive. In words, when $t_1 < t_2$, a regression of income growth on initial levels shows countries that are initially richer tend to grow more slowly. Scaling the dependent variable by $(t_2 - t_1)^{-1}$, i.e., using average income growth on the left hand side, does not alter this conclusion. Again, this apparent convergence occurs when there is, by assumption, no real convergence.

3. Dynamics and Arbitrary Signs on the Initial Level

The classical Galton fallacy above is useful for analyzing observations made at two points in time. When Y_j has interesting dynamics, it turns out that an initial levels regression can give either a strictly positive or a strictly negative coefficient, *even when the cross-section distribution remains unchanging over time.*

While the point can be made quite generally, again it is instructive to take the simplest case. For each t, let $\{Y_j(t),\ j = 1, 2, 3, \ldots\}$ be independent, and let G_t denote the cross-section distribution at time t:

$$G_t(y) \stackrel{\text{def}}{=} \text{fraction of } j \text{ such that } Y_j(t) \leq y, \qquad y \text{ in } \mathcal{R}.$$

Suppose further that for each j, the time series $\{Y_j(t),\ t = \ldots, -1, 0, 1, \ldots\}$ is zero-mean and stationary, and has finite variance and normally distributed innovations; let

$$Y_j(t) = \sum_{s=0}^{\infty} C(s)\epsilon_j(t-s), \qquad \epsilon_j \sim \mathcal{N}(0, \sigma^2),$$

be the Wold representation for Y_j. Call $\nu^2 = \sigma^2 \sum_s |C(s)|^2$. Then

$$\begin{aligned} F_j(y) &\stackrel{\text{def}}{=} (2\pi\nu^2)^{-1/2} \int_{-\infty}^{y} \exp\left(-\frac{1}{2}\zeta^2/\nu^2\right) d\zeta \\ &= F(y) \end{aligned}$$

is the unique ergodic distribution for the stochastic process $\{Y_j(t),\ t = \ldots, -1, 0, 1, \ldots\}$ for each j. Assume the number of countries is large and initialize the cross-section distributions G_t, $t \leq 0$, to equal the time-series ergodic distribution F. At each $t \geq 1$, take G_{t-s}, $s \geq 1$, as given and apply the Glivenko-Cantelli Lemma; it follows immediately that

$$G_{t+1} = G_t = G_0 = F, \qquad t \geq 1.$$

In words, the assumptions imply that the cross-section distribution of countries is time-invariant.

The cross-section initial levels regression (with $t_0 < t_1 < t_2$) is now:

$$P[Y(t_2) - Y(t_1) \mid 1,\ Y(t_0)] = \xi + \beta Y(t_0), \tag{3}$$

for some ξ and

$$\beta = g_Y(0)^{-1}\Big(g_Y(t_2 - t_0) - g_Y(t_1 - t_0)\Big),$$

with g_Y denoting the covariogram of Y_1. (Barro [1989] and Murphy, Shleifer, and Vishny [1990] have considered exactly this configuration in (t_0, t_1, t_2).) Because the cross-section distribution matches the ergodic distribution, the cross-section covariances exactly equal the corresponding unconditional time series moments. Notice that while $g_Y(0) > 0$ and $g_Y(t) \to 0$ as $t \to \infty$, intermediate values of g_Y are unrestricted. Thus, the regression coefficient β on $Y(t_0)$ can take arbitrary sign. If, for instance, $t_2 \to \infty$, β simply has the opposite sign as $g_Y(t_1 - t_0)$.

In summary, the initial levels regression coefficient has a sign that is completely uninformative for whether the cross-section distribution is converging or diverging. In the example above, the cross-section distribution is unchanging over time, yet the sign of the regression coefficient can be negative, positive, or zero.

The independence and identical distribution assumptions here play a role only in simplifying the calculations. With heterogeneity, the time-invariant cross-section distribution is a probability mixture of the different individual time series ergodic distributions. Weak forms of dependence across countries will not affect application of the Glivenko-Cantelli law. With strong dependence or small numbers of countries, the cross-section distribution will be a non-degenerate random element in the space of distributions. While the calculations then become much more difficult, the flavor of the results is unaffected. Finally, normality is used only to give an explicit form to the individual time series ergodic distribution.

4. A Possibly Correct Formulation

A dynamic panel data random effects model might be thought to justify the usual interpretation of these initial levels regressions. We will see, however, that there are serious econometric difficulties in this view.

Suppose $Y_j(t)$, $j = 1, 2, \ldots, N$, $t = 0, 1, \ldots T$, is generated by

$$Y_j(t) = X_{j1}(t) + X_{j0}(t) = \alpha_j + \theta_j t + X_{j0}(t), \qquad E\Delta X_{j0} = 0$$

$$\Longrightarrow \Delta Y_j(t) = \theta_j + \Delta X_{j0}(t), \quad E\Delta X_{j0} = 0, \tag{4}$$

and

$$\theta_j = Z_j \beta_0 + u_j, \quad E u_j Z_j = 0. \tag{5}$$

The zero expectation conditions in (4) and (5) are identifying assumptions. Equation (4) states that country j's (log) income Y_j is comprised of two components X_{j1} and X_{j0}. In the current work, X_{j1} is taken to be just a time trend $X_{j1}(t) = \alpha_j + \theta_j t$. Equation (5) is a regression that describes how growth rates θ_j vary across countries. Notice that θ_j is the growth rate of both the unobserved component X_{j1} and the observed series Y_j (since $E\Delta X_{j0} = 0$). The covariates Z_j might include measures of average education, health, openness of the economy, as well as the initial condition $Y_j(0)$.

While we have specified X_{j1} to be a time trend, more generally (X_{j1}, X_{j0}) could be simply a decomposition of Y_j into quite arbitrary stochastic permanent and transitory components (as in e.g. Quah (1990c)). This, however, would considerably complicate the discussion without introducing any new insights.

The underlying growth rate θ_j is unobservable and needs to be proxied in estimating (5). One possibility is to use:

$$\hat{\theta}_j = (t_2 - t_1)^{-1} \sum_{t=t_1+1}^{t_2} \Delta Y_j(t) = (t_2 - t_1)^{-1}(Y_j(t_2) - Y_j(t_1)). \tag{6}$$

(Another possibility is to take $\hat{\theta}_j$ to be the time trend coefficient in a least squares regression of Y_j on a constant and time: nothing essential would change in the discussion.) Thus, a regression of average growth rates (6) on Z_j should be viewed simply as an imperfect way of estimating the underlying regression (5). While $\hat{\theta}_j$ is an error-ridden measure of θ_j, it appears only on the left-hand side of the equation. Thus, it might seem that classical regression analysis suggests no significant problems.

In fact, however, the model here does not give rise to classical measurement error. Straightforward calculation shows that the least squares regression estimator β_N computed using (6), instead of the true θ_j, satisfies:

$$\sqrt{N}(\beta_N - \beta_0) = \left(N^{-1}\sum_{j=1}^{N} Z_j'Z_j\right)^{-1}\left(N^{-1/2}\sum_{j=1}^{N} Z_j'u_j\right) + \left(N^{-1}\sum_{j=1}^{N} Z_j'Z_j\right)^{-1}\left(N^{-1/2}\sum_{j=1}^{N} Z_j'\left(X_{j0}(t_2) - X_{j0}(t_1)\right)(t_2 - t_1)^{-1}\right). \quad (7)$$

Since $EZ_j'u_j = 0$ by (5), the first term gives the standard OLS zero mean normal distribution approximation for large N. However, since neither EZ_jX_{j0} nor $EZ_j\Delta X_{j0}$ are restricted, the second term dominates as N grows without bound: it diverges to plus or minus infinity. Thus, this regression yields an inconsistent estimator for β_0.[1] This effect is especially pronounced when Z_j explains both short-run and long-run dynamics in Y_j, as would be standard in real business cycle models. When Z_j is the initial condition $Y_j(0)$, the analysis of the previous section again applies. Thus, even in this setting, it is difficult to interpret the results of initial levels regressions, in terms of convergence or of divergence.

[1] Some might argue that the right conceptual experiment in (7) is to take $t_2 - t_1 \to \infty$ and then consider the approximation as N grows large. The second term in (7) might then be negligible. This is delicate: recall that in the Summers-Heston data set—that typically used in this work—$t_2 - t_1$ is at most 36 while N is about 100 so that $N \gg T$ rather than the other way round.

5. An Alternative Interpretation

An alternative interpretation of (3) and (4)–(5) of the previous sections is possible. Some unobserved common factor might cause both high growth rates θ_j and high initial levels $Y_j(t_0)$. In this view, θ_j and $Y_j(t_0)$ are jointly "endogenously" determined: (5) is a reduced form for some unspecified underlying structural model. This avoids interpreting β_0 as a structural economic parameter; the sign of β_0 in (5) nevertheless remains of interest as that is thought to indicate the validity of this hypothesis.

The Galton fallacy criticisms in Sections 3 and 4 are, of course, unaffected by this alternative interpretation. The initial levels regression coefficient says nothing about (d)-convergence regardless of whether β_0 is part of the structural or reduced forms. Significantly, section 3 implies that a particular sign on β_0 could be purely spurious from the viewpoint of the hypothesis here. In this sense, any empirical finding on the sign of β_0 turns out to be not especially informative.

6. Conclusion

We have shown that cross-section regressions of growth rates on initial levels shed no light on the validity of the convergence hypothesis in the sense of (d). It should be evident that conditioning on additional regressors does not alter this basic message. Having clarified this, it is important to emphasize what the paper does not say: it is not that there are "econometric problems" in estimating these initial level regressions. On the contrary, in every situation described above, except for the analysis in Section 4, the regressions do the right thing: they consistently estimate exactly what they are supposed to estimate. (The model of Section 4 appears to come closest to allowing the desired economic interpretation; there, however, we find significant econometric difficulties in interpreting the results from the estimation. Even if these problems could be overcome, those raised in Sections 2 and 3 would nevertheless remain.)

The difficulty therefore lies not in the econometrics but rather in the economic interpretation of these initial levels regressions: subtlety arises because researchers have provided only incomplete probability descriptions of the effects they are trying to uncover.

It would be useful to display an explicit probability model where such cross-section initial levels regressions are, in fact, sensible descriptive devices. Such a probability description would help clarify what it is that economists are using different growth models to explain. King and Robson (1989) is a useful step in this direction.

References

Barro, R. J., 1989, A Cross-Country Study of Growth, Saving, and Government, NBER Working Paper 2855, February.

Baumol, W., 1986, Productivity Growth, Convergence, and Welfare, *American Economic Review*, 76, no. 5, December, 1072-85.

DeLong, J. B., 1988, Productivity Growth, Convergence, and Welfare: Comment, *American Economic Review*, 78, no. 5, 1138-1155.

Dowrick, S. and D. Nguyen, 1989, OECD Comparative Economic Growth 1950-85: Catch-Up and Convergence, *American Economic Review*, 79, no. 5, 1010-1030.

King, M. A. and M. H. Robson, 1989, "Endogenous Growth and The Role of History," LSE Financial Markets Group Discussion Paper No. 63, August.

Murphy, K., A. Shleifer, and R. Vishny, 1990, The Allocation of Talent: Implications for Growth, GSB Univ. of Chicago Working Paper, March.

Quah, D., 1990a, Persistence in Income Disparities: I. Unit Root Random Fields, MIT mimeo, March.

Quah, D., 1990b, Persistence in Income Disparities: II. Heterogeneous Markov Chains and Duration Dependence, MIT mimeo, in preparation.

Quah, D., 1990c, Permanent and Transitory Components in Labor Income: An Explanation for 'Excess Smoothness' in Consumption, *Journal of Political Economy*, 98, no. 3, 449-475.

White, H., 1984, *Asymptotic Theory for Econometricians*, New York: Academic Press.