Why is Africa so Poor?

A Structural Model of Economic Development and Income Inequality

WPS/2001-5

David Fielding,
Department of Economics, University of Leicester,
University Road, Leicester LE1 7RH.
and CSAE
E-mail: DJF14@LE.AC.UK

Second draft: May 2000

Abstract
The paper extends existing work on inequality and economic development by estimating a cross-country structural model that identifies bi-directional relationships between income inequality and other indicators of social and economic development. Overall, lower inequality is associated with improvements in other development indicators, but this is the result of several complex interactions. The most striking feature of the structural model is the insight it provides into the reasons behind the negative "Africa dummy" in previous cross-country growth studies.

Keywords: income distribution, social and economic development
JEL categories: O1, O4

1. **The Africa Dummy, Income Inequality and Economic Development**

One of the few consistent findings in empirical cross-country studies of the causes of economic growth is that, other things being equal, African countries have lower steady state *per capita* income levels. Easterly and Levine (1997) highlight the high degrees of ethno-linguistic fractionalisation in Africa as one of the reasons for "Africa's growth tragedy". Their thesis is that ethnic homogeneity promotes the acquisition of social capital; countries lacking such homogeneity are less productive because of an absence of social capital. But, as we shall see, African countries perform more poorly *ceteris paribus,* even when development is conditioned on ethno-linguistic fractionalisation. In this paper we will argue that in order to uncover more fully the reasons for the negative Africa dummy in cross-country income regressions, it is necessary to explore the interaction between economic development and social development indicators, and in particular measures of education, health and income inequality. We will argue that although ethnic homogeneity does explain some of the variation in performance across countries, its role is small when compared with other factors.

Empirical analysis of the links between inter-household income inequality and economic development is one of the oldest genres in development economics, going back at least as far as Kuznets (1955). Interest in this topic has recently been revived by the World Bank's publication of better quality data on income distribution than has previously been available (Deninger and Squire, 1998). One of the innovations in more recent work on the correlates of income inequality is the inclusion of a wide range of development indicators, not just mean income. For example, Easterly (1999) presents results using four measures of inequality and 77 other development indicators.

One of the main criticisms of early empirical work on the link between economic development and income inequality is that causality could well run in both directions. The stage of economic development a country has reached could affect the structure of income distribution, for the reasons outlined in the Kuznets paper.[1] But the distribution of income could also affect the level of aggregate income, because credit constraints prevent low-income households from carrying out socially efficient investment plans, or because income inequality erodes social capital. While early work on income inequality was based on OLS regressions, treating income development process, and equations for indicators of inequality are estimated either with lagged income, or using an IV estimator. The natural progression from this is to model income and inequality simultaneously, in order to form a complete picture of the interaction of inequality with income and with other development indicators. Effective development policy in any country requires an understanding both of the impact of inequality on income and of income on inequality; we will see that these relationships will be particularly relevant to the causes of low income in Africa. Section 2 of the paper presents an economic model that allows for this kind of simultaneity. Section 3 discusses the result of estimating such a model on cross-country data including both African and non-African countries. Section 4 reflects on the implications of the results for social and economic development in Africa and Section 5 concludes.

2. Modelling the Development Process

In this section we describe the structure of the model estimated in Section 3. The aim is to construct a framework in which it is possible to identify the ways in which inequality interacts

---

[1] Briefly, the reasons are (i) investment opportunities in the early stage of the development process are limited to a few individuals or regions, so transition from low to middle income status involves a worsening of the income distribution; this is followed by improvements in the distribution as rents are dissipated by wider access to these opportunities. (ii) Physical capital is a complement to skilled labour and a substitute for unskilled labour, so the wage gap between skilled and unskilled grows as the capital stock increases.

with key indicators of social and economic development. The model is based on that of Fielding (1999) but extends the model by integrating income inequality into the development process.

Much of the recent work on income and inequality has used panel data to estimate regression coefficients in income and inequality equations (for example, Deninger and Squire, 1998; Easterly, 1999). In the tradition of the panel data based empirical models of growth evolving from Barro (1991), the growth-inequality relationship is investigated through regressions of the form:

$$\Delta y_i(t) = a_0 - a_1 \cdot y_i(t-1) + a_2 \cdot g_i(t) + a_3 \cdot x_i(t) + u_i(t) \quad (1)$$

and

$$g_i(t) = b_0 + f(y_i(t)) + b_1 \cdot x_i(t) + v_i(t) \quad (1a)$$

where $y_i(t)$ is *per capita* income in country i in period t, $g_i(t)$ the gini coefficient or some other index of inequality, $x_i(t)$ is a set of social and economic variables, and $\{u_i(t), v_i(t)\}$ residuals. The non-linear function f(.) reflects Kuznet's hypothesis that inequality is at first increasing, then decreasing in *per capita* income.

One major problem in the interpretation of the coefficients of equations (1) and (1a) is that slope parameters are likely to vary across countries: we cannot assume that the interaction of income and inequality follows exactly the same pattern everywhere. The goal of pooled regression analysis is then to estimate the mean of the distribution of each slope coefficient. But if there is any serial correlation in the explanatory variables, then a panel data regression that imposes common parameter values across countries will induce serial correlation in $u_i(t)$ and $v_i(t)$. As a result, the fitted coefficients will not represent consistent estimates of the cross-country average of the true parameter values (Pesaran and Smith, 1995; Lee *et al.*, 1998). The potential ways of solving this problem within the panel data framework are quite data-intensive, and we will be working with a relatively small data set containing around a hundred countries. But consistent estimates of the mean *long run* slope coefficients can be produced by averaging data for each country over time, and estimating the model on a pure cross-section.[2] This is the approach that we follow below.

The use of cross-sectional data necessitates a different solution to the simultaneity problem than is used in panel data work. Instruments must be found to identify equations for each of the development indicators, so that a structural model can be estimated. The main drawback of the cross-sectional approach is that lags can no longer be used as instruments in dealing with the endogeneity as they are in, for example, the panel data studies of Easterly (1999) and Caselli *et al.* (1996). We need to find "real" instruments in order to identify the model. The theoretical model that follows provides a basis for the selection of these instruments.

2.1 The theoretical model

Papers such as Easterly (1999) contain scores of potentially endogenous variables, and it is impossible to identify all of the interactions between them using cross-section data. Instead, we will identify equations for one measure of inequality and three other indicators of socio-economic development; the rest of the model is left in reduced form. The four endogenous variables appearing in the empirical model are *per capita* income, education, health and inequality. We will first describe a theoretical model suggesting how equations for these four

[2] This is proven in Pesaran and Smith (1995). The caveat here is that the parameters will provide estimates only of the impact of x_i on y_i. Any cross-country effects of x_j on y_i will not be captured.

variables might be identified; next we will deal with measurement issues. The theoretical model consists of nine equations. Greek characters are parameters and Roman characters are variables; endogenous variables are written in bold.

g: the degree of income inequality
y: log *per capita* income
e: a measure of the average education level
h: a measure of the average level of health
k: the log *per capita* physical capital stock
a: factor efficiency
$\mathbf{c_e}$: the cost of investment in education, net of its consumption benefits
$\mathbf{c_h}$: the cost of investment in health, net of its consumption benefits
r: the interest rate
n: the log *per capita* natural resource stock
v: a measure of ethno-linguistic diversity
m: a vector of variables capturing cultural characteristics
x: mean annual temperature

The equations are:

The determinants of the degree of income inequality:
$$\mathbf{g} = \zeta_0 + \zeta_1 \cdot v + [\zeta_2 - \zeta_2' \cdot \mathbf{y}] \cdot \mathbf{y} + \zeta_3 \cdot \mathbf{h} + \zeta_4 \cdot m + \zeta_5 \cdot \mathbf{e} \qquad (2)$$
$\zeta_1, \zeta_2, \zeta_2' > 0 > \zeta_3, \zeta_5$

The aggregate production function:
$$\mathbf{y} = \mathbf{a} + \alpha_1 \cdot \mathbf{e} + \alpha_2 \cdot \mathbf{h} + \alpha_3 \cdot \mathbf{k} + \alpha_4 \cdot n \qquad (3)$$
$1 > \alpha_1, \alpha_2, \alpha_3, \alpha_4 > 0$

The determinants of factor efficiency:
$$\mathbf{a} = \beta_0 + \beta_1 \cdot v + \beta_2 \cdot \mathbf{g} \qquad (4)$$
$\beta_0 > 0 > \beta_1, \beta_2$

A resource constraint:
$$\theta \cdot \mathbf{y} = \pi_0 + \pi_1 \cdot \mathbf{e} + \pi_2 \cdot \mathbf{h} + \pi_3 \cdot \mathbf{k} \qquad (5)$$
$1 > \pi_1, \pi_2, \pi_3 > 0 > \pi_0$

The public education decision:
$$\mathbf{e} = \ln(\alpha_1) + \mathbf{y} - \mathbf{c_e} \qquad (6)$$

The cost of education (net of consumption benefits):
$$\mathbf{c_e} = \gamma_0 - \gamma_1 \cdot v - \gamma_2 \cdot \mathbf{y} - \gamma_3 \cdot \mathbf{h} - \gamma_4 \cdot m - \gamma_5 \cdot \mathbf{g} + \mathbf{r} \qquad (7)$$
$\gamma_1, \gamma_2, \gamma_3 > 0 > \gamma_5$

The public health decision:
$$\mathbf{h} = \ln(\alpha_2) + \mathbf{y} - \mathbf{c_h} \qquad (8)$$

The cost of health (net of consumption benefits):

$$\mathbf{c_h} = \delta_0 - \delta_1 \cdot v - [\delta_2 - \delta_2' \cdot x] \cdot x - \delta_3 \cdot \mathbf{g} - \delta_4 \cdot \mathbf{y} - \delta_5 \cdot \mathbf{e} + \mathbf{r} \qquad (9)$$

$\delta_0, \delta_2, \delta_2', \delta_4, \delta_5 > 0 > \delta_3$

The private investment decision:

$$\mathbf{k} = \ln(\alpha_3) + \mathbf{y} - \mathbf{r} \qquad (10)$$

Equation (2) indicates the potential determinants of inequality. Cultural factors (m) could influence inequality through their impact on social norms, and hence public policy. m includes dummy variables for whether the country has ever been a British or French colony: Henderson (1991; 1993) argues that countries' colonial histories affect the degree of observance of civil rights and other measures of social development; this argument could be extended to the degree of observance of economic rights to a basic standard of living. m also includes the reported number of Muslims and Christians, as a fraction of the total population: Poe and Tate (1994) and Park (1987) argue that religious adherence is a major factor in social development. Ethno-linguistic diversity is included in the equation because it might increase inter-ethnic inequality, giving rise to a greater potential for inter-household inequality. A higher level of health care and education might be associated with lower inequality, because they are associated with wider access to income-augmenting activities. Finally, the quadratic term in income allows for a Kuznets Curve effect.

Equation (3) is an aggregate production function in log-linear form. *Per capita* output depends on the physical capital stock, natural resources and two human capital measures, e and h. These represent two aspects of the stock of human capital, which depreciate over time because of, for example, mortality and disease. Improvements in the stock of health and education might generate a higher *per capita* output level for two reasons. First, they can improve the inherent productivity of the labour force. Second, they may reduce fertility rates and so the fraction of the population below working age. Without an appropriate instrument for fertility we cannot identify these two effects.

Equation (4) indicates the factors determining the efficiency of resource use. Easterly and Levine (1997) find that greater ethno-linguistic diversity (v) reduces growth, and so we allow for this effect in the efficiency function. We also allow for the possibility that the degree of income inequality (g) adversely affects efficiency. This effect could work through the formation of social capital (Hardin, 1992; Galor and Zeira, 1993; Knack and Keefer, 1997), or through a link between income inequality and credit constraints (Chaterjee, 1991). In the absence of appropriate instruments, these links will remain unidentified in our model. The causal chain from inequality to economic performance is captured in reduced form, in the parameter β_2.

The economy's resource constraint is represented by equation (5) above. This equation is derived from the following investment-savings constraint. Capital letters represent levels rather than logs:

$$S = Y^{\theta} = \lambda_e \cdot E + \lambda_h \cdot H + \lambda_k \cdot K \qquad (5a)$$

S is total saving, θ the elasticity of saving with respect to income, and the λ_i are rates of depreciation for the three types of capital: steady-state investment equals the depreciation rate times the capital stock. Letting π_1 represent the share of education investment in total investment, π_2 the share of investment in health, π_3 the share of physical capital investment and

$\pi_0 = \ln(\lambda_e)\cdot\pi_1 + \ln(\lambda_h)\cdot\pi_2 + \ln(\lambda_k)\cdot\pi_3$, the constraint can be written out as equation (5).[3]

Equation (6) expresses an equality between the marginal return to investment in eduction and its perceived marginal cost, c_e.[4] Education is both a consumer good and a capital good, so c_e is to be interpreted as the marginal cost of investment net of any consumption benefits that accrue to education. Equation (7) includes the possible determinants of c_e. Income inequality appears in the equation because higher inequality could result in credit constraints that reduce education expenditure. Other factors will affect c_e through their impact on the perceived consumption benefits of eduction. When the income elasticity of demand for education-as-consumption is positive, higher y will be associated with higher e. Moreover, increases in h might lead to a higher e because education and health are complements in consumption. More ethno-linguistic diversity might increase the value of education because second-language literacy will be more important. We also allow the cultural factors, m, to affect the value placed on education. If there are no consumption benefits of education, and no credit constraints, then c_e will equal the interest rate, r.

In equation (8) the marginal return to investment in healthcare equals its marginal cost (net of consumption benefits), c_h. Equation (9) includes the factors which might determine c_h, and is analogous to equation (7). The difference between equation (9) and equation (7) is that by assumption the value of health does not depend on cultural factors; but the cost of delivering a certain level of health does depend on country size and on climatic factors: whether a country has a coastline, and the average temperature in the country, x. If extremes of temperature are unhealthy, then $c_h(.)$ will be non-monotonic in x, and the parameters δ_2 and δ_2' allow for this.

In equation (10) the marginal cost of physical capital (r) equals its marginal return. This equation could be adapted to include an efficiency term similar to equation (4) so that there is a wedge between the marginal return and the marginal cost, capturing the possibility that capital market efficiency is endogenous; but this would not alter the general form of the final equations to be estimated.

Equations (2-10) can be solved down to a partially reduced form with our four chosen variables on the left hand side of four equations: one for income (y), education (e), health (h) and inequality (g). The four equations are shown below.

Income:

$$\mathbf{y} = \{\beta_0 + \beta_1\cdot v + \beta_2\cdot\mathbf{g} + [\alpha_1 - \alpha_3\cdot\pi_1/\pi_3]\cdot\mathbf{e} + [\alpha_2 - \alpha_3\cdot\pi_2/\pi_3]\cdot\mathbf{h} + \alpha_4\cdot n\} \times \{\pi_3/[\pi_3 - \theta\cdot\alpha_3]\} \quad (11)$$

Education:

$$\mathbf{e} = \{\ln(\alpha_1/\alpha_3) - \gamma_0 + \gamma_5\cdot\zeta_0 - \pi_0/\pi_3 + [\gamma_1 + \gamma_5\cdot\zeta_1]\cdot v + [\gamma_4 + \gamma_5\cdot\zeta_4]\cdot m + [\gamma_2 + \gamma_5\cdot\zeta_6 + \theta/\pi_3]\cdot\mathbf{y} + [\gamma_3 + \gamma_5\cdot\zeta_7 - \pi_2/\pi_3]\cdot\mathbf{h}\} \div [1 - \gamma_5\cdot\zeta_5 + \pi_1/\pi_3] \quad (12)$$

Health:

$$\mathbf{h} = \{\ln(\alpha_2/\alpha_3) - \delta_0 - \pi_0/\pi_3 + \delta_1\cdot v + [\delta_2 - \delta_2'\cdot x]\cdot x + \delta_3\cdot\mathbf{g} + [\delta_4 + \theta/\pi_3]\cdot\mathbf{y} + [\delta_5 - \pi_1/\pi_3]\cdot\mathbf{e}\} \div [1 + \pi_2/\pi_3] \quad (13)$$

[3] Since the π_i are less than unity, and the $\ln(\lambda_i)$ are less than zero, π_0 is less than zero.

[4] Note that in a log-linear production function, the log of the marginal product of e is equal to the log of the average product, (y - e) plus the log of the elasticity of output with respect to e, $\ln(\alpha_1)$.

Inequality:

$$\mathbf{g} = \{\zeta_0 + \zeta_5\cdot[\ln(\alpha_1/\alpha_3) - \gamma_0 - \pi_0/\pi_3] + [\zeta_1 + \zeta_5\cdot\gamma_1]\cdot v + [\zeta_4 + \zeta_5\cdot\gamma_4]\cdot m + [\zeta_2 + \zeta_5\cdot[\gamma_6 + \theta/\pi_3] - \zeta_2'\cdot\mathbf{y}]\cdot\mathbf{y} + [\zeta_3 + \zeta_5\cdot[\gamma_7 - \pi_2/\pi_3]]\cdot\mathbf{h}\} \div [1 - \gamma_5\cdot\zeta_5] \quad (14)$$

y is identified by n.[5] h is identified by x; e and g are identified by m (representing more than one variable).[6] e is not be identified in equation (14), nor g in equation (12). Otherwise, there are three endogenous variables on the right hand side of each equation. Note that in equation (11) dy/de and dy/dh are ambiguously signed. Extra expenditure on health and education (which are both consumption and investment goods) increases productivity *ceteris paribus*. But such expenditure also crowds out physical capital investment with a long run savings = investment constraint, which leads to lower productivity. Assuming that the realised values of each dependent variable in the i^{th} country are determined by equations of the form of (11-14) plus some idiosyncratic residual effect u_i, the system can be estimated on a cross-section of countries. We will do this using the data described in the next section.

2.2 Measurement of variables in the model

Since we are going to estimate a cross-section regression, our income variable ought to represent a long-run income level. Two alternative measures are the *per capita* GDP figures reported in the Penn World Tables (Heston and Summers, 1991), averaged over time, and the wealth data presented in Dixon and Hamilton (1996). The wealth measure is an attempt to capture the current present discounted value of human, physical and natural resource capital in each country of the world, i,e., *per capita* permanent income. The tables below report the results of using the latter, but results for the former are broadly similar, and are available on request.

The stock of education might be measured in a number of ways. The approach of Barro and Lee (1993), based on years in school, delivers the most sophisticated figures available. However, the range of countries in the sample is quite small, and the authors make strong assumptions about the relative quality of schooling in different countries. This paper employs a more basic measure instead: the percentage of the adult population that is literate. This approach allows the inclusion of a wider range of countries in the sample; moreover, literacy is a direct measure of one aspect of human capital, rather than a measure based on inputs into the production of education. The data are current literacy rates taken from UNESCO (1998). The measure of health used in the model is current (log) life expectancy at birth which, along with the measure of the value of natural resources, is taken from the Dixon and Hamilton dataset.

The ethno-linguistic diversity measure comes from Krain (1997). This measures the probability that two individuals selected randomly from a country's population will speak the same native language. It is the most recent and comprehensive measure of ethno-linguistic diversity available. Temperature data (in tenths of degrees centigrade) come from Hoare (1998). For large countries, the temperature figures used are averages over several large population centres, and for small countries the temperature in the capital city; further details are available on request. Information on colonisation by Britain or France is taken from CIA (1997), and the fraction of the population that is Christian or Muslim comes from Grimes (1996).

The measure of income inequality is the gini coefficient taken from the data set of Deninger and Squire (1998), using only the "high quality" data they collate. For each country, the most recent year's gini is used. We will measure the coefficient on the interval [0,1]. With data

[5] As long as n, the value of natural resources - including land fertility - is measured accurately enough, temperature ought not to affect output independently of n.

[6] Conceivably, a country's colonial history might affect economic efficiency (for example by influencing the degree of literacy in French or English). But colonial dummies were insignificant when added to the income equation. Even with colonial dummies in the y equation, e and g are still identified by the religion variables.

drawn from these disparate sources, there are 95 countries in our data set, of which 26 are in Sub-Saharan Africa. These are listed in Appendix Table A1. The sample means and standard deviations of all continuous variables are reported in Table 1. The measures of each variable are summarised as follows.

y: log *per capita* wealth in $
e: the fraction of the adult population that is literate
h: log years' life expectancy
g: gini coefficients taken from the dataset of Deninger and Squire (1998)
n: the log *per capita* natural resource stock in $
v: Krain's measure of ethno-linguistic diversity
m: *GBR* (dummy for British colonies); *FRA* (dummy for French colonies); *CHR* (the fraction of the population that is Christian; *MUS* (the fraction of the population that is Muslim)
x: mean annual temperature in 0.1 degrees centigrade

Our regression equations also allow for the possibility that the u_i for each equation are spatially correlated: that is, that there are still unobservable characteristics in our model that affect economic development, and that are not randomly distributed across the globe. The equations include the regional dummies *EUR* (Europe), *SAM* (South and Central America), *AFR* (Africa), *MDE* (Western Asia) and *ASP* (the rest of Asia and the Pacific); the baseline intercept in the equations including regional dummies is for N. America.

Table 1: Descriptive Statistics

Sample moments

variable	*mean*	*s.d.*
ln(e/(1-e))	-0.392	2.043
y	10.41	1.703
ln(g/(1-g))	0.410	1.875
h	4.146	0.177
v	0.430	0.282
n	8.545	1.794
x	186.45	71.74
CHR	0.558	0.380
MUS	0.213	0.345

Correlation matrix

	ln(e/(1-e))	y	ln(g/(1-g))
y	0.796		
ln(g/(1-g))	-0.377	-0.259	
h	0.844	0.797	-0.284

3. Estimation Results

In this section we present the results of estimating equations (11-14) by 3SLS (using TSP 4.4). Since our measure of education (the literacy rate) cannot be normally distributed, all observations lying on the interval [0,1], the estimates of the parameters γ_i in equation (12) are based on a logistic regression, i.e.,

$$\mathbf{e} = [1 + \exp(-\Sigma\gamma_i \cdot X_i)]^{-1} + \text{error} \qquad (12a)$$

Since many of the literacy rates for industrialised countries are 99-100%, this does make a small difference to the estimated results. The same transformation is applied to the gini coefficient, which is also bounded, but here there is virtually no difference from a linear regression, since there are few observations anywhere near zero or unity. When the literacy and gini figures appear on the right hand side of an equation, they are entered as ln(e/(1-e)) and ln(g/(1-g)).

Appendix Table A2 reports the results of the unrestricted estimation of the system. These are not discussed in detail: the regressions are over-parameterised and hardly any variable is statistically significant. Table 2 reports the model specification that minimises the Akaike Information Criterion when each of the estimators is used, and it is this that we discuss in detail. The table lists estimated coefficients and associated standard errors, information criteria, equation R^2s and σs, and tests for heteroskedasticity. There is heteroskedasticity in some of the equations, and the standard errors have been corrected using the technique of White (1980). The tables also report the product of each continuous explanatory variables' coefficient and its sample standard deviation, as a rough indicator of its relative importance.

The first equation, for income, shows that increases in inequality, as measured by the gini coefficient, lead directly to lower average income levels. A one standard deviation increase in the gini (i.e., an increase of just under ten percentage points) reduces average *per capita* income by 35-40%. The estimated parameter on the gini coefficient is -1.213. There is also an indirect effect on income through life expectancy: inequality reduces life expectancy, and this reduces *per capita* income. The combination of direct and indirect effects increases magnitude of impact of the gini coefficient on income to an elasticity of -2.074. This represents and effect substantially smaller than the equilibrium effect of the gini coefficient on income implicit in Deninger and Squire (1998, Table 3, Column 1), which is equal to -15.6. This figure is the result of an OLS estimate using GDP figures for *per capita* income, the income gini not being lagged.[7]

Table 2 also shows causality running in the opposite direction, with a Kuznets Curve for income inequality illustrated in Figure 1. The curve in Figure 1 is based on the estimated coefficients on y and y^2 in the gini regression equation.[8] The turning point is very close to the sample log-mean and median *per capita* permanent income (10.41 and 10.34: in levels \$33,190 and \$31,000). At the mean value of y, d(ln(g/(1-g))/dy is equal to -0.006; this figure is not significantly different from zero. Two standard deviations below the mean, the figure is 0.275; two standard deviations above then mean, it is -0.264.

Some of the other coefficients in the income equation, indicating the impact of the exogenous explanatory variables, suggest effects that are qualitatively similar to effects found in previous studies. A one standard deviation increase in the ethno-linguistic fractionalisation measure is associated with an estimated reduction in income of 18.6%. The estimated coefficients are very close to those reported in Easterly and Levine (1997). Also, there is a strong relationship between income and natural resource endowments, with a one standard deviation increase in endowments generating a 94.1% increase in income. One striking result is that whilst increased life expectancy leads to a much higher *per capita* income level, with a one standard

---

[7] Comparison of our results with those of Deninger and Squire is complicated by the fact that their explanatory variable is g, and ours is ln(g/(1-g)). But the effect of g on y in our model is smaller than that of Deninger and Squire for all observed values of g.

[8] The curves are drawn holding all of the dummy variables in the regression equal to zero.

deviation increase in h resulting in a 174% increase in y, higher literacy is associated with lower income *per capita*. One explanation for this is that education expenditure crowds out more productive investment in physical capital. But this does not imply that too much is spent on education, since education is also a consumption good; and it is still the case that higher income levels lead *ceteris paribus* to higher levels of education (see below).

However, one major difference between our results and those of previous studies is in the role of the Africa dummy. This dummy has a significantly *positive* coefficient in the income equation in our estimates. This contrasts starkly with the negative effect in other growth regressions, for example in Easterly and Levine (1997). The difference is explained by the appearance of significant Africa effects in the gini and life expectancy equation. *Ceteris paribus*, African countries have a lower life expectancy and a poorer income distribution, which worsens economic performance. Solving out the four equations, the net impact of being in Africa (starting from the sample mean income level) would be to reduce a country's income by 242%. African countries are not unusually inefficient, given their socio-economic conditions, but they are unusually unhealthy and unusually inequitable.

The Africa dummy is not the only significant regional effect on income distribution. There is a significantly negative coefficient on the Europe dummy in the gini equation. Europe has a more equitable income distribution relative to the developing world (and relative to North America and South East Asia) than can be explained by its income level, or by any other characteristic in the model. Similarly, countries that have been British or French colonies have more equitable income distributions. The British colonisation effect is estimated to be around 17 percentage points, the French around 25.

Life expectancy is reduced by greater income inequality: a one standard deviation increase in the gini coefficient reduces life expectancy by 3.6%. Redistribution of income to poor improves the overall health of a country. Life expectancy also varies with climate: there is a significant negative quadratic relationship between health and mean annual temperature. Life expectancy is highest in temperate climates, at an average temperature of 19.3^{O}.

The variable with the largest impact on life expectancy, however, is the literacy rate. A one standard deviation increase in literacy increases life expectancy by around 17%. Better education may improve standards of hygiene, or may be associated with lower fertility (and therefore infant mortality) rates. In the absence of appropriate instruments, it is not possible to identify these effects. There is a virtuous spiral between health and education, since a higher life expectancy increases literacy rates. The impact of a one standard deviation increase in life expectancy is roughly the same as that of a similar increase in income: both result in values of $\ln(e/(1-e))$ that are around 50-60% higher.

4. What do the Results tell us about Africa?

Table 2 highlights a number of independent variables that are important in the development process: ethno-linguistic diversification, natural resource allocation, colonial history and climate. Part of the explanation for Africa’s poverty lies in these variables. But even after we have conditioned on these variables, Africa’s performance is still unusually poor with respect to health and income inequality, as indicated by the significance of the Africa dummy in the Table 2 equations for these variables. In this section we will first of all assess the magnitude of the extent to which under-development in Africa is explained by the independent factors, those that are beyond the control of any policy-maker. Then we will investigate further the possible reasons for the exceptionally poor performance of African countries as a whole with respect to health and income inequality.

Table 2: 3SLS Model Estimates (White-Corrected Standard Errors)

variable	coeff.	std. err.	|t-ratio|	p-value	coeff.×s.d.
y equation (linear)					
intercept*100	-0.34250	0.10816	3.16671	[.002]	----
ln(e/(1-e))	-0.46252	0.23353	1.98058	[.048]	-0.94
ln(g/(1-g))	-1.21267	0.36651	3.30866	[.001]	-2.27
h	9.82941	2.62768	3.74072	[.000]	1.74
v	-0.66116	0.39536	1.67232	[.094]	-0.19
n	0.52432	0.08611	6.08867	[.000]	0.94
AFR	0.82706	0.29255	2.82710	[.005]	----
MDE	-0.77779	0.34004	2.28735	[.022]	----
R^2 = 0.74577; σ = 0.86814; LM heteroskedasticity: $\chi^2(1)$ = 1.00691					
e equation (logistic)					
intercept*100	-0.13124	0.02922	4.49144	[.000]	----
y	0.38006	0.12972	2.92975	[.003]	0.65
h	2.72442	0.97748	2.78719	[.005]	0.48
MUS	0.90488	0.17821	5.07772	[.000]	0.31
GBR	-0.29143	0.16256	1.79277	[.073]	----
FRA	-0.65187	0.19775	3.29637	[.001]	----
R^2 = 0.85586; σ = 0.08784; LM heteroskedasticity: $\chi^2(1)$ = 9.57087					
g equation (logistic)					
intercept*100	0.04746	0.02160	2.19775	[.028]	----
y	0.82905	0.40814	2.03129	[.042]	1.42
y^2	-0.03854	0.01879	2.05147	[.040]	-0.11
CHR	0.31671	0.10270	3.08393	[.002]	0.12
GBR	-0.16544	0.06845	2.41714	[.016]	----
FRA	-0.24587	0.09258	2.65587	[.008]	----
EUR	-0.62100	0.09477	6.55290	[.000]	----
AFR	0.39144	0.08226	4.75857	[.000]	----
R^2 = 0.59002; σ = 0.06193; LM heteroskedasticity: $\chi^2(1)$ = 1.16641					
h equation (linear)					
intercept*100	0.33379	0.01692	19.73280	[.000]	----
ln(e/(1-e))	0.09058	0.00992	9.12935	[.000]	0.19
ln(g/(1-g))	-0.08763	0.04268	2.05305	[.040]	-0.16
x/1000	7.34944	1.71228	4.29219	[.000]	1.37
x^2/1000	-0.01901	0.00445	4.26739	[.000]	-0.66
AFR	-0.08563	0.02734	3.13222	[.002]	----
R^2 = 0.75543; σ = 0.09178; LM heteroskedasticity: $\chi^2(1)$ = 0.47681					

Schwartz-Bayesian Criterion = 0.74228; Hannan-Quinn Criterion = 0.29371
Akaike Criterion = -0.01044; Log-likelihood = 113.9841

Table 3 illustrates Africa's some geographical characteristics relative to the rest of the world: it has fewer natural resources *per capita*, it is hotter, and it suffers from greater ethno-linguistic diversification, all of which directly or indirectly affect *per capita* income. The table also illustrates the likely magnitude of these effects on income, using the coefficients of the model reported in Table 2. The final row of Table 3 indicates the percentage increase in *per capita* income above the African average that the Table 2 model predicts would result if each of the three characteristics (natural resource endowment, climate and ethno-linguistic diversity) were changed from the African average to the North American average. The figure for ethno-linguistic diversity is around 10%; the figure for natural resources is over 100%. So ethno-linguistic diversity does matter, as Easterly and Levine (1997) point out, but its effect is dwarfed by the natural resource effect. The figure for climate is negative, since average temperatures in Africa are closer to the optimum for health than in those in North America. A final point to be made about Table 3 is that although colonial history does have a significant indirect impact on economic development, Africa was not particularly heavily (or lightly) colonised by Britain or

France, relative to other parts of the world, so the colonisation variables do not explain African under-performance.

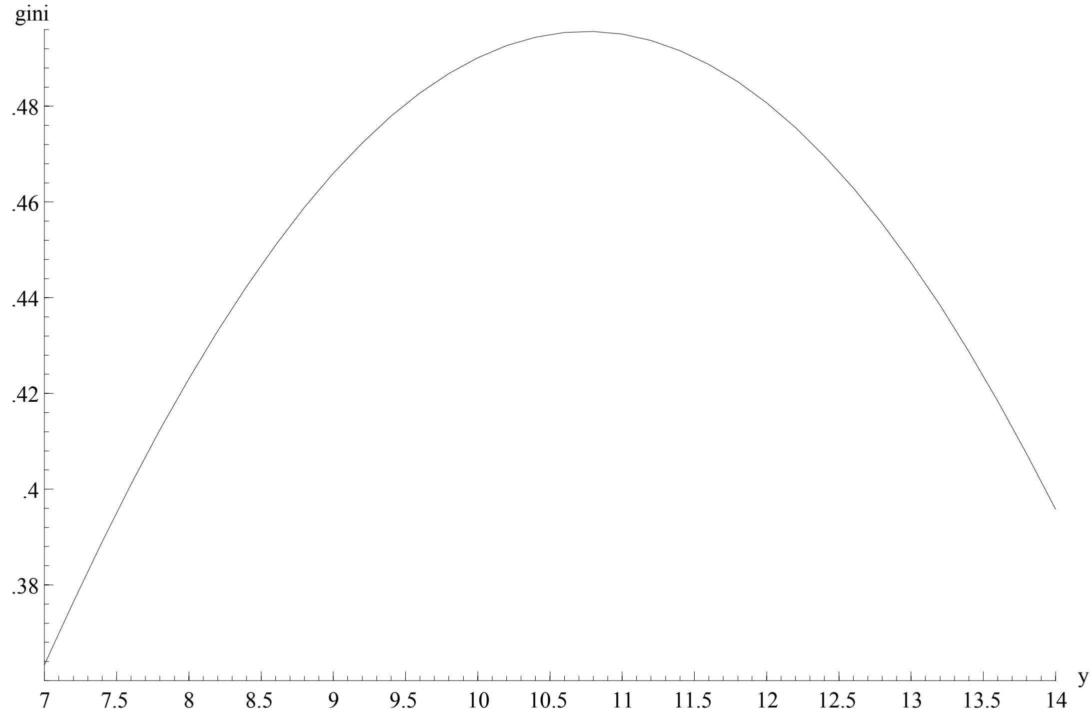

Figure 1: Estimated Kuznets Curve: Gini Coefficient as a Function of log Income

However, the significant Africa dummies in the equations for life expectancy and the gini coefficient in Table 2 indicate that these observed natural characteristics do not completely explain poor African economic performance. In the case of income inequality we have no direct evidence to bring to the question of why the Africa dummy is significant. One speculative explanation, which remains to be tested, is that war and civic strife inhibit income redistribution. Collier and Hoeffler (1998) demonstrate that Africa has been more prone to civil wars than other parts of the globe. This is partly due to low income, but may also be partly due to historical and political factors.

With respect to life expectancy, there do exist some descriptive statistics that shed light on Africa's under-performance. These are summarized in Table 4. The first part of the table reports coefficients from an OLS regression equation for the variable h. The key explanatory variables in the regression are the fraction of the population of each country with access to safe water, the fraction with access to sanitation, and the log of the number of physicians per thousand people. [9](The regression also includes independent variables from the model in Table 2, but the coefficients on these are not reported.) It is important to note that the equation does not

[9] The data for these three variables are taken from the World Bank World Development Indicators. These data are available for a sample of 95 countries that is not identical to the sample for the Table 2 regression, but which contains countries from each continent in a proportion insignificantly different from in the original sample.

necessarily capture the magnitude of the impact of, for example, an increase in the number of doctors on life expectancy, since the number of doctors might be endogenous to national income, and we know that income depends on life expectancy. However, it does at least show how life expectancy is correlated with a number of key inputs in the production of health. There is a large and significant correlation between all three of the inputs and life expectancy. The partial R^2s on access to safe water, sanitation and physicians per 1000 are, respectively, 0.10, 0.18 and 0.17. Together with the conditioning variables, they explain over 91% of the cross-country variation in life expectancy.

Table 3: Regional Means of Some Independent Variables in the Model

region	natural resources (log $pc)	climate (0.1° C)	ethnic diversity (% figure)	% of region colonised by Britain	% of region colonised by France
EUROPE	9.53	99.88	21.3	8.3	4.7
N. AMERICA	11.35	123.00	50.7	66.7	33.3
S. AMERICA	8.84	212.00	45.1	31.6	5.3
AFRICA	7.92	236.38	66.5	50.0	46.2
MID-EAST	8.19	172.00	28.2	40.0	60.0
ASIA-PAC.	7.44	217.39	38.6	50.0	11.1
Table 2 Effects (%)	179.2	-47.98	9.9		

A striking aspect of the regression equation is that the coefficient on the Africa dummy in the OLS regression (not reported in the table) is only -0.04 (compared with -0.09 in the h equation in Table 2). The t-ratio is only -1.26: *ceteris paribus*, life expectancy is not significantly lower in Africa than it is in North America. This suggests that most of the Africa dummy effect in Table 2 is associated with low levels of the three health input variables in Table 4, or of other inputs correlated with these. The second half of Table 4 lends support to this conjecture: the means of the three variables (conditioning on the set of independent variables) are significantly lower in Africa than in North America. For all three health variables the means are lower in Africa than in any other part of the World.

In other words, the poor health outcomes in Africa are closely associated with low standards with respect to key health inputs. Some of the under-performance could be related to geographical factors. The costs of providing access to safe water and sanitation are likely to depend on local geography (on aridity, for example). In the absence of data for unit costs in the provision of health inputs, it is not possible to explore the reasons for Africa's poor performance in any more quantitative detail. Nevertheless, the results here indicate that better health provision is likely to be a key condition for any future improvement in Africa's economic performance.

5. Summary and Conclusion

This is the first paper (to our knowledge) to use cross-country data to estimate simultaneously the relationship between income inequality and other indicators of development. The overall picture is that there is a correlation between reductions in inequality and improvements in development indicators such as *per capita* income, literacy and life expectancy (Table 1). Table 2 indicates that this correlation is the result of a causal relationship running from income inequality to average income and life expectancy. Inequality reduces the average material wellbeing of a country. Inequality itself is dependent on average income levels, in the form of the Kuznets Curve relationship, although the sensitivity of inequality to income is estimated to be lower than is suggested by some previous studies. At very high income levels, the negatively sloped part of the Kuznets Curve reinforces the negative correlation between inequality and other development indicators; at very low income levels the positively sloped part of the Kuznets Curve reduces the correlation.

Table 4: Significant Correlates of Life Expectancy

*Conditional correlation of **h** with health variables*

Variable	coeff.	std. err.	t ratio
safe water	0.14598	0.04739	3.081
sanitation	0.16170	0.03751	4.310
ln(doctors/1000)	0.03153	0.00764	4.127

Conditional regional means of health variables (with standard deviations)

area	safe water	sanitation	ln(doctors)
North America	0.66 (0.20)	0.66 (0.26)	0.04 (1.00)
European difference	-0.04 (0.11)	-0.13 (0.14)	-0.11 (0.55)
S. American difference	-0.17 (0.11)	-0.26 (0.14)	-0.11 (0.53)
African difference	-0.28 (0.11)	-0.30 (0.14)	-1.59 (0.53)
Mid-East difference	-0.22 (0.13)	-0.19 (0.17)	-0.92 (0.63)
Asia-Pacific difference	-0.14 (0.11)	-0.16 (0.14)	-0.78 (0.54)

The estimated model also sheds new light on the sources of poor economic performance in Sub-Saharan Africa. Part of Africa's poverty is due to exogenous geographical characteristics, including natural resource endowment and (less importantly) ethno-linguistic fractionalisation. This part of the story is more pessimistic, since there is little Africa can do to change its geography. However, there is also a source of optimism. Some of Africa's growth deficit is due to unusually poor performance with respect to health and income distribution, even when the geographical factors have been taken into account. Descriptive statistics suggest that this poor performance is at least partly due to under-expenditure on public goods, a characteristic that can be changed in the future.

References

R. Barro (1991) "Economic Growth in a Cross-Section of Countries", *Quarterly Journal of Economics*, Vol. 106, pp. 407-443

R. Barro and Lee, J-W. (1993) "International Comparisons of Educational Attainment", *Journal of Monetary Economics*, Vol. 32, pp. 363-394

F. Caselli, Esquivel, G. and Lefort, F. (1996) "Reopening the Convergence Debate: A New Look at Cross-Country Growth Empirics, *Journal of Economic Growth*, Vol. 1, pp. 363-390

S. Chatterjee (1991) "The Effect of Transitional Dynamics on the Distribution of Wealth in a Neoclassical Model of Capital Accumulation", working paper, Federal Reserve Bank of Philadelphia.

CIA (1997) *World Factbook*, Washington, DC.

P. Collier and A. Hoeffler (1998) “On Economic Causes of Civil War”, *Oxford Economic Papers*, Vol. 50, pp. 563-73

K. Deninger and Squire, L. (1998) "New Ways of Looking at Old Issues: Inequality and Growth", *Journal of Development Economics*, Vol. 57, pp. 259-287.
J. Dixon and Hamilton, K. (1996) "Expanding the Measure of Wealth", *Finance and Development*
W. Easterly (1999) "Life During Growth", mimeo, World Bank, Washington, DC.
W. Easterly and Levine, R. (1997) "Africa's Growth Tragedy: Policies and Ethnic Divisions", *Quarterly Journal of Economics*, Vol. 112, pp. 1203-1250
D. Fielding (1999) "A Structural Model of Social and Economic Development", University of Leicester Discussion Paper in Economics #99/1
O. Galor and Zeira, J. (1993) "Income Distribution and Macroeconomics", *Review of Economic Studies*, Vol. 60, pp. pp. 35-52
B. Grimes, ed. (1996) *Ethnologue: Languages of the World* (13th edition), Summer Institute of Linguistics, Dallas, TX
R. Hardin (1992) "The Street-Level Epistemology of Trust", *Analyse und Kritik*, Vol. 14, pp. 152-176
C. Henderson (1991) "Conditions Affecting the Use of Political Repression", *Journal of Conflict Resolution*, Vol. 35, pp. 120-142
C. Henderson (1993) "Population Pressures and Political Repression", *Social Science Quarterly*, Vol. 74, pp. 322-333
A. Heston and Summers, L. (1991) "The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950-1988", *Quarterly Journal of Economics*, Vol. 106, pp. 327-368
R. Hoare (1998) "World Climate", http://www.worldclimate.com
R. King and Levine, R. (1993) "Finance and Growth: Schumpeter Might be Right", *Quarterly Journal of Economics*, Vol. 108, pp. 717-737
S. Knack and Keefer, P. (1997) "Does Social Capital Have an Economic Payoff?", *Quarterly Journal of Economics*, Vol. 112, pp. 1251-1288
M. Krain (1997) "State-Sponsored Mass Murder: The Onset and Severity of Genocides and Politicides", *Journal of Conflict Resolution*, Vol. 41, pp. 331-360
S. Kuznets (1955) "Economic Growth and Income Inequality", *American Economic Review*, Vol. 45, pp. 1-28
K. Lee, Pesaran, H. and Smith, R. (1998) "Growth Empirics: A Panel Approach - A Comment", *Quarterly Journal of Economics*, Vol. 113, pp. 319-323
H. Pesaran and Smith, R. (1995) "Estimating Long-Run Relationships from Dynamic Heterogenous Panels", *Journal of Econometrics*, Vol. 68, pp. 79-113
H. Park (1987) "Correlates of Human Rights: Global Tendencies", *Human Rights Quarterly*, Vol. 9, pp. 405-413
S. Poe and Tate, C. (1994) "Repression of Human Rights to Personal Integrity in the 1980s: A Global Analysis", *American Political Science Review*, Vol. 88, pp. 853-872
UNESCO (1998) *Statistical Yearbook*, Bernan Press, Lanham, MD
H. White (1980) "A Heteroskedasticity-Consistent Covariance Matrix and a Direct Test for Heteroskedasticity", *Econometrica*, Vol. 48, pp. 721-746

Appendix Table A1: Countries in the Sample
with their Wealth Rankings in the Dixon-Hamilton Data Set and Gini Coefficients

country	wealth rank	gini	country	wealth rank	gini	country	wealth rank	gini
Ethiopia	1	0.442	Mauritania	58	0.378	Chile	128	0.5649
Nepal	2	0.3006	Philippine	59	0.45	Trinidad	129	0.4172
Malawi	4	0.62	Romania	65	0.2866	Yugoslavia	130	0.3188
Uganda	5	0.4078	Guatemala	66	0.5906	Mexico	131	0.5031
Tanzania	6	0.381	Morocco	67	0.392	Guyana	133	0.4022
Viet Nam	7	0.3571	C.A.R.	68	0.55	Puerto Rico	144	0.5086
Guinea Bissau	10	0.5612	Peru	69	0.4487	South Korea	149	0.3364
Bangladesh	12	0.2827	Ecuador	70	0.43	Portugal	152	0.3563
Niger	13	0.361	Dom. Rep.	71	0.49	Greece	153	0.3519
Burkina Faso	15	0.39	Jordan	72	0.4066	Botswana	162	0.5421
Gambia	16	0.39	Syria	75	0.4204	New Zeland	163	0.4021
Kenya	17	0.5439	Bulgaria	77	0.3442	Ireland	164	0.346
Mali	18	0.54	Lithuania	80	0.3364	Bahamas	165	0.4529
Nigeria	19	0.3747	Colombia	82	0.5132	Spain	168	0.2591
India	20	0.3202	Algeria	87	0.3873	Hong Kong	169	0.45
Madagascar	21	0.4344	Tunisia	93	0.4024	Singapore	170	0.39
China	31	0.378	Slovakia	95	0.215	Britain	171	0.324
Laos	32	0.304	Turkey	96	0.4409	Finland	172	0.2611
Pakistan	33	0.3115	Thailand	97	0.515	Italy	173	0.3219
Ghana	35	0.3391	Costa Rica	102	0.4607	Netherlands	174	0.2938
Nicaragua	37	0.5032	Guinea	104	0.404	Belgium	175	0.2692
Djibouti	39	0.381	Iran	105	0.429	Germany	178	0.322
Sri Lanka	42	0.301	Panama	109	0.5647	France	180	0.3491
Honduras	43	0.54	Jamaica	110	0.3792	U.S.A.	181	0.3794
Lesotho	44	0.5602	Brazil	113	0.596	Norway	182	0.3331
Senegal	45	0.5412	Mauritius	114	0.3669	Denmark	183	0.332
Cote d'Ivoire	46	0.38	Poland	116	0.3306	Sweden	187	0.3244
Egypt	47	0.32	Czech Rep.	117	0.2826	Japan	188	0.35
Zimbabwe	49	0.5683	Malaysia	121	0.4835	Luxembourg	190	0.2713
Indonesia	52	0.3169	Venezuela	123	0.5384	Canada	191	0.2765
Zambia	53	0.524	South Africa	124	0.623	Australia	192	0.4172
Cameroon	55	0.49	Hungary	126	0.2794			

Appendix Table A2: The Unrestricted Model

variable	coeff.	std. err.	t-ratio	p-value
y equation (linear)				
intercept	-26.12330	11.89890	-2.19544	[.028]
ln(e/(1-e))	-0.25417	0.30479	-0.83392	[.404]
ln(g/(1-g))	-1.06242	0.59345	-1.79025	[.073]
h	7.91418	2.85954	2.76764	[.006]
v	-0.85604	0.44142	-1.93928	[.052]
EUR	-0.28202	0.46669	-0.60429	[.546]
SAM	-0.18766	0.41160	-0.45593	[.648]
AFR	0.61228	0.55542	1.10237	[.270]
MDE	-0.63600	0.45627	-1.39391	[.163]
ASP	-0.06039	0.29811	-0.20257	[.839]
n	0.48959	0.09435	5.18927	[.000]
e equation (logistic)				
intercept	-11.32590	6.84444	-1.65476	[.098]
y	0.35096	0.14950	2.34756	[.019]
h	2.36730	1.88472	1.25605	[.209]
v	0.26624	0.48085	0.55368	[.580]
EUR	0.50257	0.65022	0.77292	[.440]
SAM	-0.56103	0.62398	-0.89913	[.369]
AFR	-0.40658	0.68366	-0.59470	[.552]
MDE	-0.13428	0.69432	-0.19340	[.847]
ASP	-0.21137	0.65601	-0.32221	[.747]
CHR	0.17564	0.40955	0.42886	[.668]
MUS	-0.92417	0.44105	-2.09536	[.036]
GBR	-0.25043	0.21768	-1.15045	[.250]
FRA	-0.63453	0.23684	-2.67920	[.007]
g equation (logistic)				
intercept	-0.65609	4.07290	-0.16109	[.872]
y	0.61096	0.65816	0.92829	[.353]
y^2	-0.02685	0.03042	-0.88244	[.378]
h	-0.74222	1.10200	-0.67352	[.501]
v	-0.14691	0.17097	-0.85924	[.390]
EUR	-0.49852	0.12692	-3.92795	[.000]
SAM	0.19090	0.17395	1.09743	[.272]
AFR	0.40916	0.26895	1.52133	[.128]
MDE	0.34109	0.22842	1.49322	[.135]
ASP	0.03322	0.16019	0.20735	[.836]
CHR	0.24122	0.13415	1.79817	[.072]
MUS	-0.20021	0.19830	-1.00962	[.313]
GBR	-0.14326	0.08249	-1.73662	[.082]
FRA	-0.25880	0.09234	-2.80272	[.005]
h equation (linear)				
intercept	2.72468	0.66774	4.08042	[.000]
ln(e/(1-e))	0.05895	0.04122	1.43014	[.153]
ln(g/(1-g))	-0.15604	0.10203	-1.52930	[.126]
y	0.03132	0.03442	0.90973	[.363]
v	0.02893	0.08760	0.33029	[.741]
EUR	0.06524	0.10414	0.62652	[.531]
SAM	0.08344	0.07183	1.16175	[.245]
AFR	-0.04969	0.08757	-0.56739	[.570]
MDE	-0.00983	0.10332	-0.09513	[.924]
ASP	0.01741	0.06591	0.26414	[.792]
x/1000	10.79000	0.50400	2.13982	[.032]
x^2/1000	-0.02844	0.01364	-2.08515	[.037]