Determinants of long-term economic growth redux: A Measurement Error Model Averaging (MEMA) approach

Gernot Doppelhofer
Ole-Petter Moe Hansen
Melvyn Weeks

ESCoE, May 16-17 2018

Introduction

Penn World Tables (PWT) supply Purchasing Power Parity (PPP) adjusted income data, released in vintages.

Example: Mean growth 1960-1996

PWT-vintage	6.0	6.1	6.2	6.3	7.0	7.1	8.0	8.1	9.0
Ghana	-0.41	1.21	3.31	2.13	0.99	0.08	-0.26	-0.26	-.24

Differences are caused by differing methodology, price data, national accounts data... *we don't know which is "best"*(Johnson et al., 2013).

Introduction

Uncertainty of the underlying data spills over to results from cross-country analyses of growth

In particular: which variables influence *long run growth rate*? Is it

- Institutions?
- Demographics?
- Education?
- Resource endowments?
- Geographical/climatic conditions?
- ...or other potential explanations?

Introduction

Model Averaging established method for analysing growth determinants:

- Observe *average annual growth rate 1960-1996* in few countries ($N \approx 90$)
- Have many potential covariates ($K = 67$)
- Some potential covariates highly correlated
- Don't know the "correct" model specification

Solution:

- Average across all models
- Impose shrinkage through priors to handle dimensionality.

However, results are *sensitive* to which PWT-vintage is used.

Introduction

We estimate *robust growth determinants*:

- Measurement Error (ME)
  - Take data uncertainty into account by modelling income as a *latent variable*
  - Use all PWT vintages (many data sources) to identify the distribution of "true" income
  - Flexible parametric structure to allow for differing *reliability* of data sources
- Model Averaging (MA)
  - Explicitly take into account *specification uncertainty* of the growth process
  - Use flexible, parametric setup to allow for *outliers* in growth process

Bring it all together in a *MEMA*-model.

Results

- We identify 18 variables robustly related to growth.
- Quantify reliability of data:
  - How reliable are different PWT-*Vintages*?
  - How much do we know about income in different *countries*?
- Detect *outliers* in the growth process that behave very different from most countries.

- Results are robust to alternative priors, specifications etc. once we take *uncertainty about income measurement* (PWT) into account.

Literature

- Model averaging in growth regressions
  - Sala-i Martin (1997)
  - Fernandez, Ley and Steel (2001)
  - Sala-i Martin, Doppelhofer and Miller (2004)
- "Noise" in the Penn World Tables (PWT)
  - Johnson et al. (2013)
  - Deaton and Heston (2010)
- Measurement error in general
  - Hausman (2001)
  - Lubotsky and Wittenberg (2006)
- Model averaging *and* "noise" in PWT:
  - Ciccone and Jarociński (2010)

Measurement error model

We use $i = 1, ...N$ countries in analysis.

Have eight measurements of income through PWT-vintages, $v = 1, ..., 8$

Need to extract true income in the initial (1960) and end (1996) period.

How can we identify true income?

Measurement error model

Measurement equations:

$$y^{I}_{v,i} = a_v + y^{I}_i + \sigma^{I}_{v,i}\varepsilon^{I}_{v,i} \tag{1}$$

$$y^{E}_{v,i} = a_v + y^{E}_i + \sigma^{E}_{v,i}\varepsilon^{E}_{v,i} \tag{2}$$

Where we assume

$$\begin{aligned} \varepsilon^{I}_{v,i} &\sim N(0,1) \\ \varepsilon^{E}_{v,i} &\sim N(0,1) \end{aligned} \tag{3}$$

- Country i, vintage v, initial I, end E
- $y^{I}_{v,i}, y^{E}_{v,i}$ are *observed* income, ${y_i}^{I}$ and ${y_i}^{E}$ denote the *truth*
- a_v is a vintage specific level fixed effect
- $\sigma^{I}_{v,i}$ and $\sigma^{E}_{v,i}$ are scale parameters

Modelling variance of measurement error (ME)

$$\sigma^I_{v,i} = \sqrt{\omega^N_i \omega^V_v}\sigma^I$$
$$\sigma^E_{v,i} = \sqrt{\omega^N_i \omega^V_v}\sigma^E$$

Where:

- $\left(\sigma^I\right)^2, \left(\sigma^E\right)^2$ are *average variance* of ME in 1960 and 1996
- $\omega^V_1, ...\omega^V_V$ is ME-variance of each vintage *relative to average*
- $\omega^N_1, ...\omega^N_N$ is ME-variance of each country *relative to average*

Note: $1 \geq \omega^i_j \geq 0, \forall i, j$, and $\sum_{j \in J} \omega^i_j = 1$

Measurement error (ME) variance priors

Used *scaled Dirichlet*-distribution to capture relative variance.

$$\left(\omega_1^V, \ldots \omega_V^V\right) / V \sim Dir\left(\Omega_1^V, \ldots \Omega_V^V\right)$$
$$\left(\omega_1^N, \ldots \omega_N^N\right) / N \sim Dir\left(\Omega_1^N, \ldots \Omega_N^N\right)$$
$$\sigma^I \sim U(0, 1000)$$
$$\sigma^E \sim U(0, 1000)$$

Constants $\Omega_1^V, \ldots \Omega_V^V, \Omega_1^N, \ldots \Omega_N^N$ allow us to remain a priori agnostic of measurement error - or can impose e.g. a given vintage as *true*.

Model Averaging

$$\frac{y_i^E - y_i^I}{T_1 - T_0} = \alpha + \sum_{k=1}^{K} x_{k,i}\beta_k\gamma_k + \sigma_i\varepsilon_i \qquad (4)$$

- LHS is average growth
- β_k is the coefficient for variable k
- σ_i is a scale parameter
- ε_i is an *i.i.d.* standard normal error term.
- γ_k is a *binary* variable, indicating whether variable k is included in the regression.

Model Averaging

Use the standard Zellner-prior for coefficients:

$$\beta|\sigma^2, M \sim N(0, \sigma^2 \mathbf{V}_{0j}) \tag{5}$$

Where

$$\mathbf{V}_{0M} = (g_0 \mathbf{X}'_M \mathbf{X}_M)^{-1} \tag{6}$$

with factor of proportionality g_0, and $\mathbf{X}_M$ is the matrix of covariates that are included in model M.and $g_0 = N^{-1}$.

$$\sigma \sim U(0, 1000) \tag{7}$$

Note: Initial income is one of the potential regressors!

Model Averaging

We follow the BACE-prior, and place independent Bernoulli priors on the γ_k, with prior inclusion probability $7/67$:

$$\gamma_k \sim Bernoulli\left(\frac{7}{67}\right) \tag{8}$$

Results robust to richer specification of prior inclusion probability and Zellner shrinkage factor.

Outliers

The parameter σ_i^2 is the unexplained variance of growth in each country. Allow for heteroskedasticity using Dirichlet-trick:

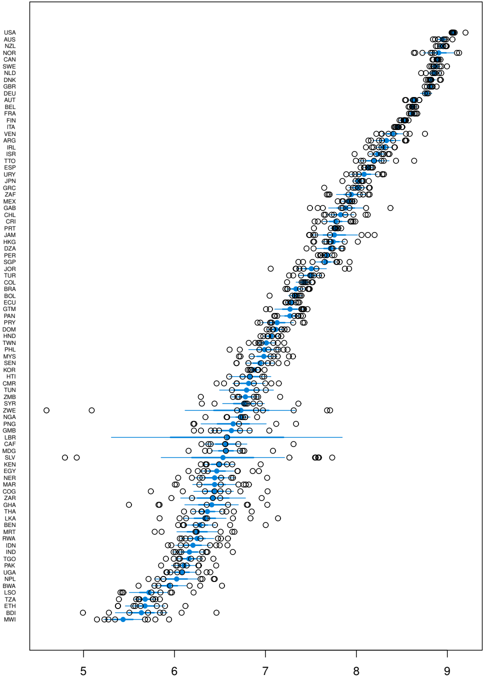
$$\sigma_i = \sqrt{\omega_i}\sigma$$
$$(\omega_1, ..., \omega_N) / N \sim Dir(\Omega_1, ..., \Omega_N)$$
$$\sigma \sim U(0, 1000)$$

Hyperconstants $\Omega_1, ..., \Omega_N$ allow us to remain a priori agnostic of outliers, or e.g. impose homoscedasticity.

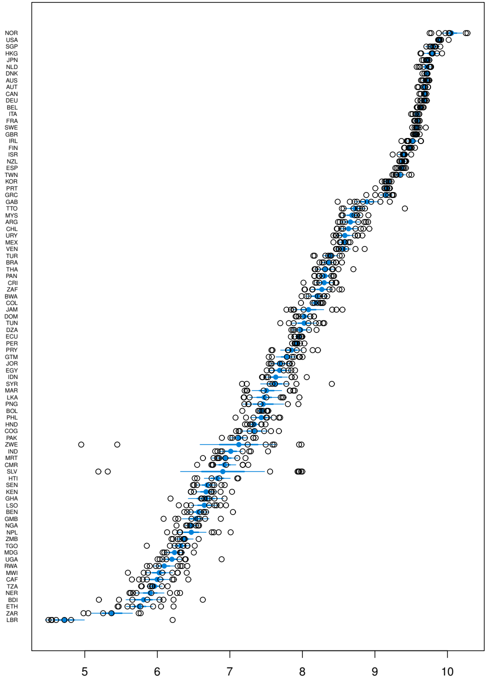
Roadmap of results

1. Fit ME-model to obtain estimates of *latent income* and *reliability of measurements in PWT-vintages and countries*.
2. Estimate full MEMA-model.

Log GDP PC
1960



Log GDP PC 1996

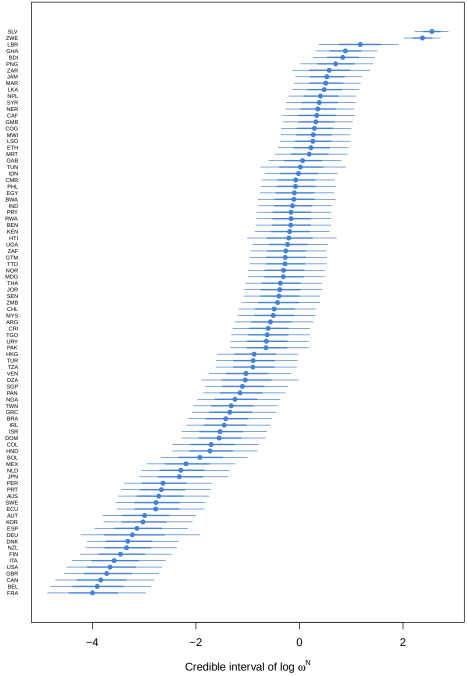


Relative measurement error in PWT-vintages

PWT 6.0	PWT 6.1	PWT 6.2	PWT 6.3	PWT 7.0	PWT 7.1	PWT 8.0	PWT 8.1
2.11	1.01	1.05	0.9	0.89	0.87	0.59	0.58
(0.22)	(0.14)	(0.14)	(0.12)	(0.13)	(0.13)	(0.11)	(0.1)

Posterior means and standard deviations.

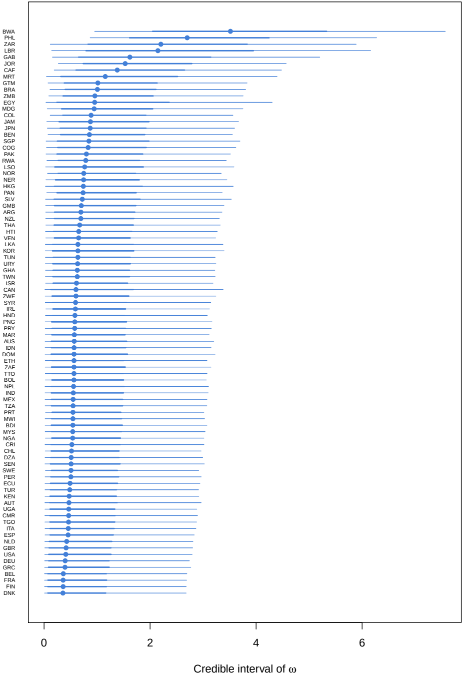
Rel. var of ME in countries



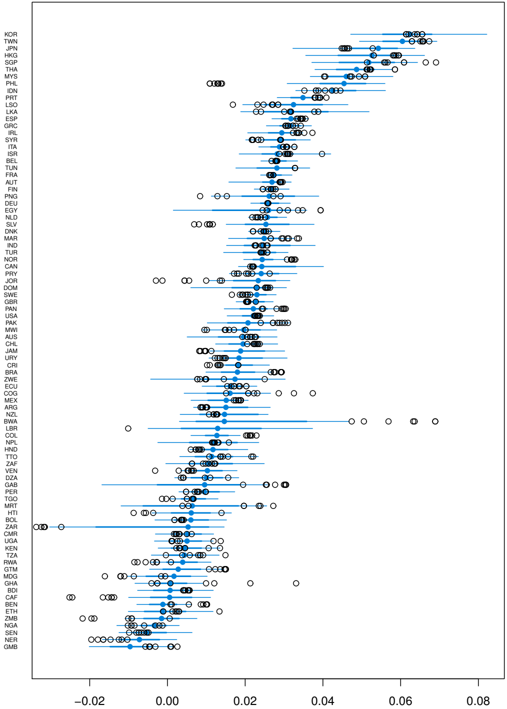
Posterior inclusion probabilities

Outlier model	Homoscedastic	Dirichlet
ME-variance by country	Random, agnostic	Random, agnostic
ME-variance by vintages	Random, agnostic	Random, agnostic
Initial Income	97.4	99.8
East Asian Dummy	81.2	90.5
Primary Schooling Enrollment	89.4	90.3
Fraction of Tropical Area	45.8	51.1
Air Distance to Big Cities	30.0	43.6
Investment Price	58.3	34.4
Life Expectancy	24.6	34.2
Population Density	77.8	34.0
Sub-Saharan Africa Dummy	25.3	31.8
Population Coastal Density	37.3	26.6
Primary Exports	41.1	20.8
Latin American Dummy	8.7	20.1
Fraction Muslim	22.7	18.1
Malaria Prevalence	8.5	16.6
Fraction Confucian	19.0	14.8
Openness 1965-74	10.5	13.0
Landlocked Country Dummy	5.3	12.4
Outward Orientation	3.9	11.1
Fraction Population In Tropics	19.9	7.8
Years Open 1950-94	14.0	5.9
Political Rights	15.6	5.1
Fraction Buddhist	13.9	4.7
Ethnolinguistic Fractionalization	10.6	3.1
Fraction GDP in Mining	22.5	2.7

Rel. var of countries' *growth*



Predicted and realised growth



Main results

	Cond. mean	Std.d.	Sign certainty	PIP
Initial Income	-1.1E-02	2.7E-03	100.0	99.8
East Asian Dummy	2.2E-02	5.5E-03	100.0	90.5
Primary Schooling Enrollment	3.4E-02	8.6E-03	100.0	90.3
Fraction of Tropical Area	-1.3E-02	3.6E-03	99.7	51.1
Air Distance to Big Cities	-1.7E-06	5.5E-07	99.7	43.6
Investment Price	-7.8E-05	2.7E-05	99.5	34.4
Life Expectancy	7.1E-04	2.9E-04	99.6	34.2
Population Density	2.1E-05	7.4E-06	99.3	34.0
Sub-Saharan Africa Dummy	-1.5E-02	6.4E-03	99.5	31.8
Population Coastal Density	6.5E-06	2.3E-06	99.3	26.6
Primary Exports	-1.5E-02	5.8E-03	99.0	20.8
Latin American Dummy	-1.2E-02	5.1E-03	97.5	20.1
Fraction Muslim	1.4E-02	5.7E-03	98.8	18.1
Malaria Prevalence	-1.3E-02	5.7E-03	98.0	16.6
Fraction Confucian	4.7E-02	2.2E-02	98.3	14.8
Openness 1965-74	8.7E-03	4.1E-03	98.2	13.0
Landlocked Country Dummy	-7.2E-03	3.3E-03	98.5	12.4
Outward Orientation	-4.4E-03	2.1E-03	98.2	11.1

Summary: MEMA

- We find 18 variables *robustly* related to growth
  - Model uncertainty
  - Measurement error
  - Outliers
  - Parametric specification
- Robustness checks
  - Alternative outlier classification Link
  - Random model size and g-factor Link
  - Data sources and priors Link

References

Ciccone, Antonio, and Marek Jarociński. 2010. "Determinants of economic growth: will data tell?" *American Economic Journal: Macroeconomics*, 2(4): 222–246.

Deaton, Angus, and Alan Heston. 2010. "Undecountries PPPs and PPP-based national accounts." *American Economic Journal: Macroeconomics*, 2(4): 1–35.

Fernandez, Carmen, Eduardo Ley, and Mark FJ Steel. 2001. "Model uncertainty in cross-country growth regressions." *Journal of applied Econometrics*, 16(5): 563–576.

Geweke, John. 1993. "Bayesian treatment of the independent student-t linear model." *Journal of Applied Econometrics*, 8(S1): S19–S40.

Hausman, Jerry. 2001. "Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left." *Journal of Economic Perspectives*, 15(4): 57–67.

Hoeting, Jennifer, Adrian E Raftery, and David Madigan. 1996. "A method for simultaneous variable selection and outlier identification in linear regression." *Computational Statistics & Data Analysis*, 22(3): 251–270.

Johnson, Simon, William Larson, Chris Papageorgiou, and Arvind Subramanian. 2013. "Is newer better? Penn World Table revisions and their impact on growth estimates." *Journal of Monetary Economics*, 60(2): 255–274.

Ley, Eduardo, and Mark FJ Steel. 2009. "On the effect of prior assumptions in Bayesian model averaging with applications to growth regression." *Journal of applied econometrics*, 24(4): 651–674.

Liang, Feng, Rui Paulo, German Molina, Merlise A Clyde, and Jim O Berger. 1998. "Mixtures of g priors for Bayesian variable selection." *Journal of the American Statistical Association*, 103(481): 410–423.

Lubotsky, Darren, and Martin Wittenberg. 2006. "Interpretation of Regressions with Multiple Proxies." *Review of Economics and Statistics*, 88(3): 549–562.

Sala-i Martin, Xavier. 1997. "I just ran 2 million regressions." *American Economic Review (Papers and Proceedings)*, 87(2): 178183.

Sala-i Martin, Xavier, Gernot Doppelhofer, and Ronald I Miller. 2004. "Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach." *American Economic Review*, 813–835.

Binary outlier classification Link

Following Hoeting, Raftery and Madigan (1996):

$$p\left(\sigma_i \epsilon_i | \varpi_i, \rho, \sigma\right) = (1 - \varpi_i) N\left(0, \sigma^2\right) + \varpi_i N\left(0, \rho\sigma^2\right) \quad (9)$$

$$\begin{aligned} \varpi_i &\sim Bernoulli(.1) \\ \rho - 1 &\sim Exp(.1) \end{aligned} \quad (10)$$

$$\sigma_i^2 | \varpi_i, \sigma, \rho = (1 - \varpi_i)\sigma^2 + \varpi_i \rho \sigma^2 \quad (11)$$

Geweke outlier classification Link

Follow Geweke (1993) as opposed to Dirichlet-weighting of country variances in the growth model:

$$\begin{aligned} \nu &\sim exp\left(25^{-1}\right) \\ \nu/\sigma_i &\sim \boldsymbol{\chi}\left(\nu\right) \end{aligned} \tag{12}$$

Outlier Robustness checks Link

Outlier model	Dirichlet	Binary	Geweke
ME-variance by country	Random, agnostic	Random, agnostic	Random, agnostic
ME-variance by vintages	Random, agnostic	Random, agnostic	Random, agnostic
Initial Income	99.8	99.9	99.8
East Asian Dummy	90.5	97.9	94.3
Primary Schooling Enrollment	90.3	82.8	88.7
Fraction of Tropical Area	51.1	29.8	41.6
Air Distance to Big Cities	43.6	36.0	43.0
Investment Price	34.4	17.6	23.0
Life Expectancy	34.2	43.6	36.1
Population Density	34.0	17.0	22.3
Sub-Saharan Africa Dummy	31.8	41.4	39.3
Population Coastal Density	26.6	18.5	22.7
Primary Exports	20.8	27.1	24.1
Latin American Dummy	20.1	29.4	24.9
Fraction Muslim	18.1	21.9	21.6
Malaria Prevalence	16.6	37.0	27.6
Fraction Confucian	14.8	12.8	15.9
Openness 1965-74	13.0	12.3	17.0
Landlocked Country Dummy	12.4	38.3	28.8
Outward Orientation	11.1	10.7	13.9
European Dummy	9.6	20.3	16.7
Fraction Speaking Foreign Language	6.9	12.7	10.6

Random model size and g-factor Link

Model size:

$$\theta \sim Beta\left(1, \frac{60}{7}\right) \qquad \gamma_k \sim Bernoulli\left(\theta\right) \tag{13}$$

See Ley and Steel (2009) for a discussion. Benchmark MEMA model, we use $g_0 = N^{-1}$. Following Liang et al. (1998):

$$\frac{1}{1+g_0} \sim Beta\left(1, N^{-1}\right) \tag{14}$$

Model size and g-factor Link

ME-variance by country	Random, agnostic	Random, agnostic
ME-variance by vintages	Random, agnostic	Random, agnostic
Outlier model	Dirichlet	Dirichlet
PWT-vintages	6.0 - 8.1	6.0 - 8.1
Prior model size	Fixed	Random
Prior Zellner g-shrinkage	Fixed	Random
Initial Income	99.8	99.5
East Asian Dummy	90.5	90.5
Primary Schooling Enrollment	90.3	95.0
Fraction of Tropical Area	51.1	58.6
Air Distance to Big Cities	43.6	53.7
Investment Price	34.4	42.3
Life Expectancy	34.2	34.9
Population Density	34.0	55.7
Sub-Saharan Africa Dummy	31.8	29.9
Population Coastal Density	26.6	37.1
Primary Exports	20.8	25.1
Latin American Dummy	20.1	17.1
Fraction Muslim	18.1	23.9
Malaria Prevalence	16.6	13.0
Fraction Confucian	14.8	19.2
Openness 1965-74	13.0	19.0
Landlocked Country Dummy	12.4	17.8
Outward Orientation	11.1	20.6
Real Exchange Rate Distortions	10.0	17.1
European Dummy	9.6	10.6
Fraction Population In Tropics	7.8	11.9
Fraction Speaking Foreign Language	6.9	11.0
Fertility	6.2	10.9
Hydrocarbon Deposits	3.8	12.9

Data and sensitivity Link

In benchmark MEMA, we use PWT6.0-PWT8.1. We also estimate model with PWT 9.0, in addition to other vintages. However, do this in two ways:

1. Exact same model as benchmark MEMA
2. Impose "newer is better"-prior for measurement error

$$\begin{aligned} \left(\tilde{\omega}_1^V, ...\tilde{\omega}_V^V\right)/V &\sim Dir\left(\Omega_1^V, ...\Omega_V^V\right) \\ \omega_1^V, ..., \omega_V^V &= \text{SORT}\left(\tilde{\omega}_1^V, ...\tilde{\omega}_V^V\right) \end{aligned} \tag{15}$$

Relative measurement error in PWT-vintages Link

Posterior mean of relative variance per vintage:

PWT vintage	Unrestricted	Newer is better
PWT 6.0	2.02	2.12
PWT 6.1	0.93	1.12
PWT 6.2	0.95	1.01
PWT 6.3	0.82	0.93
PWT 7.0	0.94	0.88
PWT 7.1	0.92	0.83
PWT 8.0	0.61	0.74
PWT 8.1	0.54	0.70
PWT 9.0	1.28	0.67

Robustness: Data and priors Link

ME-variance by country	Random, unrestricted	Random, unrestricted	Random, unrestricted
PWT-vintages	6.0 - 8.1	6.0 - 9.0	6.0 - 9.0
Outlier model	Dirichlet	Dirichlet	Dirichlet
Vintage ME-variance	Unrestricted	Newer is better	Unrestricted
Initial Income	99.8	99.6	99.8
East Asian Dummy	90.5	90.1	90.5
Primary Schooling Enrollment	90.3	80.8	87.6
Fraction of Tropical Area	51.1	44.6	49.4
Air Distance to Big Cities	43.6	38.5	42.7
Investment Price	34.4	34.4	35.0
Life Expectancy	34.2	38.9	34.8
Population Density	34.0	28.4	31.4
Sub-Saharan Africa Dummy	31.8	29.4	30.5
Population Coastal Density	26.6	21.5	24.5
Primary Exports	20.8	13.7	17.0
Latin American Dummy	20.1	20.7	20.1
Fraction Muslim	18.1	16.3	18.5
Malaria Prevalence	16.6	27.0	19.1
Fraction Confucian	14.8	14.2	14.4
Openness 1965-74	13.0	11.7	13.1
Landlocked Country Dummy	12.4	27.5	17.0
Outward Orientation	11.1	7.8	9.1
European Dummy	9.6	11.0	10.2