

## CROSS-COUNTRY REGRESSIONS:

### WHICH VARIABLES REALLY MATTER?

"Significant" coeff on RHS variable could mean:

- 1) Variable is indeed correlated with Y/L or growth but direction of causality?
- 2) Omitted variable bias - variable is correlated with omitted variables that are really related to Y/L

So why not just add more variables to RHS?

Number of countries in PWT, 1960:  $\approx 150$

Number of potential RHS variables:  $> 60$

Can't put them all on RHS at once.

"significant"  
in published  
papers

Try this:

small number of RHS variables, but  
all (or some) possible combinations of  
possible RHS variables.

Look for variables that are always (or usually)  
significant.

## WHICH VARIABLES REALLY MATTER? (cont.)

(2)

Papers that I. this:

Levine & Renelt, AER 1992

Sala-i-Martin, AER 1997

"I Just Ran Two Million Regressions"  
AER 2004

Remember: even if you find some variables  
with "robust" relationship to  $Y/L$ ,  
you still don't know direction of causality.

# WHICH VARIABLES REALLY MATTER?

(3)

Salai-i-Martin, "I Just Ran..."

Run variables in small sets and assume some variables matter: initial income 1960

primary-school enrollment 1960 } human  
life expectancy 1960 } capital

L145  
% Growth 1960  
- end year

R145  
those three variables plus 3 (out of 59)

TABLE 1—MAIN RESULTS OF REGRESSIONS  
(DEPENDENT VARIABLE = GROWTH)

Independent variable	(i) $\beta$	(ii) SD	(iii) CDF <sup>a</sup>
Equipment investment	0.2175	0.0408	1.000
Number of years open economy	0.0195	0.0042	1.000
Fraction Confucian	0.0676	0.0149	1.000
Rule of law	0.0190	0.0049	1.000
Fraction Muslim	0.0142	0.0035	1.000
Political rights Latin America dummy	-0.0026	0.0009	0.998
Sub-Saharan Africa dummy	-0.0115	0.0029	0.998
Civil liberties	-0.0121	0.0032	0.997
Revolutions and coups	-0.0029	0.0010	0.997
Fraction of GDP in mining	-0.0118	0.0045	0.995
SD black-market premium	0.0353	0.0138	0.994
Primary exports in 1970	-0.0290	0.0118	0.993
Degree of capitalism	-0.0140	0.0053	0.990
War dummy	0.0018	0.0008	0.987
Non-equipment investment	-0.0056	0.0023	0.984
Absolute latitude	0.0562	0.0242	0.982
Exchange-rate distortions	0.0002	0.0001	0.980
Fraction Protestant	-0.0590	0.0302	0.968
Fraction Buddhist	-0.0129	0.0053	0.966
Fraction Catholic	0.0148	0.0076	0.964
Spanish colony	-0.0089	0.0034	0.963
	-0.0065	0.0032	0.938

Looking across all possible combinations, probability coeff is not zero

WHICH VARIABLES...?

(4)

Sala-i-Martin et al., "Determinants of Long-term Growth"

Do not assume any variables matter

run variables in large sets as well as small sets ("model size")

What matters with 7 variables in RHS? ("base line")

TABLE 2—BASELINE ESTIMATION FOR ALL 67 VARIABLES

Rank	Variable	Posterior inclusion probability (1)	Posterior mean conditional on inclusion (2)	Posterior s.d. conditional on inclusion (3)	BACE sign certainty probability (4)	OLS p-value (5)	OLS sign certainty probability (6)	Fraction of regressions with $ tstat  > 2$ (7)
1	East Asian dummy	0.823	0.021805	0.006118	0.999	0.505	0.999	0.99
2	Primary schooling 1960	0.796	0.026852	0.007977	0.999	0.155	0.999	0.96
3	Investment price	0.774	-0.000084	0.000025	0.999	0.032	0.999	0.99
4	GDP 1960 (log)	0.685	-0.008538	0.002888	0.999	0.387	0.999	0.30
5	Fraction of tropical area	0.563	-0.014757	0.004227	0.997	0.466	0.997	0.59
6	Population density coastal 1960's	0.428	0.000009	0.000003	0.996	0.767	0.996	0.85
7	Malaria prevalence in 1960's	0.252	-0.015702	0.006177	0.990	0.515	0.010	0.84
8	Life expectancy in 1960	0.209	0.000808	0.000354	0.986	0.761	0.014	0.79
9	Fraction Confucian	0.206	0.054429	0.022426	0.988	0.377	0.988	0.97
10	African dummy	0.154	-0.014706	0.006866	0.980	0.589	0.980	0.90
11	Latin American dummy	0.149	-0.012758	0.005834	0.969	0.652	0.969	0.30
12	Fraction GDP in mining	0.124	0.038823	0.019255	0.978	0.305	0.978	0.07
13	Spanish colony	0.123	-0.010720	0.005041	0.972	0.507	0.028	0.24
14	Years open	0.119	0.012209	0.006287	0.977	0.826	0.023	0.98
15	Fraction Muslim	0.114	0.012629	0.006257	0.973	0.478	0.973	0.11
16	Fraction Buddhist	0.108	0.021667	0.010722	0.974	0.460	0.974	0.90
17	Ethnolinguistic fractionalization	0.105	-0.011281	0.005835	0.974	0.991	0.974	0.52
18	Government consumption share 1960's	0.104	-0.044171	0.025383	0.975	0.344	0.025	0.77
19	Population density 1960	0.086	0.000013	0.000007	0.965	0.815	0.965	0.01
20	Real exchange rate distortions	0.082	-0.000079	0.000043	0.966	0.835	0.034	0.92
21	Fraction speaking foreign language	0.080	0.007006	0.003960	0.962	0.474	0.962	0.43

high value means that adding this variable to RHS improves fit a lot.

average value of coefficient if you do put it on RHS

probability we know the sign of the coeff; probability its

not zero or negative (if we think it's positive), or not zero or positive (if...)

results of regression with all 67 variables on RHS