



Econ Journal Watch
Scholarly Comments on Academic Economics
Volume 1, Issue 2, August 2004

COMMENTS

Does Economic Performance Correlate with Big Government? <i>Peter Gordon and Lanlan Wang</i>	192-221
Kuznets Curveball: Missing the Regional Strike Zone <i>Jeff Edwards and Anya McGuirk</i>	222-234
Response to Edwards and McGuirk, <i>Jih Y. Chang and Rati Ram</i>	235-243
Reply to Chang and Ram, <i>Jeff Edwards and Anya McGuirk</i>	244-259
Statist Quo Bias, <i>Daniel B. Klein</i>	260-271
Response to Klein, <i>Cass R. Sunstein</i>	272-273
Reply to Sunstein, <i>Daniel B. Klein</i>	274-276
Rejoinder to De Alessi, <i>Michael Kremer</i>	277-278
DO ECONOMISTS REACH A CONCLUSION?	
Licensing Doctors: Do Economists Agree? <i>Shirley Svorny</i>	279-305

INTELLECTUAL TYRANNY OF THE STATUS QUO

The Development Set: The Character of the *Journal of Development Economics*
2002, *Susan Anderson and Peter Boettke* [306-318](#)

INVESTIGATING THE APPARATUS

Institutional Ties of the *Journal of Development Economics* Authors and Editors,
Daniel B. Klein, with Therese DiCola [319-330](#)

ECONOMICS IN PRACTICE

Size Matters: The Standard Error of Regressions in the *American Economic*
Review, *Stephen T. Ziliak and Deirdre N. McCloskey* [331-358](#)

CHARACTER ISSUES

Preference Falsification in the Economics Profession, *William L. Davis*
[359-368](#)

Learning to Lose a Leg: Casualties of PhD Economics Training in
Stockholm, *Anne D. Boschini, Matthew J. Lindquist, Jan Pettersson, and Jesper*
Roine [369-379](#)

CORRESPONDENCE

[380](#)



*Econ Journal Watch,
Volume 1, Number 2,
August 2004, pp 192-221.*

Does Economic Performance Correlate with Big Government?

PETER GORDON AND LANLAN WANG*

A COMMENT ON: R. LA PORTA, F. LOPEZ-DE-SILANES, A. SHLEIFER, R. VISHNY. 1999. THE QUALITY OF GOVERNMENT. *JOURNAL OF LAW, ECONOMICS, & ORGANIZATION* 15(1): 222-279.

[ABSTRACT, KEYWORDS, JEL CODES](#)

FOR OVER A DECADE ECONOMISTS HAVE WORKED TO DEVELOP accurate measurements of economic freedom—a leading effort can be found online at www.freetheworld.com. Researchers are now using this work to address the most important economic question of them all: What accounts for differences in the wealth of nations? In fact most of a recent issue of the *Cato Journal* (2003) is devoted to papers that include new tests of the hypothesis that economic growth and economic freedom are positively associated. They corroborate the expected relationship.

Adam Smith discussed the importance of division of labor, the maturity of markets, and improved transportation as determinants of economic growth. But more importantly he put forward a theory *in terms of policy and institutional variables*. He argued English prosperity was due to “the general liberty of trade,” the “equal and impartial administration of justice” (Smith [1776], 576), “the private frugality and good conduct of individuals . . . protected by law and allowed by liberty” (Smith [1776], 329). As for the size

* Both: School of Policy, Planning and Development, University of Southern California. We wish to thank an anonymous referee and USC Professor Yongheng Deng and USC grad student Xudong An for valuable assistance and insights that helped us to improve the paper in many ways.

of government, Smith favored the government provision of certain basic public goods, but he generally supported provision by local independent authorities that received fees from users. He regarded high taxes and big government to be at odds with “the obvious and simple system of natural liberty” (Smith [1776], 651) and an “impertinent obstruction” (508) to economic development.

A Smithian resurrection is found in the works of such writers as Peter Bauer (1976) and Deepak Lal (1999), in the *Economic Freedom of the World* by James Gwartney and Robert Lawson (2003), and in the work of “New Institutionalists” (e.g., Haber, North, and Weingast 2003). Furthermore, like Adam Smith, some researchers have also recognized that causality runs in both directions. The character of the political system is itself subject to the development of the economy (Rapaczynski 1996).

Inquiry into the true reasons behind the differences in development has prompted statistical studies of the relationship between economic development, political institutions, and the size of public sectors. Some of these (Goldsmith 1995, Dawson 2003, Farr, Lord, and Wolfenbarger 1998, Powell 2003, Vega-Gordillo and Alvarez-Arce 2003) have found a positive relationship between economic growth and economic freedom. A study by Campos and Nugent showed that governance played a prominent role in improving development performance (Campos and Nugent 1999, 449). It is less clear whether political freedom by itself has economic consequences. Kirmanoglu (2003) found, “more political freedom seems to generate higher economic growth” (Kirmanoglu 2003, p. 7) in some countries. But other studies (Vega-Gordillo and Alvarez-Arce 2003, and Kurzman, Werum, and Burkhart 2002) indicated no strong statistical evidence in support of the relationship between economic development and political freedom; neither could a causal relationship between political freedom and economic freedom be found by Farr, Lord, and Wolfenbarger (260). On the other hand, the most recent IMF report (2003) and studies by Acemoglu, Johnson, and Robinson (2002), Bajic and Belgrade (2003), and Rodrik and Subramanian (2003) showed that the quality of institutions has a significantly positive relationship with economic development, vindicating the impression that countries with *respect for the development of political institutions* tend to grow faster and enjoy higher levels of prosperity.

As for the relationship between size of public sector and economic development, Agell, Lindh, and Ohlsson (1997), Ayal and Karras (1998), and Nelson and Singh (1998), did not find statistically significant relationships between the rate of economic growth and the size of the public sector. However, Knack, and Keefer (1995), Barro (1997), Gwartney,

Holcombe, and Lawson (1998), and Folster and Henrekson (2001) have provided evidence that the relationship between growth and public expenditures, “may be more robustly negative than it first appears” (Folster, Henrekson 2001, 1514).

LA PORTA, LOPEZ-DE-SILANES, SHLEIFER, AND VISHNY ON SIZE OF GOVERNMENT

A particularly interesting and wide-ranging discussion of these possible relationships is included in “The Quality of Government” by La Porta, Lopez-de-Silanes, Shleifer, and Vishny, in *The Journal of Law, Economics, & Organization* (1999). The authors note that “government shapes the economic life of a country in a variety ways” (225), and, “economic development itself creates a demand for good government” (230). Thus they acknowledge the simultaneous relationship between economic growth and political institutions. Yet, as the discussants of the paper, Zucker and Darby (1999) note, La Porta *et al.* do not test this suggested simultaneous relationship—even though the La Porta *et al.* argument points this way.¹ Instead, they define and try to explain the following dimensions of government performance.

- *Interference with the Private Sector*, which is concerned with the quality of regulation, security of property rights, and marginal tax rate.
- *Efficiency of Government*, which is indicated by corruption, bureaucratic delays, tax compliance, and average relative wages of government officials.
- *Output of Public Goods*, measured by infant mortality, school achievement, illiteracy rate, and infrastructure quality.

¹ It should be noted that except for the studies by Dawson (2003), Farr, Lord, and Wolfenbarger (1998), Vega-Gordillo and Alvarez-Arce (2003), Kirmanoglu (2003) who used Granger causality tests to address question of causality between the factors, the other studies mentioned so far have merely suggested “a partial correlation” (Dawson 2003, 480) between various measures of economic growth, government performance, and related institutions.

- *Size of Government*, which they infer from government consumption, transfers and subsidies, SOEs, and employment in public sectors.

La Porta *et al.* seek to explain these dimensions of government performance via three exogenous variables expected to have significant influences. The suggested explanatory variables are:

- *Legal Origin* (English Common Law, French Commercial Code, German Commercial Code, Scandinavian Commercial Code, Socialist/Communist laws).
- *Religion* (the percentage of the population of each country that belong to the three most widespread religions, namely Roman Catholic, Protestant, or Muslim).
- *Ethnolinguistic Fractionalization* (a complex index of ethnic heterogeneity).

The exogenous variables were found to have systematic influences on the quality of government. Ethnolinguistically homogeneous countries, common law countries, and countries with lower proportions of Catholic or Muslim religions have better governments (265).

La Porta *et al.*'s numerous linear regressions also reveal that, "rich nations have better governments than poor ones." They add that, "we will use the term 'good' in this article to stand for good-for-economic-development" (223). This makes perfect sense because development is the ultimate test of policy effectiveness.

Given the soundness of their finding, it would be correct to caution, as La Porta *et al.* do, that, "larger governments tend to be the better performing ones" (222), and "identifying big government with bad government can be highly misleading" (266).

In light of the tantalizing questions about causality relationships and the influence of the size of government on economic development raised by La Porta *et al.*, we tested the causal relationships between *three* dimensions: *Economic Development*, *Political Institutions*, and *Scope of Public Sector*, by estimating two simultaneous equation models. The indicators used were created via an aggregation procedure described in the appendix; we derived indicators that roughly correspond to the La Porta *et al.* groupings of dependent variables.

Table 1: Variables and Indices

PLSV: The Quality of Government			Gordon and Wang			
Dimensions of the Quality of Government		Variables		Dimensions of Development	Variables	
Economic Development	Independent Variables	Log GNP per capita	Exogenous Variables		Ethnolinguistic fractionalization	
Determinants		Latitude			Legal origin	
		Ethnolinguistic fractionalization			Religion	
		Legal origin				
		Religion				
Interference with Private Sector	Dependent Variables	Property rights index	Endogenous Variables	Political Institutions	Property rights index	
		Business regulation index			Regulation index	
		Top tax rate			Political rights index	
		Civil liberties index				
Political Rights		Democracy index		Scope of Public Sector	Transfer and subsidies as % of GDP	
		Political rights index			Govt. consumption as % of GDP	
Size of Public Sector		Transfers and subsidies as % of GDP			SOEs in the economy	SOEs in the economy
		Govt. consumption as % of GDP			Public sector employ. as % of population	
Output of Public Goods		Log of infant mortality		Economic Development	Infant mortality	
		Log of school achievement			GDP per capita	
		Illiteracy rate			Urbanization	
		Infrastructure quality			Imports	
					Exports	
Efficiency	Corruption					
	Bureaucratic delays					
	Tax compliance					
	Average government wages/GDP per capita					

The data processed in our analysis were for 47 countries for five periods. We did not know which sample was used by La Porta *et al.* but we utilized their published data where possible. Their data appendix lists many more countries than they report were actually used in their tests. Data details are discussed in our Appendix Tables: A1-A4.

A comparison of La Porta *et al.* groupings and our indicators is shown in Table 1. Based on the revised set of indicators, the following three variables are presumed to be interrelated: 1) *Political Institutions*, approximates economic freedom, 2) *Scope of Public Sector*, is self-explanatory, and 3) *Economic Development*, also self-explanatory.

The results of our statistical tests challenge La Porta *et al.*'s finding that governments with better performing economies are larger and collect higher taxes.

STATISTICAL FINDINGS

We show versions of the simultaneous model that La Porta *et al.* suggested, but chose not to estimate. Model 1 suggests a simultaneous relationship between *Economic Development*, *Political Institutions*, and *Scope of Public Sector*. The three equations are identified by adding the value of the five-year lagged dependent variable in each equation. This approach rests on the assumption that lagged variables are "predetermined" (Theil 1971, 436)—and that, in this case, the five-year changes in these variables were sufficient for them to qualify as exogenous. Rapid change in the latter part of the twentieth century gives some credence to this view (Campos and Nugent 1999, Shirley 2003²). Model 2 is similar, but adds La Porta *et al.*'s exogenous variables to each equation. Findings from the estimations of both models do not support the conclusions about the size of government reported by La Porta *et al.*—admittedly, they used slightly different indicators to describe the size of government (the differences are noted in Table 1).

Both of our models show that economic freedom, as indicated by *Political Institutions*, is significantly positive in explaining *Economic Development*.

² Mary M. Shirley cites this as the view of the foreign aid community.

Table 2: Results of Statistical Estimations

	Dependent Variable					
	Model 1			Model 2		
	Economic Development(<i>n</i>)	Political Institutions(<i>n</i>)	Scope of Public Sector(<i>n</i>)	Economic Development(<i>n</i>)	Political Institutions(<i>n</i>)	Scope of Public Sector(<i>n</i>)
Intercept	-2.218 (-0.61)	1.51 (0.45)	6.57 (3.11)	159.9 (0.4)	347.4 (0.99)	-43.36 (-0.16)
Economic Development(<i>n</i>)		0.112 (4.43)	-0.02 (-1.08)		0.095 (3.59)	-0.023 (-1.06)
Political Institutions(<i>n</i>)	0.14 (2.90)		0.019 (0.54)	0.281 (3.81)		0.005 (0.10)
Scope of Public Sector(<i>n</i>)	0.058 (0.65)	0.06 (0.78)		0.017 (0.17)	0.055 (0.62)	
Economic Development (<i>n-1</i>)	0.893 (31.93)			0.876 (30.02)		
Political Institutions(<i>n-1</i>)		0.773 (20.76)			0.67 (14.14)	
Scope of Public Sector(<i>n-1</i>)			0.811 (16.83)			0.77 (15.09)
Ethno-linguistic fraction				3.24 (1.00)	-5.12 (1.88)	1.07 (0.5)
Legal origin				-1.712 (-1.80)	0.72 (0.86)	-0.263 (-0.41)
Protestant				-1.64 (-0.42)	-3.36 (-0.96)	0.527 (0.2)
Catholic				-1.62 (-0.41)	-3.38 (-0.96)	0.532 (0.2)
Muslim				-1.56 (-0.40)	-3.44 (-0.98)	0.509 (0.19)
Other Religions				-1.61 (-0.41)	-3.40 (-0.97)	0.51 (0.19)
Adj R-sq	0.94	0.86	0.66	0.94	0.86	0.65
Pr > F	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001

* Numbers in parentheses are t-value for the coefficients. Bolded numbers indicate that results are statistically significant, meaning that they could not arise from chance with a 95 percent level of confidence.

Economic Development, in turn, fosters economic freedom. *Size of Public Sector* does not significantly affect either one.

These interpretations are based on the standard 95 percent level of statistical significance cut-off, meaning that the results are unlikely to come from chance associations.

The La Porta *et al.* exogenous variables tested in Model 2 do *not* show a statistically significant influence on any of the endogenous variables. However, this does not necessarily mean that legal origin and religions are not important for economic development, because the substance of these two exogenous variables could be embedded in the current institutional arrangements of the society³. This incomplete exogeneity is also noted by La Porta *et al.* (233). Simple OLS tests (see discussion of Table 3, below) consider this relationship.

Results in Table 3 show four La Porta *et al.*-type OLS tests for the endogenous variables developed in this study. Compared with the La Porta *et al.* results, the influence from the exogenous variables on the *Political Institutions*, *Scope of Public Sector*, and *Economic Development* are similar. Countries with more democratic legal origins show freer economic institutions, and also tend to be related with superior economic development, and smaller size of public sectors. On the other hand, Ethnolinguistic Fractionalization, which is positively related with the *Scope of Public Sector*, “has a very consistent adverse effect” (La Porta *et al.*, 245) on *Political Institutions* and *Economic Development*. For the religions, Catholic and Muslim tend to negatively affect the *Economic Development* and *Political Institutions*.

³ Our 47 observations (see Table A2 in the appendix) include countries with English Law, French Law, German Law, Scandia law, and countries with large percentages of Muslims. The geographic locations of these countries are in Africa, Europe, North America, South America, south Asia, southeast Asia, and south Pacific. institutions, public sectors, and economic development across the countries could capture some impact from different religions and legal origins.

Table 3: OLS Estimation of Determinants of Dimensions of Development

	Dependent Variable					
	Model 1			Model 2		
	Economic Develop.	Political Institutions	Scope of Public Sector	Economic Develop.	Political Institutions	Scope of Public Sector
Intercept	507.5 (0.45)	279.59 (0.52)	82.34 (0.22)	78.66 (7.73)	41.08 (8.48)	26.62 (7.98)
Ethno-linguistic fraction	-58.75 (8.58)	-35.52 (-10.90)	8.79 (3.92)	-59.85 (-9.68)	-36.13 (-12.29)	8.65 (4.27)
Legal origin	3.51 (1.39)	6.39 (5.31)	-0.98 (-1.19)	3.817 (1.60)	6.56 (5.78)	-0.95 (-1.21)
Muslim	-4.43 (-0.39)	-2.52 (-0.46)	-0.52 (-0.14)	-0.15 (-1.97)	-0.136 (-3.73)	0.03 (1.25)
Catholic	-4.26 (-0.37)	-2.29 (-0.42)	-0.49 (-0.13)	0.017 (0.23)	0.09 (2.69)	0.063 (2.62)
Protestant	-4.06 (-0.36)	-2.21 (-0.41)	-0.56 (-0.15)	0.2 2(2.97)	0.168 (4.77)	-0.002 (-0.08)
Other	-4.28 (-0.38)	-2.38 (-0.44)	-0.56 (-0.15)			
Adj R-sq	0.41	0.62	0.14	0.41	0.62	0.14
Pr > F	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
	Model 3			Model 4		
	Economic Development	Political Institutions	Scope of Public Sector	Economic Development	Political Institutions	Scope of Public Sector
Intercept	94.66 (10.78)	53.29 (12.41)	26.46 (9.38)	80.09 (15.7)	52.15 (21.06)	34.33 (20.6)
Ethno-linguistic fraction	-64.12 (-10.49)	-39.39 (-13.17)	8.69 (4.42)	-65.44 (-10.69)	-39.49 (-13.3)	9.4 (4.7)
Legal origin	2.73 (1.14)	5.73 (4.88)	-0.93 (-1.21)	5.9 (3.22)	5.98 (6.73)	-2.65 (-4.42)
Muslim	-0.289 (-4.65)	-0.24 (-7.94)	0.03 (1.64)	-0.208 (-4.35)	-0.235 (-10.51)	-0.011 (-0.72)
Catholic	-0.119 (-2.03)	-0.009 (-0.33)	0.06 (3.41)			
Protestant						
Other						
Adj R-sq	0.39	0.58	0.14	0.38	0.58	0.10
Pr > F	<.0001	<.0001	<.0001	<.0001	<.0001	0.003

For Table 3:

Numbers in parentheses are the t-values for estimated coefficients.

Four models are formulated as:

$$\text{Model 1: } Y = \alpha + b_1 \text{Eth} + b_2 \text{Legal} + b_3 \text{Muslm} + b_3 \text{Cath} + b_4 \text{Pr ote} + b_5 \text{Oth}$$

$$\text{Model 2: } Y = \alpha + b_1 \text{Eth} + b_2 \text{Legal} + b_3 \text{Muslm} + b_3 \text{Cath} + b_4 \text{Pr ote}$$

$$\text{Model 3: } Y = \alpha + b_1 \text{Eth} + b_2 \text{Legal} + b_3 \text{Muslm} + b_3 \text{Cath}$$

$$\text{Model 4: } Y = \alpha + b_1 \text{Eth} + b_2 \text{Legal} + b_3 \text{Muslm}$$

DISCUSSION AND FURTHER RESEARCH

Development is a systemic evolution that is not limited to the improvement of economic performance *per se*. It also requires and prompts improvement of political institutions and performance of public sectors. Inspired by the La Porta *et al.* discussion, results shown here—which, contrary to La Porta *et al.*, find no effect on economic development from the size of government—represent an effort to identify simultaneous relationships between three variables, *Political Institutions*, *Scope of Public Sector*, and *Economic Development*. Not only is the beneficial effect of a large public sector cast into doubt, but the interdependencies between three fundamental endogenous variables are tested and the links that are found to matter are plausible. In simultaneous equation models, *Political Institutions* that indicate economic freedom positively affect *Economic Development*, which, in turn, promotes economic freedom. Also, the three indices tested relate to the La Porta *et al.* exogenous variables in ways that corroborate their OLS results, suggesting that our indices do reflect the labels we have given them.

Much remains to be done, including:

1. Improved measurement of the three endogenous variables: *Economic Development*, *Political Institutions* and *Scope of Public Sector*. La Porta *et al.* classified the observable variables into four categories: *Interference with Private Sector*, *Efficiency of Government*, *Output of Public Goods*, and the *Size of Government*. However, quantitative measurement for these four is missing in their analysis. In some other studies, for example, the recent Brookings Child Well-Being Index, and some

index values in the Index of Economic Freedom, are derived by using unweighted average value from the related indicators.⁴ The index value for the three new variables developed in this study took the correlation between the observable indicators into consideration and tried to eliminate double-counting. The shortcoming of this measurement, however, is that it could reduce the importance of the proportion of variance for each of the observable variable shared by all remaining variables. There might be some better way to not only solve the double-counting problem, but also to emphasize the derived index as well.⁵

2. Due to the few available time periods (five here), a dynamic simultaneous equations model—perhaps a better way to capture lag influences on different variables—could not be estimated. In future work with better data, Granger-causality tests between the variables, especially between *Political Institutions* and *Economic Development* might illuminate these matters.

3. Future work should consider past differences across countries—a point noted by Hayek (1978), who wrote, “a high rate of growth can sometimes testify to bad policies of the past rather than good policies of the present” (189), and by Olson (1996), who noted, “poorer countries that adopt relatively good economic policies and institutions enjoy rapid catch-up growth” (20). The coefficients for a hypothesized causal relationship between *Political Institutions*, *Economic Development*, and *Scope of Public Sector* might differ between poor countries and richer ones.

In any event, we regard La Porta *et al.* as a stimulating and valuable contribution, yet at this point our priors about big government remain.

$$Y = \sum_{i=1}^n x_i / n$$

⁴ The formula is:

⁵ Factor analysis was also attempted, but the result did not provide plausible classifications. The possible explanation for this is probably due to the correlations between the indicators for *Political Institutions*, *Scope of Public Sector*, and *Economic Development*.

APPENDIX

Data

The variables used are indicators of *Political Institutions*, *Economic Development* and *Scope of Public Sector*, available from different sources for different years. Data for 1980, 1985, 1990, 1995, and 2000 were collected. Combining a large number of data sets that have different numbers of observations and restricted by data comparability across different years left us 47 observations. Table A1 shows the availability of observations. Table A3 lists the mean value and standard deviation of each indicator for 47 countries. The definitions and sources for all the variables used in this research are summarized as follows:

Endogenous indicators

Property rights index: A rating of property rights in each country (on a scale from 1 to 10). “The more protection private property receives, the higher the score” (La Porta *et al.* 1999, 235). This index is based on five sub-factors: (1) “judicial independence: the judiciary is independent and not subject to interference by the government or parties in disputes;” (2) “impartial court: a trusted legal framework exists for private businesses to challenge the legality of government actions or regulation;” (3) “protection of intellectual property;” (4) “military interference in rule of law and the political process;” (5) “integrity of the legal system.” This index value comes from the “Index of Economic Freedom.”⁶

Business regulation index: This rating index is based on five sub-factors as well. 1) the presence of price controls; 2) whether administrative procedures are an important obstacle to starting a new business; 3) time taken up by government bureaucracy; 4) whether it is generally easy to start a new business; 5) whether the “irregular, additional payments connected with import and export permits, business licenses, exchange controls, tax assessments, police protection, or loan applications”⁶ are not common. It also ranges from 1 to 10. Higher scores mean that regulations are

⁶ Gwartney, J. G., and Lawson, R. A. (2003), *Economic Freedom of the World 2003 Annual Report*, available from <http://www.freetheworld.com/2003/1EFW2003ch1.pdf>, p24-32. Data are available from <http://www.freetheworld.com/2003/EFW2003Dataset.xls>

straightforward and applied uniformly to all businesses and the regulations are less of a burden to business. Index value is from the “Index of Economic Freedom.”

Political rights index: This index comes from "Freedom in the World,"⁷ published annually by Freedom House since 1978. Political rights enable people to participate freely in the political process. It includes the right to vote and compete for public office and to elect representatives who have a decisive vote on public policies. The rating ranges from 1 to 7. The higher the rating, the less freedom.

Civil liberties: This index also comes from “Freedom in the World.”⁸ It indicates the freedom to develop opinions, institutions, and personal autonomy without interference from the state. The rating ranges from 1 to 7. The higher the rating, the less civil rights the public can expect.

Transfers and subsidies: Transfers and subsidies as a share of GDP. The rating ranges from 1 to 10. Lower ratings refer to a larger proportion of transfers and subsidies from the GDP. The data source is the “Index of Economic Freedom.”

Government consumption: General government consumption as a share of total consumption. The rating ranges from 1 to 10 and is negatively related to government consumption. The data source is the “Index of Economic Freedom.”

SOEs in the economy: These data are also from the “Index of Economic Freedom”. Information on the number, composition, and share of output supplied by State-Operated Enterprises (SOEs), and government investment as a share of total investment were used to construct 0-to-10 ratings. Countries with more government enterprises and government investment received lower ratings.

Infant mortality: This value comes from the *World Development Indicators 2003* (World Bank 2003). It refers to the infant mortality rate per 1,000 live births.

⁷ Gastil, R. D., *Freedom in the World: Political Rights and Civil Liberties*, New York: Freedom House, 1978, 1981, 1983-1984-1985/1986, 1987/1988, 1990/1991, 1992/1993-1993/1994, 1996/1997-1999/2000, 2001/2002-2003.

GDP per capita: This is an adjusted GDP value based on constant 1995 US dollars. Data are from *World Development Indicators 2003*.

Urbanization: This is a measure for the percent of the urban population as a fraction of the country's total population. Data are from *World Development Indicators 2003*.

Imports/Exports of goods and services: The percentage of imports/exports of goods and services in GDP. Data are from *World Development Indicators 2003*.

*Exogenous variables*⁸

Ethnolinguistic fractionalization:⁹ This is the average value of five different indices, ranging from 0 to 1. “The five component indices are: 1) index of ethnolinguistic fractionalization in 1960, which measures the probability that two randomly selected people from a given country will not belong to the same ethnolinguistic group (the index is based on the number and size of population groups as distinguished by their ethnic and linguistic status); 2) probability of two randomly selected individuals speaking different languages; 3) probability of two randomly selected individuals do not speak the same language; 4) percent of the population not speaking the official language; and 5) percent of the population not speaking the most widely used language” (La Porta *et al.* 1999, 238).

Religion: “(T)he percentage of the population of each country that belonged to three most widely spread religious formation:” (1) Roman Catholic; (2) Protestant; (3) Muslim. The data are available for 1990-1995 and are also from La Porta *et al.* (1999).

Legal Origin: Five possible legal originations are identified: (1) English Common law; (2) French Commercial Code; (3) German Commercial

⁸ Data source for exogenous variables comes from: La Porta *et al.* (1999).

⁹ Original data source: Easterly and Levine (1997). The sources of the components of the average index are: (1) Atlas Narodov Mira (1964); (2) Muller (1964); (3) Roberts (1962); (4) and (5) Gunnemark (1991), see page 238 in La Porta *et al.* (1999).

Code; (4) Scandinavian Commercial Code; and (5) Socialist/Communist laws. The source is La Porta *et al.* (1999).

Data Analysis

Of the variables used in this study, the property rights index, the business regulation index, the political rights index and the civil liberties index are mainly concerned with the political policies which define the “rules of game,” or in other words, define the *Political Institutions* relevant for socioeconomic activities. Transfers and subsidies, government consumption, SOEs could be the proxies for the *Scope of Public Sector*, similar to the definition of “*Size of Government*” in La Porta *et al.* The performance of *Economic Development* could be inferred from GDP per capita, urbanization, the infant mortality rate, exports and imports. In order to specify these three variables: *Political Institutions*, *Scope of Public Sector*, and *Economic Development*, a summary index¹⁰ is constructed as follows:

$$A_{cn} = \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m (1 - |\gamma_{n,ij}|/2) \times (X_{cn,i} + X_{cn,j}) \quad i \neq j$$

where,

A_{cn} ---represents the calculated index value for latent factors in country c and in specific time n . Latent variables refer to

¹⁰ We have defined a weighted index. Gwartney and Lawson (2003) acknowledge the problems of defining a proper index." Over the years, we have struggled with how to assign weights to various components and areas to construct a summary index. After experimenting with three different weighting schemes in the first edition, we finally settled on using principle component analysis to assign weights statistically. Although principle component analysis has the advantage of being value-neutral, it does create problems of its own. In particular, when two components are closely correlated, as often happens, then principle component analysis tends to assign low weights to one or both of these components. In essence, the principle component analysis wants to drop out the interrelated components. But, we often want to include these componentseven if they are correlated with other components in order to offset measurement error in the data and to increase the number of countries we can rate. We, therefore, decided to return to using a simple average to combine the components into area ratings and the area ratings into summary ratings in the 2002 report. Although this edition of the index uses averages, we do not mean to imply that all components and areas of economic freedom are equally important in whatever sense. Readers who want to reweigh the components and areas to suit themselves are invited to do so" (Gwartney and Lawson 2003, 22).

Political Institutions, Scope of Public Sector, and Economic Development.

i, j ---refers the number of observable variables.

Economic Development has 5 observable variables, including: GDP per capita; infant mortality¹¹; urbanization; exports and imports;

Political Institutions includes 4 observable variables:

business regulation index; property rights index; political rights index and civil liberties;

Scope of Public Sector includes 3 observable variables:

transfer and subsidies; SOEs; and government consumption shares.

$\gamma_{n,ij}$ ---represents the correlation value in time n between each two observable variables X .

$X_{cn,i}$, $X_{cn,i}$ ---represents the normalized value¹² in country c and in specific time n for observable variables.

The correlation value for the original data and for three newly generated index values is displayed in the Table A4 and Table A5. From Table A4, most of the correlation values are statistically significant, especially, the correlation between political rights and civil liberties (0.897), between GDP per capita and infant mortality rate (0.831), between GDP per capita and property rights (0.802) is noteworthy. As for the exogenous variables, Muslim religion and ethnolinguistic fraction displayed significant and negative correlation with the regulation index, property rights index, political rights, and civil liberties. This may indicate that predominance of Muslim religion or higher ethnolinguistic fraction could come with less economic freedom. Legal origins¹³ has positive correlation with the regulation and property rights indices, which suggests that common law could be associated with the better political institutions. From Table A5, there are three interesting findings:

¹¹ The value for infant mortality rate, political rights index, and civil liberties index is inversed before aggregation in order to make higher value indicate better health situation, stronger civil liberties and political freedom.

¹² The value for indicators is normalized for 1-10 scales (GDP per capita value is normalized based on its natural log value).

¹³ In this paper, English common law is assigned value 4, German Commercial Code and Scandinavian Commercial Code are assigned value 3, French Commercial Code is assigned 2, and Socialist/Communist laws are assigned value 1.

1. *Economic Development*, *Political Institutions*, and *Scope of Public Sector* have significant path dependencies. For example, the correlation between the *Scope of Public Sector* in year 2000 and in 1995 is 0.923. We can also see that the longer the time gaps, the smaller the correlation, which might be an indication that there could exist other factors besides performance itself coming into play during the passage of time. We can also observe that the correlation between *Political Institutions* and *Economic Development* across time persists more strongly by itself—the decrease of correlation with elapsed time is smaller than that of *Scope of Public Sector*.
2. Another interesting observation is that the correlation between *Economic Development* and *Political Institutions* is stronger than the correlation between *Scope of Public Sector* and *Political Institutions* or *Economic Development*.
3. For the exogenous variables, Muslim and Ethnolinguistic fraction have strong negative correlations with *Political Institutions*, while the Protestant religion has a positive correlation with *Political Institutions*. Legal origin seems to have weaker correlations with *Political Institutions*.

Based on these observations, two simultaneous equation models are suggested to estimate the significance of each factor's influence.

$$\begin{cases} P_n = \alpha_p + \gamma_p P_{n-1} + \beta_{p1} I_n + \beta_{p2} E_n \\ I_n = \alpha_i + \gamma_i I_{n-1} + \beta_{i1} P_n + \beta_{i2} E_n \\ E_n = \alpha_e + \gamma_e E_{n-1} + \beta_{e1} P_n + \beta_{e2} I_n \end{cases} \quad (2)$$

$$\begin{cases} P_n = \alpha_p + \gamma_p P_{n-1} + \beta_{p1} I_n + \beta_{p2} E_n + \lambda_{11} Eth + \lambda_{12} Legal + \lambda_{13} Pr\ ote + \lambda_{14} Cath + \lambda_{15} Muslm + \lambda_{16} Oth \\ I_n = \alpha_i + \gamma_i I_{n-1} + \beta_{i1} P_n + \beta_{i2} E_n + \lambda_{21} Eth + \lambda_{22} Legal + \lambda_{23} Pr\ ote + \lambda_{24} Cath + \lambda_{25} Muslm + \lambda_{26} Oth \\ E_n = \alpha_e + \gamma_e E_{n-1} + \beta_{e1} P_n + \beta_{e2} I_n + \lambda_{31} Eth + \lambda_{32} Legal + \lambda_{33} Pr\ ote + \lambda_{34} Cath + \lambda_{35} Muslm + \lambda_{36} Oth \end{cases} \quad (3)$$

where,

- n : refers to time, measured in five-year period,
- P : refers to *Scope of Public Sector*,
- I : refers to *Political Institutions*,
- E : refers to *Economic Development*,
- Eth : refers to *Ethnolinguistic Fraction* value,
- $Legal$: *Legal Origin*,
- $Pr\ ote$, $Cath$, $Muslm$, and Oth : refers proportion value for Protestant, Roman Catholic, Muslim and Other religions.

α -- are regression intercepts

β, λ --are regression coefficients

The reason for using a simultaneous equation model in this research is to try to capture the causal forces between the variables based on available data sources. More specifically, what is interesting about the simultaneous equation model is that the variable of *Political Institutions* is one of the suspected causes of another variable like *Economic Development*; at the same time, *Economic Development* is one of the suspected causes of *Political Institutions*. Based on theoretical arguments from institutional economics, the simultaneous equation model might be able to provide a more comprehensive understanding of the relationship between the factors. In this investigation, models (2) and (3) assume that all the countries have the same path dependency pattern in terms of the relationship between *Economic Development*, *Political Institutions*, and *Scope of Public Sector*. Regression results are displayed in Table 2.

From Table 2, the results shown for Model 1 and Model 2 corroborate the idea that *Economic Development*, *Political Institutions*, and *Scope of Public Sector* have significant path dependencies. For example, in Model 1, improvement of one unit on *Political Institutions* in year $n-1$ would cause a 0.77 unit of improvement on *Political Institutions* in year n .

For cross sectional influence, we see that *Political Institutions* has an important and positive effect on *Economic Development*. The influence of *Political Institutions* on *Economic Development* in Model 1 and Model 2 is 0.14 and 0.28 respectively. On the other hand, *Economic Development* in both models shows that one unit improvement on economic performance could well induce 0.1 unit of institutional change. *Scope of Public Sector* doesn't significantly affect either one.

Table A1: Data Availability

Variable	Data Availability, No. of Countries				
	Year 1980	Year 1985	Year 1990	Year 1995	Year 2000
Business Regulation index ¹	103	106	118	123	123
Property Rights index ¹	90	110	111	123	123
Political Rights index ³	109	112	115	123	123
Civil Liberties ³	109	112	115	123	123
Transfers and Subsidies ¹	85	90	96	103	99
SOE ¹	113	114	114	122	123
Government Consumption Share ¹	113	115	120	122	122
GDP per Capita (US Dollars '95) ²	105	108	115	115	113
Infant Mortality ²	116	116	116	116	116
Urbanization ²	115	116	115	115	115
Exports of Goods and Services ²	101	104	115	115	111
Imports of Goods and Services ²	103	104	115	115	111
Ethnolinguistic Fractionalization ⁴	107	107	107	107	107
Religion ⁴	123	123	123	123	123
Legal Origin ⁴	119	119	119	119	119
1. Gwartney and Lawson (2003): http://www.freetheworld.com/2003/EFW2003Dataset.xls					
2. World Bank, World Development Indicators 2003, Washington: World Bank.					
3. Raymond D. Gastil, Freedom in the world : political rights and civil liberties, New York : Freedom House, 1981, 1983-1984-1985/1986, 1990/1991, 1992/1993-1993/1994, 1996/1997-1999/2000, 2001/2002-2003).					
4. La Porta <i>et al.</i> (1999)					

Table A2: 47 Observations Used in This Study

COUNTRIES	Region	Legal Origin	Protestant	Catholic	Muslim	Other
Australia	Australia	English	23.5	29.6	0.2	46.7
Austria	Europe	German	6.5	88.8	0.6	4.1
Belgium	Europe	French	0.4	90	1.1	8.5
Brazil	S. America	French	4	87.8	0.1	8.1
Cameroon	West Africa	French	19.1	35	22	24.9
Canada	N. America	English	29.6	46.6	0.6	23.2
Chile	S. America	French	1.9	82.1	0	16
Colombia	S. America	French	0.9	96.6	0.2	2.3
Costa Rica	S. America	French	5.8	90.5	0	3.7
Denmark	Europe	Scandin	95.2	0.6	0.2	4
Domin. Rep	S. America	French	1.4	96.6	0	2
Ecuador	S. America	French	1.9	96.4	0	1.7
Egypt	Middle East	French	0.2	0.2	81.8	17.8
Finland	Europe	Scandin	93.1	0.1	0	6.8
France	Europe	French	2.4	76.4	3	18.2
Greece	Europe	French	0.1	0.4	1.5	98
Guatemala	S. America	French	4.9	94	0	1.1
Iceland	Europe	Scandin	96.6	0.7	0	2.7
India	South Asia	English	1.1	1.3	11.6	86
Indonesia	SE Asia	French	4.8	2.7	43.4	49.1
Israel	Middle East	English	0.2	1	8	90.8
Italy	Europe	French	0.4	83.2	0.1	16.3
Jordan	Middle East	French	0.3	1.7	93	5
Kenya	East Africa	English	19.3	26.4	6	48.3
Malaysia	South Asia	English	1.4	2.8	49.4	46.4
Mexico	S. America	French	1.2	94.7	0	4.1
Morocco	N. Africa	French	0	0.2	99.4	0.4
Netherlands	Europe	French	42.4	42.6	1	14
New Zealand	Australian	English	37.9	18.7	0	43.4
Norway	Europe	Scandin	97.8	0.3	0.1	1.8
Pakistan	South Asia	English	0.8	0.5	96.8	1.9
Panama	S. America	French	5.2	85	4.5	5.3
Philippines	SE Asia	French	3.8	84.1	4.3	7.8
Portugal	Europe	French	1.1	94.1	0	4.8
South Africa	S. Africa	English	39	10.4	1.3	49.3
Spain	Europe	French	0.1	96.9	0	3
Sri Lanka	South Asia	English	0.4	6.8	7.2	85.6
Sweden	Europe	Scandin	68.4	1.4	0.1	30.1
Switzerland	Europe	German	43.2	52.8	0.3	3.7
Trinidad & Tob	S. America	English	13.2	35.8	6.5	44.5
Tunisia	N. Africa	French	0	0.1	99.4	0.5

Turkey	Europe	French	0	0.1	99.2	0.7
U. Kingdom	Europe	English	16.1	13.1	1.4	69.4
U. States	N. America	English	43.6	30	0.8	25.6
Uruguay	Middle Af.	French	1.9	59.5	0	38.6
Zambia	Middle Af.	English	31.9	26.2	0.3	41.6
Zimbabwe	Middle Af.	English	21.4	14.4	0.9	63.3
Data source: La Porta, et al. (1999) Appendix B.						

**Table A3: Average and Standard Deviation
(47 Observations Used in This Study)**

	2000		1995		1990	
	Ave.	Std	Ave.	Std	Ave.	Std
Regulation Index	6.31	0.93	6.19	1	5.6	0.68
Property Rights	6.67	2.04	6.58	1.8	5.92	2.06
Political Rights	2.45	1.97	2.49	1.84	2.55	1.84
Civil Liberties	2.7	1.55	2.91	1.73	2.6	1.6
Trans & Sub.	7.14	2.2	6.89	2.54	7.02	2.38
SOEs	5.7	2.77	5.23	2.64	4.66	2.54
Govt. Consump.	5.65	2.58	5.46	2.3	5.42	2.28
GDP Per Capita	12925.05	13743.55	11374.6	12109.25	10800.39	11789.91
Mortality	25.05	27.36	27.7	27.48	31.31	28.05
Urbanization	66.01	19.34	64.26	19.91	61.99	20.51
Exports	36.72	19.77	32.22	15.16	29.69	14.39
Imports	37.16	16.93	33.47	15.54	31.29	16.05
	1985		1980			
	Average	Std	Average	Std		
Regulation Index	5.55	0.87	5.49	0.83		
Property Rights	5.75	1.78	5.64	1.77		
Political Rights	2.64	1.75	2.85	1.85		
Civil Liberties	2.89	1.82	2.91	1.77		
Trans & Sub.	6.87	2.32	7.15	2.16		
SOEs	3.89	2.54	3.13	2.68		
Govt. Consump.	6.58	2.12	7.2	2.08		
GDP Per Capita	9770.4	10606.31	9165.67	9779.44		
Mortality	37.44	31.62	43.99	35.55		
Urbanization	59.91	21.29	57.86	22.13		
Exports	28.79	12.96	28.35	13.02		
Imports	30.21	14.4	32.29	15.03		

Data source: see Table A1.

Table A4: Correlations, All Indicators Used in This Study

	REGU	PROP RIGT	Political rights	civil liberty	TRASF SUB	SOE	GOV COS
REGUL	1						
PROP RIGT	.441 **	1					
Political rights	.355 **	.667 **	1				
Civil liberty	.370 **	.695 **	.897 **	1			
TRASF SUB	-.136 *	-.683 **	-.618 **	-.655 **	1		
SOE	.433 **	.315 **	.388 **	.376 **	-.188 **	1	
GOV COS	-.192 **	-.491 **	-.329 **	-.358 **	.604 **	-.073	1
GDPC	.352 **	.802 **	.736 **	.776 **	-.793 **	.381 **	-.509 **
MORTA	.424 **	.649 **	.695 **	.688 **	-.640 **	.397 **	-.456 **
URBAN	.326 **	.621 **	.597 **	.631 **	-.630 **	.360 **	-.499 **
EXP	-.003	.099	.000	.090	-.217 **	-.105	-.181 **
IMP	.039	.020	-.065	-.016	-.126	-.138 *	-.276 **
ETHOF	-.176 **	-.389 **	-.502 **	-.531 **	.494 **	-.218 **	.231 **
LAW	.270 **	.180 **	.076	.091	.001	-.096	-.224 **
Protestnt	.277 **	.516 **	.317 **	.422 **	-.364 **	.192 **	-.385 **
Catholic	.035	-.040	.212 **	.200 **	-.033	.208 **	.260 **
Muslim	-.243 **	-.353 **	-.511 **	-.528 **	.347 **	-.303 **	.112
OTHER	-.052	-.066	-.032	-.104	.020	-.139 *	-.098

Table A4: Correlations (cont.)

	GDPC	MORTA	URBAN	EXP	IMP	ETHOF	LAW
GDPC	1						
MORTA	.831 **	1					
URBAN	.795 **	.701 **	1				
EXP	.210 **	.247 **	.179 **	1			
IMP	.086	.189 **	.114	.837 **	1		
ETHOF	-.584 **	-.604 **	-.618 **	.000	-.049	1	
LAW	.012	.002	-.102	.047	.000	.384 **	1
Protestnt	.498 **	.313 **	.296 **	.141 *	-.019	-.121	.305 **
Catholic	.084	.151 *	.149 *	-.074	-.141 *	-.196 **	-.506 **
Muslim	-.412 **	-.438 **	-.292 **	.004	.205 **	.060	-.152 *
OTHER	-.154 *	-.028	-.177 **	-.046	-.021	.333 **	.573 **
	Protestant	CATHOLIC	MUSLIM	OTHER			
Protestnt	1						
Catholic	-.392 **	1					
Muslim	-.301 **	-.469 **	1				
OTHER	-.130 *	-.458 **	-.192 **	1			

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Table A5: Correlations, Three Index Values Used in This Study

	SCP00	SCP95	SCP90	SCP85	SCP80	INT00	INT95
SCP00	1						
SCP95	.923 **	1					
SCP90	.860 **	.920 **	1				
SCP85	.764 **	.805 **	.885 **	1			
SCP80	.567 **	.644 **	.619 **	.556 **	1		
INST00	-.232	-.280	-.222	-.165	-.445 **	1	
INST95	-.228	-.297 *	-.230	-.152	-.466 **	.957 **	1
INST90	-.145	-.225	-.146	-.094	-.413 **	.903 **	.923 **
INST85	-.252	-.333 *	-.246	-.188	-.443 **	.853 **	.868 **
INST80	-.144	-.251	-.192	-.156	-.384 **	.783 **	.805 **
ECO00	-.319 *	-.324 *	-.212	-.237	-.499 **	.724 **	.728 **
ECO95	-.374 **	-.376 **	-.269	-.290 *	-.538 **	.718 **	.732 **
ECO90	-.412 **	-.415 **	-.315 *	-.326 *	-.571 **	.730 **	.747 **
ECO85	-.466 **	-.478 **	-.389 **	-.383 **	-.627 **	.776 **	.794 **
ECO80	-.460 **	-.479 **	-.408 **	-.417 **	-.620 **	.742 **	.752 **
ETHOF	.218	.240	.136	.164	.240	-.502 **	-.520 **
LAW	-.106	-.184	-.277	-.235	-.088	.138	.167
Protestnt	-.224	-.263	-.254	-.047	-.372 **	.484 **	.515 **
Catholic	.210	.267	.361 *	.334 *	.101	.141	.169
Muslim	.009	-.018	-.039	-.083	.211	-.565 **	-.559 **
OTHER	-.074	-.083	-.200	-.324 *	-.003	-.043	-.120

Table A5: Correlations (cont.)

	INT90	INT85	INT80	ECO00	ECO95	ECO90	ECO85
INST90	1						
INST85	.936 **	1					
INST80	.853 **	.921 **	1				
ECO00	.744 **	.708 **	.639 **	1			
ECO95	.717 **	.686 **	.617 **	.989 **	1		
ECO90	.705 **	.673 **	.594 **	.968 **	.986 **	1	
ECO85	.732 **	.709 **	.631 **	.936 **	.956 **	.978 **	1
ECO80	.662 **	.651 **	.573 **	.896 **	.930 **	.955 **	.976 **
ETHOF	-.609 **	-.550 **	-.437 **	-.561 **	-.563 **	-.582 **	-.545 **
LAW	.023	.161	.234	-.077	-.025	-.055	.002
Protestant	.404 **	.414 **	.424 **	.293 *	.328 *	.329 *	.409 **
Catholic	.228	.132	.125	.110	.074	.085	.048
Muslim	-.507 **	-.480 **	-.522 **	-.254	-.259	-.255	-.334 *
OTHER	-.150	-.057	-.006	-.162	-.142	-.163	-.102
	ECO80	ETHF	LAW	Protestnt	Catholic	Muslim	OTHER
ECO80	1						
ETHOF	-.486 **	1					
LAW	.075	.384 **	1				
Protestnt	.426 **	-.121	.305 *	1			
Catholic	.012	-.196	-.506 **	-.392 **	1		
Muslim	-.351 *	.060	-.152	-.301 *	-.469 **	1	
OTHER	-.049	.333 *	.573 **	-.130	-.458 **	-.192	1

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

REFERENCES

- Acemoglu, D., Johnson, S. and Robinson, J. A.** 2002. Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution. *The Quarterly Journal of Economics* 117(4): 1231-1294.
- Agell, J., Lindh, T., Ohlsson, H.** 1997. Growth and the Public Sector: A Critical Review Essay. *European Journal of Political Economy* 13(1): 33-52.
- Ayal, E. B., and Karras, G.** 1998. Components of Economic Freedom and Growth: An Empirical Study. *Journal of Developing Areas*. 32(3): 327-338.
- Bajic, V., Belgrade, IEN.** 2003. Building Institutions—Essential Determinant of Development. Online:
<http://www.icn.bg.ac.yu/02-06%20full%20text.pdf> (cited: June 2004).
- Bardhan, P.** 1993. Symposium on Democracy and Development. *The Journal of Economic Perspectives* 7(3): 45-49.
- Barro, R. J.** 1997. *Determinants of Economic Growth: A Cross-Country Empirical Study*. Cambridge, MA: The MIT Press.
- _____. 2000. Rule of Law, Democracy, and Economic Performance. In *2000 Index of Economic Freedom*, ed. O'Driscoll, G. P., Holmes, K. R., and Kirkpatrick, M. New York: Dow Jones Co.
- Bauer, P.** 1976. *Dissent on Development*. Cambridge, MA.: Harvard University Press.
- Berggren, N.** *The Benefits of Economic Freedom: A Survey*. Online:
http://www.independent.org/tii/content/pubs/review/tir82_berggren.html (cited: June 2004).
- Campos, N. F., and Nugent, J. B.** 1999. Development Performance and the Institutions of Governance: Evidence from East Asia and Latin America. *World Development* 27(3): 439-452.
- Davis, L., North, D.** 1970. Institutional Change and American Economic Growth: A First Step Towards A Theory of Institutional Innovation. *The Journal of Economic History* 30(1): 131-149.
- Dawson, J. W.** 2003. Causality in the Freedom-Growth Relationship, *European Journal of Political Economy* 19(3): 479-495.

- De Soto.** 2000. *The Mystery of Capital*. New York: Basic Books.
- Farr, W. K., Lord, R. A., and Wolfenbarger, J. L.** 1998. Economic Freedom, Political Freedom, and Economic Well-Being: A Causality Analysis. *Cato Journal* 18(2): 247-262.
- Folster, S., Henrekson, M.** 1999. Growth and the Public Sector: A Critique of the Critics. *European Journal of Political Economy* 15(2): 337-358.
- Folster, S., and Henrekson, M.** 2001. Growth Effects of Government Expenditure and Taxation in Rich Countries. *European Economic Review* 45(8): 1501-1520.
- Gastil, R. D.** 1978-2003. *Freedom in the World: Political Rights and Civil Liberties*. New York: Freedom House.
- Goldsmith, A. A.** 1995. Democracy, Property Rights and Economic Growth. *The Journal of Development Studies* 32(2): 29-44.
- Gwartney, J. G., and Lawson, R. A.** 2003. *Economic Freedom of the World: 2003 Annual Report*. Online: <http://www.freetheworld.com/2003/1EFW2003ch1.pdf> (cited: June 2004).
- Gwartney, J. G., and Lawson, R. A.** 2002. *Economic Freedom of the World: 2002 Annual Report*. Vancouver: The Fraser Institute.
- Gwartney, J. G., Holcombe, R. G., and Lawson, R. A.** 1998. The Scope of Government and the Wealth of Nations. *Cato Journal* 18(2): 163-190.
- Haber, S., North, D. C., Weingast, B. R.** 2003. If Economists Are So Smart, Why Is Africa So Poor? *Wall Street Journal*, July 30th.
- Haan, D. J., and Sturm, J. E.** 2003. Does More Democracy Lead to Greater Economic Freedom? New Evidence for Developing Countries. *European Journal of Political Economy* 19(3): 547-563.
- Hayek, F. A.** 1973. *Law, Legislation and Liberty, Volume 1: Rules and Order*, Chicago: The University of Chicago Press.
- Hayek, F. A.** 1978. Competition as a Discovery Procedure. In *New Studies in Philosophy, Politics, Economics, and the History of Ideas*. Chicago: University of Chicago Press, 179-190.
- International Monetary Fund.** 2003. *World Economic Outlook* (Chapter 3). Online: <http://www.imf.org/external/pubs/ft/weo/2003/01/> (cited: June 2004).

- Johnson, B. T., Holmes, K. R., and Kirkpatrick, M.** 1999. *1999 Index of Economic Freedom*. Washington, D.C: The Heritage Foundation and Dow Jones & Company, Inc.
- Kirmanoglu, H.** 2003. Political Freedom and Economic Well-Being: A Causality Analysis. Online: http://www.ecomod.net/conferences/ecomod2003/ecomod2003_papers/Kirmanoglu.pdf (cited: June 2004).
- Knack, S., and Keefer, P.** 1995. Institutions and Economic Performance: Cross-Country Test Using Alternative Institutional Measures. *Economics and Politics* 7(3): 207-227.
- Kurzman, C., Werum, R., and Burkhart, R. E.** 2002. Democracy's Effect on Economic Development: A Pooled Time-Series Analysis, 1951-1980. *Studies in Comparative International Development* 37(1): 3-33.
- La Porta, R., Lopez-de-Silanes, F., Shleifer, A., Vishny, R.** 1999. The Quality of Government. *The Journal of Law, Economics, & Organization* 15(1): 222-279.
- Lal, D.** 1999. *Unintended Consequences*. Cambridge: The MIT Press.
- Nelson, M., and Singh, R. D.** 1998. Democracy, Economic Freedom, Fiscal Policy and Growth in LDCs: A Fresh Look. *Economic Development and Cultural Change* 46(4): 677-696.
- North, D. C.** 1990. *Institutions, Institutional Change, and Economic Performance*, Cambridge: Cambridge University Press.
- O'Driscoll, G. P., Holmes, K. R., and Kirkpatrick, M.** 2000. *2000 Index of Economic Freedom*. New York: Dow Jones Co.
- Olson, Mancur Jr.** 1996. Big Bills Left on the Sidewalk: Why Some Nations are Rich, and Others Poor. *Journal of Economic Perspectives* 10(2): 3-24.
- Powell, B.** 2003. Economic Freedom and Growth: the Case of the Celtic Tiger. *Cato Journal* 22(3): 431-448.
- Rapaczynski, A.** 1996. The Role of the State and the Market in Establishing Property Rights. *The Journal of Economic Perspectives* 10(2): 87-103.
- Rodrik, D., and Subramanian, A.** 2003. The Primacy of Institutions, *Finance and Development* 40(2): 31-34.

- Shirley, M. M.** 2003. *What Does Institutional Economics Tell Us About Development*. A paper presented at ISNIE, Budapest, Hungary.
- Smith, A.** [1776] 1937. *The Wealth of Nations*. New York: Modern Library.
- Stocking, G. W.** 1959. Institutional Factors in Economic Thinking. *The American Economic Review*. 49(1): 1-21.
- Theil, H.** 1971. *Principles of Econometrics*. New York: Wiley.
- Vega-Gordillo, M., and Alvarez-Arce, L.** 2003. Economic Growth and Freedom: A Causality Study. *Cato Journal* 23(2): 199-215.
- World Bank.** 2003. *World Development Indicators 2003*. Washington: World Bank.
- Zucker, L. G., Darby, M. R.** 1999. Comment on "The Quality of Government": Why Do Some Governments Have Better Institutions Than Others? *The Journal of Law, Economics, & Organization* 15(1): 280-282.

ABOUT THE AUTHORS



Peter Gordon is a Professor in the University of Southern California's School of Policy, Planning and Development and in its Department of Economics. He is also Director of USC's Master of Real Estate Development program. Gordon's research interests are in applied urban economics. He is the co-editor (with David Beito and Alexander Tabarrok) of *The Voluntary City* (The University of Michigan Press, 2002). Gordon is the co-editor of *Planning and Markets*, an all-electronic refereed journal (www-pam.usc.edu). Gordon and his colleagues have developed the Southern California Planning Model, which they apply to the study of economic impacts at a detailed spatial and sectoral scales. His recent papers are at www-rcf.usc.edu/~pgordon. He has consulted for local, state and federal agencies, the World Bank, the United Nations and many private groups. Gordon received the Ph.D. from the University of Pennsylvania. His email address is: pgordon@almaak.usc.edu.



Lanlan Wang is a Ph.D student in the University of Southern California's School of Policy, Planning and Development. Wang's previous studies include a Master of Science in Regional Economic Development at Lanzhou University (Gansu Province, China). Wang's research interests are in property rights and institutions. Her email address is lanlanwa@usc.edu.



*Econ Journal Watch,
Volume 1, Number 2,
August 2004, pp 222-234.*

Kuznets Curveball: Missing the Regional Strike Zone

JEFF EDWARDS* AND ANYA MCGUIRK**

A COMMENT ON: JIH Y. CHANG AND RATI RAM. 2000. LEVEL OF DEVELOPMENT, RATE OF ECONOMIC GROWTH, AND INCOME INEQUALITY. *ECONOMIC DEVELOPMENT AND CULTURAL CHANGE* 48(4): 787-799.

[ABSTRACT, KEYWORDS, JEL CODES](#)

THE OBJECTIVE OF THIS PAPER IS TO EMPHASIZE THE NEED FOR adhering to the standard normal and identically distributed (NID) assumptions that are the basis of proper statistical inference in Ordinary Least Squares econometric models. The testing of and adherence to these assumptions are particularly important with small samples. In this paper we offer a statistical critique of a study by Chang and Ram (2000). The inferences they draw from their models are incorrect because of a violation of the assumption of identically distributed errors. Using graphical analysis and descriptive statistics, we were able to determine that significant regional fixed effects resulted in the invalidation of their conclusions that the Kuznets hypothesis holds, and that high growth countries maintain a lower level of income inequality over all income levels than do low growth countries. We add to this critique by commenting on the data used, as well as, the way Chang and Ram force the data through a somewhat arbitrary theoretical form.

* Department of Economics, Texas Tech University.

** Department of Statistics and Department of Agricultural and Applied Economics, Virginia Tech.

According to Simon Kuznets (1955) there should exist an inverted-U correlation between income inequality and the level of economic development (usually measured as per capita GDP). The inverted-U correlation suggests that for low levels of economic development there will exist low levels of inequality in per capita GDP, and inequality will increase as income increases. However, as incomes reach a particular level, i.e., as the economy develops past a certain point, the relationship reverses and income inequality decreases. Kuznets empirically documented this relationship for nations transitioning from agricultural to industrial economies.

The most common measure of income inequality is the Gini coefficient developed by Corrado Gini (1936). If the area between the line of perfect equality and the Lorenz curve is A , and the area beneath the Lorenz curve is B , then the Gini coefficient is $A/(A+B)$.¹ The Gini coefficient ranges from 0% (perfectly even distribution) to 100% (perfectly concentrated distribution). In terms of income inequality, a relatively low Gini coefficient reflects low levels of income inequality between the rich and the poor, while a relatively high Gini coefficient reflects high levels of inequality.

THE MODELS, THE DATA, AND THE DUPLICATION OF CHANG AND RAM

In their paper *Level of Development, Rate of Economic Growth, and Income Inequality*, Chang and Ram evaluate the Kuznets hypothesis using a common functional form that includes an additive growth term. They estimate the following regression:

$$\text{Ineq}_i = a_1 + b_1(\text{LY}_i) + c_1(\text{LY}_i)^2 + d(\text{GY}_i) + u_{1i} \quad (1)$$

where Ineq_i is the Gini coefficient for country i , LY_i is the natural log of income per capita, GY_i is the rate of growth of income per capita, and u_{1i} is a disturbance term that is assumed to be normal and identically distributed (NID). Multi-country cross-sectional studies prior to Chang and Ram lead to ambiguous evidence regarding the relationship between income inequality and the additive growth term in a pooled sample. To investigate this issue, Chang and Ram stratify their multi-country data by growth rate:

¹ See Appendix A.

high (2% and above), medium (between zero and 2%), and low growth countries (zero or negative growth). In their final regression they actually throw out the medium growth countries and estimate the following:

$$\begin{aligned} \text{Ineq}_i = & a_2 + b_2(\text{LY}_i) + c_2(\text{LY}_i)^2 + a_{22}\text{D} + b_{22}(\text{D}*\text{LY}_i) \\ & + c_{22}(\text{D}*(\text{LY}_i)^2) + u_{2i} \end{aligned} \quad (2)$$

where D takes the value of 1 for high growth countries and 0 for low growth countries.

The data that Chang and Ram use come from a variety of sources. The Gini coefficient measuring income inequality consists of the “high quality” observations from the Deininger and Squire (1996) data set “. . . for or around the year 1985” (790). The per capita GDP variable comes from the Penn World Tables (PWT), version 6.1 constructed by Heston, Summers, and Aten (2002). On the other hand, the growth variable comes from the United Nations Human Development Report (1993) and is averaged over the years 1980-90. The choice of this data set brings forth several interesting issues.

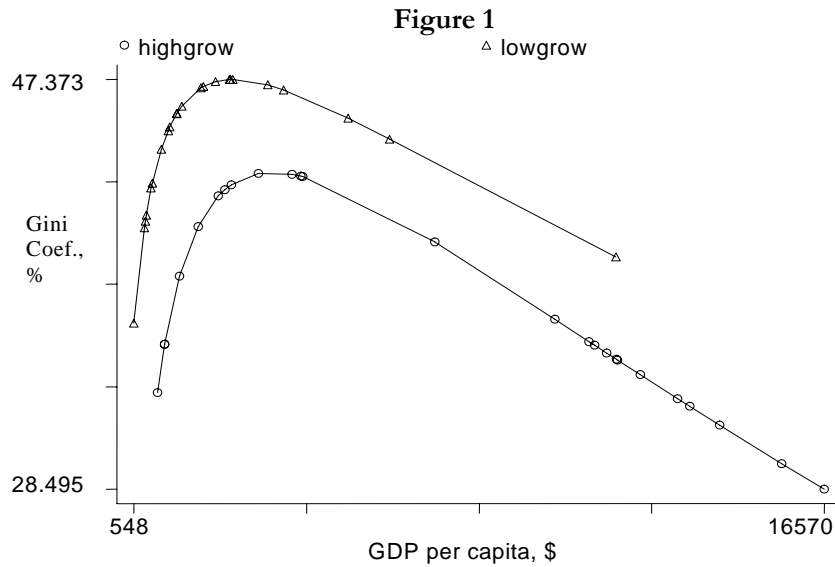
First, even though the Deininger and Squire data set is carefully constructed, with special care taken to obtain 'high quality' observations, the data on the Gini coefficient comes from many different sources, possibly leading to systematic measurement error in this series, e.g., a less democratic government source may want to report the coefficient as low as possible while a more liberal source may want to report the coefficient as high as possible. Second, it is unclear why growth in per capita GDP is averaged over a 10 year period when we are trying to measure the contemporaneous effects of per capita GDP on the Gini coefficient. One reason for the averaging may be to smooth the business cycle effects in the series, however, the same argument can then be used for the contemporaneous Gini data as well as per capita GDP, even though it is well-known that the time rate of any measured time series is ‘noisier’ than the levels variable measured over time. Third, it seems that using consistent data sources would be preferable to using many sources. For instance, the growth rate in per capita GDP could have easily been calculated from the PWT, but was not. It can be assumed that most data will have particular flaws with the way it was collected or calculated. Given these flaws, data should be taken from the same source if at all possible. Doing this will help control for the possibility of different biases resulting from data attained from different sources.

Despite the fact that there are these potential problems with the data used in Chang and Ram's paper, we now replicate their results for model (2) as shown below:

Table 1
Chang and Ram's Results of Model (2)

	Constant	LY	LY ²	D	D*LY	D*LY ²
Estimate	-212.774	65.366*	-4.106*	-193.792	43.875	-2.529
(t-statistics)	(-1.62)	(1.92)	(-1.87)	(-0.89)	(0.82)	(-0.77)
The number of observations is 48						
The unadjusted R ² is 0.40						
The adjusted R ² is 0.33						
*P≤0.10						

When graphed, the estimates of income inequality for the high and low growth countries take the following form:²



² Chang and Ram actually use predetermined values of real GDP per capita starting at 750 and increasing to 17,000, while we use the *observed* values of per capita GDP. The resulting graphs are almost identical.

The conclusion that is drawn by Chang and Ram is that high growth countries have lower levels of income inequality over all levels of income than do low growth countries. However, as we will demonstrate below, the model above is incorrectly specified because of differences in mean levels of income inequality across regions. Because relevant regional fixed effects have been omitted, Chang and Ram's results are biased and inconsistent.

THE STATISTICAL CRITIQUE, RESPECIFIED RESULTS, AND THE CONCLUSION

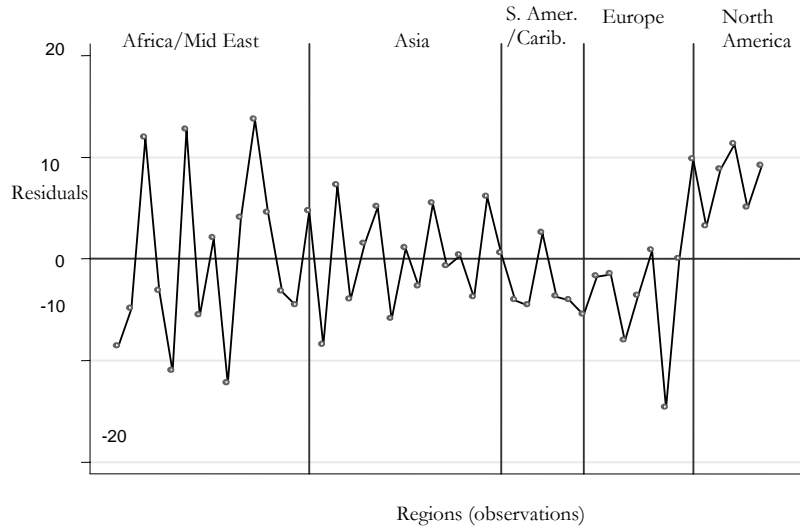
The correction of mean heterogeneity through the inclusion of fixed effects is critical to drawing accurate statistical inference from the estimated parameters. The errors of any Ordinary Least Squares regression, such as the ones used here, are assumed to be normal and identically distributed. Mean heterogeneity causes the assumption of identical distribution to break down resulting in biased and inconsistent estimators—hence, no inference can be drawn.

A close analysis of the Chang and Ram results led us to suspect the need for regional fixed effects. Figure 2 best illustrates this statistical evidence.³ This figure, which shows a plot of the residuals from the Chang and Ram model ordered by geographical region, indicates a downward structural shift in the residuals for Europe, an upward structural shift for North America, while the other residuals are fairly evenly distributed around the mean of zero.⁴ The inclusion of fixed regional effects, rather than some other sort of fixed effect, can also be justified from a theoretical point of view—one might expect that for political and socio-economic reasons, inequality in incomes may spill over from one nation to another within a particular geographical region.

³ These regions are delineated according to the Penn World Tables version 6.1, which separates countries into 15 regions, however, because of the limited number of observations, testing all of them would greatly decrease our degrees of freedom and can be akin to data mining. Furthermore, as Figure 2 depicts, this detailed delineation is not needed.

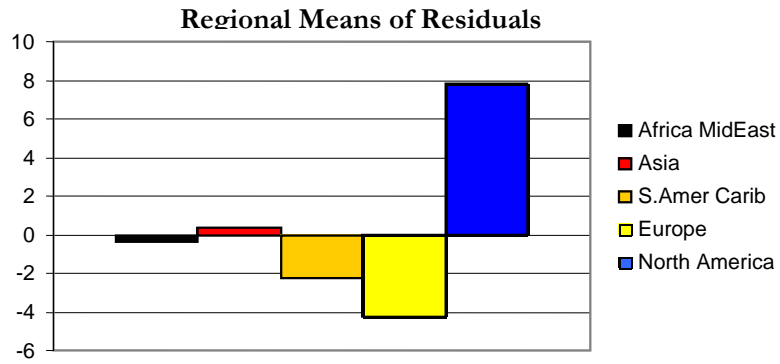
⁴ It could be argued that there exists slight heterogeneity in the South America/Caribbean region, however, with South America as the control group, the p-value for the significance of the Caribbean was 0.909, and an f-test of the joint significance of Africa and Middle East, Asia, and the Caribbean returned a p-value of 0.998. These statistics indicate no statistically significant difference in the South America/Caribbean region from the regions of Africa and Middle East, or Asia.

Figure 2



In order to emphasize the differences in the means of the residuals across regions, Figure 3 charts these means as they deviate from zero. We see that estimation of model (2) results in residual means that vary from -0.3379 for the Africa and Middle East region to 7.8344 for North America.

Figure 3



When dummy variables are included for the regions of North America, Asia, Africa and the Middle East, Europe, and the Caribbean (South America is the control group), an F-test of the joint significance of the regional intercept shifts yields a p-value of 0.0188 (the F-statistic is 3.13).⁵ To save degrees of freedom, we drop the insignificant regional dummy variables and keep Europe and North America. The resulting estimates are summarized below:

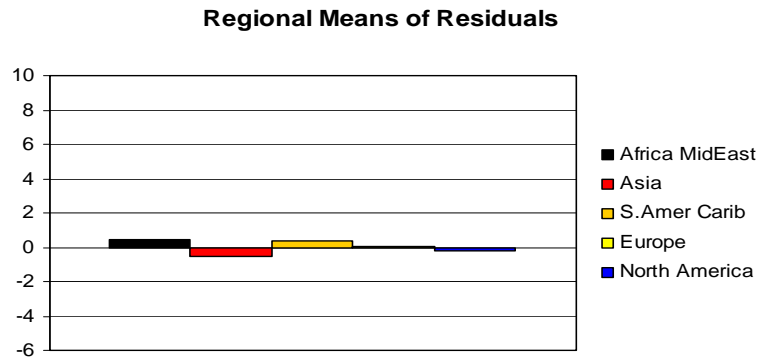
**Table 2:
Respecified Results**

	Constant	LY	LY ²	D
Estimate (t-statistic)	-143.350 (-1.242)	47.678 (1.595)	-3.018 (-1.571)	-349.76* (-1.775)
	D*LY	D*LY ²	Europe	N. Amer.
Estimate (t-statistic)	81.692* (1.672)	-4.757 (-1.581)	-6.147* (-1.804)	8.793* (2.837)
The number of observations is 48 The unadjusted R ² is 0.58 The adjusted R ² is 0.51 * P ≤ 0.10				

As we can see above, LY and LY² are no longer statistically significant at the 10% level, whereas they were under the specification in Table 1 (the t-statistic cutoff is 1.684 for a 90% confidence interval with 40 degrees of freedom). Furthermore, our mean heterogeneity results suggest that the relationship between the Gini coefficient versus incomes (comparable to those of Chang and Ram) will vary by region—i.e., the intercepts will differ. In addition, the residuals of this new model show no un-modeled systematic differences by regions.

⁵ We could have used a Chow-type test here to test for structural change, however, to get a reliable test statistic, we would need a much larger data set due to the design of the Chow test.

Figure Four



However, there is a slight problem with the estimates in Table 2.⁶ The region of North America includes the countries of Canada, Mexico, the United States, Honduras, Guatemala, and Panama. This is problematic because the average gini coefficient of the United States and Canada is 35.03, while the average gini coefficient of Mexico, Honduras, Guatemala and Panama is 55.06. To correct for this, we separated the United States and Canada from the rest—redefining the North American dummy variable into two separate dummy variables. The results are that the dummy variable controlling for the United States and Canada had a t-statistic of -0.213, while the dummy variable controlling for Mexico, Honduras, Guatemala, and Panama was 3.394. We dropped the former dummy variable and kept the latter and then re-estimated the model.

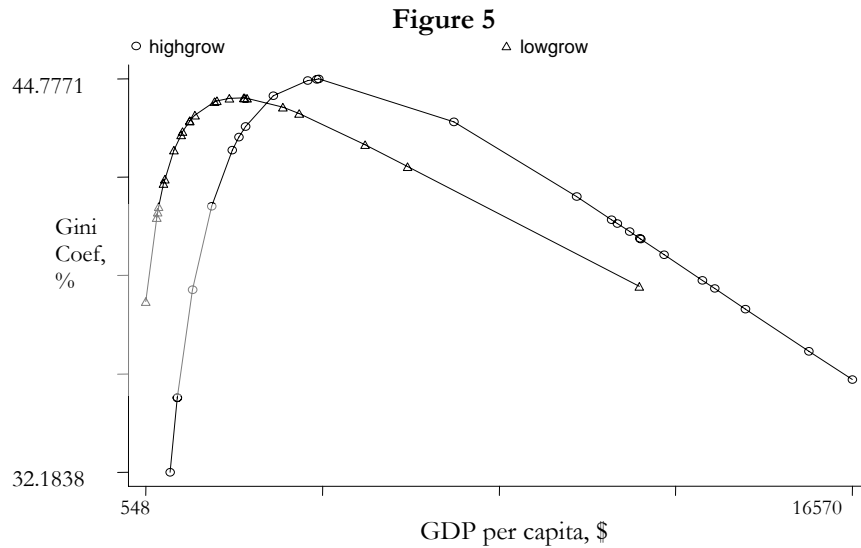
⁶ This observation was graciously contributed by the authors of the original paper in correspondence prior to the submission of this paper to *Econ Journal Watch*.

Table 3:
Respecified Results

	Constant	LY	LY ²	D
Estimate	-143.350	42.097	-2.675	-260.863
t-statistic	(-1.242)	(1.453)	(-1.438)	(-1.436)
	D*LY	D*LY ²	Europe	N. Amer.
Estimate	59.027	-3.310	-9.483*	11.568*
t-statistic	(1.314)	(-1.200)	(-3.079)	(3.435)

The number of observations is 48
 The unadjusted R² is 0.61
 The adjusted R² is 0.54
 * P ≤ 0.10

After re-estimation only the dummies for Europe and the four Central American countries are significant and none of the coefficients for GDP are significant. Therefore, the Kuznets hypothesis is invalid. However, if we assume the hypothesis does hold, and force the data through Chang and Ram's quadratic functional form from equation (2), we can graph the relationship as in Figure 1 assuming the countries are all in the control group (we subtracted out the estimated effects of the two dummy variables from the estimates of the gini coefficient).



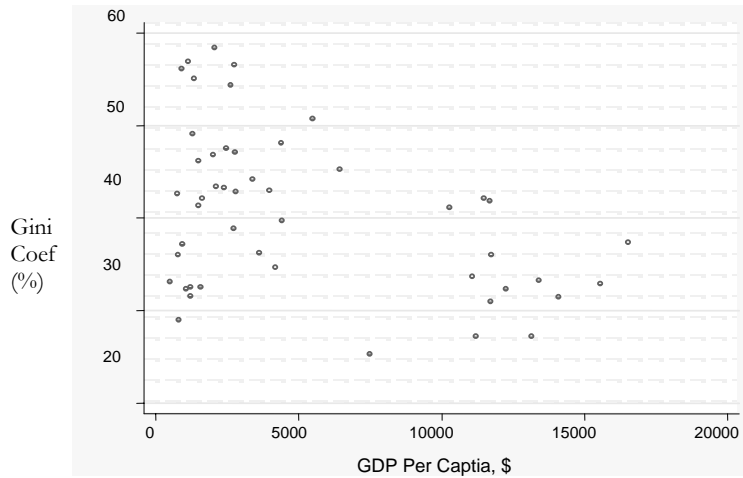
What we find in this new graph is that for low levels of GDP per capita, high growth countries have less income inequality than low growth countries, and for medium to high levels of GDP per capita, low growth countries have less income inequality than high growth countries. When important regional heterogeneity is modeled, the conclusion of Chang and Ram that higher growth countries have lower levels of inequality no longer holds for all income levels. The economic explanation of this crossing pattern would be beyond the scope of this paper, but is moot given that the GDP terms are both statistically insignificant and, given the multiple problems of the entire exercise, economically insignificant.

What has just been depicted is a complete structural breakdown of Chang and Ram's results. This breakdown results from the violation of two critical assumptions—the assumption that the Kuznets hypothesis holds in the first place, and the assumption that the errors corresponding to the regression of model (2) are normal and identically distributed. In order to address the former, we must first address the latter.

We have seen that the identically distributed assumption of the errors breaks down with regard to regionally-specific fixed effects. Including the dummy variables of Europe and (what we define as) Central America, results in the statistical insignificance of all per capita GDP variables in Table 3. In fact, only the two dummy variables are statistically significant. This is in contrast to Chang and Ram's results where at least the Kuznets hypothesis holds for the low growth countries, but not for the high growth countries. One might argue that the regional dummy variables are simply proxies for the high and low growth countries, however, this line of reasoning is flawed. The flaw lies in the simple fact that the complete set of high or low growth countries are not embodied in the regional dummies, and furthermore, the intercept dummy for the high growth countries in Table 1 was highly insignificant. This indicates that there never was a statistically significant structural difference between high and low growth countries.

The first assumption—that the Kuznets hypothesis holds at all—is much easier to address. Given the fact that the structural dummy for the high growth countries is statistically insignificant in both the Chang and Ram regression (Table 1) and in our final regression (Table 3), a simple plot of the Gini coefficient on per capita GDP would tell us if there is any inverted-U pattern to estimate.

Figure 6

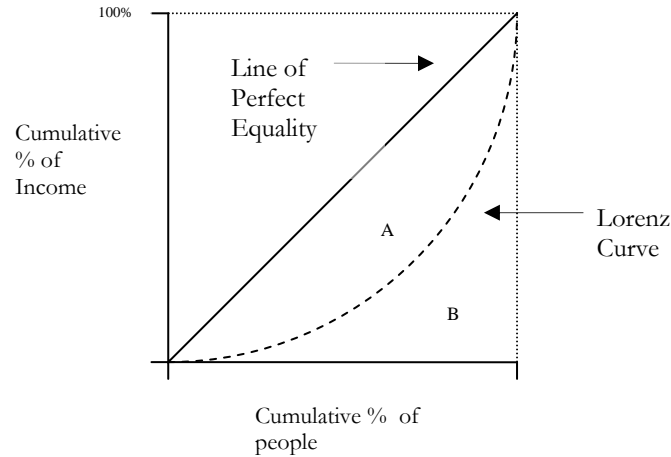


The appearance of any quadratic correlation between the Gini coefficient and per capita GDP is not apparent in Figure 6. In fact, if pressed, there would almost seem to be a somewhat linear, negative correlation between the two. We must remember that the reason why Figures 1 and 5 depict the typical inverted-U pattern is simply because the coefficients to the squared GDP variables happen to be negative (even though not statistically significant) and because *the data was forced through an equation that produces a quadratic-type curve*. It is also unclear why this type of quadratic structure was chosen among many others that produce the same shape. Randolph and Lott (1993) address seven possible forms, with supporting results, for the inverted-U pattern of the Kuznets hypothesis.

In conclusion, in Chang and Ram's paper, not only is the data set troublesome and the functional form ad hoc, but the forcing of the data through the chosen functional form when there are apparent mean heterogeneity problems is statistically improper.

APPENDIX A

Graph of the Gini coefficient and the Lorenz curve with regard to income inequality:



REFERECNES

- Chang, Jih Y., and Rati Ram.** 2000. Level of Development, Rate of Economic Growth, and Income Inequality. *Economic Development and Cultural Change* 48(4): 787-799.
- Gini, Corrado.** 1936. On the Measure of Concentration with Special Reference to Income and Wealth. Abstracts of Papers Presented at the Cowles Commission Research Conference on Economics and Statistics, Colorado College Publications.
- Klaus Deininger and Lyn Squire.** 1996. A New Data Set Measuring Income Inequality. *World Bank Economic Review* 10(3): 565-591.
- Kuznets, Simon.** 1955. Economic Growth and Income Inequality. *American Economic Review* 45: 1-28. need issue number.
- Randolph, Susan, and William F. Lott.** 1993. Can the Kuznets Effect Be Relied on to Induce Equalizing Growth? *World Development* 21(5): 829-840.

Heston, Alan, Robert Summers, and Bettina Aten. 2002. The Penn World Table Version 6.1. Center for International Comparisons at the University of Pennsylvania. October.

United Nations Development Program. 1993. *Human Development Report*, 188-89.

ABOUT THE AUTHORS



Jeffrey Edwards is an assistant professor of economics at Texas Tech University. He holds a Ph.D. in economics from Virginia Tech with a major concentration in econometrics. His research interests are empirically broad from testing probabilistic hypotheses of empirical models, to growth and development, to water resource management. He is also an assistant editor for the *Journal of Economics*, and currently sits on the advisory panel of the first international water resources conference sponsored by Texas Tech.

His email address is jeffrey.edwards@ttu.edu.



Anya McGuirk is a full professor at Virginia Tech. She holds a joint appointment with the Departments of Statistics, and Agricultural and Applied Economics. Her primary areas of research are applied econometrics and microeconomics. Her most recent works include studies of the restrictions imposed when modeling error autocorrelation and the nature of the implications of multicollinearity in regression analysis. McGuirk also teaches econometrics and microeconomic theory at the graduate and undergraduate levels. Anya obtained her MSc. as a Marshall Scholar at the University of Reading, England. She received her Ph.D. from Cornell University and her dissertation, "Technological Change and Agricultural Growth in Punjab, India," received the AAEE Outstanding Dissertation Award (1989). While McGuirk enjoys her research and teaching, her real loves are fly-fishing and family.

[GO TO RESPONSE BY JIH CHANG AND RATI RAM](#)



Response to Edwards and McGuirk: Income Level, Economic Growth, and Inequality: Flawed Methodology and Inaccurate Inference

JIH Y. CHANG* AND RATI RAM**

THE COMMENT BY EDWARDS AND MCGUIRK (2004) HAS numerous significant flaws. These include: (1) an incomplete understanding of the paper by Chang and Ram (2000), (2) lack of familiarity with development data sets and the vast literature on cross-country studies of inequality and growth, and (3) an inadequate grasp of the basics of econometric model specification and the elementary principles of generating predicted values from regression estimates. We first briefly note a few aspects that might not be considered major, and then explain the fatal flaw that renders its main conclusion inaccurate.

NONFATAL PROBLEMS

Kuznets hypothesis

The Comment carries several types of misunderstandings about our work. In particular, despite our explicit statements (Chang and Ram 2000, 790-791), Edwards and McGuirk do not recognize the point that testing of Kuznets's hypothesis was a minor and secondary part of our paper—our

* Financial Aid Office, Illinois State University.

** Department of Economics, Illinois State University.

principal objective was to compare evolution of income inequality in low-growth and high-growth countries. Neither the functional form used nor the actual structure of the income-inequality relation, is critical for a comparison of the income-inequality profiles in high- and low-growth contexts. Therefore, long discussions about the functional form and repeated references to the validity, or lack of validity, of the hypothesis (and the cheap title) tend to obscure the main issue. However, since a large part of the Comment refers to, or discusses, the hypothesis, we note the following points that refute what Edwards and McGuirk say.

1. We explicitly stated that “Presence or absence of the inverted-U . . . is secondary to the main purpose of our study,” and that the “evidence in favor of Kuznets’s hypothesis [in Table 1 of our original paper] is not the main point of our work” (Chang and Ram 2000, 790, 791).
2. The functional form chosen by us to relate income with inequality is not “arbitrary” or “ad hoc,” but is probably the most widely used specification in the literature. Besides its usage in numerous studies, the survey by Adelman and Robinson (1989, 958) noted that researchers on the subject “generally assumed that the relationship . . . is quadratic in the log of per capita GNP.” The terminology used in the Comment about the functional form reflects its lack of familiarity with the literature.
3. Before asserting that our work contains a “critical . . . assumption that the Kuznets hypothesis holds,” Edwards and McGuirk could have read our observation (790) that our equation “does not necessarily presume that there is an inverted-U pattern. Presence or absence of the inverted-U in either group is determined by the data.” Perhaps even more important, it should be evident that presence or absence of an inverted-U does *not* affect comparisons of income-inequality profiles in the two growth groups.
4. In the above context, it seems absurd to suggest that forcing the data through our specification produces an inverted-U. It should be elementary to see that, depending on the signs and magnitudes of the parameter estimates, the quadratic specification can reveal the relation between income and inequality to be linear, inverted-U, regular-U, monotonically increasing, or monotonically decreasing. Such quadratics have been used in numerous studies to *test* and not to force an inverted-

U, or any other pattern. In fact, the quadratic specified by us has been used to show that there is an inverted-U relation, a regular-U relation, or no relation between the variables. There are scores of studies in the literature where this very specification yields evidence *against* the Kuznets hypothesis. By way of a quick example, consider the following two subsample estimates from the data used in the present work (t-statistics are in parentheses).

$$\text{Low-growth cases: Gini} = -121.44 + 42.10 \text{ LRY} - 2.68 \text{ LRYSQ} + 11.57 \text{ CAMER}$$

$$(N=23, \text{RSQ}= 0.43) \quad (-0.97) \quad (1.30) \quad (-1.29) \quad (3.08)$$

$$\text{High-growth cases: Gini} = -382.31 + 101.13 \text{ LRY} - 5.99 \text{ LRYSQ} - 9.48 \text{ EUROPE}$$

$$(N=25, \text{RSQ}= 0.60) \quad (-3.03) \quad (3.34) \quad (-3.33) \quad (-3.49)$$

Thus, our quadratic specification produces strong evidence in favor of the hypothesis in one subsample and against the hypothesis in the other with almost identical sample size. It is not the quadratic specification that produces a Kuznets (or any other) curve— what one sees depends on the data. The subsample estimates indicate two additional points. First, it should be evident that, contrary to what the Comment suggests, the structure of the income-inequality relation does differ markedly in the two groups. Second, the low-growth group has no member from “Europe,” and the high-growth group contains no country from “Central America,” indicating that the regional dummies might be surrogates for low- and high-growth scenarios and could distort considerably the low- and high-growth classification.

5. It is not obvious which observations are covered in Edwards and McGuirk’s Figure Six, nor what is meant by “The appearance of any quadratic correlation between the Gini coefficient and per capita GDP is not apparent.” However, the following estimates may be instructive relative to the observations in the Comment about the absence of Kuznets-curve in our data (t-statistics are in parentheses).

Full sample: (N=65, RSQ= 0.46)

$$\text{Gini} = -214.06 + 64.09 \text{ LRY} - 3.97 \text{ LRYSQ} - 7.57 \text{ Europe} + 11.80 \text{ CAMER}$$

$$(-2.94) \quad (3.56) \quad (-3.60) \quad (-2.59) \quad (3.30)$$

Excluding medium-growth cases: (N=48, RSQ= 0.54)

$$\text{Gini} = -123.01 + 41.23\text{LRY} - 2.55\text{LRYSQ} - 8.94 \text{ Europe} + 12.46 \text{ CAMER}$$

(-1.66) (2.23) (-2.23) (-2.94) (3.79)

The estimates provide unmistakable evidence in favor of the hypothesis even with the regional dummies, and this is what the *data* show, and not what the specification forces on the data. As noted, there are numerous studies where the same specification yields evidence against the hypothesis.

Data

Edwards and McGuirk's observations about the Deininger-Squire compilation reflect their lack of familiarity with the strong and weak points of inequality data sets. Similarly, their suggestion about growth rates being based on PWT indicates that they do not realize that PWT data are in international dollars and do not constitute as good a basis for intracountry growth rates as domestic currency numbers. Also, Edwards and McGuirk do not appear to have noticed that we (Change and Ram 2000, 799 n23) had explicitly said that the patterns are similar for different growth periods and inequality measures.

Specification search

While making a big issue about the properties of the error terms, Edwards and McGuirk start with six regional groups that are defined geographically and have little obvious relevance to cross-country distribution of income inequality. After stating that the five regional dummies are jointly significant, they drop three "to save degrees of freedom," and finally truncate "North America" to retain only Guatemala, Honduras, Mexico, and Panama. Edwards and McGuirk show little awareness of the potentially serious hazards of such specificational searches in terms of pretest bias and omitted-variable problems due to the deletion of individually "insignificant" variables. The poverty of the model Edwards and McGuirk choose (in their Table Three) is reflected in the fact that it shows the two regional dummies as the only significant determinants of cross-country distribution of income inequality.

FATAL ERROR

After having strenuously argued that the inclusion of regional dummies is “critical,” it is amazing that Edwards and McGuirk offer the plots in Figure Five by way of a comparison of predicted income-inequality profiles in low-growth and high-growth cases. The plots are drawn by ignoring the parameters for the regional dummies and by assuming that “countries are all in the control group.” In other words, the plots are drawn on the assumption that the constant term for all countries is the same, which is a blatant contradiction of their major claim. Thus, the plots are inaccurate relative to Edwards and McGuirk’s own estimates (in their Table Three), and the conclusion drawn from the plots is wrong even in terms of the model estimates chosen.

Although Edwards and McGuirk do not state what countries are included in “Europe,” we were able to replicate almost exactly (except for the constant term) estimates of their Table Three by treating this group as consisting of Denmark, Finland, Germany, Italy, Luxembourg, Norway, Spain, and U.K. These estimates are shown in Table 1.

Table 1:
Estimates of the model proposed in the Comment
that includes two regional dummies

Constant	LY	LYSQ	D	D*LY	D*LYSQ	Europe	CAMER
-121.445	42.098	-2.675	-260.864	59.028	-3.311	-9.483	11.568
(-1.09)	(1.45)	(-1.44)	(-1.44)	(1.31)	(-1.20)	(-3.08)	(3.44)

t-statistics are in parentheses
N=48, R-square is 0.61, Dependent Variable is Gini

Table 2 shows the value of predicted Gini for each country in the low-growth and high-growth groups on the basis of the estimates in our Table 1 here. We follow Edwards and McGuirk in using the actual income data to generate the predicted value of Gini for each country, and the countries are listed in the ascending order of GDP per capita. Figure 1 here plots the predicted inequality profiles for the two groups. Several points may be noted from Table 2 and Figure 1.

Table 2

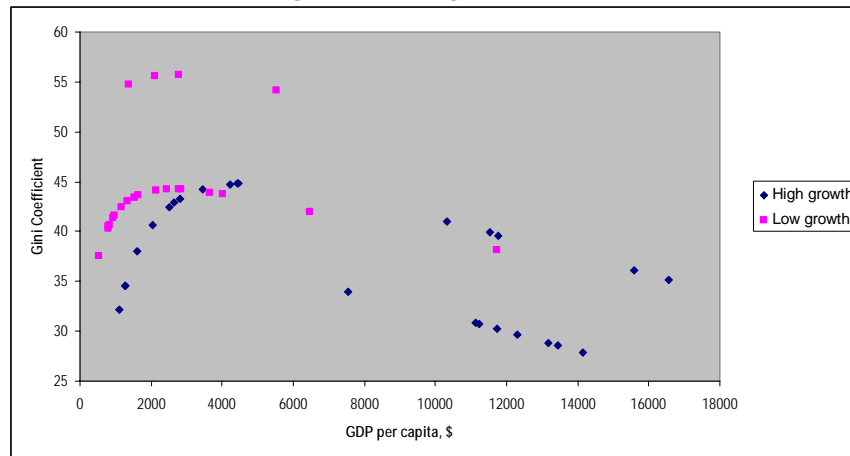
Predicted value of Gini for each country in high-growth and low-growth groups					
LOW-GROWTH CASES			HIGH-GROWTH CASES		
County	GDP pc	Predicted Gini	Country	GDP pc	Predicted Gini
Uganda	548	37.6	India	1092	32.2
Mauritania	788	40.3	China	1262	34.6
Ghana	811	40.5	Pakistan	1262	34.6
Rwanda	834	40.7	Indonesia	1610	38.0
Lesotho	949	41.4	Sri Lanka	2040	40.7
Nigeria	973	41.6	Thailand	2510	42.5
Zimbabwe	1182	42.5	Botswana	2662	42.9
Cameroon	1342	43.0	Seychelles	2811	43.3
Honduras	1381	54.7	Turkey	3441	44.2
Philippines	1542	43.4	S. Korea	4217	44.7
Cote d' Ivo.	1545	43.4	Malaysia	4420	44.8
Bolivia	1658	43.6	Mauritius	4462	44.8
Guatemala	2104	55.6	Spain	7536	33.9
Dom. Republic	2157	44.1	Singapore	10316	41.0
Jamaica	2443	44.2	Italy	11115	30.8
Algeria	2769	44.2	U.K.	11237	30.7
Panama	2785	55.7	Hong Kong	11520	39.9
Peru	2838	44.2	Finland	11746	30.2
Jordan	3649	43.9	Japan	11771	39.6
Iran	4027	43.7	Germany	12302	29.7
Mexico	5524	54.2	Luxembourg	13175	28.8
Venezuela	6483	42.0	Denmark	13449	28.6
Tri. & Tobago	11738	38.1	Norway	14144	27.9
			Canada	15589	36.1
			U.S.A.	16570	35.2

First, the plot in Edwards and McGuirk's Figure Five is inaccurate since it does not use the regional-dummy parameter estimates. Figure 1 here shows the correct plot of predicted Gini for every country in each group based on the estimates in Tables 1 and 2 here.

Second, the overall scenario in Figure 1 here is very different from that suggested by Edwards and McGuirk, and is almost identical with that shown by Chang and Ram (2000, 794). Of the 48 cases, there is only one (Trinidad and Tobago) where a low-growth country has smaller predicted

inequality than a high-growth country at a similar income level. Therefore, the Chang-Ram inference, that a high-growth country is likely to have smaller income inequality than a low-growth country at the same income level, is true even on the basis of Edwards and McGuirk’s model that includes two regional dummies after a highly selective procedure.

Figure 1:
Predicted values of Gini at various income levels in high- and low-growth cases



Third, unlike our original Figure 1 (Chang and Ram 2000, 794), the predicted inequality profiles in Figure 1 of this reply are not smooth, because inequality now depends, not merely on income, but also on regional location, and, in either growth regime, two countries with the same income level would have different values of predicted Gini, if these are located in different regions. It was such a truncation of predicted income-inequality profiles that Chang and Ram sought to avoid by working with the simple Kuznets quadratic that contains only income terms and that has been used in scores of cross-country studies of income inequality. Edwards and McGuirk’s Figure Five portrays smooth income-inequality profiles by assuming the same constant (intercept) term for all regions, which is a blatant contradiction of its own “insight” about the need for regional dummies. In fact, the title of the Comment seems to apply more to what it does to generate predicted inequality than to anything that Chang and Ram (2000) might have done.

Last, despite the truncation of the inequality profiles, Figure 1 here supports our important observation (Chang and Ram 2000, 793, 795) that the high-growth advantage in terms of smaller inequality is particularly large for countries at lower income levels, relative to whom the entire discussion is more meaningful.

CONCLUSION

Putting aside Edwards and McGuirk's largely irrelevant observations about the Kuznets-quadratic and the degree of validity of the Kuznets-hypothesis in our data, and even overlooking their highly selective choice of the regional dummies, the estimates of their own specification generate a picture of predicted income-inequality profiles that is remarkably similar to that shown by Chang and Ram (2000) and is very different from the profiles displayed in the Comment, profiles which are inaccurate relative to Edwards and McGuirk's own specification. Therefore, nothing in Edwards and McGuirk's Comment alters the Chang-Ram conclusion that at any income level a high-growth economy is likely to experience smaller inequality than a low-growth country. As a secondary matter, we note that, while carrying a denigrating tone relative to the Chang-Ram work, the Comment assumes the air of a superior understanding of development data sets, models of cross-country inequality and the related literature, and econometric methodology, although, as our brief observations might indicate, the opposite seems more likely.

REFERENCES

- Adelman, Irma, and Sherman Robinson.** 1989. Income Distribution and Development. In *Handbook of Development Economics*, Vol. 2, ed. Hollis Chenery and T.N. Srinivasan. Amsterdam: North-Holland, 949-1003.
- Chang, Jih Y., and Rati Ram.** 2000. Level of Development, Rate of Economic Growth, and Income Inequality. *Economic Development and Cultural Change* 48(4): 787-799.
- Edwards, Jeff, and Anya McGuirk.** 2004. Kuznets Curveball: Missing the Regional Strike Zone. *Econ Journal Watch* 1(2): 222-234.

ABOUT THE AUTHORS

JIH Y. CHANG is Assistant Director in the Financial Aid Office, at Illinois State University, Normal, IL.

RATI RAM is Distinguished Professor of Economics at Illinois State University, Normal, IL.

[GO TO REPLY BY JEFF EDWARDS AND ANYA MCGUIRK](#)



Reply to Chang and Ram: Statistical Adequacy and the Reliability of Inference

JEFF EDWARDS* AND ANYA MCGUIRK**

FIRST, WE WOULD LIKE TO APOLOGIZE TO JIH CHANG AND RATI Ram for the tone of our Comment; we in no way meant to be denigrating. When submitting the paper we had no clue that such offense would be taken. On receiving the response from Chang and Ram and rereading what we said, we can see why offense was taken. If we could reword our Comment as it goes to electronic press, we would.

Rather than continue the mud slinging which we initiated, we would like to step back and give some perspective on our Comment and attempt to clarify our main points, which for the most part were misinterpreted by Chang and Ram. We then reflect on the whole discussion and consider what, if anything, Chang and Ram could conclude even if their analysis were statistically justified. A simple look at some confidence intervals indicates that nothing at all can be concluded regarding the differences found in the inequality-income profiles of high- and low- growth countries. The model we ultimately propose based on statistical grounds also finds no statistically significant differences between high- and low- growth countries.

We begin by noting that like many other economists, we are skeptical of, and disappointed with, the current state of published empirical work in economics. Empirical results often seem to be sensitive to model

* Department of Economics, Texas Tech University.

** Department of Statistics and Department of Agricultural and Applied Economics, Virginia Tech.

specification, few theories have ever been abandoned based on evidence from regression analysis, and conflicting results seem to co-exist happily (some even utilizing the same data). In their *EDCC* paper, Chang and Ram illustrate that the inequality and growth literature is not immune from these problems: “the overall empirical evidence does not show a clear picture” and “studies that did consider the growth-inequality nexus have reported somewhat contradictory results” (Chang and Ram 2000, 787). Unfortunately, this is status quo in almost all areas of economic research.

Many have recognized this unsatisfactory state of affairs. For example, almost 30 years ago, Leontief noted, “In no other field of empirical inquiry has so massive and sophisticated statistical machinery been used with such indifferent results” (1971, 3). A multitude of reasons have been given for the apparent unreliability and non-decisiveness of empirical evidence. At least some of the problem can be attributed to the non-systematic way in which applied econometric studies are carried out.¹ Pagan (1984) cynically describes the “typical” approach to econometrics.

Four steps almost completely describe it: a model is postulated, data gathered, a regression run, some t-statistics or simulation performance provided and another ‘empirical regularity was forged.’ (Pagan 1984, 103)

Although the state of affairs described by Pagan is, we hope, somewhat exaggerated, it is true that the way econometrics is currently taught provides little guidance in how to systematically choose a model or to fix (re-specify) empirical models if statistical or theoretical “problems” with the initial specification are encountered.

The probabilistic reduction approach, proposed and championed by Spanos (1986, 1995, and 1999) provides a well-needed systematic alternative to the approach characterized by Pagan. The foundation of Spanos’ approach is the principle of *statistical adequacy*. This principle asserts that to evaluate any theory empirically, that theory must be viewed in the context of a valid statistical model—a model whose underlying assumptions are appropriate for the data being analyzed. Otherwise, test statistics will not have their anticipated distributions (and associated error probabilities), and inferences drawn using those statistics will be unreliable and likely misleading.

¹ We could go on about other aspects of this problem, but we must limit the scope of this response. See for example, Leamer (1983), Spanos (1995), Tomek (1993, 1997), and Hendry (1993).

Note that “inferences drawn” include both inferences regarding hypotheses formally tested, as well as, *predictions* made from the model.

Our comment on the Chang and Ram article was written to illustrate potential problems and pitfalls with the way empirical work is often carried out and to illustrate that “statistical” results may not be as clear-cut as they first seem if one digs a little deeper, and examines more carefully the statistical properties of their data, and the relevance of the model estimated.

On the basis of the regression model we re-examine in our comment, Chang and Ram test Kuznets’s hypothesis (though not their primary focus) *and* examine simulated/predicted gini coefficient profiles for high- and low growth regimes. The quality—reliability and properties—of the inferences drawn from *both* these activities depends on the validity of the statistical assumptions underlying the model estimated for the data they are using.

Like many researchers, Chang and Ram chose to estimate the simplest, most widely used model in this field of work. This simple model selection criterion is very common in applied work. Unfortunately the popularity and simplicity of a regression model does not guarantee that the model assumptions are reasonable for the data being analyzed and, thus, that the analysis carried out using the model will lead to reliable inferences. It is in this sense that we claimed that the model estimated by Chang and Ram is *ad hoc*. They simply chose the model, subjected the model to the data, and drew conclusions regarding the Kuznets hypothesis using the reported t-statistics, and regarding the differences between high- and low-growth countries, using predictions from the model. Without attempting to assess whether the statistical assumptions underlying this model are valid for the data chosen, we have no way of assessing whether or not, a la Pagan, ‘another empirical regularity was forged.’

The readily testable assumptions underlying Chang and Ram’s model:

$$INEQ_i = a_3 + b_3(LY_i) + c_3(LY_i)^2 + a_{33}D + b_{33}(D \times LY_i) + c_{33}[D \times (LY_i)^2] + u_{3i}, \quad [1]$$

which forms the basis of their analysis, are:

- (i) functional form is correct: $E(u_{3i}) = 0$
- (ii) $Var(u_{3i}) = \sigma^2$: homoskedasticity-- $Var(u_{3i})$ is not a function of the regressors
- (iii) $(a_3, b_3, c_3, a_{33}, b_{33}, c_{33}, \sigma^2)$ are constant over the index i
- (iv) $E(u_{3i}u_{3j}) = 0 \forall i \neq j$, and

$$(v) \quad u_{3i} \sim N(\cdot, \cdot).^2$$

Although not mentioned in our Comment, we performed a battery of misspecification tests designed to examine the appropriateness of each of these assumptions.³ Note that in order to evaluate whether the parameters varied over the index i (assumption iii), and whether the errors were autocorrelated (assumption iv), we ordered the data three different ways—by growth rate, by location (region), and by GDP.⁴ The results of these tests suggested that almost all of the testable assumptions underlying the simple model above seemed reasonable for the data being analyzed. The one assumption found to be clearly inappropriate for these data was parameter stability; misspecification tests and graphical analysis both indicated that the intercept of the regression differed by region (see our Comment for details).⁵ When equation [1], supplemented with regional dummies, was subjected to the complete battery of misspecification tests, we found no evidence against any of the testable model (statistical) assumptions. Because, based on the misspecification tests we conducted, the new model is adequate from a statistical standpoint, we can use it to examine any hypotheses or theories we are interested in and for prediction—there are well-known results about the relevant test statistics and the properties of these forecasts, etc.⁶

An F-test of the significance of all the regional dummies (base was South America) indicated that, as a group, these dummies were significantly different from zero (p-value 0.004)—hence the intercept differs by region.⁷

² Note assumptions (i)-(v) are encompassed by $u_{3i} \sim NIID(0, \sigma^2)$ --the notation used in our Comment. We used this notation so we could avoid the details we now feel we should reveal to make our general point.

³ We performed both the individual and joint misspecification tests suggested in McGuirk, Driscoll, and Alwang (1993), and McGuirk, Driscoll, Alwang, and Huang (1995).

⁴ To shorten our discussion we write GDP rather than real GDP per capita, though we are referring to the latter.

⁵ Violation of the parameter stability assumption implies OLS estimators are biased and inconsistent.

⁶ Unfortunately we can never conclude that this is “*the correct*” model to use—it is possible, for example, that another researcher, conducting a different set of misspecification tests, could find evidence against the model assumptions. It is also possible that, say, another model, with a different functional form could pass the same battery of tests conducted here. In this case, we would try to distinguish between these models, to determine which is most appropriate for these data using other criteria.

⁷ Note that the F-test reported here is different than the one reported in the Comment. For the F-test here we have the North American Countries split into USA/Canada and Central

However, when all the dummies were included, the only variables significant at the 5% level were the Central American (CA) and European (E) dummies; the variables LY and LY² were significant at the 10% level. This evidence, in conjunction with the graphical evidence, led us to test whether or not we even needed to include the non-significant regional variables at all. An F-test of the hypothesis that all but the CA and E dummies were equal to zero yielded: $F(3,38)=0.101$ ($p\text{-value}=0.96$) and, thus, we dropped the remaining regional effects. The main effect of deleting these extra regional variables was an increase in the t-statistics associated with the variables involving the high-growth dummy. When this simpler model was subjected to the complete battery of misspecification tests, we found no evidence against any of the testable model (statistical) assumptions. Thus, we concluded that this simpler model was also adequate from a statistical point of view (see Table 1).⁸

Table 1

Edwards and McGuirk Results							
Constant	LY	LY ²	D	D*LY	D*LY ²	Europe	C. Amer.
-121.45	42.10	-2.68	-260.86	59.03	-3.31	-9.48*	11.57*
(-1.09)	(1.45)	(-1.44)	(-1.44)	(1.31)	(-1.20)	(-3.08)	(3.44)
t-statistics in parenthesis							
The number of observations is 48							
The unadjusted R ² is 0.61							
The adjusted R ² is 0.54							
* P ≤ 0.10							

The question that remains is whether or not all of this additional analysis actually changes substantively any of the conclusions of Chang and Ram. Chang and Ram's test of the Kuznets hypothesis and the income-inequality profiles analyzed were obtained using a model whose assumptions were not all reasonable for the data being analyzed—thus,

America. In the Comment we refer to the test using the un-split group. Though the regression with the un-split group also seemed to be statistically adequate, we realized, after communicating with Chang and Ram, that splitting this group not only made sense theoretically but also improved the evidence for the model assumptions. For a discussion regarding the issues raised by Chang and Ram regarding 'specification search' or data mining and when it is a legitimate practice, see Spanos (2000).

⁸ This is the model reported in Table 3 of our Comment except—as pointed out by Chang and Ram—we made a typographical error when reporting our constant. Our results are identical to those reported by Chang and Ram.

there is the potential that we will be misled by their model results.⁹ For completeness, the final Chang and Ram model is summarized in Table 2.

Table 2

Chang and Ram Final Results					
Constant	LY	LY ²	D	D*LY	D*LY ²
-212.77 (-1.62)	65.37* (1.92)	-4.11* (-1.87)	-193.79 (-0.89)	43.88 (0.82)	-2.529 (-0.77)
t-statistics in parenthesis The number of observations is 48 The unadjusted R ² is 0.40 The adjusted R ² is 0.33 * P ≤ 0.10					

We begin by comparing the forecasts of the two models, as this was the primary focus of the *EDCC* paper (see Chang and Ram’s Response). In figure 5 of our Comment we attempted to illustrate how our model predictions for high- and low-growth countries differed from those in Chang and Ram. To make this plot, we subtracted out the European (E) and Central American (CA) regional differences and then graphed the model predictions. Thus, as stated in our comment, we graphed our model predictions for all the *other* high- and low-growth countries, and for E and CA countries *after* having subtracted out their specific regional effects.¹⁰ This is a very different exercise than *not* modeling the regional differences at all. If we ignored the regional effects we would get the Chang and Ram predictions.

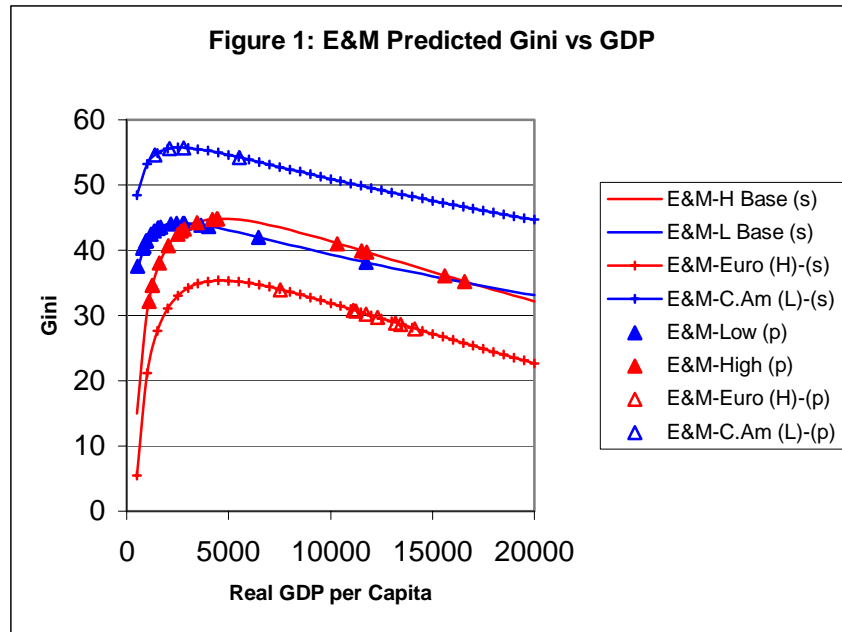
We now make another attempt to illustrate the differences in predictions. As pointed out by Chang and Ram, the 8 E countries are all high-growth and the 4 CA countries are low-growth. Thus, our model essentially has 2 high-growth groups (of 8 and 17 countries respectively) and 2 low-growth groups (of 4 and 19 countries respectively) and as a consequence we get (smooth) prediction lines for all 4 groups. The E and

⁹ Chang and Ram are right in their Response when they say that the presence or absence of the inverted U does not affect comparisons of the predicted income-inequality profiles. The quality and reliability of the different predictions (and thus, comparisons)—since they are based on a regression model—*will* depend on the relevance of the model’s underlying statistical assumptions for the data being analyzed.

¹⁰ In hindsight, we should have just graphed the simple predictions for our base groups. The resulting graph would have been *identical* to the one presented, except it would have had 8 fewer triangles and 4 fewer dots.

CA country predictions differ from their particular base group predictions by a simple vertical shift.

In Figure 1, we illustrate our predictions for these 4 groups, with the predictions associated with actual GDP levels designated with a triangle.¹¹ Figure 2 illustrates the comparable predictions from the Chang and Ram model. Obviously their predictions for the E and CA countries are captured in their high- and low-growth groups respectively.

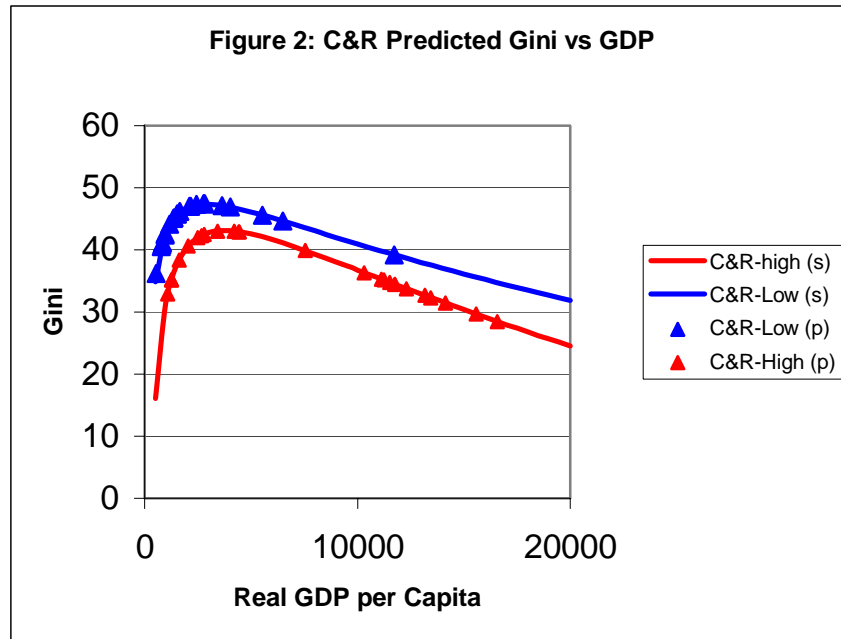


There are several points we would like to make regarding these figures and the associated results. First, by allowing for E and CA regional effects, our model predicts little difference in the inequality-GDP relationship between the 17 high- and 19 low-growth base countries. Based on our model results (Table 1), these differences are, *at best*, marginally significant from a statistical point of view.¹² Further, the differences

¹¹ Indeed the base group lines in Figure 1 are identical to the 2 lines graphed in Figure 5 of our Comment. Note also that our new Figure 1 is essentially Figure 1 in Chang and Ram's Response augmented with predictions of the 4 groups at all income levels.

¹² Prediction confidence intervals (derived later) are needed to assess the significance of these differences.

illustrated in the figure seem economically insignificant, particularly in comparison to those predicted by Chang and Ram, except, perhaps, at the very lowest GDP levels. In contrast to the Chang and Ram predictions, our model also does not predict that higher growth countries have lower inequality at all income levels for these base group countries.



Second, our predictions for CA and E *are* consistent with the story painted by Chang and Ram in the sense that the low-growth CA countries do have higher predicted levels of inequality relative to the high growth countries with similar GDP levels. Similarly, the high-growth E countries have much lower predicted levels of inequality compared to the (few) low-growth countries with similar GDP levels. However, the fact that we needed the 2 regional fixed effects, *in addition* to the growth dummy variables, indicates that it must be more than just differences in speed of growth that explain the positions of these CA and E curves. The increase in the individual t-statistics on the high-growth dummy variables in our model relative to those of Chang and Ram seems to provide evidence that the inclusion of the CA and E dummies has not detracted from our ability to capture differences between the growth groups and that there is something

else driving these results. Certainly, additional research is needed to understand what these regional variables are actually capturing.

We now shift to consider whether or not the data provide support for the Kuznets hypothesis (though this is a minor focus for Chang and Ram). Before doing so, it is instructive to look at the data we are attempting to model.

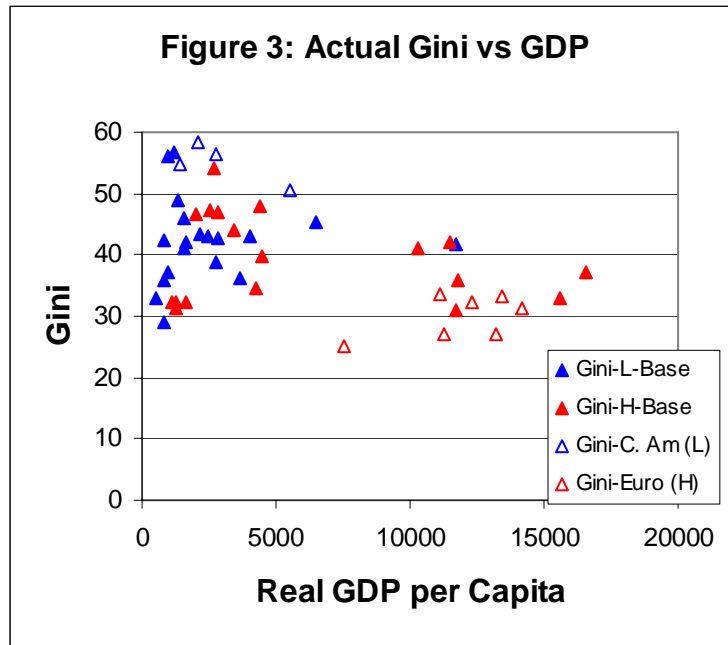


Figure 3 provides a cross-plot of the gini coefficient data versus GDP; we delineate the data by group to facilitate our discussion.¹³ There are several interesting things that can be learned from this graph—without using any statistics. First, the low-growth countries almost all have very low GDP levels, and at any given (low) level of GDP, there is a strikingly large (economically significant) range of gini coefficients. The high-growth countries, on the other hand, have a much larger range of GDP levels, and

¹³ Figure 6 in our Comment is the graph of these same data; unfortunately, in the Comment we used the same symbol (dot) for all the observations. As a consequence, we confused Chang and Ram and misled ourselves somewhat. By delineating the data by group we can see more clearly what is happening with these data (see text).

though there is also a range of gini coefficients at each level of GDP, it is not too hard to visualize that an inverted U (Kuznets-curve) *could* possibly fit these data.¹⁴ This simple “reading” of the data contradicts some of the findings of both the Chang and Ram and Edwards and McGuirk models. The Chang and Ram model seems to support the hypothesis for low-growth countries, while the evidence is much weaker for the high-growth countries. Chang and Ram attribute the weaker high-growth country evidence to collinearity, and conclude that both group results support the Kuznets hypothesis. Our model, on the other hand, seems to provide no evidence in its favor for either group. More evidence regarding the relevance of the Kuznets hypothesis is obtained in the Chang and Ram Response (point 4), when they estimate 2 separate regressions for the high- and low-growth groups. This time they find evidence that the Kuznets hypothesis holds for the high-growth group and not for the low group; these new findings are consistent with Figure 3. Given that the separate regressions estimated are less restricted versions of our statistically adequate model, we know that the t-statistics obtained should be distributed student’s-t as the model assumptions seem relevant for the data.¹⁵ The puzzle then is why these new separate regression results seem to differ from those of our statistically adequate model. After considering this question, we realized that the best, most direct, way to examine the Kuznets hypothesis for the two groups is to estimate the following (equivalent) form of the basic model:

$$INEQ_i = a_3 D_L + b_3 (LY_i \times D_L) + c_3 (LY_i^2 \times D_L) + a_{33} D_H + b_{33} (D_H \times LY_i) + c_{33} [D_H \times LY_i^2] + u_{3i},$$

where D_L (D_H) is a dummy variable equal to 1 for the low (high) growth countries and zero otherwise. Of course, our formulation includes the two regional dummies and the Chang and Ram formulation does not. The estimation results are presented in Table 3.

¹⁴ The left leg of the inverted U would begin at the (GDP, Gini) point of approximately (1200, 32) or so, reach a maximum at approximately (2500, 51) and then decrease (much more slowly than it rose) through the remaining red triangles.

¹⁵ Under point 5 in their Response, Chang and Ram estimate two more regression models and talk about evidence for the Kuznets hypothesis. Of course nothing should be concluded from these models until their statistical validity is assessed. If Chang and Ram really believe that the income-inequality relationship differs by growth rate, these regressions will, by definition, be misspecified and thus, should not be used to test any hypothesis.

Table 3

Edwards and McGuirk Model Results by Growth Group							
D _L	D _L × LY	D _L × LY ²	D _H	D _H × LY	D _H × LY ²	Euro	CA
-121.4 (-1.09)	42.10 (1.45)	-2.67 (-1.44)	-382.3* (-2.67)	101.12* (2.95)	-5.99* (-2.94)	-9.48 (-3.08)	11.57 (3.44)
t-statistics in parenthesis The number of observations is 48 The unadjusted R ² is 0.61 The adjusted R ² is 0.54 * P ≤ 0.10							
Chang and Ram Model Results by Growth Group							
D _L	D _L × LY	D _L × LY ²	D _H	D _H *LY	D _H *LY ²		
-212.77 (-1.62)	65.37* (1.92)	-4.11* (-1.88)	-406.6* (-2.36)	109.24* (2.64)	-6.64* (-2.71)		
t-statistics in parenthesis The number of observations is 48 The unadjusted R ² is 0.40 The adjusted R ² is 0.33 * P ≤ 0.10							

Indeed, our results are very close to those obtained from the separate regressions in the Chang and Ram Response. This is reassuring, as the only difference between the two specifications is that our pooled regression assumes that the conditional variance is the same for both groups. The Table 3 results indicate that the Edwards and McGuirk and Chang and Ram models both reveal evidence of a Kuznets inverted U curve for the high-growth country data. Thus, we were wrong in concluding that our results did not support the hypothesis for this group.¹⁶ Further, it is reassuring that our eyes did not deceive us as we looked at the data in Figure 3. Interestingly, we find more support for Kuznets than Chang and Ram for the high-growth countries—even with our regional dummies. Again, this suggests that by including the regional dummies we are not detracting from our ability to pick up differences by growth rate.

For the low-growth countries, the Edwards and McGuirk model indicates that there is little to no support for the Kuznets hypothesis, while the Chang and Ram model indicates support. In terms of the Kuznets hypothesis, this is the main difference between the two model formulations.

¹⁶ Thank goodness the examination of the Kuznets hypothesis was only a minor part of Chang and Ram's paper!

Once again, it is reassuring that our statistically adequate model results agree with what we see in Figure 3.

Before concluding the Kuznets discussion, we should note that the fact that we find evidence for Kuznets in the high-growth group and not in the low-growth group does not necessarily imply that the structure of the income-inequality profiles differ significantly by growth level. After looking at the data in figure 3, we believe that our model is unable to find a significant relationship between gini and GDP for the low-growth group, and a significant difference between the high- and low-growth countries, because of the tremendous “noise” in the low-growth data (the huge range of gini coefficients over a relatively small range of GDP).

To summarize, thus far we can say the following: Based on a model that was found to be statistically adequate, we find—using point estimate/predictions only—that high-growth countries do not necessarily have lower inequality levels for a given GDP than low-growth countries. Further, in contrast to what we claimed in our comment, there appears to be evidence of the Kuznets hypothesis for high-growth countries and not for low-growth countries. With the exception of the high-growth country evidence for the Kuznets hypothesis, these findings contradict those of Chang and Ram and corroborate the findings of our Comment.

Before concluding this Reply we want to address the final aspect of our complaints that Chang and Ram ‘force’ the data through a quadratic functional form and that their functional form/model is *ad hoc*. The first aspect of this criticism has been the focus of the Reply up to this point: you cannot just specify a model, fit the data, and run some t-tests—you must investigate the statistical validity of this model. The second aspect—the overall lack of significance of most of the variables in the models and their impact on the predictions—has not been addressed. What we were thinking is this: *if*, for example, the income terms are really not significantly different from zero, it is not really fair to use the (insignificant) parameter estimates to predict the relationship between gini and income and make a big deal out of these predictions. For example, suppose there really is no relationship between gini and income at all—the true parameters are zero. If we use a quadratic model, say, to capture the relationship between gini and income our parameter estimates will never be 0.0 (with probability 1)—and any predictions from this model will indicate some sort of quadratic relationship

between gini and income.¹⁷ Notice that like Chang and Ram, our model, though statistically adequate, also has very few significant variables; the parameter estimates on the two regional fixed effects are the only ones statistically different than 0 (at $\alpha=0.05$; see Table 1). While it is legitimate to use our model to show how Chang and Ram's conclusions, based on point predictions, regarding high- and low-growth countries, no longer seem to hold, we cannot conclude much else using these results until we consider confidence intervals on these predictions. The impact of the lack of significant variables and fit of the overall model will show up in these confidence intervals.

We begin this last investigation, by simply ignoring the issues raised above regarding the statistical adequacy of the Chang and Ram model. That is, we take the Chang and Ram results at face value and derive confidence intervals for the low and high-growth country predictions presented in Figure 1 and Table 3 of their *EDCC* paper. In Figure 4 we re-graph their predictions along with the associated 95% confidence intervals.¹⁸ As illustrated, the 95% confidence intervals are huge; the low-growth predictions are very near the middle of the high-growth confidence intervals and the high-growth predictions are very near the middle of the low-growth confidence intervals. The conclusions of Chang and Ram:

the estimates and the simulations show a statistically significant and quantitatively substantial structural difference between the two groups, and the high-growth scenario is characterized by lower inequality at all income levels. Moreover, the high-growth advantage seems particularly large at low-income levels, (Chang and Ram 2000, 795)

are not substantiated with their own empirical evidence.

¹⁷ Yes, we do know that one can get different shapes to this curve depending on the parameter signs and magnitude—we might be idiots but not total idiots—Chang and Ram misinterpreted our complaints.

¹⁸ The prediction confidence intervals were made using the usual formula for forecast error variance (see, for example, Davidson and MacKinnon 2004, 104).

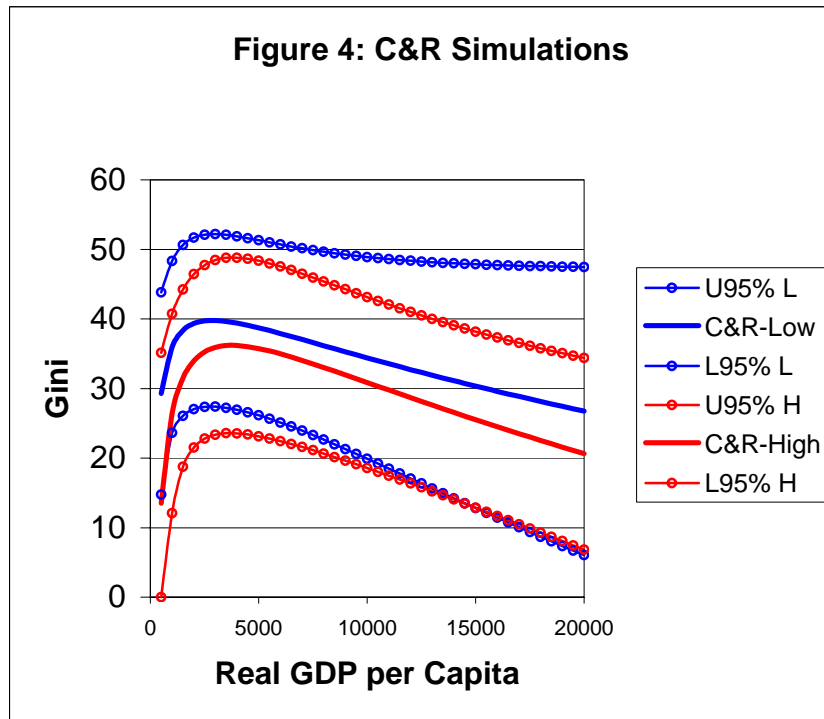


Figure 4 illustrates quite clearly that one cannot talk about differences between the high- and low-growth countries with any confidence. The intervals obtained with our model, though we may have more confidence in their statistical validity, are almost as large as those of Chang and Ram. Thus, based on these data and our models, we can only conclude that there is no significant difference between the predicted high- and low-growth country inequality-GDP profiles.

REFERENCES

- Chang, Jih Y., and Rati Ram.** 2000. Level of Development, Rate of Economic Growth, and Income Inequality. *Economic Development and Cultural Change* 48(4): 787-799.
- Chang, Jih Y., and Rati Ram.** 2004. Response to Edwards and McGuirk: Income Level, Economic Growth, and Inequality: Flawed Methodology and Inaccurate Inference. *Econ Journal Watch* 1(2): 235-243.
- Davidson, R. and James G. MacKinnon.** 2004. *Econometric Theory and Methods*. Oxford: Oxford University Press.
- Edwards, Jeff, and Anya McGuirk.** 2004. Kuznets Curveball: Missing the Regional Strike Zone. *Econ Journal Watch* 1(2): 222-234.
- Hendry, David F.** 1993. *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers.
- Leamer, E. E.** 1983. Let's Take the Con out of Econometrics. *American Economic Review* 73(1): 31-43.
- Leontief, W.** 1971. Theoretical Assumptions and Nonobserved Facts. *American Economic Review* 61(1): 1-7.
- McGuirk, Anya M., Paul J. Driscoll, and Jeff Alwang.** 1993. Misspecification Testing: A Comprehensive Approach. *American Journal of Agricultural Economics*. 75(4): 1044-1055.
- McGuirk, Anya M., Paul J. Driscoll, Jeffrey Alwang, and Huilin Huang.** 1995. System Misspecification Testing and Structural Change in the Demand for Meats. *Journal of Agricultural and Resource Economics*. 20(1): 1-21.
- Pagan, A.** 1984. Model Evaluation by Variable Addition. In *Econometrics and Quantitative Economics*, ed. D.F. Hendry and K.F. Wallis. Oxford: Blackwell, 103-133.
- Spanos, Aris.** 1986. *Statistical Foundations of Econometric Modelling*. Cambridge: Cambridge University Press.
- Spanos, Aris.** 1995. On Theory Testing in Econometrics: modeling with non-experimental data. *Journal of Econometrics* 67(1): 189-226.

- Spanos, Aris.** 1999. *Probability Theory and Statistical Inference: Econometric modeling with observational data*. Cambridge: Cambridge University Press.
- Spanos, Aris.** 2000. Revisiting Data Mining: 'Hunting' with or without a License. *Journal of Economic Methodology* 7(2): 231-264.
- Tomek, William G.** 1993. Confirmation and Replication in Empirical Econometrics: A Step Toward Improved Scholarship. *American Journal of Agricultural Economics* 75(special 75th Anniversary Issue): 6-14.
- Tomek, William G.** 1997. *Structural Econometric Models: Past and Future*. Havlicek Memorial Lecture, Department of Agricultural Economics, Ohio State University.



*Econ Journal Watch,
Volume 1, Number 2,
August 2004, pp 260-271.*

Statist Quo Bias

DANIEL B. KLEIN*

A COMMENT ON: RICHARD H. THALER AND CASS R. SUNSTEIN. 2003. LIBERTARIAN PATERNALISM. *AMERICAN ECONOMIC REVIEW (AEA PAPERS AND PROCEEDINGS)* 93(2): 175-79.

[ABSTRACT, KEYWORDS, JEL CODES](#)

LIBERTARIANISM (OR CLASSICAL LIBERALISM) POSES SERIOUS intellectual challenges to conventional political sensibilities and institutions. Libertarianism says that occupational licensing, central banking, the postal monopoly, drug prohibition, and government schooling shouldn't exist. In establishment forums, people rarely address directly the libertarian challenge, but it lurks in the background. Both sympathizers and opponents speak and behave in ways that are significantly influenced by the libertarian challenge.

When establishment actors do openly address libertarianism the performance is often curious. We hear that libertarians favor the cash nexus, perfect competition, social atomism, social Darwinism, Republicans, and profits for McDonalds and GM. We also hear that libertarians assume that people are rational.

Consider Richard Thaler and Cass Sunstein's "Libertarian Paternalism," a brief, invited conference paper included in the May 2003 issue of the *American Economic Review (Papers and Proceedings)*. They begin:

* Santa Clara University, Santa Clara, CA.

Many economists are libertarians and consider the term “paternalism” to be derogatory. Most would think that the phrase libertarian paternalism is an oxymoron. The modest goal of this essay is to encourage economists to rethink their views on paternalism. (Thaler and Sunstein 2003, 175)

The authors describe libertarian paternalism as the taking of actions in which “no coercion is involved” and “with the goal of influencing the choices of affected parties in a way that will make those parties better off” (175). One of their primary examples is the placement of dessert in a cafeteria:

Consider the problem facing the director of a company cafeteria who discovers that the order in which food is arranged influences the choices people make. . . . Putting the fruit before the desserts is a fairly mild intervention. A more intrusive step would be to place the desserts in another location altogether, so that diners have to get up and get a dessert after they have finished the rest of their meal. This step raises the transaction costs of eating dessert, and according to standard economic analysis the proposal is unattractive: it seems to make dessert-eaters worse off and no one better off. But once self-control costs are incorporated, we can see that some diners would prefer this arrangement, namely, those who would eat a dessert if it were put in front of them but would resist temptation if given a little help. (Thaler and Sunstein 2003, 175, 177)

I gather that, in my job as a teacher, I practice libertarian paternalism when I make certain readings required rather than optional. Thaler and Sunstein do not discuss policy issues that do involve coercion like drug prohibition, anti-smoking laws, and Social Security levies. The paper is actually about things like dessert placement.

But it seems odd to drag the terms “libertarian” and “paternalism” into matters like dessert placement. What they speak of could be more accurately, if less provocatively, addressed using such terms as benevolence, discipline, delegation, propriety, help, cooperation, and so on. The terms “libertarian” and “paternalism” reside naturally in discussions of *political* affairs and particularly in issues involving coercive government policy. The *paternalism* entry of *The Blackwell Encyclopaedia of Political Thought* begins: “In

modern use the term usually refers to those laws and public policies *which restrict the freedom* of persons in order that their interests may be better served” (Weale 1991, 367-68, italics added). It seems to me that Thaler and Sunstein pull “libertarian” and “paternalism” out of their normal context just to create an oxymoronic gimmick. If Thaler and Sunstein were to proceed with this kind of gimmick, we could anticipate the following papers:

“Libertarian Socialism”: A paper pointing out that sometimes patrons at a Chinese restaurant may decide to order a number of dishes and share them in common.

“Libertarian Communism”: A paper pointing out that sometimes adults such as the Jesuits at Santa Clara University live and eat and work together within a nexus of private property and voluntary association.

“Libertarian *Dirigisme*”: A paper pointing out that executives centrally plan the basic skeleton of activities at the worksite.

“Libertarian Interventionism”: A paper pointing out that condominium associations involve restrictions on members’ activities.

“Libertarian Repression”: A paper pointing out that, as described in Schelling (1984), individuals control their own behavior by repressing, subduing, or annihilating certain subordinate selves.

To motivate “Libertarian Paternalism,” Thaler and Sunstein say that “the anti-paternalistic fervor expressed by many economists” is based on a false assumption:

The false assumption is that people always (usually?) make choices that are in their best interest. (Thaler and Sunstein 2003, 175)

Thus, the authors suggest that libertarian sensibilities are based on a belief that, free of coercion, man acts in his (expected) best interest. Thaler and Sunstein cite the Surgeon General’s report on obesity, academic studies of retirement-savings decisions, and various experimental studies to reject the idea that man always acts in his best interest.

Thaler and Sunstein do not quote or mention a single libertarian thinker.¹ Since the eighteenth century, the classical liberal tradition has, indeed, had a rationalistic strain that posits man as a highly integrated will that acts purposively through time. That tradition has sometimes badly over-emphasized such ideas, even by claiming foundationalist or axiomatic status for them.² But this hyper-rationalistic strain, which corners itself into saying that we cannot make any ethically relevant distinction between the consumer surplus from crack-cocaine and the consumer surplus from orange juice, is really just a strain. Virtually all classical liberals since the eighteenth century would readily say otherwise.

Many works in the classical liberal tradition have started with the idea that man is a creature highly prone to developing bad habits and prejudices, to feeling remorse, and to engaging in self-reform, that man is a creature trying to figure out who he is and should be. Many works argue that liberty accords people ownership of their story, including their errors and vices, and thereby allows them to learn the contours of action, experience, and consequence; that liberty makes experience more meaningful; that liberty leaves the one who is most naturally concerned to act on his knowledge and to negotiate the limitations of his knowledge; that liberty engenders the institutions, practices, and attitudes that heighten the individual's discovery of the worthiness of actions and of ways to refine his habits and control his impulses; that liberty affirms the dignity³ of the

¹ In the longer version of the paper (Sunstein and Thaler 2003), they bring up Mill in a footnote and explain why autonomy issues are not considered in the paper.

² In the past century, these views are represented notably by the praxeology tradition of Ludwig von Mises and especially the libertarian polymath Murray Rothbard. Mises (1966, 11-29) sets out purposive human action as a scientific foundation, but the formulation leaves doors wide open for saying that a lot of human behavior is not "purposive action." The subjunctive condition—in so far as man is an integrated will acting through time—is effectively dropped by Rothbard (1956 and 1962), who blithely goes on to make claims that seem to say that in all cases non-coerced human behavior advances that person's (expected) interest (e.g., 1956, 250). He and his followers have indeed used such 100-percent claims in erecting moral axioms that purportedly justify libertarianism, and have made statements to the effect that there is no ethically relevant difference between the consumer surplus from crack-cocaine and the consumer surplus from orange juice. For example, Walter Block (1976, 42) writes that, in reducing the market price, "the much reviled drug dealer . . . must be considered the heroic figure." Incidentally, Rothbard (1956, 225) highlights Irving Fisher and Frank A. Fetter as early American economists who maintained that man always acts so as to advance his (expected) interest.

³ I encourage experimentalists and model-builders to try to devise experiments and models to help define, measure, and operationalize dignity. However, we should recognize that their inability to do so would not mean that dignity is not important or meaningful. Some keys are not under those lamp-posts.

human being. Older works that develop such themes include John Milton (1644), Wilhelm von Humboldt ([1791]), T. B. Macaulay (various works), John Stuart Mill ([1859]), Lysander Spooner ([1875]), Auberon Herbert (1978), Albert Jay Nock ([1924]), Frank H. Knight (1951, esp. 13, 17-18), and Friedrich Hayek (1948a and 1967a). More recent libertarian writings in this vein include Buchanan (1979), Szasz (1963), Murray (1988), Klein (1992, 1997).⁴ On libertarianism not being *100 percent* opposed to coercive paternalism, see Hospers (1980).

Like the idea that man maximizes his material payoffs, the idea that he always chooses in his best interest in a strong sense is a canard. In a polar opposite sense, he never chooses in his best interest. Both poles are silly abstractions, and all the really meaningful discourse lies between them, recognizing that the expression “best interest” is probably not in our better interest.

Yet some people suppose that other people—who are rarely ever identified—actually believe the idea that man always chooses in his best interest. The canard probably arises from the journal-article practice among Neoclassical economists of treating the human being as a mathematical apparatus. That practice is of a piece with the scientific intellectual tendencies that Hayek devoted his life to combating. In his essay “Individualism: True and False,” Hayek propounds what he calls true individualism, “which regards man not as highly rational and intelligent but as a very irrational and fallible being, whose individual errors are corrected only in the course of a social process, and which aims at making the best of a very imperfect material” (1948a, 9). Hayek explains that the Scottish enlightenment philosophers did not view human behavior as optimization: “It would be nearer the truth to say that in their view man was by nature lazy and indolent, improvident and wasteful, and that it was only by the force of circumstances that he could be made to behave economically or carefully to adjust his means to his ends” (11). Hayek adds: “[T]he famous presumption that each man knows his interests best . . . is neither plausible nor necessary for the [true] individualist’s [policy] conclusions” (15).

Ronald Coase concurs with Hayek’s reading of the Scottish enlightenment and Hayek’s assessment of the libertarian position. Coase concluded his essay, “Adam Smith’s View of Man” ([1976]) with the following sentences:

⁴ Of course, there are other arguments. One is the constitutional focal point. Who is to say what people’s “best interest” really is? Once the government coerced people for their own good, the “protections” multiplied, fulfilling Tocqueville’s prophesy of paternalistic despotism (see the last seven chapters of *Democracy in America*). A firm stand would consecrate a barrier against the Pandora’s Box of coercive “help.”

Smith would not have thought it sensible to treat man as a rational utility-maximiser. He thinks of man as he actually is: dominated, it is true, by self-love but not without some concern for others, able to reason but not necessarily in such a way as to reach the right conclusion, seeing the outcomes of his actions but through a veil of self-delusion. No doubt modern psychologists have added a great deal, some of it correct, to this eighteenth-century view of human nature. But if one is willing to accept Adam Smith's view of man as containing, if not the whole truth, at least a large part of it, realisation that his thought has a much broader foundation than is commonly assumed makes his argument for economic freedom more powerful and his conclusions more persuasive. (Coase [1976], 116)

Thus, in pinning hyper-rationalism on libertarianism, Thaler and Sunstein neglect that the premier libertarian theorists Adam Smith and Friedrich Hayek made human ignorance, error, and frailty the cornerstone of their thought.

In twentieth-century economics, a sustained effort to make a place in economics for error and self-reproach has come from those working in the Hayekian tradition. They incorporate ideas of error and misinterpretation into the economic theory of the market process⁵ and into the theory of government behavior.⁶ Those libertarians have worked to incorporate human frailty into economic ideas.

⁵ E.g., Kirzner (1979). Kirzner's project has been to make entrepreneurship a working part of economic theory. In his system, the theoretical inverse of entrepreneurship is error. Kirzner sometimes associates self-reproach, remorse or regret with error. For example, he uses the term reproach seven times in three pages (1979, 128-30), and writes of blaming oneself for having erred as essential to error (1985, 56). Klein (1999) employs a multiple-self framework and explicitly makes self-reproach (or would-be self-reproach) the essential feature of error (65-69).

⁶ Hayek ([1967b], 184), Ikeda (1997 and 2003), Boettke and Lopez (2002), and Klein (1994). In a related vein, the libertarian economist Bryan Caplan has parsed "rational ignorance" in politics, to distinguish "rational irrationality," the idea being that many people are led by incentives not only to ignore information but also to adopt foolish political interpretations (Caplan 2001). That Caplan would describe mass foolishness as "rational," however, as well as leave us wondering about the seemingly implied distinction between "rational irrationality" and "irrational irrationality," leaves his work a curious mixture of traditions.

And then there is the work on how the demand for quality and safety *assurance* creates opportunities to profit (in a broad sense) by supplying assurance.⁷ Far from assuming that people never err, this literature takes as its starting point people's vulnerabilities and frailties, and their demand to find agents, knowers, and middlemen whom they can trust to help them with decisions.

So the immediate problem in "Libertarian Paternalism" is the fatuity of its declared motivation. Very few libertarians have maintained what Thaler and Sunstein suggest they maintain, and, indeed, many of the leading theorists have worked with ideas in line with what Thaler and Sunstein have to say about man's nature. Thaler and Sunstein are forcing an open door.

But a careful reading of Thaler and Sunstein's paper detects a deeper problem that tells the real story.

Above I provided a quotation about dessert placement in the company cafeteria. Immediately following the sentences quoted above, Thaler and Sunstein continue:

To the extent that the dessert location is not hard to find, and no choice is forbidden, this approach meets libertarian muster. (Thaler and Sunstein 2003, 177)

Thaler and Sunstein make no elaboration. They seem to be saying that if the cafeteria director were to put the desserts in a hard-to-find place, or to eliminate desserts altogether, then this form of paternalism would no longer meet libertarian muster. Elsewhere Thaler and Sunstein repeatedly associate libertarianism with the absence of coercion. In the quoted sentence they are either departing from that association or they mean to suggest that it would be coercive to make hard to find or eliminate the desserts. That suggestion is of course at variance with the commonly understood distinction between voluntary and coercive action. Unless there is a pretty clear obligation to have desserts conveniently available, the dessert policy, whatever it be, is not a matter of coercion—just as not displaying fresh flowers would not be coercion. If Thaler and Sunstein believe otherwise, they must be employing some other, idiosyncratic distinction between coercive and noncoercive actions that would hold the absence or removal of dessert to be a form of coercion. I doubt that they could make that distinction at all coherent.

⁷ See the papers and bibliography in Klein (1997 and 2002), and Hayek (1948b, 97).

It is likely that the authors simply wish to reject the distinction between voluntary and coercive action upon which the very idea of libertarianism is based. For Sunstein's part, he makes it amply clear in his book with Stephen Holmes (Holmes and Sunstein 1999) that he rejects the formulations that are basic to libertarian ideas. Holmes and Sunstein hold that the nation-state is a voluntary organization (210), that the rules issuing from the government are voluntarily entered into like a corporation's bylaws (217), that libertarian ideas of liberty are just "fairy tales" (216), and that, as the title of their first chapter declares, "All Rights Are Positive." In their social democrat worldview, all things are owned, fundamentally and ultimately, by the government, and any decentralized exercise of property rights or contract is undertaken by its authorized delegation. "Private property [is] a creation of state action," (66) "laws [enable property holders] to acquire and hold what is 'theirs'" (230). The quotation marks on "theirs" are there to tell us: That car you drive everyday is really the property of the state; we just let you think it's yours.

With the Sunstein of Holmes and Sunstein (1999) in mind, it is no great surprise that in discussing dessert policy Thaler and Sunstein muddy the focal distinction between voluntary and coercive action. They can use "libertarian" and "paternalism" in discussing dessert policy because they reject the key distinctions between governmental and nongovernmental affairs. My reading of what is really going on in "Libertarian Paternalism" is that they use the term libertarian disingenuously. I believe they seek to dispose of libertarianism by speaking in a way that upsets libertarianism's key semantics, but without making clear that that is what they are doing, much less challenging those semantics directly.

During the last third of the nineteenth century in Britain the terms "freedom," "liberty," and "liberal" were quite deliberately transformed, under the influence of Hegelian idealism, by thinkers like T.H. Green, A. Toynbee, B. Bosanquet, D.G. Ritchie, J.A. Hobson, and L.T. Hobhouse and came to mean something at odds with what they had meant for Mill, Spencer, Bright, Gladstone, and Morley. Once subverted, classical liberalism is now reinvigorated by libertarianism. Now we see efforts to muddy the term "libertarian." A Federal Reserve Bank conference in 2003 arranged for Thaler and Sunstein to present a longer version of their paper and for Alicia H. Munnell of Boston College to comment. Professor Munnell, who has lately coauthored a Brookings book on retirement-savings plans (Munnell and Sunden 2003), entitles her remarks "A Non-libertarian Paternalist's Reaction." She strongly endorses Thaler and Sunstein, adds evidence of people acting ignorantly or foolishly in their

retirement-savings decisions, and explicitly propounds both the paternalist and redistributive rationales for mandatory Social Security levies. (Incidentally, she refers to mandatory Social Security levies as “*real* paternalism” (her emphasis, 7).) She ends her comment: “[Thaler and Sunstein’s] ‘libertarian paternalism’ approach may serve as a bridge between the libertarians on the right and the New Deal traditionalists on the left, and that bridge could help rebuild the national consensus on social and economic policy.” I am sure that this “bridge” would never approach the libertarian shore—it is of course nonsensical to speak of bridging libertarianism and the New Deal. Rather, the establishment social democrats would like to keep real libertarians away from the deliberations yet declare that libertarianism has been represented in the new “consensus.”

What, then, are we to make of “Libertarian Paternalism”? I suggest that Thaler and Sunstein suffer from deep biases. They sense the challenge of libertarian ideas, but react in a way that is anchored in the political status quo and their own commitments to certain ideological ideas and values.

REFERENCES

- Block, Walter.** 1976. *Defending the Undefendable*. New York: Fleet Press.
- Boettke, Peter J. and Edward J. Lopez.** 2002. Austrian Economics and Public Choice. *Review of Austrian Economics* 15(2/3): 111-119.
- Buchanan, James M.** 1979. Natural and Artfactual Man. In his *What Should Economists Do?* Indianapolis, Ind.: Liberty Fund, 93-112.
- Caplan, Bryan.** 2001. Rational Ignorance vs. Rational Irrationality. *Kyklos* 54(1): 3-26.
- Coase, Ronald H.** ([1976]) 1994. Adam Smith’s View of Man. In his *Essays on Economics and Economists*. Chicago: University of Chicago Press, 95-116.
- Hayek, Friedrich A.** 1948a. Individualism: True and False. Chapter 1 of his *Individualism and Economic Order*. Chicago: University of Chicago Press, 1-32.
- Hayek, Friedrich A.** 1948b. The Meaning of Competition. Chapter 5 of his *Individualism and Economic Order*. Chicago: University of Chicago Press, 92-106.

- Hayek, Friedrich A.** 1967a. The Moral Element in Free Enterprise. Chapter 16 in his *Studies in Philosophy, Politics and Economics*. Chicago: University of Chicago Press, 229-236.
- Hayek, Friedrich A.** 1967b. The Intellectuals and Socialism. Chapter 12 in his *Studies in Philosophy, Politics and Economics*. Chicago: University of Chicago Press, 178-194.
- Herbert, Auberon.** 1978. *The Right and Wrong of Compulsion by the State and other Essays*, ed. E. Mack. Indianapolis: Liberty Fund.
- Holmes, Stephen and Cass R. Sunstein.** 1999. *The Cost of Rights: Why Liberty Depends on Taxes*. New York: Norton.
- Hospers, John.** 1980. Libertarianism and Legal Paternalism. *Journal of Libertarian Studies* 4(3): 255-265.
- Humboldt, Wilhelm von.** [1791] 1969. *The Limits of State Action*, ed. and trans. J.W. Burrow. Cambridge, UK: Cambridge University Press.
- Ikeda, Sanford.** 1997. *Dynamics of the Mixed Economy: Toward a Theory of Interventionism*. London: Routledge.
- Ikeda, Sanford.** 2003. How Compatible are Public Choice and Austrian Political Economy? *Review of Austrian Economics* 16(1): 63-75.
- Kirzner, Israel M.** 1979. Economics and Error. Chapter 8 in his *Perception, Opportunity, and Profit: Studies in the Theory of Entrepreneurship*. Chicago: University of Chicago Press, 120-136.
- Kirzner, Israel M.** 1985. *Discovery and the Capitalist Process*. Chicago: University of Chicago Press.
- Klein, Daniel B.** 1992. Go Ahead and Let Him Try: A Plea for Economic Laissez-Faire. *Inquiry: An Interdisciplinary Journal of Philosophy* 35(1): 3-20.
- Klein, Daniel B.** 1994. If Government Is So Villainous, How Come Government Officials Don't Seem Like Villains? *Economics & Philosophy* 10: 91-106.
- Klein, Daniel B.** ed. 1997a. *Reputation: Studies in the Voluntary Elicitation of Good Conduct*. Ann Arbor: University of Michigan Press.
- Klein, Daniel B.** 1997b. Liberty, Dignity, and Responsibility: The Moral Triad of a Good Society. *The Independent Review: A Journal of Political Economy* 1(3): 325-351.

- Klein, Daniel B.** 1999. Discovery and the Deepself. *Review of Austrian Economics* 11: 47-76.
- Klein, Daniel B.** 2002. The Demand for and Supply of Assurance. In *Market Failure or Success: The New Debate*, ed. T. Cowen and E. Crampton. Cheltenham, UK: Edward Elgar, 172-192.
- Knight, Frank H.** 1951. The Role of Principles in Economics and Politics. *American Economic Review* 41(1): 1-29
- Mill, John Stuart.** [1859] 1961. On Liberty. Reprinted in *The Utilitarians*. New York: Doubleday.
- Milton, John.** [1644] 1949. Aeropagitica: A Speech for the Liberty of Unlicensed Printing, to the Parliament of England. In *The Portable Milton*, ed. D. Bush. New York: The Viking Press, 151-205.
- Mises, Ludwig von.** 1966. *Human Action: A Treatise on Economics*. New Haven: Yale University Press.
- Munnell, Alicia H.** 2003. A Non-libertarian Paternalist's Reaction to 'Libertarian Paternalism is Not an Oxymoron.' Comment delivered at the Federal Reserve Bank of Boston's 48th Annual Conference, Chatham MA, June 8-10.
- Munnell, Alicia H. and Annika Sunden.** 2003. *Coming Up Short: The Challenge of 401(k) Plans*. Washington, DC: The Brookings Press.
- Murray, Charles.** 1988. *In Pursuit: Of Happiness and Good Government*. New York: Simon and Schuster.
- Nock, Albert Jay.** [1924] 1991. On Doing the Right Thing. Reprinted in *The State of the Union: Essays in Social Criticism by Albert Jay Nock*, ed. C.H. Hamilton. Indianapolis: Liberty Fund, 317-325.
- Rothbard, Murray N.** 1956. Toward a Reconstruction of Utility and Welfare Economics. In *On Freedom and Free Enterprise: Essays in Honor of Ludwig von Mises*, ed. M. Sennholz. Princeton, NJ: D. Van Nostrand, 224-262.
- Rothbard, Murray N.** 1962. *Man, Economy, and State*. Princeton, NJ: D. Van Nostrand.
- Schelling, Thomas C.** 1984. The Intimate Contest for Self-Command. In his *Choice and Consequence*. Cambridge, MA: Harvard University Press, 57-82.

- Spooner, Lysander.** [1875] 1992. *Vices Are Not Crimes: A Vindication of Moral Liberty*. Reprinted in *The Lysander Spooner Reader*. San Francisco: Fox & Wilkes, 25-52.
- Sunstein, Cass R. and Richard H. Thaler.** 2003. "Libertarian Paternalism is Not an Oxymoron." *University of Chicago Law Review* 70(4), Fall: 1159-1202.
- Szasz, Thomas.** 1963. *Law, Liberty, and Psychiatry*. New York: Collier Books.
- Thaler, Richard H. and Cass R. Sunstein.** 2003. Libertarian Paternalism. *American Economic Review (Papers and Proceedings)* 93(2): 175-179.
- Weale, Albert.** 1991. Paternalism. *The Blackwell Encyclopaedia of Political Thought*, ed. D. Miller et al. Oxford: Basil Blackwell, 367-368.

ABOUT THE AUTHOR



Daniel B. Klein is associate professor of economics at Santa Clara University, associate fellow at the Ratio Institute (Stockholm), and editor of *Econ Journal Watch*. He teaches courses in economic principles and policy, and has received several teaching commendations. He is the director of the Civil Society Institute at Santa Clara University. In 2004 The Fund for the Study of Spontaneous Orders awarded him a Spontaneous Order Prize for his research on dignity, the "deepself," and the demand and supply of assurance. His email address is dklein@scu.edu.

[GO TO RESPONSE BY CASS SUNSTEIN](#)



Response to Klein

CASS R. SUNSTEIN*

LIBERTARIAN PATERNALISTS INSIST ON FREEDOM OF CHOICE; they do not want to foreclose options (Sunstein and Thaler 2003). At the same time, they believe that, in many contexts, planners cannot help influencing choosers, even if they aspire to neutrality. If influences are inevitable, shouldn't planning be undertaken with some awareness of its effects? Libertarian paternalists believe that the answer is clearly yes—and that planners in the private and public sectors should explore approaches that lead people toward welfare-increasing outcomes while also leaving them free to choose.

As far as I can see, Daniel Klein (2004) has no quarrel with libertarian paternalism. He suggests that Richard Thaler and I are focused on cafeteria design rather than policy issues. But as Klein knows, we refer to many policy issues, including employee savings, consumer protection, social security reform, and employment discrimination. Klein claims that we must be making some "idiosyncratic" distinction between voluntary and coercive action. But it is not idiosyncratic to distinguish between approaches that respect freedom of choice and those that do not. Klein suspects "that the authors simply wish to reject the distinction between voluntary and coercive action upon which the very idea of libertarianism is based" (Klein 2004, 267). But that very distinction is pivotal to our argument, which opposes libertarian paternalism to nonlibertarian varieties, and which endorses the former over the latter.

Klein protests that many libertarians emphasize human fallibility. Of course, he is correct. Among our primary targets are libertarian economists

* Law School and Department of Political Science, University of Chicago.

who emphasize human rationality—there are many of them, and they have a great deal of influence. But our argument is strengthened, not undermined, by the suggestion from Hayek and Coase, among others, that people are not perfect calculating machines. In short, Klein seems to me to have nothing to say against libertarian paternalism.

REFERENCE

Klein, Daniel B. 2004. Statist Quo Bias. *Econ Journal Watch* 1(2): 260-271.

Sunstein, Cass R., and Richard Thaler. 2003. Libertarian Paternalism Is Not an Oxymoron. *University of Chicago Law Review* 70: 1159-1202.

ABOUT THE AUTHOR



Cass R. Sunstein is Karl N. Llewellyn Distinguished Service Professor at the University of Chicago, Law School and Department of Political Science. A member of the American Academy of Arts and Sciences, he has taught at Columbia and Harvard and served as a member of the President's Advisory Committee on the Public Service Obligations of Digital Television Broadcasters. Professor Sunstein is author of a number of books, including *Legal Reasoning and Political Conflict* (1996), *Free Markets and Social Justice* (1997), *Risk and Reason* (2002), *Punitive Damages: How Juries Decide* (2002), *Why Societies Need Dissent* (2003), and *The Second Bill of Rights* (2004).

[GO TO REPLY BY DANIEL KLEIN](#)



Reply to Sunstein

DANIEL B. KLEIN*

TERMS CARRY IDEAS. WHEN TERMS ARE MUDDIED, THE IDEAS they carry suffer.

The cogent idea of liberty or freedom—the idea as understood by everyone in the tradition of Locke, Hume, Smith, the American Founders, the Abolitionists, and so on—is that each is free to do with his property (including his person) and to enter into agreements as he sees fit, provided he does not tread on others’ property (including contracted rights). Social Security taxes and consumer protection laws (with concomitant enforcement) tread on property and freedom of contract, and hence on freedom; they are coercive. If anyone other than the government tried it, everyone would cry bloody murder. Imagine if your neighbor threatened his own violence against you for employing someone at a wage rate he deemed to be too low. Everyone would recognize it as coercion. In contrast, the rules at the Weight Watchers club do not tread on anyone’s property; they are not coercive. Libertarianism is the political persuasion that government coercion should be vastly reduced.

Yet Thaler and Sunstein (2003: 177) imply that making the dessert at a cafeteria hard to find would not “meet libertarian muster.” That’s wrong. Making the desserts hard to find *would* “meet libertarian muster.” Thaler and Sunstein warp “libertarian,” partly by identifying it with “rationality.”

They put “libertarian” to work in a context in which it simply does not belong. You might watch Jay Leno, or you might watch David Letterman. Neither activity would fail to “meet libertarian muster.” Saying that watching Jay Leno meets libertarian muster is like say that birds flying north meets libertarian muster. Libertarianism does not speak to birds

* Department of Economics, Santa Clara University.

flying north v. flying some other direction, nor Leno v. Letterman, nor hard-to-find desserts v. easy-to-find desserts.

The term “paternalism” is only seriously used in its political sense. An encyclopaedia defines “paternalism” as “public policies which restrict the freedom of persons in order that their interests may be better served” (Weale 1991, 367). By calling both government coercion and cafeteria decisions “paternalism,” Thaler and Sunstein disparage the notion that one is a restriction of freedom while the other is not. As I emphasized in my Comment, denying the natural, classical idea of freedom is openly propounded in *The Cost of Rights* (Holmes and Sunstein 1999).

Sunstein’s Response elides my challenge to the “libertarian muster” sentence and persists in muddying terms:

But as Klein knows, we refer to many policy issues, including employee savings, consumer protection, social security reform, and employment discrimination. Klein claims that we must be making some "idiosyncratic" distinction between voluntary and coercive action. But it is not idiosyncratic to distinguish between approaches that respect freedom of choice and those that do not. (Sunstein 2004, 272)

Again he lumps together voluntarily entered rules and coercive government rules as instances of rules that do not “respect freedom of choice.”

And in the Response, Sunstein writes that the Thaler-Sunstein approach “opposes libertarian paternalism to nonlibertarian varieties, and . . . endorses the former over the latter” (272). Sunstein is playing nice, but really is dodging the criticism. By “nonlibertarian varieties” Sunstein here means, not drug prohibition and the like, but cases like the cafeteria that eliminates dessert options that people would otherwise select. This “non-libertarian paternalism” is discussed by Sunstein and Thaler in their *University of Chicago Law Review* article at pages 1185-86.

Suggesting that the cafeteria or company policy violates freedom weakens the very idea of freedom, and thus weakens the claim that Social Security levies etc. violate freedom. That is precisely Sunstein’s strategy in *The Cost of Rights*. Page 210 of that book says that people pay Social Security taxes voluntarily. Likewise, in their *University of Chicago Law Review* article, Sunstein and Thaler (2003: 1188) gives as an example of libertarian paternalism mandatory “cooling off” laws (e.g., for door-to-door sales)—a

clear case of coercion, as such laws threaten aggression against vendors and would-be vendors who have not tread on anyone's property.

By adding "libertarian" to "paternalism," Thaler and Sunstein make it seem like they affirm the libertarian distinction between voluntary and coercive. They use "libertarian"—a political call to depoliticize—to counteract the political meaning of "paternalism." But "libertarian paternalism" is like "voluntary coercion." Sunstein and Thaler concoct new meanings—of "voluntary" and "coercion" as much as of "libertarian" and "paternalism."

The bastardization "libertarian paternalism" upsets people's understanding of those terms in their critical function: highlighting the coercive nature of government intervention. In espousing government intervention, Sunstein would advance honest discourse by admitting that it is coercive and defending it as such.

REFERENCES

- Holmes, Stephen and Cass R. Sunstein.** 1999. *The Cost of Rights: Why Liberty Depends on Taxes*. New York: Norton.
- Sunstein, Cass R.** 2004. Response to Klein. *Econ Journal Watch* 1(2): 272-273.
- Sunstein, Cass R. and Richard H. Thaler.** 2003. Libertarian Paternalism is Not an Oxymoron. *University of Chicago Law Review* 70(4): 1159-1202.
- Thaler, Richard H. and Cass R. Sunstein.** 2003. Libertarian Paternalism. *American Economic Review (Papers and Proceedings)* 93(2): 175-179.
- Weale, Albert.** 1991. Paternalism. *The Blackwell Encyclopaedia of Political Thought*, ed. D. Miller et al. Oxford: Basil Blackwell, 367-368.



*Econ Journal Watch,
Volume 1, Number 2,
August 2004, pp 277-278.*

Rejoinder to De Alessi

Michael Kremer*

CONTINUATION OF EXCHANGE BETWEEN MICHAEL DE ALESSI
AND MICHAEL KREMER FROM THE APRIL 2004 ISSUE OF *EJW*.

[De Alessi's Comment on Kremer and Morcom](#)

[Kremer's Response](#)

[De Alessi's Reply](#)

I WOULD JUST LIKE TO NOTE A FEW THINGS IN RESPONSE TO
De Alessi's reply.

First, he criticizes the paper as not saying anything about the real world. Our argument is that in the real world prices of storable goods like ivory are influenced by expectations of the future. De Alessi presents no evidence that this is incorrect, rather he simply asserts that the paper is not about the real world.

Second, he argues that loss of habitat is a greater threat to endangered species than is poaching. That is correct, but irrelevant. As we note in the paper, many species are also threatened by poaching. If one is trying to write a comprehensive book about endangered species then one should spend more time on the most important cause of endangerment than on the second or third most important cause. In a journal article, however, it is perfectly appropriate to write an article that seeks to add knowledge on what may be the second or third most important cause of endangerment. If one were to take De Alessi's argument further we should not write any

* Department of Economics, Harvard University.

MICHAEL KREMER

articles about endangered species at all because endangered species might not even be the main environmental problem that our society is facing.

De Alessi argues that we are advocating devaluing species. That is incorrect. We are in favor of creating a situation in which species would be valued as long as they do not become extinct but the value of poached body parts would decline if the species became extinct. This would improve incentives.

Our assumption about open access is a positive assumption about the situation under which many elephants currently live. It is not a normative one. Factually, many elephants in Africa are not fully protected against poaching.

ABOUT THE AUTHOR



Michael Kremer is Gates Professor of Developing Societies in the Department of Economics at Harvard University. His research interests include development economics and macroeconomics.

His email address is: mkremer@fas.harvard.edu.



*Econ Journal Watch,
Volume 1, Number 2,
August 2004, pp 279-305.*

DO ECONOMISTS REACH A CONCLUSION?

MEDICAL LICENSING

Licensing Doctors: Do Economists Agree?

SHIRLEY SVORNY*

[Abstract, Keywords, JEL Codes](#)

IN THE UNITED STATES, STATE LEVEL BOARDS DICTATE RULES for physician licensure and discipline.¹ Would-be physicians must complete an approved medical training program and pass a standardized test. Scope-of-practice laws prohibit other health professionals from offering similar services. Given the resources involved in licensing doctors, taxpayers might be surprised to learn that the link between licensing and service quality is tenuous at best. In fact, economists who have examined the market for physician services generally view medical licensing as a constraint on the efficient combination of inputs and a drag on innovations in health care and medical education.

LICENSING AND PHYSICIAN QUALITY

Shapiro (1986) argued that the assumption of complementarity between human capital investment and physician service quality is “critical to the efficacy of licensing” (844). Yet critics of licensure reject the idea that approved programs of education and training assure competent care.

* Department of Economics, California State University, Northridge.

¹ The Federation of State Medical Boards documents state licensing requirements, as well as characteristics of state boards and regulations regarding discipline (www.fsmb.org).

Currently the process for ensuring physician quality relies wholly on graduation from an approved medical school and the passing of a licensing examination . . . no reexamination is required. . . . State licensing boards are responsible for monitoring physicians' behavior. . . . Unfortunately, this approach for assuring physician quality and competence is completely inadequate. (Feldstein 1994, 189)

Licensure does not restrict physicians to practice in a particular area of medicine. (In the United States, it is not against the law for an ophthalmologist to perform heart surgery.) Furthermore it is hard to argue that passing a standardized exam . . . offers much information about physician competence or success. (Svorny 2000, 303)

licensing laws supported by the AMA left physicians free to practice medicine according to any system of therapeutics they chose once they had obtained a license to practice. (Goodman 1980, 7)

consumers were (and still are) not as well protected from unqualified and unethical practitioners as they have been led to believe. (Feldstein 1999, 395)

It is hard to argue that quality care is the objective as, in many cases, licensing laws exclude individuals for reasons unrelated to their professional competence.

Residency requirements for foreign-trained physicians . . . continue to exceed the requirements for graduates of U.S. medical schools. (Seldon, et al. 1998, 820)

Researchers question whether state medical boards effectively oversee medical professionals.

few resources have been devoted to monitoring the quality of practice after an individual has been licensed. Applicants to medical school are given close scrutiny. The

budgets for the state medical boards suggest that the licensed practitioners are not. (Benham 1991, 81)

In the United States, state license revocation efforts have been subject to criticism for avoiding (admittedly difficult) issues of physician competence and for focusing, instead, on physicians who prescribe narcotics inappropriately to others or who abuse drugs and/or alcohol themselves. (Svorny 1992, 32)

Medical boards in many states do not even specify incompetence as grounds for disciplinary action. (Gaumer 1984, 398)

To many, it is clear that the current system of regulation does not assure quality care.

Though no one is suggesting that eliminating licensure and other requirements will reduce . . . negative outcomes, it is clear that regulation does not assure quality care. (Folland, Goodman, and Stano 2001, 358)

The available research does not suggest that existing systems of regulation have effectively controlled initial or subsequent competency of professionals. (Gaumer 1984, 406)

[There is evidence that mandatory programs of continuing education are] burdensome and fruitless. (Gaumer 1984, 399)

The performance of the medical profession, state regulatory agencies, and the malpractice system in protecting patients against negligent physicians has been inadequate. (Feldstein 1999, 397)

An important point is that, perversely, licensing can reduce the quantity and quality of health care.

Restrictive licensing can . . . result in declines in the quality of received services in that there may be (1) self-substitution of inferior products and/or services . . . ; (2) decreases in the average per capita service time rendered; for example, short, hurried, delayed office visits with a harried physicians; (3) differential geographic availability as numbers are reduced and the remaining members of the profession can choose their locations with more discretion, such as doctor shortages in rural areas; and (4) increased waiting time for provision of a service where delay in service entails expense for the buyers. (Carroll and Gaston 1979, 2)

The existing system results in some persons receiving no care, or being treated by individuals without any medical training (family members, neighbors, friends). (Fuchs 1986, 19)

Kugler and Sauer (2004) suggest the high costs of obtaining a license may deter talented individuals from pursuing state approval to practice, reducing physician service quality. Examining the earnings of immigrant physicians in Israel, they find a negative selection bias due to the cost of re-licensing. By negative selection, they mean, “immigrant physicians who acquire a license have lower intrinsic earnings potential (in the absence of a license) than those who do not” (Kugler and Sauer 2004, 5).

Benham (1991) pointed to yet another way in which licensing restrictions reduce service quality.

a clear consequence of licensure is to inhibit the production of information concerning the comparative performance of practitioners and hospitals. This in turn reduces the incentive to introduce innovations that would facilitate comparative evaluations and improve quality control. (Benham 1991, 89-90)

Phelps (1997) does not share this view. He sees licensure as a “guarantee” of minimum quality that may increase the production of information.

On net, one cannot say that licensure is necessarily a benefit or a harm to consumers. The potential gains in quality information may outweigh any costs from monopolization. Indeed, it may even be that the market operates more competitively because the “minimum quality” guarantee that licensure produces may increase consumer willingness to search for lower prices. (Phelps 1977, 243)

Of course, such a “guarantee” might result from certification rather than licensure. The relative merits of the two schemes are discussed below.

COMMENTS ON THE EFFICIENT PRODUCTION OF PHYSICIAN SERVICES

A concern is that physician licensure limits innovation in medical markets and flexibility in hospital and other institutional staffing.

On innovation

I am persuaded that licensure has . . . retarded technological development both in medicine itself and in the organization of medical practice. (Friedman 1962, 158)

it must be kept in mind that the various licensure laws . . . have rarely been designed to keep up with the rapidly changing organizational and technical innovations that are potentially feasible in health-care delivery. The preponderant view—certainly among health economists—is that physicians have not even begun to exploit the productivity potential actually within their reach. (Reinhardt 1975, 231, 233)

Medical licensure has had a deleterious effect on the quality of medical care by sharply reducing heterogeneity in the practice of medicine. . . . Even in those areas

where promising innovations have arisen, medical licensure laws have restricted, or threatened to restrict, their application. (Goodman 1980, 36)

It is my view that economists have concentrated excessively on the indirect effects of barriers to entry and too little on such issues as restrictions on innovation, excessive training requirements. (Benham 1980, 24)

On scope-of-practice limitations

It is now widely accepted that more extensive task delegation in medical practice would be in society's interest. . . . Constraints widely believed to inhibit more efficient use of health manpower are the various licensure laws governing the practice of medicine in this country. (Reinhardt 1975, 229, 231)

Many studies . . . show that the quality of care would not suffer if licensure policies were selectively liberalized allowing mid-level practitioners to perform some tasks not reserved only for . . . physicians. (Gaumer 1984, 397)

Scope-of-practice rules limit medical professionals' career mobility (Gaumer 1984). Licensing statutes preclude the informal transitions that occur in other industries as individuals gain expertise over time.

On inefficient training requirements

The American Medical Association has the power to control the *costs* of medical training as well as the number obtaining that training. By making it more costly to become a physician . . . the profession may insure that all incomes rise while the expected returns "at the margin" remain normal. The profession, in other words, is influenced to make medical education as inefficient as possible. (Lindsay 1973, 346)

licensing can cause applicants to overinvest in education and formal training. . . . If these attributes do not improve productivity the investments are wasted socially. (Dorsey 1980, 433)

GENERAL AGREEMENT THAT LICENSING ENFORCES CARTEL BEHAVIORS

Economists see state licensing as a source of cartel power among physician groups. Kessel (1958 and 1970) pointed out that licensing requirements increase returns for existing practitioners at consumers' expense. He was especially concerned that graduation from an American Medical Association-approved medical school was a condition for admission to state licensing exams—allowing organized medicine to control entry to the very market it served (1958, 283).

Folland, Goodman, and Stano (2001) note that “organized medicine historically exerted considerable influence over the supply of trained physicians” (354). Given more recent empirical evidence, however, they express doubt about the continued role of medical professionals in limiting medical school enrollment.

data in recent decades indicate that medical school enrollments are responsive to market forces . . . continuing to view medical education as controlled by a monolithic or conspiratorial medical profession is somewhat implausible. (Folland, Goodman, and Stano 2001, 354)

Still, they write of:

the questionable effects of licensure on quality...and the anticompetitive effects of restrictions on entry and restrictions on the scope of practice of potential competitors. (Folland, Goodman, and Stano 2001, 358)

Zweifel (1991) listed several factors that favor medical groups (over other professions) in acting as a cartel. Except perhaps for elective surgeries

on the wealthiest clients, medicine is characterized by a lack of international competition for its product. Also, what is sold is a personal service that cannot be resold, making it difficult to undermine price discrimination. Finally, licensure allows control of market entry.

The view that licensure facilitates cartel-like behavior has been expressed over and over.

[organized medicine] has, for over 100 years, sought and obtained special privileges from government. These special privileges take the form of restrictions on free competition in the marketplace. (Young 1987, 2)

In granting sole authority to the boards to issue licenses, society has, in effect, given considerable power to organized medicine to restrict the supply of physicians and to influence the pattern of medical care for the benefit of the profession. (Rayack 1982, 393)

The ability of the profession to influence medical school admissions and licensure exams, as well as their resistance to legal delegation of more routine tasks to other health professionals, has certainly helped perpetuate their economic advantage. (Burstein and Cromwell 1985, 77)

The establishment of limits on the use of physician extenders is yet another method physicians employ to protect their economic interests. (Santerre and Neun 2000, 423)

Economists have, for some time, suspected that occupational licensure operates as a legally sanctioned cartelization device, restricting entry . . . and restraining competition. . . . Excessive limits . . . can result in monopoly rents for members of the profession and higher prices and fewer services for consumers. (Martin 1980, 143-144)

The AMA has not only controlled the supply of medical school spaces in the United States . . . but also has worked to assure that state licensing statutes require graduation from AMA-accredited schools. Foreign entry has also been

curtailed by restrictive licensing laws as well as a strict federal immigration policy. (Noether 1986, 233)

coupled with findings that consumers are rarely, if ever, involved in the process, and that the resulting regulations do in fact raise prices and decrease the availability of services, the evidence supporting the self-interest model of regulation is substantial. (Begun, Crowe, and Feldman 1981, 250)

The emphasis in terms of quality is always on the training of *entering* physicians and not on those currently practicing in the profession. It is in the economic interests of current practitioners . . . they will receive higher prices and higher incomes. . . . If the medical profession was concerned primarily with quality rather than with monopoly power, there would be at least some emphasis by the profession on the quality of care provided by practicing physicians. (Feldstein 1999, 383, 386)

By the 1950's, organized medicine had achieved virtually all of its political goals . . . [one of which was] to control entry into the medical profession and to suppress competition for physicians' services. . . . The most important consequence of the control of medical education by organized medicine . . . was that physicians acquired the power to reduce the supply of medical services and increase their incomes. (Goodman and Musgrave 1992, 137, 147).

As time passed, restrictions were expanded to cover . . . advertising, price cutting, and other conduct considered to be "unprofessional." Clearly licensing laws serve not only to protect patients but also to limit the number of practitioners, thus protecting physicians from would-be competitors. (Henderson 2002, 107)

The [American Medical Association] has been . . . vigorous in attacking health practitioners who are widely considered

legitimate but who represent a competitive threat to the members of the medical profession. (Rayack 1982, 405)

Mainstream physicians benefit from a more restrictive regulatory regime governing practitioners of alternative therapies. . . . Licensing laws that reduce the supply of such alternative services harm consumers by rendering such low cost options unavailable while these laws also appear to create rents for mainstream physicians. (Anderson, et al., 2000, 497)

the medical profession as a whole must ultimately bear responsibility for the nature of these laws and their effect on resource allocation within the health-care sector. (Reinhardt 1975, 232)

Of course, given the nature of the political system in the United States, it is likely that both consumer and physician interests influence regulatory outcomes. In empirical tests, Leffler (1978) found that variations in licensing laws across states could be explained by consumer demand for quality. However, two studies that examined the relative influence of physicians and consumers found that physician interests dominate regulatory outcomes.

on the margin, the licensure restrictions in practice in 1965 increased entry costs by more than they reduced consumers' costs of generating quality assurance in the market for physician services. . . . The implication is that professional or special interests dominated consumer interests in the setting of licensure requirements. (Svorny 1987, 507)

[With respect to the regulation of certified nurse midwives] . . . supply-side (quantity reducing) effects dominate the demand-side (quality assurance and quantity enhancement) effects . . . it appears that regulation of this type of service has detrimental consumer welfare effects. In a time when many medical service delivery systems are in chaos, the advantages to deregulation of such fundamental activities should not be minimized. (Adams, et al. 2003, 673)

The continued influence of the American Medical Association is attributed to effective lobbying and few challenges from consumers. Milton Friedman wrote:

The groups that think they have a special interest . . . are concentrated groups to whom the issue makes a great deal of difference. The public interest is widely dispersed. In consequence . . . producer groups will invariably have a much stronger influence . . . [than the] widely spread consumer interest. (Friedman 1962, 143)

Weingast (1980), alone, takes exception to the idea that physician interests will dominate the political decision-making process.

While the producers probably have superior organization, once the issue enters election campaigns, further coordination by the diffuse nonproducers is not needed: all they need to do is vote. (Weingast 1980, 90)

However, if it is costly to assess the likely vote of each candidate on every issue, nonproducers may face higher costs of influencing outcomes, bringing us back to the more commonly held belief that professional groups are likely to dominate public policy.

DO INFORMATION ASYMMETRIES JUSTIFY LICENSING DOCTORS?

Some economists take the position that information asymmetries justify government intervention in medical labor markets. As Evans puts it,

the essence of the professional relationship is that the consumer does not know what he needs before service, nor does he know afterward whether he was adequately served. (Evans 1980, 250)

The counter position is that word-of-mouth and physician referrals provide guidance, as do other mechanisms, such as institutional reputation.

people do not . . . choose physicians by picking names at random from a list of licensed physicians. (Friedman 1962, 158)

even in many situations labeled 'emergency' the consumer has in principle a considerable amount of power over what can be done to him (including whether or not he chooses to be an 'emergency' case) and which physician he chooses in order to obtain advice. (Pauly 1980, 43)

Since a consumer has generally recognized the existence of a problem, he can presumably recognize its diminution. This ability to evaluate quality *ex post*, even if the evaluation is only approximate, provides checks on low-quality sellers, through both liability laws and the consumer's ability to shop elsewhere if quality is poor. (Beales 1980, 128)

EXAMINING THE EXCEPTIONS

Arrow's 1963 paper on medical care is frequently cited in support of physician licensure.

The general uncertainty about the prospects of medical treatment is socially handled by rigid entry requirements. These are designed to reduce the uncertainty in the mind of the consumer as to the quality of product insofar as this is possible. I think this explanation, which is perhaps the naïve one, is much more tenable than any idea of a monopoly seeking to increase incomes. No doubt restriction on entry is desirable from the point of view of the existing physicians, but the public pressure needed to achieve the restriction must come from deeper causes. (Arrow 1963, 966)

However, in a footnote, Arrow acknowledged the difficulty of assuring quality through the regulatory licensing of professionals. As to the ability of licensing to reduce uncertainty about quality, he wrote:

How well they achieve this end is another matter. R. Kessel points out to me that they merely guarantee training, not continued good performance as medical technology changes. (Arrow 1963, 966)

In addition, Arrow expressed a number of other concerns about state licensing.

Both the licensing laws and the standards of medical-school training have limited the possibilities of alternative qualities of medical care . . . [that might] appeal to different tastes and incomes. (Arrow 1963, 953).

[restrictions on entry to the field have] constituted a direct and unsubtle restriction on the supply of medical care. (Arrow 1963, 955)

The licensing laws . . . exclude all others from engaging in any one of the activities known as medical practice. As a result, costly physician time may be employed at specific tasks for which only a small fraction of their training is needed, and which could be performed by others less well trained and therefore less expensive. (Arrow 1963, 957)

the present all-or-nothing approach could be criticized as being insufficient with regard to complicated specialist treatment, as well as excessive with regard to minor medical skills. (Arrow 1963, 967)

Despite how often it is cited in favor of physician licensing, Arrow's article offers little to public policy makers trying to decide if licensing makes sense. He catalogs the ways in which health care markets depart from perfect competition and presumes that some sort of government intervention would improve upon a market outcome. He hedges this conclusion with numerous comments on the inefficiencies of licensing

regulations in the United States. Arrow does not make a clear case for or against licensing doctors.

Weingast (1980) who, like Arrow, asserts that information asymmetries are a problem for consumers of health care, argued that state licensure is not the solution. He wrote: “[the] political solution to the market inadequacies fails for precisely the same reason the market failed in the first place—the informational asymmetries” (93).

In their 1989 paper, Graddy and Nichol express the belief that information asymmetries present in health care markets require some degree of regulation. Like Weingast, they do not support licensure as it exists.

Consumers should be protected from incompetent providers, but should otherwise be able to choose among different price/quality options which may satisfy individual preferences. (Graddy and Nichol 1989, 614)

Graddy and Nichol do not specify the ways in which they would revise existing licensing statutes to increase choice in medical markets, but other economists, quoted below, have some ideas.

LICENSURE VS. CERTIFICATION

Many economists prefer certification to licensure.

Under certification buyers have a wider range of choices . . . they can buy low-quality goods or services if they wish. (Leland 1980, 283)

The case for licensure presumably rests on the proposition that the consumer is a poor judge of the quality of medical care and therefore needs guidance. . . . Assuming this to be true, the need for guidance could be met by voluntary *certification* rather than compulsory licensure. . . . Under a certification system patients would be free to choose the level of expertise that they wanted, including uncertified practitioners. (Fuchs 1986, 19)

As long as certified personnel are available, economic models suggest consumers will not gain as a result of replacing certification with mandatory licensure unless there are some sort of problems with market failure which go beyond difficulties simply in identifying qualified personnel. (White 1987, 32)

Even if entry controls do improve quality, that improvement can be purchased far more cheaply via certification of professionals, rather than through licensing. Consumers would then be able to choose between high-quality, high-priced services and lower quality at a lower price. (Beales 1980, 140)

The efficiency case for licensing can be made only under restrictive conditions—when market failure cannot be remedied by private exchange (such as by certification and advertising) as costlessly as it can be remedied by government identification and the outlawing of incompetent and unscrupulous practitioners. No study has yet been produced by the economics profession that makes a case, on cost-benefit grounds, for the licensing of any profession. (Elzinga 1980, 114)

Two arguments for licensure over certification carry little weight with economists. The first is that consumers need the government to make decisions for them. This does not go over well, as most economists recognize that the government can not begin to speak for the tastes and preferences of millions of individuals—the private market allows the expression of those tastes and preferences. It is this variation in tastes, in fact, that generally leads economists to favor a system of registration or certification, under which consumers make choices for themselves.

A second argument in support of licensure over certification is that there are externalities associated with the consumption of low quality physician services (Moore 1961). The issue here is that, if a consumer purchases incompetent care and a contagious disease is misdiagnosed, others will suffer. In the United States, however, a bigger problem appears to be people who do not purchase care at all. Eliminating licensure would

make care cheaper and more available, encouraging many of those who do not currently seek care to do so.

A final justification for licensure over certification suggests that licensure reduces agency costs in the market for physician services. By restricting entry and, therefore, increasing the profitability of medical practice, licensure creates incentives for physicians to act with the best interests of their patients in mind. With greater profitability, physicians have more to lose if they engage in malfeasance. This efficiency wage argument for licensure over certification is outlined by Svorny (1987). The incentive effects of a loss upon malfeasance have been mentioned elsewhere.

licensing may serve to protect consumers . . . by providing an asset, namely the license itself, that may be seized in the event of negligent performance. (Shapiro 1986, 861)

As one would expect, the more ethical wealth that one must forego as a result of being discovered reducing quality, the less likely is the reduction in quality. (Blair and Kaserman 1980, 194)

However, no one has yet made the case that information costs are sufficiently high enough to justify the inefficiencies associated with government intervention over those of the market.

PROPOSALS FOR CHANGE

As the above discussion indicates, at the least, economists favor reducing the power of the American Medical Association over state licensure. In that vein, Rayack argues to replace profession dominated licensure boards by “responsible administrative agencies.”²

The social acceptance of licensing in medicine indicates a general belief in the desirability of providing protection of

² The economics literature on regulatory capture suggests that shifting power from the AMA to administrative agencies reporting to state legislatures would continue to leave the process subject to AMA influence.

the consumer through the maintenance of standards. . . [a moderate] approach is possible whereby AMA power can be curbed and at the same time socially acceptable medical standards can be maintained. (Rayack 1982, 425)

Also, to reduce the influence of medical professionals, Blair and Kaserman suggest separating the functions of the board into regulating product quality and regulating competition.

in situations requiring self-regulation by members of the profession . . . the attainment of a socially optimum outcome may require the existence of more than one regulatory body with a separation of goals: self-regulation of product quality with external regulation of competitive practices. (Blair and Kaserman 1980, 197)

Gaumer (1984) advocates “reforms of the exclusionary and self-serving aspects of credentialing” (410) by means of changes in administrative processes and practice constraints. He expressed support for efforts to group related health professionals together on state boards to internalize “state manpower planning concerns, interoccupational conflicts, and service delivery productivity losses” (409).

Also to reduce the influence of physicians over licensing decisions, Svorny and Toma (1998) suggest shifting the source of state medical board funding away from physician fees to the state legislature. Controlling for other factors, they find that the influence of physicians over board actions is less in states where boards are funded by the state legislature.

Feldstein proposed “specific-purpose licensure.”

Specific-purpose licensure would mean that not all physicians would need to take the same educational training, training in some specialties would take a much shorter period of time. . . .When a physician wants an additional specific-purpose license, he or she could receive additional training and then take the qualifying exam for that license. The training requirements for entering the medical profession would be determined not by the medical profession, but by the demand for different types of physicians and the least-cost manner of producing them. (Feldstein 1994, 190-191)

In the third edition of his health economics textbook, Feldstein (1999) refers to this as “task” licensure, making the point that licensing physicians to perform specific tasks would lower the cost of a medical education (396).

Even Evans (1980), who advocated “an extensive regulatory web . . . to constrain the [health care] industry” (263) opposed state licensing regulations in which physician groups are given power to influence entry and practice patterns. Evans argued that the “collective self-regulation of processes of service production, as well as of the economic behavior of professional firms, must be weakened or removed” (260). He imagined that “public regulation of a more sophisticated type would still be needed to substitute for the quality control provided by self-regulation” (259).

Not one of these proposals, however, can be clearly put in the camp of significantly liberalizing the regulation of physicians. Their tenuous joint premise is that the actions of a reconfigured regulatory arrangement would be an improvement over the current situation and an improvement over market outcomes.

Others favor liberalization of medical licensing regulation, such as a switch from licensure to simple certification.

Licensing regulations can be quite restrictive . . . certification and registration systems represent lower degrees of regulation. More research into classifying degrees of manpower regulation, and matching those with the need for regulation, would be fruitful. (Begun and Feldman, 1990, 97)

[Under a system of certification rather than licensure] If we are wrong and no consumers want lower quality at lower prices, the substitution of certification for licensure would have no effects—the market would effectively make certification mandatory, much as licensing does today. The risks of certification are therefore very low. Combined with output monitoring, the risk can be reduced even further. (Beales, 1980, 140)

there could be several grades or categories, and periodic recertification would be more practicable (and less threatening) than periodic relicensure. (Fuchs 1986, 19)

Based on their findings of negative selection in licensing status among immigrant physicians in Israel, Kugler and Suaer (2004) suggest a direction for future policy.

The policy implication...is that lowering the direct costs of acquiring a license may raise physician quality. (Kugler and Sauer 2004, 28)

Seldon, et al. (1998) advocated supply side efforts to resolve the problems in the market for physician services.

The government could increase market competition by encouraging increased admissions into medical schools . . . by loosening visa restrictions imposed on foreign-trained physicians . . . [and by encouraging] the use of primary-care providers such as nurse practitioners. (Seldon et al., 1998, 820)

Goodman and Musgrave (1992) expressed support for shifting control over purchasing health care from third-party payers to consumers:

Some physicians do abuse patients and payers by overbilling. A smaller number do practice bad medicine . . . If patients controlled their health care dollars and were more involved in medical decisions, there would

undoubtedly be fewer instances of overbilling and unnecessary procedures. (297-298)

Professor Friedman's prescription for medical markets is straightforward: "licensure should be eliminated as a requirement for the practice of medicine" (1962, 158).

Feldstein (1999) echoes this sentiment:

It appears that reliance on a competitive health care market might well be the most useful approach to improving physician performance and providing consumers with the necessary information to make informed choices. (Feldstein 1999, 397)

In the course of the last thirty years, the emergence of health maintenance organizations and commercial interests in health care have changed the market for physician services dramatically, leading some economists to have even more confidence in private markets as opposed to government regulation.

One potential benefit of increased commercialization of medicine is in [the] area of quality control. The threatened loss of institutional reputation because of poor quality controls would provide incentives to monitor systematically and to alter practices when appropriate. (Benham 1991, 90)

[Changes in] knowledge about quality of medical care and ability to monitor quality . . . [and] the more extensive activities by purchasers of care . . . are likely to diminish the relative importance of licensure as we know it today. (Ginsberg and Moy 1992, 33)

Svorny (2003) identified changes in liability and technology that make licensing regulations increasingly redundant to market forces:

It is reasonable to ask whether . . . it makes sense to preserve licensing restrictions and disciplinary activities. The advent of computer technology and innovative

software programs have made information on physicians and practice patterns available to health care providers and their patients. Because liability for physician malpractice has shifted, hospitals, health maintenance organizations, insurers . . . who do not take advantage of the new technology to check physicians' qualifications are open to costly judgments in court. Prescription fraud can be reduced by means of electronic tracking. For all these reasons, it becomes ever more difficult to justify state licensing and the continued funding of state medical boards. (Svorny 2003, 155)

CONCLUSION

Despite the wide reach of medical licensing in health care production through its impact on the nature and cost of care, it has been all but ignored in debates over health care reform. As the above discussion indicates, many economists view licensing as a significant barrier to effective, cost efficient health care. State licensing arrangements have limited innovations in physician education and practice patterns of health professionals.

Some states have moved to reform their scope-of-practice laws, suggesting a direction for other reform-minded states. This includes an expanded scope of practice for paraprofessionals, allowing them to take on some tasks previously restricted to physicians. In many states paraprofessionals have been allowed to work fairly independently and are permitted to prescribe medication.

Consumers would benefit from a regulatory environment in which health care provider organizations and hospitals are free to employ health manpower in flexible ways and medical training is offered in a variety of forms. A rigid four-year curriculum is not necessarily the only good way to train physicians for a variety of tasks. Nurses and other health professionals, whose skills develop, can be moved sequentially into increasingly difficult practice situations without having to sit in classes that ostensibly "assure" their knowledge of appropriate practice patterns.

There are many ways to train competent health care providers. The existing, rigid rules severely limit entry and constrain health care providers from innovations in manpower use that could increase services and lower health care costs in the United States.

REFERENCES

- Adams, A. Frank, III, Robert B. Ekelund, Jr., and John D. Jackson.** 2003. Occupational Licensing of a Credence Good: The Regulation of Midwifery. *Southern Economic Journal* 69(3): 659-675.
- Anderson, Gary M., Dennis Halcoussis, Linda Johnston, M.D. and Anton D. Lowenberg.** 2000. Regulatory Barriers to Entry in the Healthcare Industry: The Case of Alternative Medicine. *The Quarterly Review of Economics and Finance* 40(4): 485-502.
- Arrow, Kenneth J.** 1963. Uncertainty and the Welfare Economics of Medical Care *American Economic Review* 53(5): 941-973.
- Beales, J. Howard, III.** 1980. The Economics of Regulating the Professions. In *Regulating the Professions*, ed. Roger D. Blair and Stephen Rubin. Lexington, Massachusetts: Lexington Books, 125-142.
- Begun, James W., Edward W. Crowe, and Roger Feldman.** 1981. Occupational Regulation in the States: A Causal Model. *Journal of Health Politics, Policy and Law* 6(2): 229-254.
- Begun, James W. and Roger Feldman.** 1990. Policy and Research on Health Manpower Regulation: Never Too Late to Deregulate? *Advances in Health Economics and Health Services Research* 11(1): 79-110.
- Benham, Lee.** 1980. The Demand for Occupational Licensing. In *Occupational Licensure and Regulation*, ed. Simon Rottenberg. Washington, DC: American Enterprise Institute, 13-25.

- Benham, Lee.** 1991. Licensure and Competition in Medical Markets. In *Regulating Doctors' Fees: Competition, Benefits, and Controls Under Medicare*, ed. H.E. Frech III. Washington, DC: American Enterprise Institute, 75-90.
- Blair, Roger D. and David L. Kaserman.** 1980. Preservation of Quality and Sanctions within the Professions. In *Regulating the Professions*, ed. Roger D. Blair and Stephen Rubin. Lexington, Massachusetts: Lexington Books, 185-198.
- Burstein, Philip L. and Jerry Cromwell.** 1985. Relative Incomes and Rates of Return for U.S. Physicians. *Journal of Health Economics* 4(1): 63-78.
- Carroll, Sidney L. and Robert J. Gaston.** 1979. State Occupational Licensing Provisions and Quality of Service: The Real Estate Business. *Research in Law and Economics* 1: 1-13.
- Dorsey, Stuart.** 1980. The Occupational Licensing Queue. *The Journal of Human Resources* 15(3): 424-434.
- Elzinga, Kenneth G.** 1980. The Compass of Competition for Professional Services. In *Regulating the Professions*, ed. Roger D. Blair and Stephen Rubin. Lexington, Massachusetts: Lexington Books, 81-86.
- Evans, Robert G.** 1980. Professionals and the Production Function: Can Competition Policy Improve Efficiency in the Licensed Professions? In *Occupational Licensure and Regulation*, ed. Simon Rottenberg. Washington, DC: American Enterprise Institute, 225-264.
- Feldstein, Paul J.** 1994. *Health Policy Issues: An Economic Perspective on Health Reform*. Ann Arbor, Michigan: Health Administration Press.
- Feldstein, Paul J.** 1993. *Health Care Economics*. Albany, New York: Delmar Publishers.
- Feldstein, Paul J.** 1999. *Health Care Economics*, 5th Ed. Albany, NY: Delmar Publishers.
- Folland, Sherman, Allen C. Goodman, and Miron Stano.** 2001. *The Economics of Health and Health Care*, 3rd Ed. City of publication: Prentice Hall.
- Friedman, Milton.** 1962. *Capitalism and Freedom*. Chicago: University of Chicago Press.

- Fuchs, Victor R.** 1986. *The Health Economy*. Cambridge, MA: Harvard University Press.
- Gaumer, Gary L.** 1984. Regulating Health Professionals: A Review of the Empirical Literature. *Milbank Memorial Fund Quarterly Health and Society*. 62(3): 380-416.
- Ginsberg, Paul B. and Moy, Ernest.** 1992. Physician Licensure and the Quality of Care. *Regulation* Fall: 32-39.
- Goodman, John C.** 1980. *The Regulation of Medical Care: Is the Price Too High?* San Francisco: Cato Institute.
- Goodman, John C. and Gerald L. Musgrave.** 1992. *Patient Power: Solving America's Health Care Crisis*. Washington, DC: Cato Institute.
- Graddy, Elizabeth and Michael B. Nichol.** 1989. Public Members on Occupational Licensing Boards: Effects on Legislative Regulatory Reforms. *Southern Economic Journal* 55(3): 610-625.
- Henderson, James W.** 2002. *Health Economics and Policy*, 2nd Ed. Cincinnati: South-Western.
- Kessel, Reuben A.** 1958. Price Discrimination in Medicine. *Journal of Law and Economics* 1: 20-53.
- Kessel, Reuben A.** 1970. The A.M.A. and the Supply of Physicians. *Law and Contemporary Problems* 35(2): 267-283.
- Kugler, Adreiana D. and Robert M. Sauer.** 2004. Doctors Without Borders? Re-licensing Requirements and Negative Selection in the Market for Physicians. Unpublished manuscript. June 24.
- Leffler, Keith.** 1978. Physician Licensure: Competition and Monopoly in American Medicine. *Journal of Law and Economics* 21(1): 165-186.
- Leland, Hayne E. Quacks.** 1979. Lemons and Licensing: A Theory of Minimum Quality Standards. *Journal of Political Economy* 87(6): 1328-1346.
- Leland, Hayne E.** 1980. Minimum-Quality Standards and Licensing in Markets with Asymmetric Information. In *Occupational Licensure and Regulation*, ed. Simon Rottenberg. Washington, DC: American Enterprise Institute, 264-284.

- Lindsay, Cotton M.** 1973. Real Returns to Medical Education. *The Journal of Human Resources* 8(3): 331-348.
- Martin, Donald L.** 1980. Will the Sun Set on Occupational Licensing? In *Occupational Licensure and Regulation*, ed. Simon Rottenberg. Washington, DC: American Enterprise Institute, 142-154.
- Moore, Thomas G.** 1961. The Purpose of Licensing. *Journal of Law and Economics* 4: 91-117.
- Noether, Monica.** 1986. The Effect of Government Policy Changes on the Supply of Physicians: Expansion of a Competitive Fringe. *Journal of Law & Economics* 29(2): 231-262.
- Pauly, Mark.** 1980. *Doctors and Their Workshops*. Chicago: The University of Chicago Press.
- Phelps, Charles E.** 1997. *Health Economics*, 2nd Ed. Addison Wesley Educational Publishers.
- Rayack, Elton.** 1982. The Physician Services Industry. In *The Structure of American Industry*, 6th Ed., ed. Walter Adams. New York: Macmillan Publishing Company, Inc., 388-426.
- Reinhardt, Uwe E.** 1975. *Physician Productivity and the Demand for Health Manpower*. Cambridge, MA: Ballinger Publishing Company.
- Santerre, Rexford E. and Stephen P. Neun.** 2000. *Health Economics: Theories, Insights, and Industry Studies*. Fort Worth: Dryden Press.
- Seldon, Barry J., Chulho Jung and Robert J. Cavazos.** 1998. Market Power Among Physicians in the U.S., 1983-1991. *The Quarterly Review of Economics and Finance* 38(4): 799-824.
- Shapiro, Carl.** 1986. Investment, Moral Hazard, and Occupational Licensing. *Review of Economic Studies* 53(5): 843-862.
- Svorny, Shirley.** 1987. Physician Licensure: A New Approach to Examining the Role of Professional Interests. *Economic Inquiry* 25(3): 497-509.
- Svorny, Shirley.** 1992. Should We Reconsider Licensing Physicians? *Contemporary Policy Issues* 10(1): 31-38.

- Svorny, Shirley.** 1993. Advances in Economic Theories of Medical Licensure. *Federation Bulletin: The Journal of Medical Licensure and Discipline* 80(1): 27-32.
- Svorny, Shirley.** 2000. Licensing, Market Entry Regulation. In *Encyclopedia of Law and Economics, Vol. III, The Regulation of Contracts*, ed. Boudewijn Bouckaert and Gerrit De Geest. Cheltenham, UK: Edward Elgar, 296-328.
- Svorny, Shirley.** 2003. Medical Licensing; Existing Public Policy and Technological Change. In *The Half Life of Policy Rationales; How New Technology Affects Old Policy Issues*, ed. Fred E. Foldvary and Daniel B. Klein. New York: New York University Press, 144-157.
- Svorny, Shirley and Eugenia Toma.** 1998. Entry Barriers and Medical Board Funding Autonomy. *Public Choice* 97(1-2): 93-106.
- Weingast, Barry R.** 1980. Physicians, DNA Research Scientists, and the Market for Lemons. In *Regulating the Professions*, ed. Roger D. Blair and Stephen Rubin. Lexington, MA: Lexington Books, 81-86.
- White, William.** 1987. The Introduction of Professional Regulation and Labor Market Conditions: Occupational Licensure of Registered Nurses. *Policy Sciences* 20(1): 27-51.
- Young, S. David.** 1987. *The Rule of Experts: Occupational Licensing in America*. Washington, DC: The Cato Institute.
- Zweifel, Peter.** 1991. Protecting the Medical Profession. In *Regulating Doctors' Fees: Competition, Benefits, and Controls Under Medicare*, ed. H.E. Frech III. Washington, DC: American Enterprise 91-96.

ABOUT THE AUTHOR



Shirley Svorny is chair of the economics department at California State University, Northridge. She holds a Ph.D. in economics from UCLA. From 1981 to 1984, she was engaged in policy analysis for the Getty Oil Company. In 1986-87, she managed an industry risk group at Security Pacific Bank. She was a Milken Institute Affiliated Scholar and served as director of the San Fernando Valley Economic Research Center at Cal State Northridge. She has published articles in *Economics of Education Review*, *Contemporary Economic Policy*, *Urban Affairs Review*, *Public Choice*, *Regional Science and Urban Economics*, *Cato Journal*, *Applied Economics*, *The Journal of Medical Licensure and Discipline*, *The Energy Journal*, *Economic Inquiry*, and the *Journal of Labor Research*. Her opinion articles have appeared in the *Los Angeles Times* and the *Los Angeles Daily News*. Her research interests are in the areas of urban, labor, and health economics. Her E-mail address is shirley.svorny@csun.edu.



*Econ Journal Watch,
Volume 1, Number 2,
August 2004, pp 306-318.*

INTELLECTUAL TYRANNY OF THE STATUS QUO



The Development Set: The Character of the *Journal of Development Economics* 2002

SUSAN ANDERSON AND PETER BOETTKE*

[ABSTRACT, KEYWORDS, JEL CODES](#)

BOOK REVIEWERS KNOW THAT IT IS A CHALLENGE TO REVIEW A collection of 12 separately-authored essays. Here, we have been asked to read and comment briefly on the character of the 62 dense, technical articles in the six 2002 issues of the *Journal of Development Economics* (vols. 67, 68, and 69). Started in 1974, the journal is published by Elsevier and is conventionally regarded as one of the top field journals in development economics.

There is no way to discuss more than a few in any detail. We have read the entire set, but we have not attempted to really master the material. We approached the material with critical priors that, to our mind, were essentially confirmed, but we try not to let our treatment become cavalier or dismissive. Unfortunately, it is perhaps inevitable that a brief and unfavorable gloss on the character of a set of 62 articles will come off as cavalier.

Around the mid-20th century, economics strived to develop, in the phraseology of Francis Bator, “institutionally antiseptic theory.” The strong trend was toward ahistorical, acontextual, and uniformly mathematical

* Both: Department of Economics, George Mason University.

economics. In the past 25 years, however, there has been somewhat of a counter-trend, reflected by the development of law and economics, public choice, positive political economy, property rights economics, the new economic history, and new institutionalism. More broadly, there are now vibrant critical movements of diverse impetus and ideological flavor challenging the scientific status of business-as-usual formalism.

The approach that interprets real world institutions in the context of candid practical policy questions—the Smithian approach—is widely pursued in the books and authentic policy discourse on comparative economics and development economics. The collapse of state socialism in Eastern and Central Europe, and the growing recognition of the failure of development planning, have broken up the official Development thinking of the 1960s and 1970s and paved the way for a new political economy approach to comparing economies and their economic development (see Djankov, et. al., 2003). At long last, the common-sensical ideas of early critics like Peter Bauer (1972 and 1981) and Deepak Lal (1983 and 1987) are coming to be widely accepted in books and authentic policy discourse. For example, William Easterly's widely read book *The Elusive Quest for Growth: Economists' Adventures and Misadventures in the Tropics* (2002) is quite critical of what we here call the "Development Set" mentality—the mentality of agency officials and experts trying to feel as though they know enough to justify their agency's programs.¹

Yet, in some "leading" (i.e., high citation impact) academic field journals, the formalistic tendencies still dominate. At the *Journal of Development Economics*, we observe the academic manufacture of formalistic crafts. The two dominant modes are the usual ones of academic economics, namely, equilibrium model building and statistical significance. Very few papers pursue a plain, detailed institutional narrative that makes sense of the policy history, developments, and alternatives. In the development field, economists are in a unique position to try out big ideas and learn from cross country comparisons, yet Development Set character and academic formalism conspire to make the *Journal of Development Economics* rather boring.

In the field of development economics, thinking is dominated by researchers at the leading development assistance agencies—the World Bank, International Monetary Fund, United Nations, etc. As the accompanying article by Klein with DiCola shows, a significant majority of the 2002 *JDE*

¹ The expression "Development Set" comes from Ross Coggins satirical poem by that title, variations of which can easily be found on the Internet.

authors and Contributing Editors have ties to at least one such agency—by present or past employment, funding, publication, or research presentation. From a sociological point of view, it is plain that the “invisible college” of mainstream academic development economics centers on the missions and practices of the large governmental development agencies. These agencies aim to study civil society, poverty reduction, institutional infrastructure, health and human services, etc., and thus many articles in development journals can be seen as following this research agenda.

But if the research agenda is led by the development agencies, the modes of discourse pursued by the *JDE* articles are framed by academic officialdom. For example, in the February 2002 issue, we find two papers addressing the effect of political instability on economic growth. Nauro F. Campos (University of Newcastle and William Davidson Institute) and Jeffery Nugent (University of Southern California) argue that when we look at the impact of instability on economic growth using a Granger-causality framework we find that there is no evidence of the hypothesized negative relationship. Political instability does not Granger-cause poor economic growth. The other paper by Mark Gradstein (Ben-Gurion University, Israel) uses a model to show that the more socially polarized and unstable a society is the less likely it is to adopt growth enhancing policy rules. Instability obstructs policy reforms that would promote investment and thus growth. Both articles are more concerned with working through the formal statistical scheme than with developing a full, sensible answer to the question about the connections between political stability and growth.

There are at least three common aspects of apparatchism in scholarship: (1) the field is formulated in a way that, even when a piece is critical in particular matters, it nonetheless affirms the mission, tasks, and importance of certain big international agencies, (2) these big governmental agencies directly engage academic researchers, often holding out money and the prestige of high-level influence and policymaking to the “top” researchers in the field, and (3) there is a collective conservatism with regard to ideas, due to the risk of alienating the established institutions of funding, so policy discussion tends to be toothless.

Table 1

Vol., Issue	Model no Measure	Measure no Model	Model and Measure	Institutions, History, Public Policy Considered	Total
Vol. 67, Issue 1	4	7	0	0	11
Vol. 67, Issue 2	3	3	3	1	9
Vol. 68, Issue 1	3	5	3	1	11
Vol. 68, Issue 2	4	5	2	1	11
Vol. 69, Issue 1	9	4	1	0	14
Vol. 69, Issue 2	3	1	2	1	6
Totals	26	25	11	4	62
Percentages	42%	40%	18%	6%	100%

Apparatchism combined with the inherent intellectual conservatism of the academy means that most of the research published in a field will, itself, reflect merely tinkering at the existing margins of what is currently scientifically acceptable. Much of the empirical work amounts to little more than the documenting of the warts and moles of current realities. In a field like development economics where the fresh voices see that much of the old, establishment thinking missed the boat and resulted in misery, such tinkering and wart-documenting simply perpetuates bad thinking and doesn't make room for relevant research and bold ideas.

Here, Easterly's book is noteworthy, because Easterly is one of the Associate Editors at the *Journal of Development Economics* (there were 22 Associate Editors in 2002). Easterly documents failed efforts in development assistance since WWII, including the records of public policy with regard to investment gap, schooling, population control, and infrastructure investment. His conclusion is that the failure is a result of a disregard by donor nations of the incentives that recipient nations face. 'Incentives Matter' is the mantra of economics and Easterly's book echoes this refrain from start to finish. His criticisms of development policy generated controversy and eventually he was dismissed from the World Bank. However, it is important to realize that Easterly's criticisms are

directed at the incentives that exist in the country receiving the aid, and not the agencies responsible for providing the aid. In short, his work demonstrates that the thinking behind development assistance must be re-examined in light of the incentive issues. This is an important message, but we need to look at not only the incentives that actors in the developing world face, but also the incentives that researchers in the international agencies entrusted with devising development policy face.²

TWO EXAMPLES OF BOLD ANTI-APPARATCHIK LINES OF RESEARCH

Schooling

One example of the sort of contribution that can be made when rejecting the economic establishment in favor of discovering real answers is found in the work done in recent years on private schooling in the developing world. One of the more amazing facts being uncovered is the vast extent to which private schooling is being substituted for poorly functioning and, in many instances, blatantly corrupt government schools³. This research flies in the face of the standard literature in development economics that argues that schooling is required for the human capital development that enables democracy and economic development, and typically looks to and endorses *government* schooling. As Easterly's work documents (2002, 72-73), the standard argument cannot be sustained in the wake of the empirical evidence showing that, as investment in government schooling has increased, economic growth has declined in many countries. The reason for this decline is: (a) poor incentives for the schools to teach useful information to the students, and (b) poor incentives for graduates of the school system to employ their skills in the wealth creating private sector as opposed to becoming government workers. Too often, government schools are highly integrated to the social apparatus of a highly politicized

² A first-step in this direction was made by Ostrom, et. al. (2002), where the incentives of donors and recipients and the agencies responsible for managing that relationship are examined via institutional analysis.

³ See Tooley and Dixon (2003) and Tooley (1999). Research by scholars associated with the E G West Centre at the University of Newcastle on private schooling is currently being done under Tooley's supervision in India, China, Nigeria, and Ghana.

society. But if schools teach basic skills in language and mathematics, and at higher levels science and technology, then we see a positive correlation between schooling and economic growth. A major reason why private schools are emerging in Africa and India in the poorest-of-the-poor sections is because parents see that education provides a way to a better future for their children only if they are taught useful skills such as English language, math, science, and technology. The public schools are failing, so alert ‘edupreneurs’ are filling the void and providing educational services for a fee that poor parents scrape together to provide a better possible future for their children.

But you will not read much about private schooling in the *Journal of Development Economics*.⁴ The private schooling movement upsets the Development Set’s sensibilities; private schooling manifests social improvement in forms that are bottom-up, informal, and highly local—all aspects that the big development agencies are ill suited to know about, appreciate, participate in, or take credit for.⁵ The dominant opinion is well reflected in Nobel Laureate Amartya Sen’s reaction to the recent study done in India. “No developed country had educated itself using private schools,” he said (Waldman 2003, A1). The private schooling movement upsets academic modes of discourse because the methodologies used involved researchers in the field, interviewing teachers, students, and parents. As described by Tooley and Dixon (2003), such research has uncovered practices previously unrecorded—even deliberately hidden, in the effort to avoid taxation, regulation, and state control.⁶ There are no good official statistics on private schooling because one of its *raison d’être* is to escape the notice of the state sector. Thus, to rely solely on official data would make any academic study on schooling blind to the phenomena of private schooling. The field research was designed to uncover this unrecorded activity and, thus, never looked at aggregate economic indicators of human capital, but instead, looked at incentives and the pursuit of happiness in a commonsensical way.⁷

⁴ There are two articles in the JDE from 1974 to 2002 comparing public and private schooling. Bedi, Arjun S. and Ashish Garg (2000), and Cox, Donald and Emmanuel Jimenez (1990). Both articles use data sets (on earnings and academic achievement, respectively) to conclude that private schooling offers advantages over public.

⁵ We should note that much of the research on the private schooling movement has been funded by the World Bank and reported through Bank publications.

⁶ See, also, Dixon (forthcoming).

⁷ This style of research should also remind the reader of the work done by Hernando de Soto on property holdings in poor countries. De Soto has become famous, even an intellectual celebrity among policy makers and news media, but academics often dismiss his

There are three articles in the 2002 issues which deal with education more broadly. In the June 2002 issue, Arjun Singh Bedi (Institute of Social Studies, The Hague) and John H.Y. Edwards (Tulane University) examine the impact of school quality on educational returns. The authors use a unique data set from Honduras, and improved estimation techniques, to find that school quality has an important positive effect on labor market earnings. They conclude that their work “support[s] the idea that school quality should be an important aspect of educational policy” (Bedi and Edwards 2002, 183). The paper also includes background on the rationale for public provision of education, and in doing so gives us insight into the traditional mindset: “A purely private educational system cannot function efficiently without perfect capital markets and capital market imperfections are likely to be especially severe in developing countries” (Bedi and Edwards 2002, 159). The paper does not include any evaluation of this pronouncement, nor any criticisms. It would have been equally relevant to make the following pronouncement: *A governmental educational system cannot function efficiently without perfect benevolence, omniscience, and non-distortionary taxation, and government imperfections are likely to be especially severe in developing countries.* But we do not find any assertion of that sort.

The October issue contains two articles dealing with the supply and demand for education. Sudhanshu Handa (Inter-American Development Bank) accepts, as given, the “development imperative” of raising primary school attendance and creates an econometric model using a data set from Mozambique, in order to analyze the factors affecting the supply and demand of primary schooling in developing countries. The second article from the October issue, by Arjun S. Bedi (Institute of Social Studies, The Hague) and Jeffery H. Marshall (Stanford University), accepts that “investments in education are widely recognized as a key component of a

work as not rigorous. In the 2001 *Journal of Economic Literature*, Christopher Woodruff wrote a review essay on *The Mystery of Capital*. Woodruff provides a number of reasons for his doubts about the insight that de Soto offers, and some of these are worth serious consideration by those who look to de Soto for intellectual inspiration, but for the most, part Woodruff is concerned that de Soto has had such an impact in the policy world when he has “scarcely published an article in an academic journal,” and his methods and results have been challenged by justly skeptical academics. Woodruff admits that there is much to what de Soto has to say, but exactly how much and how valuable, he concludes, will have to wait until we have better data, and better analysis of the data. Overall, the amount of serious academic discussion of de Soto’s work starting with his *The Other Path* has been far less than the attention he has received outside of academics. We contend that this dismissal can be attributed, in part, to de Soto’s rejection of both the standard model-and-measure mentality of the academic establishment and the top-down mentality of the Development Set.

country's development strategy" and focuses only on the demand for schooling—the factors which underlie the decision of school attendance. A regression using a data set from Honduras provides the authors with the conclusion that “notwithstanding the limited supply side analysis, our paper shows that a more fruitful approach towards raising school attendance may lie not in reducing opportunity costs or enhancing school inputs but in policies that target the supply of schooling” (Bedi and Marshall 2002, 152). We are left guessing what the authors have in mind by “policies that target the supply of schooling”.

Economic Freedom

A much more prominent, even monumental, example of bold research in the field of development economics that unfortunately has not been incorporated into the literature published in the journal is the movement to measure Economic Freedom (Gwartney and Lawson 2004). We regard the Economic Freedom indices to be one of the most important intellectual developments in the practical application of economic reasoning in the past 15 years. James Gwartney and Robert Lawson presented a paper on the impact of the Economic Freedom of the World Index at a regional meeting of the Mont Pelerin Society in September 2003. In the paper they note that while it takes time for new ideas and new tools to penetrate the professional economic literature, references and journal citations to the EFW index have been growing substantially, and from a wide cross-section of researchers, not just economists who favor liberty. Yet, the most recent issue we were able to consult as this article went into publication, the June 2004 issue, there has been only one article in the history of the *Journal of Development Economics* that mentions or references any economic freedom index.⁸

At a May 2004 conference at Princeton University celebrating the contributions of P. T. Bauer, including presentations by Nobel Prize economists James Buchanan, Douglass North, and Amartya Sen, a question was put to Sen to describe the difference in development economics in 1964 and in the year 2004. Sen puzzled over an answer for a minute and gave a measured response that emphasized the relative weight that scholars put on the ability of the market to generate social order and prosperity, as

⁸ See Smith, Douglas. (2001)

opposed to power discrepancies, privilege, and conflict. In 1964, the relative weighting led to work that concentrated on class conflict and the unequal distribution of income. In 2004, Sen argued, economists have become more sensitive to the ability of markets to generate social harmony. Still, while there is impressionistic evidence that this shift in the relative weighting has occurred, and it is certainly evident in the broader intellectual climate of opinion, it is not obvious by looking at the *Journal of Development Economics*.

Looking further than the 2002 issues and sampling the journal from 1975-2002, one can definitely see change in the field of development. Scholarship published in the *JDE* over this larger span does demonstrate that researchers in the field are beginning to look at micro-level institutions, rather than just macro relationships. This change is not limited to the *JDE*. The same evolution can be seen in the publications of the development institutions as well. Recent World Bank publications⁹ include statements on the importance of the environment for investment, and the consequent need to have strong property rights and rule of law. This noted, it is reasonable to conclude that most of the articles published in the *Journal of Development Economics* 2002 are influenced more by the agenda of the field of economics in general than by development institutions. The articles revolve around disproving or enhancing past research through more complex modeling, or running regressions with a new data set. A reader is often confronted with the question of whether the authors are trying to find solutions to growing economies, alleviating poverty, and other World Bank goals, or simply looking to get published. A PhD in economics should not be seen as a license to be boring, and yet much—not all—of what goes on in this field is just that—uninteresting exercises in modeling and measuring.¹⁰

THE TWO PAPERS WE LIKED BEST

Leo Sleuwaegen (Catholic University of Leuven, Belgium) and Micheline Goedhuys (Erasmus University Rotterdam, Netherlands) in their

⁹ For example, Panaritis, Elena, Marc A. Weiss, and Alven H. Lam, ed. (2001)

¹⁰ Perhaps more accurately we should not say uninteresting because the exercises are interesting to other economists, or at least we are trained to behave as if they were interesting to us while we are in graduate school and during the probationary period of being an assistant professor.

article “Growth of firms in developing countries, evidence from Cote d’Ivoire” examine why existing theories on the growth of firms fail to explain the dynamics that lie beneath the observed structure in developing countries. They concluded that the reason is the existing theories do not consider the institutional features of poorly developed markets. The paper searches for the relevant institutional elements in data of manufacturing firms in Cote d’Ivoire. It then tests the validity of the empirical analysis using survey data on actual growth barriers as perceived by owners and managers of firms.

“Natural resources, rent seeking and welfare” by Ragnar Torvik (Norwegian University of Science and Technology) confronts the issue of how some countries continue to remain poor despite an abundance of natural resources. Previous literature on the subject has focused on the possibility of increasing returns to scale and shifting of factors away from more productive sectors. Torvik builds on more recent literature that emphasizes the role of rent seeking. Including entrepreneurship and the possibility of rent seeking within a simple four sector economy model, Torvik finds that a greater amount of natural resources can increase the number of entrepreneurs engaged in rent seeking, lowering the number of entrepreneurs running productive firms. Rent seeking can explain why it is possible that more natural resources can actually lead to lower welfare.

CONCLUDING REMARK

Development economics deals with multiple countries through time. It needs to pay close attention to the institutional details of the countries under examination.

The *Journal of Development Economics* 2002 reveals that while some of the institutional issues are getting play, the field has not thrown off its “Development Set” character. Instead of getting a field that varies methodological approaches and formulations on a case-by-case basis, we observe conventional methods addressing traditional “development economics” ideas, notably the *dirigisme* of diligent public servants, but in a more tentative and watered-down form.

Appendix A: Article classifications

[Attached excel spreadsheet](#)

REFERENCES

- Bauer, Peter T.** 1972. *Dissent on Development*. Cambridge: Harvard University Press.
- Bauer, Peter T.** 1981. *Equality, the Third World, and Economic Delusion*. Cambridge: Harvard University Press.
- Bedi, Arjun S. and Ashish Garg.** 2000. The effectiveness of private versus public schools: the case of Indonesia. *Journal of Development Economics* 61: 463-494.
- Bedi, Arjun S. and John H.Y. Edwards.** 2002. The impact of school quality on earnings and educational returns—evidence from a low-income country. *Journal of Development Economics* 68: 157-185.
- Bedi, Arjun S. and Jeffery H. Marshall.** 2002. Primary school attendance in Honduras. *Journal of Development Economics* 69: 129-153.
- Campos, Nauro F. and Jeffery Nugent.** 2002. Who is afraid of political instability? *Journal of Development Economics* 67: 157-172.
- Cox, Donald and Emmanuel Jimenez.** 1990. The relative effectiveness of private and public schools: Evidence from two developing countries. *Journal of Development Economics* 34: 99-121.
- Djankov, S., E. Glaeser, R. La Porta, and F. Lopez-de-Silanes.** 2003. The New Comparative Economics. *Journal of Comparative Economics* 31(4): 595-619.
- De Soto, Hernando.** 1989. *The Other Path: the invisible revolution in the Third World*. New York: Harper & Row.
- De Soto, Hernando.** 2000. *The Mystery of Capital: why capitalism triumphs in the West and fails everywhere else*. New York: Basic Books.
- Dixon, Pauline.** Forthcoming. The regulation of private schooling serving low-income families in India. *Review of Austrian Economics*.

- Easterly, William.** 2002. *The Elusive Quest for Growth: Economists' Adventures and Misadventures in the Tropics*. Cambridge: MIT Press.
- Gradstein, Mark.** 2002. Rules, stability, and growth. *Journal of Development Economics* 67: 471-484.
- Gwartney, James and Robert Lawson.** 2003. The Impact of the Economic Freedom of the World Index. Paper presented at the Regional Meeting of the Mont Pelerin Society, Chattanooga, TN.
- Gwartney, James and Robert Lawson.** 2004. *Economic Freedom of the World: 2004 Annual Report*. Vancouver: Fraser Institute.
- Handa, Sudhanshu.** 2002. Raising primary school enrolment in developing countries: The relative importance of supply and demand. *Journal of Development Economics* 69: 103-128.
- Klein, Daniel B. with Therese DiCola.** 2004. Institutional Ties of *Journal of Development Economics* Authors and Editors. *Econ Journal Watch* 1(2): 319-330.
- Lal, Deepak.** 1983. *The Poverty of 'Development Economics'* London: Institute of Economic Affairs. [A Revised and Expanded edition appeared in 1997.]
- Lal, Deepak.** 1987. Markets, Mandarins, and Mathematicians. *Cato Journal* 7(1): 43-70.
- Ostrom, Elinor, et. al.** 2002. *Aid, Incentives, and Sustainability*. Stockholm, Sweden: SIDA.
- Panaritis, Elena, Marc Weiss, and Alven H. Lam, eds.** 2001. Do Property Rights Matter? An Urban Case Study from Peru. In *The Globalization Index*. Washington, DC: World Bank.
- Smith, Douglas.** 2001. International evidence on how income inequality and credit market imperfections affect private saving rates. *Journal of Development Economics* 64:103-127.
- Sleuwaegen, Leo and Micheline Goedhuys.** 2002. Growth of firms in developing countries, evidence from Côte d'Ivoire. *Journal of Development Economics* 68: 117-135.
- Tooley, James.** 1999. *The Global Education Industry*. London: Institute for Economic Affairs.

Tooley, James and Pauline Dixon. 2003. *Private Schools for the Poor: a case study from India*. Reading: The Center for British Teachers.

Torvik, Ragnar. 2002. Natural resources, rent seeking and welfare. *Journal of Development Economics* 67: 455-470.

Waldman, Amy. 2003. India's Poor Bet Precious Sums on Private Schools. *The New York Times*, 15 November, A1.

Woodruff, Christopher. 2001. Review of the book *The Mystery of Capital* by Hernando de Soto. *Journal of Economic Literature* 39(4): 1215-1223.

AUTHOR INFORMATION



Susan Anderson is currently pursuing a Ph.D. in economics at George Mason University, focusing in the area of Institutional economics. She is a Social Change Project Graduate Fellow at the Mercatus Center and this summer will be part of a team doing field work in Botswana as part of the Mercatus Center's [Global Prosperity Initiative](#).



Peter Boettke is a professor of economics at George Mason University and the research director of the Global Prosperity Initiative at the Mercatus Center. His home page is: <http://www.gmu.edu/departments/economics/pboettke/index.html>.



*Econ Journal Watch,
Volume 1, Number 2,
August 2004, pp 319-330.*

INVESTIGATING THE APPARATUS

Institutional Ties of *Journal of Development Economics* Authors and Editors

DANIEL B. KLEIN, WITH THERESE DICOLA*

[Abstract, Keywords, JEL Codes](#)

IN SPRING 2003, I RECRUITED THERESE DICOLA, THEN A SANTA Clara University senior, to research the curriculum vitas of the people involved in the six 2002 issues of the *Journal of Development Economics*. An author who published two articles in the 2002 issues would generate two “authorships.” Authorships are not weighted for co-authorship, so an article with two authors would create two authorships. Altogether there were 124 authorships. We also investigated the 26 editorial officers—we collected resumes and CVs online. For each individual we searched for employment, consultancy, grant, publication, and presentation ties (through 2003) to the major development agencies the World Bank, International Monetary Fund, The Inter-American Development Bank, the Asian

* Department of Economics, Santa Clara University.

Note on authorship: Although two authors are shown above, this paper is written in the first-person singular. Daniel Klein led the investigation and wrote up the paper. He is the “I” here. Therese DiCola was chiefly responsible for collecting the data.

Development Bank, U.S. AID, the United Nations, and so on. We also searched on the individual's name at each organization website.

Regarding the 124 authorships, we find that 75 percent have ties to the "Big 8" official development institutions, and 84 percent to official development institutions more broadly defined. Of the 26 editorial officers, we find that all have ties to the Big 8, all but one with ties of employment, consultancy, or grant. These numbers understate the actual extent of such ties, because for a few names we were unable to obtain a CV.

The results of this study are not surprising. Before presenting the findings, I offer some remarks that are not meant as interpretation of the facts, but rather *as presentation of the investigation*.

WHY SHOW INSTITUTIONAL TIES OF RESEARCHERS?

There is no scandal in development economists having ties to development institutions. I do not mean to suggest that institutional ties corrupt the otherwise honest efforts of researchers, nor that such ties impose narrow bounds on research or policy conclusions.

For example, in the field of transportation economist, Gabriel Roth has led research and understanding of how property rights and freedom can enhance the provision of highways, streets, parking, taxi services, and urban transit. Roth has had a long and productive career, including nineteen years (1967 to 1986) at the World Bank. The Bank generally supported his investigations, with his research culminating in such market-oriented works as the World Bank book *The Private Provision of Public Services in Developing Countries* (Oxford University Press, 1987).

The corpus of World Bank research shows diversity in method, findings, and policy ideas. For many academic researchers, the World Bank and the IMF are just other stops on the seminar circuit, like any normal university department.

Still, we should be concerned that the apparatus will neglect or elide fundamental criticism of the institutions. The Congress-appointed Meltzer Commission (2000: Exe. Sum.) said the IMF's "system of short-term crisis management is too costly, its responses too slow, its advice often incorrect, and its efforts to influence policy and practice too intrusive." The Commission continued: "High cost and low effectiveness characterize many development bank operations as well. The World Bank's evaluation of its

own performance in Africa found a 73% failure rate. Only one of four programs, on average, achieved satisfactory, sustainable results. In reducing poverty and promoting the creation and development of markets and institutional structures that facilitate development, the record of the World Bank and the regional development banks leaves much room for improvement.” Yet the Commission does not come out in favor of abolishing any of the seven institutions investigated.

Figures like Peter Bauer and Deepak Lal developed broad criticisms of foreign aid etc., and those criticisms have been continued in recent years. Specific institutions are criticized for being slow to act, too political, and taking the wrong actions (Stiglitz 2002; for a wide collection see McQuillan and Montgomery 1999). Critics on the left criticize them also especially for interfering with local customs and sovereignty, promoting markets and privatization, and damaging the environment and the position of women (e.g., Danaher 1994). Classical liberals criticize them also for politicizing and clouding public affairs and strengthening and expanding the national government, and for giving free markets, privatization, etc. a bad name (e.g., Bandow and Vásquez 1994; Boettke 1994; Anderson 2004). Others criticize the “Lords of Poverty” for sheer apparatchism, including bureaucracy, waste, cynicism, and exploitation of largesse (Hancock 1989; Irwin 1990). Outside of “the development set” there are many voices calling for abolition of one or more of the institutions (e.g., Anna Schwartz 1998).

Highlighting, as I do here, the money and status that lie behind research is not meant as an exposé. There is no realm of purely independent discourse, unsullied by interests and “biases.” Research is costly and the bills are always paid by someone; discourse is always organized by someone. Beliefs and culture are profoundly characterized by quasi-irreversibility of investment, status-quo bias, non-rivalrous enjoyment, non-excludability, and network effects, so no one should have any illusions about how we would have a pristine Eden of culture if it weren’t for certain evil institutions.

Still, it is proper to document the ties between research and official institutions, if only to provide a touchstone for discussion. Although the World Bank has employed Gabriel Roth and others like him, we should bear in mind that, whatever they might think privately, they generally do not publicly express their views about the institution’s character. There are probably people at the World Bank and IMF who would favor the ending of taxpayer support and any form of governmental privilege, but they probably do not air such views. Former World Bank insider James B.

Burnham (1994: 75) reports that “a surprising amount of [internal analysis] is straightforward and candid in describing and analyzing why bank projects fail,” and former employees like William Easterly of the World Bank (and of the *JDE* editorial board) sometimes come out with sharp criticism. But all such criticism rarely questions whether the institutions should retain their official status and continue to enjoy governmental privileges. The *fundamental* criticism comes mostly from organizations like the Cato Institute and the Hoover Institution, and from independent authors and publishers, not the academic-government complex illuminated here.

Works in the sociology of science, such as *The Intellectual and Social Organization of the Sciences* by the British sociologist Richard Whitley, suggest that scientific communities and cultures are coordination problems that exhibit network externalities, are prone to lock-in, and, especially, can be determined in character and purpose by elite official agencies. Whitley writes:

Co-ordination thus is not just a technical matter of integrating specialist contributions to common goals but involves the organization of programmes and projects in terms of particular priorities and interests. It is a political activity which sets research agenda, determines the allocation of resources, and affects careers in reputational organizations and employment organizations. (Whitley 1984: 88-89)

Although institutional ties do not impose narrow conclusions, it is more than reasonable to suppose that such ties close off, or at least diminish, methods, ideas, and policies that are inhospitable to the missions, purposes, and dignity of the institutions themselves. As Whitley writes:

Within general and sometimes rather diffuse objectives, then, scientists pursue reputations and decide on the significance of results by assessing their contributions to reputational goals. Implicitly or explicitly it is usually assumed that these goals are consonant with, and contributing to, the general mission of the employment organization and funding agencies. . . . By delegating much control over goals and processes to scientists pursuing reputations, funding agencies and government departments have substantially expanded the public

science system and also gained a considerable degree of influence over it. (pp. 297-98)¹

Individuals highly critical of the institutions, maybe believing they should be depoliticized and degovernmentalized as much as possible, are not likely to advance within a field in which such institutions play an integral role. Scientific institutions and cultures exhibit self-reinforcing mechanisms, including screening and selection, self-sorting, preference falsification, intellectual elision, cynicism, and belief adaptation (Kuran 1995).

Richard Whitley's theories of the organization of science and his warnings about the undue influence of governmental institutions are developed mainly in the context of the hard sciences in Britain, where certain "science policy" institutions are central. A U.S. analog is the National Institute of Health for the medical sciences. In development economics, however, the official institutions are more numerous, so the organization of the field might be polycentric. However, insofar as the various institutions treated here are alike in character and mission, and belong to one integrated social network, they might function as a sort of association and exert the kind of influence that Whitley describes for more centralized systems. We should learn more about the development institutions' degree of diversity in character and approach, and about their degree of social integration.

Journal of Development Economics, started in 1974 and published by Elsevier, has been ranked as a leading field journal in development economics (e.g., Kalaitzidakis et al 2002). As such, it helps to define the field. With such close ties to the large governmental agencies, it is unlikely to treat adversarial ideas and judgments. In particular, it is unlikely to be hospitable to spontaneous order economics, an engine of criticism of those agencies. An *Econ Journal Watch* article by Susan Anderson and Peter Boettke (2004) interprets the character of *Journal of Development Economics*.

¹ For an introduction to Whitley's ideas, see Siler 2003.

EMPIRICAL FINDINGS

Again, an author who published two articles in the 2002 issues would generate two “authorships;” authorships are not weighted for co-authorship, so an article with two authors would create two authorships. The six 2002 issues (Vols. 67-69) contained 124 authorships.² For each individual, we sought information about ties to the official development institutions. We distinguished five types of tie: employment, consultancy, grant,³ publication⁴, and presentation.⁵ We drew primarily from three types of sources: (1) resumes and curriculum vitas available on the Web, (2) acknowledgements found in publications and working papers, and (3) the website search-engine of each institution. When we found, say, information about the individual having a consultancy with the World Bank (past or current), we entered the information as a “C” into the appropriate column of a comprehensive Excel file; we did not attempt to measure or record *extent*. That is, an individual who consulted once and an individual who consulted 10 times would both just have a “C” in the World Bank column (as well as possibly other tie indicators). The Excel file with individual’s name redacted is appended here as an Internet link. We will make available the original (un-redacted) version to anyone who will use it discreetly for research and confirmation purposes.

Because the demarcations between employment, consultancy, and grant are vague, here we present those as a single group. They are the primary forms of tie. On top of those are people *without an employment, consultancy, or grant tie to the relevant institution but with publication or presentation ties*. Here the publication and presentation ties are presented together. The darker tops represent those who have *only* such ties. (In most cases, those with the primary forms of tie also have publication and presentation ties).

It should be noted that our data probably understates the true portion of authorships with ties. For several individuals we could not obtain a substantial CV or resume. It should also be noted that a number of the

² The 124 includes all pieces except the Erratum (in Volume 69, Issue 1).

³ The first three categories do not include the honoraria and expense reimbursement that normally accompany presentation and sometimes accompany publication.

⁴ Publication with, say, the World Bank, takes the form of an article, chapter, or book published by the Bank as a stand alone report, a conference volume, a working paper, or a journal article (in *The World Bank Economic Review*).

⁵ Presentation with, say, the World Bank means that the individual’s work has been presented at a Bank conference, meeting, or internal seminar.

JDE authors were junior colleagues who simply have not yet had much opportunity to develop ties to external institutions of any kind.

Figure 1
Known Ties to Official Institutions among 124 JDE Authorships

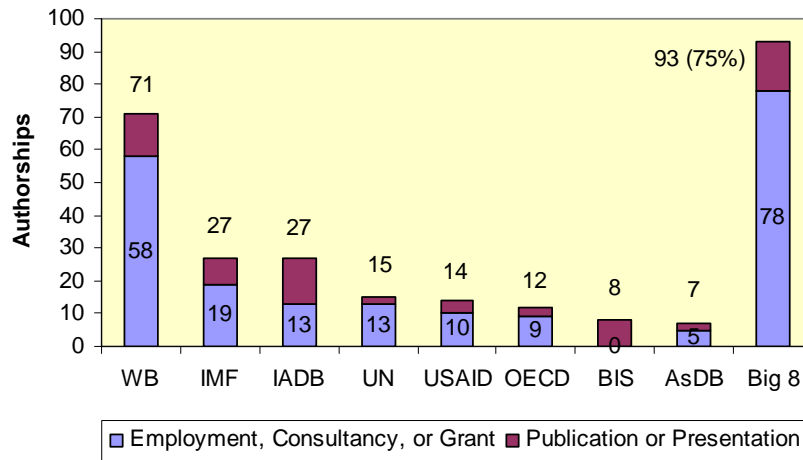


Figure 1 shows the results for what we term the “Big 8”: World Bank, International Monetary Fund, Inter-American Development Bank, United Nations, U.S. Agency for International Development, Organization for Economic Co-operation and Development, Bank for International Settlements, and Asian Development Bank. These are simply the official development institutions which showed the most ties. The African Development Bank, for example, does not have on its website an internal site-search engine, and for that reason alone might have not made it into our Big 8. Figure 1 shows that 78 of the 124 authorships (or 63 percent) had employment, consultancy, or grant ties to at least one of the Big 8, and that 93 (or 75 percent) had some form of tie.

Figure 2
Cumulative Ties to Official Institutions among 124 *JDE* Authorships

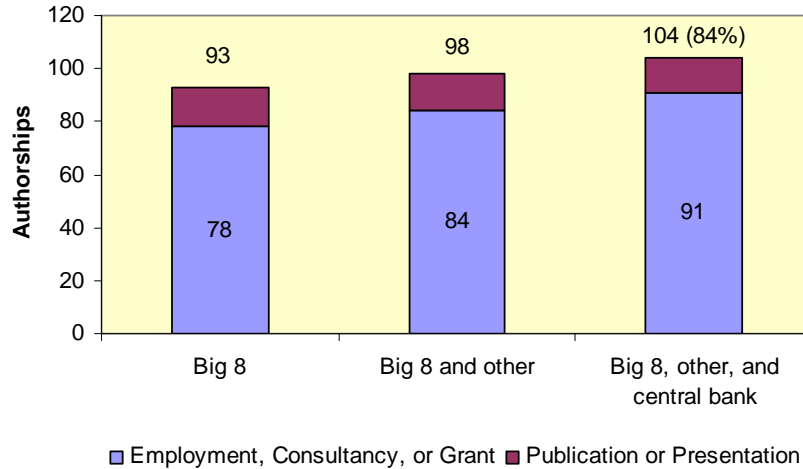
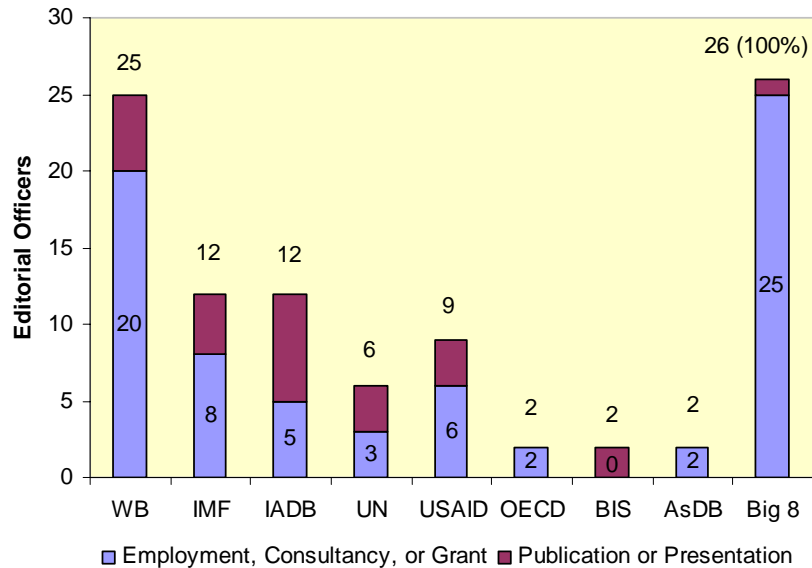


Figure 2 expands the set of official institutions, first by adding an “other” category. “Other” includes the following: WTO/GATT, Canadian International Development Agency, Swedish International Development Agency, German Marshall Fund, European Bank for Reconstruction and Development, European Investment Bank, Korean Development Trust, Ethiopian Ministry of Planning and Economic Development, Planning Institute of Jamaica, East-West Center, Centre de Recherche en Economie du Developpement, and International Food Policy Research Institute. The final bar in the figure then adds national central banks; the authors showed ties to the national central banks of the United States, Australia, Belgium, Canada, Chile, England, Israel, Italy, Japan, Mexico, New Zealand, Papua New Guinea, Portugal, and Spain. Ties for the cumulative set of institutions reach 104, or 84 percent of authorships.

Figure 3
Known Ties to Official Institutions of 26 JDE Editorial Officers



We did the same investigation for the 2002 editorial officers (one Editor, two Co-Editors, one Book Review Editor, and 22 Associate Editors). As shown in Figure 3, 25 of the 26 editorial officers have employment/consultancy/grant ties to at least one of the Big 8, and the 26th has publication/presentation ties. The editors' ties to other institutions and central banks can be gleaned from the Excel sheet, but one should not assume that information found there provides a complete account of actual ties; it surely understates the true extent.

CONCLUSION

Our investigation establishes that the authors and editors of *Journal of Development Economics* have extensive ties to official policy, loan, and grant making institutions dealing with the developing countries. How such ties

affect the character and intellectual content of the field of development economics is another question. It is a question worthy of serious and candid exploration.

APPENDIX 1

[Excel file](#): Institutional ties of 2002 *JDE* authorships and editorial officers, with individual names redacted. (We will make available the original (un-redacted) version to anyone who will use it discreetly for research and confirmation purposes.)

REFERENCES

- Anderson, Robert E.** 2004. *Just Get Out of the Way: How Government Can Help Business in Poor Countries*. Washington DC: Cato Institute.
- Anderson, Susan and Peter J. Boettke.** 2004. The Development Set: The Character of *Journal of Development Economics* 2002. *Econ Journal Watch* 1(2): 306-318.
- Bandow, Doug and Ian Vásquez, eds.** 1994. *Perpetuating Poverty: The World Bank, the IMF, and the Developing World*. Washington DC: Cato Institute.
- Boettke, Peter J., ed.** 1994. *The Collapse of Development Planning*. New York: New York University Press.
- Burnham, James B.** 1994. Understanding the World Bank: A Dispassionate Analysis, Ch. 4, pp. 75-85 in Doug Bandow and Ian Vásquez, eds. *Perpetuating Poverty: The World Bank, the IMF, and the Developing World*. Washington DC: Cato Institute.
- Danaher, Kevin, ed.** 1994. *50 Years is Enough: The Case Against the World Bank and the International Monetary Fund*. Boston: South End Press.

- Easterly, William.** 2002. *The Elusive Quest for Growth: Economists' Adventures and Misadventures in the Tropics*. Cambridge: MIT Press.
- Hancock, Graham.** 1989. *Lords of Poverty: The Power, Prestige, and Corruption of the International Aid Business*. New York: Atlantic Monthly Press.
- Irwin, Michael H. K.** 1990. [Banking on Poverty: An Insider's Look at the World Bank](#). *Cato Foreign Policy Briefing*, No. 3. September 20.
- Kalaitzidakis, Pantelis, Theofanis P. Mamuneas, and Thanasis Stengos.** 2002. Ranking of Academic Journals and Institutions in Economics. *Journal of the European Economics Association* 1(6): 1346-66.
- Kuran, Timur.** 1995. *Private Truths, Public Lies: The Social Consequences of Preference Falsification*. Cambridge, MA: Harvard University Press.
- McQuillan, Lawrence J. and Peter C. Montgomery.** 1999. *The International Monetary Fund -- Financial Medic to the World? A Primer on Mission, Operations, and Public Policy Issues*. San Francisco: Hoover Institution Press.
- Meltzer Commission** (International Financial Institutions Advisory Commission). March 8, 2000. [Report](#). Washington DC: US Congress.
- Roth, Gabriel.** 1987. *The Private Provision of Public Services in Developing Countries*. New York: Oxford University Press.
- Schwartz, Anna J.** 1998. [Time to Terminate the ESF and the IMF](#). Foreign Policy Briefing #48. Cato Institute, August 26.
- Siler, Kyle.** 2003. [The Social and Intellectual Organization and Construction of Economics](#). *Post-autistic economics review*, issue no. 22, 24 November, article 3.
- Stiglitz, Joseph E.** 2002. *Globalization and Its Discontents*. New York: W.W. Norton.
- Whitley, Richard.** 1984. *The Intellectual and Social Organization of the Sciences*. Oxford: Oxford University Press.

ABOUT THE AUTHORS



Daniel B. Klein is associate professor of economics at Santa Clara University, associate fellow at the Ratio Institute (Stockholm), and editor of *Econ Journal Watch*. He is author (with Adrian Moore and Binyam Reja) of *Curb Rights: A Foundation for Free Enterprise in Urban Transit* (Brookings Institution, 1997); editor of *Reputation: Studies in the Voluntary Elicitation of Good Conduct* (University of Michigan Press, 1997); and co-editor (with Fred Foldvary) of *The Half-Life of Policy Rationales: How New Technology Affects Old Policy Issues* (New York University Press, 2003). His email address is dklein@scu.edu.

Therese DiCola is an underwriter for St. Paul Travelers, San Francisco. She graduated with a Bachelors of Science in Economics from Santa Clara University in June 2003.



*Econ Journal Watch,
Volume 1, Number 2,
August 2004, pp 331-358.*

Economics in Practice

Size Matters: The Standard Error of Regressions in the *American Economic Review*

STEPHEN T. ZILIAK* AND DEIRDRE N. MCCLOSKEY**

[Abstract](#), [Keywords](#), [JEL Codes](#)

Sophisticated, hurried readers continue to judge works on the sophistication of their surfaces. . . . I mean only to utter darkly that in the present confusion of technical sophistication and significance, an emperor or two might slip by with no clothes.

Annie Dillard, *Living by Fiction*
New York: Harper and Row, 1988 ed., 31.

EIGHT YEARS AGO, IN "THE STANDARD ERROR OF REGRESSIONS," we showed how significance testing was used during the 1980s in the leading general interest journal of the economics profession, the *American*

* Roosevelt University, Chicago.

** University of Illinois at Chicago and Erasmus University at Rotterdam.

This paper also appears in the *Journal of Socio-Economics*: Ziliak, Stephen T. and Deirdre N. McCloskey. 2004. Size Matters: The Standard Error of Regressions in the *American Economic Review*. *Journal of Socio-Economics* 33(4) September.

<http://www1.elsevier.com/homepage/sae/econworld/econbase/soceco/frame.htm>

We thank, for their amazed attention to the present paper, audiences at Baruch College (CUNY), the University of Colorado at Boulder, the Georgia Institute of Technology, the University of Georgia, the University of Illinois at Chicago, the annual meeting of the Eastern Economic Association (2003), and the ICAPE Conference on "The Future of Heterodox Economics" (University of Missouri-Kansas City). Cory Bilton, David McClough, and Noel Winter provided excellent research assistance. Our special thanks go to William Kruskal.

Economic Review (McCloskey and Ziliak 1996). The paper reported results from a 19-item “questionnaire” applied to all of the full-length papers using regression analysis. Of the 182 papers 70% did not distinguish statistical significance from policy or scientific significance—that is, from what we call “economic significance” (Question 16, Table 1, 105). And fully 96 percent misused a statistical test in some (shall we say) significant way or another. Of the 70% that flatly mistook statistical significance for economic significance, further, again about 70% failed to report even the magnitudes of influence between the economic variables they investigated (1996, 106). In other words, during the 1980s about one half of the empirical papers published in the *AER* did not establish their claims as *economically* significant.

Some economists have reacted to our finding by saying in effect, “Yes, we know it’s silly to think that fit is the same thing as substantive importance; but *we* don’t do it: only bad economists do.” (Such as, it would seem, the bad ones who publish in the *AER*, an implied evaluation of our colleagues that we do *not* accept.) And repeatedly in the several score of seminars we have given together and individually on the subject since 1996 we have heard the claim that “After the 1980s, the decade you examined in your 1996 paper, best practice improved.”

All the better econometricians we have encountered, of course, agree with our point in substance. This is unsurprising, since the point is obviously true: fit is not *the same thing* as scientific importance; a merely statistical significance cannot substitute for the judgment of a scientist and her community about the largeness or smallness of a coefficient by standards of scientific or policy oomph. As Harold Jeffreys remarked long ago, to reject a hypothesis because the data show “large” departures from the prediction “requires a quantitative criterion of what is to be considered a large departure” (Jeffreys 1967, 384, quoted in Zellner 1984, 277n). Just so. Scientific judgment requires quantitative *judgment*, not endlessly more machinery. As lovely and useful as the machinery is, at the end, having skillfully used it, the economic scientist needs to *judge* its output. But the economists and calculators reply, “Don’t fret: things are getting better. Look for example at this wonderful *new* test I have devised.”

We are very willing to believe that our colleagues have since the 1980s stopped making an elementary error. But like them we are empirical scientists. And so we applied the same 19-item questionnaire of our 1996 paper to all the full-length empirical papers of the *next* decade of the *AER*, just finished, the 1990s.

Significance testing violating the common sense of first-year statistics

and the refined common sense of advanced decision theory, we find here, is not in fact getting better. It is getting worse. Of the 137 relevant papers in the 1990s, 82% mistook statistically significant coefficients for economically significant coefficients (as against 70% in the earlier decade). In the 1980s 53% had relied exclusively on statistical significance as a criterion of importance at its first use; in the 1990s 64% did.

“SIGNIFICANCE TESTING IS A CHEAP WAY TO GET MARKETABLE RESULTS”

William Kruskal, an eminent statistician long at the University of Chicago, an editor of the *International Encyclopedia of the Social Sciences*, and a former president of the American Statistical Association, agrees. “What happened?” we asked him in a recent interview at his home (William Kruskal 2002). “Why did significance testing get so badly mixed up, even in the hands of professional statisticians?” “Well,” said Kruskal, who long ago had published in the *Encyclopedia* a devastating survey on “significance” in theory and practice (Kruskal 1968a), “I guess it's a cheap way to get marketable results.”

Bingo. Finding statistical significance is simple, and publishing statistically significant coefficients survives at least that market test. But cheap *t*-tests, becoming steadily cheaper with the Moore's-Law fall in computation cost, have in equilibrium a marginal scientific product equal to their cost. Entry ensures it. In the 1996 paper we discussed the history of statistical versus economic significance. Viewed from the sociology and economics of the discipline the notion of statistical significance has been a smashing success. Many careers have prospered on testing, testing, testing (as David Hendry likes to put it). But intellectually the testing has been a disaster, as indeed Edgeworth had warned at the dawn.¹ He corrected Jevons, who had concluded that a “3 or 4 per cent” difference in the volume of commercial bills is not economically important: “[b]ut for the purpose of science, the discovery of a difference in condition, a difference of 3 per cent and much less may well be important” (Edgeworth 1885, 208). It is easy to see why: a statistically *insignificant* coefficient in a financial

¹ Edgeworth (1885, 187), we believe, is the first source of the word “significance” in a context of hypothesis testing. Our earlier paper claimed erroneously that John Venn was first (McCloskey and Ziliak 1996, 97; see Baird 1988, 468). Anyway, the 1880s: for some purposes not a meaningful difference.

model, for example, may nonetheless give its discoverer an edge in making a fortune; and a statistically *significant* coefficient in the same model may be offset in its exploitation by transactions costs. Statistical significance, to put it shortly, is neither necessary nor sufficient for a finding to be *economically* important. Yet an overwhelming majority of economists, we have shown for the 1980s and now again still more for the 1990s, believe statistical significance *is* necessary; and a simple majority believe it is sufficient.

Economists are skeptics, members of the tribe of Hume. But Ronald Aylmer Fisher (1890-1962), who codified the usage we are objecting to, was a rhetorical magician (as Kruskal once noted, the inventor of such enchanting phrases as “efficiency” and “analysis of variance”; “significance” was older). Long-lived and persistent, he managed to implant for example a “rule of 2” in the minds of economic and other scientists. Listen, for example, as Fisher computes for the masses in 1925 a first test of significance in his *Statistical Methods for Research Workers*:

The value for which $P=.05$, or 1 in 20, is 1.96 or nearly 2; it is *convenient* to take this point as a limit in judging whether a deviation is to be considered significant or not. Deviations exceeding twice the standard deviation are *thus formally regarded as significant*. Using this criterion we should be led to follow up a false indication only once in 22 trials, even if the statistics were the only guide available. Small effects will still escape notice if the data are insufficiently numerous to bring them out, but no lowering of the standard of significance would meet this difficulty. (Fisher [1925], 42; emphasis added)

Notice how a standard of “convenience” rapidly became in Fisher’s prose an item to be “formally regarded.” With Fisher there’s no loss function. There’s no thinking beyond the statistic. We are “to take this point as a limit.” Fisher’s famous and influential book nowhere confronts the difference between scientific and substantive significance (123-124, 139-140, concerning soporific drugs and algae growth). He provided (and then stoutly defended for the rest of his long life against the decision-theoretic ideas of Neyman, Pearson, and Wald) the cheapest way to get marketable results.

Our policy recommendation is this: that the profession adopt the standards set forth 120 years ago by Edgeworth, and in the years intervening

by a small but distinguished list of dissenters from the mechanical standard of 5% (and no loss function about it).

**PRACTICE HAS IMPROVED IN A FEW WAYS, BUT NOT
IN THE CRUCIAL MATTER OF SIGNIFICANCE**

Table 1 reports the results distinguished by decade, the 319 full-length papers using regression from January 1980 to December 1999. (We have at hand the whole population, not a sample; the urn of nature is poured out before us; unlike many of our colleagues, therefore, we will refrain from calculating statistics relevant only to inference from *samples* to a population, such as the “statistical significance” of the differences between the two decades.) Like Table 1 in McCloskey and Ziliak (1996) Table 1 here ranks in ascending order each item of the questionnaire according to “Percent Yes.” A “yes” means that the paper took what every statistical theorist since Edgeworth (with the significant exception of R. A. Fisher) has regarded as the “correct” action on the matter. For example, in the 1980s 4.4% of the papers considered the power of the tests (and we do not believe it accidental that every paper considering power also considered “a quantitative criterion of what is to be considered a large departure.”) That is, 4.4% did the correct thing by considering also the probability of a Type II error. In the 1990s 8% did. That’s an encouraging trend.

Table 1
The *American Economic Review* Had Numerous Errors
In the Use of Statistical Significance, 1980-1999

Survey Question	Percent Yes in 1990s	Percent Yes in 1980s
Does the paper...		
8. Consider the power of the test?	8.0	4.4
6. Eschew reporting all standard errors, <i>t</i> -, <i>p</i> -, and <i>F</i> - statistics, when such information is irrelevant?	12.4	8.3
16. Consider more than statistical significance decisive in an empirical argument?	18.2	29.7
11. Eschew “sign econometrics,” remarking on the sign but not the size of the coefficient?	19.0	46.7

14. Avoid choosing variables for inclusion solely on the basis of statistical significance?	25.5	68.1
15. Use other criteria of importance besides statistical significance after the crescendo?	28.5	40.7
10. Eschew “asterisk econometrics,” the ranking of coefficients according to the absolute value of the test statistic?	32.8	74.7
17. Do a simulation to determine whether the coefficients are reasonable?	35.0	13.2
7. At its first use, consider statistical significance to be one among other criteria of importance?	36.5	13.2
19. Avoid using the word “significance” in ambiguous ways?	37.2	41.2
9. Examine the power function? ^a	45.5	16.7
18. In the conclusions, distinguish between statistical and economic significance?	52.6	30.1
13. Discuss the scientific conversation within which a coefficient would be judged large or small?	54.0	28.0
2. Report descriptive statistics for regression variables?	66.4	32.4
1. Use a small number of observations, such that statistically significant differences are not found merely by choosing a very large sample?	67.9	85.7
12. Discuss the size of the coefficients?	78.1	80.2
5. Carefully interpret the theoretical meaning of the coefficients? For example, does it pay attention to the details of the units of measurement, and to the limitations of the data?	81.0	44.5
4. Test the null hypotheses that the authors said were the ones of interest?	83.9	97.3
3. Report coefficients in elasticities, or in some other useful form that addresses the question of “how large is large”?	86.9	66.5

Source: All the full-length papers using tests of statistical significance and published in the *American Economic Review* in the 1980s (N=182) and 1990s (N=137). Table 1 in McCloskey and Ziliak (1996) reports a small number of papers for which some questions in the survey do not apply.

Notes: ^a Of the papers that mention the power of a test, this is the fraction that examined the power function or otherwise corrected for power.

The change in practice is more easily seen in Tables 2 and 3, which isolate improvement and decline. In the 1980s only 44.5% of the papers paid careful attention to the theoretical and accounting meaning of the regression coefficients (Question 5). That is, in the 1980s the reader of an empirical paper in the *AER* was nearly 6 times out of 10 left wondering how to interpret the economic meaning of the coefficients. In the 1990s the share taking the correct action rose to 81%, a net improvement of about 36 percentage points. (This is what we mean by oomph: a big change, important for the science.) Similarly, the percentage of papers reporting units and descriptive statistics for regression variables of interest rose by 34 percentage points, from 32.4% to 66.4% (Question 2). And gains of more than 20 percentage points were made in the share of papers discussing the scientific conversation in which a coefficient would be judged large or small, the share of papers keeping statistical and economic significance distinct in the "conclusions" section, and the share of papers doing a simulation to determine whether the estimated coefficients are reasonable. (Our definition of "simulation" is broad. It includes papers that check the plausibility of the regression results by making, for example, Harberger-Triangle-type calculations on the basis of descriptive data. But a paper that uses statistical significance as the sole criterion for *including* a coefficient in a later simulation is coded "No," which is to say that it does *not* do a simulation *to determine whether the coefficients are reasonable*.)

These few gains are commendable. Whether they are scientifically significant is something only we scientists can judge, in serious conversation with each other (for example: that 8% rather than 4% consider power is nice, but still leave 92 percent of the papers risking high levels of a Type II error). In almost every question (that is, in all except perhaps Question 5 concerning the interpretation of theoretical coefficients, in which the improvement approaches levels that most people would agree are good practice) the improved levels of performance are still less than impressive. For example, in the 1990s *two-thirds* of the papers did not make calculations to determine whether the estimated magnitude of the coefficients made sense (Question 17)—only a third, we found, had simulated the effect of their coefficients with at least the elementary force of Ec 1. Skepticism of alleged effect is by contrast normal practice in sciences like chemistry and physics. (By the way, we have found by examining *The Physical Review* that physicists approximately never use tests of statistical significance; so too, in the magazine *Science*, the chemists and geologists; many biologists reporting their results in *Science* are less clear-minded on the matter; and in their own

journals the medical scientists, like the social scientists, are hopelessly confused about substantive error as against sampling error. Bald examples of this last may be found in the technical notes enclosed with medicines such as Rogaine.)

Table 2
The Economic Significance of the *American Economic Review*
Has in Some Regards Improved
 (Measured by Net Percentage Point Difference, 1980-1999)

Survey Question	Percent Yes in 1990s	Net Improvement since 1980s
Does the paper...		
5. Carefully interpret the theoretical meaning of the coefficients? For example, does it pay attention to the details of the units of measurement, and to the limitations of the data?	81.0	+36.5
2. Report descriptive statistics for regression variables?	66.4	+34.0
9. Examine the power function? ^a	45.5	+28.8
13. Discuss the scientific conversation within which a coefficient would be judged large or small?	54.0	+26.0
18. In the conclusions, distinguish between statistical and economic significance?	52.6	+22.5
17. Do a simulation to determine whether the coefficients are reasonable?	35.0	+21.8

Notes: ^a Of the papers that mention the power of a test, this is the fraction that examined the power function or otherwise corrected for power.

Milton Friedman, from 1943 to 1945, was a statistician for the Statistical Research Group of the Division of War Research at Columbia University (there is still a non-parametric test named after him). Listen to his experience with statistical vs. substantive significance:

One project for which we provided statistical assistance was the development of high-temperature alloys for use as the lining of jet engines and as blades of turbo superchargers—alloys mostly made of chrome, nickel, and other metals. . . . Raising the temperature a bit increases substantially the efficiency of the turbine, turbo supercharger,

or jet engine I computed a multiple regression from a substantial body of data relating the strength of an alloy at various temperatures to its composition. My hope was that I could use the equations that I fitted to the data to determine the composition that would give the best result. On paper, my results were splendid. The equations fitted very well [note: statistically; with high R^2] and they suggested that a hitherto untried alloy would be far stronger than any existing alloy. . . . The best of the alloys at that time were breaking at about ten or twenty hours; my equations predicted that the new alloys would last some two hundred hours. Really astounding results! . . . So I phoned the metallurgist we were working with at MIT and asked him to cook up a couple of alloys according to my specifications and test them. I had enough confidence in my equations to call them F1 and F2 but not enough to tell the metallurgist what breaking time the equations predicted. That caution proved wise, because the first one of those alloys broke in about two hours and the second one in about three. (Friedman 1985, quoted in Freidman and Schwartz 1991, 48-49)

Friedman learned that statistical significance is not the same as metallurgical significance.

The core confusion over the meaning of significance testing is reported in Table 3. One problem, which is often taken to be our main objection (it is not, though bad enough on its own), is that statistical *nonsignificance* is nonpublic. In the 1990s only one fourth of the papers avoided choosing variables *for inclusion* (pretests, that is) solely on the basis of statistical significance, a net *decline* in best practice of fully 43 percentage points (Question 14). As Kruskal put it in his 1968 article,

Negative results are not so likely to reach publication as are positive ones. In most significance-testing situations a negative result is a result that is not statistically significant, and hence one sees in published papers and books many more statistically significant results than might be expected. . . . The effect of this is to change the interpretation of published significance tests in a way that is hard to analyze quantitatively. (1968a, 245)

The response to Question 14 shows that economists made it hard in the 1990s to analyze quantitatively, in Kruskal's sense, the real-world relevance of their "significant" results. It's the problem of searching for significance, which numerous economists have noted, in cynical amusement or despairing indignation, is encouraged by the incentives to publish.

"Asterisk econometrics," the ranking of coefficients according to the absolute value of the test statistic, and "sign econometrics," remarking on the sign but not the size of coefficient, were widespread in the 1980s. But they are now a plague. Eighty-one percent of the papers in the 1990s engaged in what we called "sign econometrics" (in the 1980s 53% did [Question 11]). In their paper "Tax-based Test of the Dividend Signaling Hypothesis" Bernheim and Wantz (June 1995, 543) report that "the coefficients [in four regressions on their crucial variable, high-rated bonds] are all negative However, the estimated values of these coefficients," they remark, "are not statistically significant at conventional levels of confidence." The basic problem with sign econometrics, and with the practice of Bernheim and Wantz, can be imagined with two price elasticities of demand for, say, insulin, both estimated tightly, one at size -0.1 and the other at -4.0. Both are negative, and would both be treated as "success" in establishing that insulin use responded to price; but the policy difference between the two estimates is of course enormous. Economically (and medically) speaking, for most imaginable purposes -0.1 is virtually zero. But when you are doing sign econometrics you ignore the *size* of the elasticity, or the *dollar effect* of the bond rating, and say instead, "the sign is what we expected."

Sign econometrics is worse when the economist does not report confidence intervals. Perhaps because they were not trained in the error-regarding traditions of engineering or chemistry, economists seldom report confidence intervals. Thus Hendricks and Porter, on "The Timing and Incidence of Exploratory Drilling on Offshore Wildcat Tracts" (June 1996, p. 404): "In the first year of the lease term, the coefficient of HERF is positive, but not significant. This is consistent with asymmetries of lease holdings mitigating any information externalities and enhancing coordination, and therefore reducing any incentive to delay." Yet the reader does not know how much "HERF"—Hendricks' and Porter's Herfindahl index of the dispersion of lease holdings among bidders at auction—contributed to the probability the winners would then engage in exploratory oil drilling. In *Life on the Mississippi* Mark Twain noted that "when I was born [the city of] St. Paul had a population of three persons; Minneapolis had just a third as

many" (390). The sign is what a St.-Paul-enthusiast would want and expect. But the sign gives no guidance as to whether a size of 1 is importantly different from 3. No oomph.

Table 3
...But the Essential Confusion of Statistical and Economic
Significance is Getting Worse
 (Measured by Net Percentage Difference, 1980-1999)

Survey Question	Percent Yes in 1990s	Net Decline since 1980s
Does the paper...		
14. Avoid choosing variables for inclusion solely on the basis of statistical significance?	25.5	-42.6
10. Eschew "asterisk econometrics," the ranking of coefficients according to the absolute value of the test statistic?	32.8	-41.9
11. Eschew "sign econometrics," remarking on the sign but not the size of the coefficient?	19.0	-27.7
1. Use a small number of observations, such that statistically significant differences are not found merely by choosing a very large sample?	67.9	-17.8
4. Test the null hypotheses that the authors said were the ones of interest?	83.9	-13.4
15. Use other criteria of importance besides statistical significance after the crescendo?	28.5	-12.2
16. Consider more than statistical significance decisive in an empirical argument?	18.2	-11.5
7. At its first use, consider statistical significance to be one among other criteria of importance?	36.5	-10.8

Source: All the full-length papers using tests of statistical significance and published in the *American Economic Review* in the 1980s (N=182) and 1990s (N=137). Table 1 in McCloskey and Ziliak (1996) reports a small number of papers for which some questions in the survey do not apply.

About two-thirds of the papers *ranked* the importance of their estimates according to the absolute values of the test statistics, ignoring the estimated size of the economic impact (Question 10). In other words,

asterisk econometrics (which is what we call this bizarre but widespread practice), became in the 1990s a good deal more popular in economics (it has long been popular in psychology and sociology), increasing over the previous decade by 43 percentage points. Bernanke and Blinder (1992), Bernheim and Wantz (1995), and Kachelmeier and Shehata (1992), for example, published tables featuring a hierarchy of p -, F -, and t -statistics, the totems of asterisk econometrics (p905, 909; 547; 1130). The asterisk, the flickering star of *, has become a symbol of vitality and authority in economic belief systems. Twenty years ago Arnold Zellner pointed out that economists then (in a sample of 18 articles in 1978) never had “a discussion of the relation between choice of significance levels and sample size” (one version of the problem we emphasize here) and usually did not discuss *how far* from 5% the test statistic was: “there is room for improvement in analysis of hypotheses in economics and econometrics” (Zellner 1984, 277-80). Yes.

What is most distressing about Table 3, however, is the rising conflation of statistical and economic significance, indicated by the responses to Questions 16 and 7. Our main points are:

- 82% of the empirical papers published in the 1990s in the *American Economic Review* did not distinguish statistical significance from economic significance (Question 16). In the 1980s, 70% did not—scandalous enough (McCloskey and Ziliak 1996, 106).
- At the first use of statistical significance, typically in the “Estimation” or “Results” section, 64% in the 1990s did not consider *anything but* the size of the test statistics as a criterion for the inclusion of variables in future work. In the 1980s, 53%—11 percentage points fewer—had done so (Question 7, 106).

FOLLOWING THE WRONG DECISION RULE HAS LARGE SCIENTIFIC COSTS

Of course, not everyone gets it wrong. The *American Economic Review* is filled with examples of superb economic science (in our opinion most of the papers can be described this way—even though most them, we have

seen, make elementary mistakes in the use of statistical significance; in other words, we do *not* accept the opinion of one eminent econometrician we consulted, who dismissed our case by remarking cynically that after all such idiocy is to be expected in the *AER*). Table 4 reports the author rankings by economic significance, in five brackets. If a paper chose between 15 and 19 actions correctly, as Gary Solon's paper did (June 1992), then it is in the top bracket, the best if not perfect practice. If the paper chose between 6 and 8 actions correctly, as Gary Becker, Michael Grossman, and Kevin Murphy did (June 1994), then it is in the fourth bracket, second to last.

Joshua D. Angrist does well in his "The Economic Returns to Schooling in the West Bank and Gaza Strip" (Dec 1995 1065-1087). "Until 1972," Angrist writes, "there were no institutions of higher education in these territories. Beginning in 1972. . . higher education began to open in the West Bank. Previously, Palestinian residents of the territories had to obtain their advanced schooling abroad. But by 1986, there were 20 institutions granting post-high school degrees in the territories. As a consequence, in the early and mid 1980's, the labor market was flooded with new college graduates. This paper studies the impact of this dramatic influx of skilled workers on the distribution of wages in the occupied territories" (1064). In a first regression Angrist estimates the magnitude of wage premia earned by Israelis and Palestinians who work in Israel:

The first column of Table 2 shows that the daily wage premium for working in Israel fell from roughly 18 percent in 1981 to zero in 1984. Beginning in 1986, the Israel wage premium rose steeply. By 1989, daily wages paid to Palestinians working in Israel were 37 percent higher than local wages, nearly doubling the 1987 wage differential. The monthly wage premium for working in Israel increased similarly. These changes parallel the pattern of Palestinian absences from work and are consistent with movements along an inelastic demand curve for Palestinian labor. (1072)

The reader is told magnitudes. She knows the oomph.

Yet even Angrist falls back into asterisk econometrics. On page 1079 he is testing alternative models, and emphasizes that:

The alternative tests are not significantly different in five out of nine comparisons ($p < 0.02$), but the joint test of

coefficient equality for the alternative estimates of θ leads to rejection of the null hypothesis of equality. (1079)

To which his better nature would say, “So?”

David Zimmerman, in his “Regression Toward Mediocrity in Economic Stature” (1992), and especially the well-named Gary Solon, in his “Intergenerational Income Mobility in the United States” (1992), have set an admirable if rare standard for the field. Line by line Solon asks the question “How much?” and then gives an answer. How much, he wonders, is a son’s economic well-being fated by that of his father? The sign, the star, the sign-and-the-star-together, don’t tell. Previous estimates, observes Solon, had put the father-son income correlation at about 0.2 (394). A new estimate, a tightly fit correlation of 0.20000000001***, would say nothing new of *economic* significance. And a poorly fit correlation with the “expected sign” would say nothing. Nothing at all. Solon’s attempts at a new estimate, on pages 397 to 405, refer only once to statistical significance (404). Instead, Solon writes 18 paragraphs on *economic* significance: why he believes the “intergenerational income correlation in the United States is [in fact] around 0.4” (403) and how the higher correlation changes American stories about mobility. Solon’s paper is three standard deviations above the average of the *AER*.

“Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania” by David Card and Alan B. Krueger (1994a), falls far below the median for cogency in statistical testing, though well *above* the median in other features of scientific seriousness. Card and Krueger designed their own surveys, collected their own data, talked on the telephone with firms in their sample, and visited firms that did and did not respond to their survey, all of which is most unusual among economists, and seems to have raised scientific standards in the field. It matches the typical procedure in economic history, for example, or the best in empirical sociology and experimental physics. Their sample was designed to study prices, wages, output, and employment in the fast food industry in Eastern Pennsylvania and Western New Jersey before and after New Jersey raised its minimum wage above the national and Pennsylvania levels. On pages 775-776 of the article (and pages 30-33 in their widely cited book [1994b]), Card and Krueger report their crucial test of the conventional labor market model.

Table 4
Author Rankings by Economic Significance

Number Yes, that is, Good, in the 19-Question Survey of the 1990s [Year and Month of Publication in Brackets]		
15-19:Yes	12-14:Yes	
Solon [9206] Zimmerman [9206] Goldin [9109] Craig and Pencavel [9212] Anderson and Holt [9712] Ransom [9303] Allen [9203] Ausubel [9012]	Simon [9812] Angrist and Evans [9806] Berk, Hughson, and Vandezande [9609] Myagkov and Plott [9712] Gordon and Bovenberg [9612] Angrist [9512] Gilligan [9212] Hoover and Sheffrin [9203] Benhabib and Jovanovic [9103] Angrist [9006] Cecchetti, Lam, and Mark [9006] Baker and Benjamin [9709] Paxson [9203] Blank [9112] Froot and Obstfeld [9112]	
9-11:Yes	6-8:Yes	Less than 6:Yes
Brainerd [9812] Calomiris and Mason [9712] Morrison and Schwartz [9612] Landers, Rebitzer, and Taylor [9606] Guiso, Jappell, and Terlizzese [9603] Borjas [9506] Kaminsky [9306] Calvo and Leiderman [9203] Fair and Shiller [9006] Sauer and Leffler [9003] Schachar and Nalebuff [9906] Craft [9812] Dyck [9709] Genesove and Meyer	Mendelson, Nordhaus, Shaw [9409] Fernald [9906] Gali [9903] Murray, Evans, and Schwab [9809] Alesina and Peroti [9712] Harrigan [9709] Dorwick and Quiggin [9703] Chevalier and Scharfstein [9609] Levin, Kagel, and Richard [9606] Trefler [9512] Feldstein [9506] Mark [9503] Ashenfelter and Krueger [9412] Gale and Scholz [9412] Cohen [9306]	Frankel and Romer [9906] Kroszner and Stratman [9812] Bernard and Jones [9612] Munnell, Tootell, Browne, and McEneaney [9603] Attanasio and Browning [9512] Marin and Schnitzer [9512] Chevalier [9506] Currie and Thomas [9506] Bronars and Grogger [9412] Kroszner and Rajan [9409]

[9706]	Altonji, Hayashi, and Kotlikoff [9212]	Kim and Singal [9306]
Pontiff [9703]	Bernanke and Blinder [9209]	Bronars and Deere [9303]
Rosenszweig and Wolpin [9412]	Card [9009]	Kashyap, Stein, and Wilcox [9303]
Currie and McConnell [9109]	Aitken and Harrison [9906]	Falvey and Gemmell [9112]
Hendry and Ericsson [9103]	Levine and Zervos [9806]	Keeley [9012]
Pitt, Rosenzweig, and Hassan [9012]	Blonigen [9706]	Ramey and Ramey [9512]
Berry, Levinsohn, and Pakes [9906]	Hines [9612]	Hamermesh and Biddle [9412]
Yano and Nugent [9906]	Henderson [9609]	Keane [9309]
Ham, Sveinar, and Terrell [9812]	Laitner and Juster [9609]	Grossman [9209]
Hallock [9809]	Grinblatt, Titman, and Wermers [9512]	Cukierman, Edwards, and Tabellini [9206]
Rajan and Zingales [9806]	Lemieux, Fortin, and Frechette [9403]	Wolak and Kolstad [9106]
Ichnowski, Shaw, and Prennushi [9706]	Hanes [9309]	Keane and Runkle [9009]
Nalbantian and Schotter [9706]	Blundell, Pashardes, and Weber [9306]	Roberts and Tybout [9709]
Wilhelm [9609]	Kachelmeier and Shehata [9212]	Engel and Rogers [9612]
Fuchs [9603]	Wolff [9106]	Besley and Case [9503]
Rotemberg and Woodford [9603]	Hardouvelis [9009]	Levine and Renelt [9209]
Griliches and Cockburn [9412]	Wright [9009]	Trejo [9109]
James [9309]	Card and Krueger [9409]	Brainard [9709]
Forsythe, Nelson, Neumann, and Wright [9212]	Burman and Randolph [9409]	Bernheim and Wantz [9506]
Stratman [9212]	Palfrey and Prisbrey [9712]	
Lin [9203]	Peek and Rosengren [9709]	
Viscusi and Evans [9006]	Levitt [9706]	
	Cardia [9703]	
	Hamilton [9703]	
	Foster and Rosenweig [9609]	
	Hendricks and Porter [9606]	
	Ayers and Siegelman [9506]	
	Jones and Kato [9506]	
	Meyer, Viscusi, and Durbin [9506]	
	Fuhrer and Moore [9503]	
	Shea [9503]	
	Becker, Grossman, and Murphy [9406]	
	Persson and Tabellini [9406]	
	Alogoskoufis and Smith [9112]	
	Fair and Dominguez [9112]	

The chief prediction of the conventional model is that full-time equivalent employment in New Jersey relative to Pennsylvania would fall following the increase in the New Jersey minimum wage. Specifically Card and Krueger's null hypothesis says that the difference-in-difference is zero—that “change in employment in New Jersey” minus “change in employment in Pennsylvania” should equal zero if as they suppose the minimum wage is *not* oomphul. If they find that the difference-in-difference is zero (other things equal), then New Jersey gets the wage gains without loss of employment: a good thing for workers. Otherwise, New Jersey employment under the raised minimum wage will fall, perhaps by a lot: a bad thing for workers.

Yet Card and Krueger fail to test the null they claim. Instead they test two distinct nulls, “change in employment in New Jersey = zero” and (in a separate test) “change in employment in Pennsylvania = zero.” In other words, they compute *t*-tests for each state, examining average full-time equivalent employment before and after the increase in the minimum wage. But they do not test the (relevant) difference-in-difference null of zero. Card and Krueger report on page 776 a point estimate suggesting employment in New Jersey increased by “0.6” of a worker per firm (from 20.4 to 21; rather than falling as enemies of the minimum wage would have expected). Then they report a second point estimate suggesting that employment in Pennsylvania fell by 2.1 workers per firm (from 23.3 to 21.2). “Despite the increase in wages,” they conclude from the estimates, “full-time equivalent employment *increased* in New Jersey relative to Pennsylvania. Whereas New Jersey stores were initially smaller, employment gains in New Jersey coupled with losses in Pennsylvania led to a small and statistically insignificant interstate difference in wave 2” (776; their emphasis). The errors are multiple: Card and Krueger run the wrong test (testing the wrong null, by the way, was less common in the *AER* during the 1980s [Table 1, Question 4]); they “reject” a null of zero change in employment in New Jersey, having found an average difference, estimated noisily at $t = 0.2$, of 0.6 workers per firm; they do not discuss the power of their tests, though the Pennsylvania sample is larger by a factor of 5; they practice asterisk econometrics (with a “small and statistically insignificant interstate difference”); and yet they emphasize *acceptance* of their favored alternative, with italics. Further attempts to measure with multiple regression analysis the size of the employment effect, the price effect, and the output effect, though technically improved, are not argued in terms of economic significance. That's the main point, after all: how big is big.

The cost of following the wrong decision rule is especially clear in "An Empirical Analysis of Cigarette Addiction" by Gary Becker, Michael Grossman, and Kevin Murphy (June 1999; you can see that we are anxious not to be accused of making our lives easy by picking on the less eminent economic scientists). Sign econometrics and asterisk econometrics decide nearly everything in the paper, but most importantly the "existence" of addiction.

Our estimation strategy is to begin with the myopic model. We then test the myopic model by testing whether future prices are significant predictors of current consumption as they would be in the rational-addictive model, but not under the myopic model (403). . . . According to the parameter estimates of the myopic model presented in Table 2, cigarette smoking is inversely related to current price and positively related to income.

And then: "The highly significant effects of the smuggling variables (ldtax, sdimp, and sdexp) indicate the importance of interstate smuggling of cigarettes."

But as Kruskal put it, echoing Neyman and Pearson from 1933, "The adverb 'statistically' is often omitted, and this is unfortunate, since statistical significance of a sample bears no necessary relationship to possible subject-matter significance of whatever true departure from the null hypothesis might obtain" (Kruskal 1968a, 240). At $N =$ about 1,400 with high power they can reject a nearby alternative to the null—an alternative different, *but trivially different*, from the null (at high sample sizes, after all s/\sqrt{N} approaches zero: all hypotheses are rejected, and in mathematical fact, without having to look at the data, you know they will be rejected at any pre-assigned level of significance). Yet they conclude that "the positive and significant past-consumption coefficient is consistent with the hypothesis that cigarette smoking is an addictive behavior" (404). It's sign econometrics, with policy implications. When sign econometrics meets asterisk econometrics the mystification redoubles:

When the one-period lead of price is added to the 2SLS models in Table 2, its coefficient is negative and significant at all conventional levels. The absolute t ratio associated with the coefficient of this variable is 5.06 in model (i), 5.54 in model (ii), and 6.45 in model (iii). These results

suggest that decisions about current consumption depend on future price. They are inconsistent with a myopic model of addiction, but consistent with a rational model of this behavior in which a reduction in expected future price raises expected future consumption, which in turn raises current consumption. While the tests soundly reject the myopic model, [and so forth]. (404)

Eventually they report (though never interpret) the estimated magnitudes of the price elasticities of demand for cigarettes. But their way of finding the elasticities is erroneous. Cigarette smoking may be addictive. But Becker, Grossman, and Murphy have not shown why, or how much. (They are, incidentally, inferring individual behavior from state-wide data; sociologists call this the ecological fallacy.) Perhaps what they have shown is that statistics play multiple roles.

There are some other roles that activities called “statistical” may, unfortunately, play. Two such misguided roles are (1) to sanctify or provide seals of approval (one hears, for example, of thesis advisors or journal editors who insist on certain formal statistical procedures, whether or not they are appropriate); (2) to impress, obfuscate, or mystify (for example, some social science research papers contain masses of undigested formulas [or tests of significance] that serve no purpose except that of indicating what a bright fellow the author is). (Kruskal 1968b, 209)

Table 5 shows what happens if statistical significance is the only criterion of importance at first use. In a large number of cases, if only statistical significance is said to be of importance as its first use, then statistical significance tends to decide the entire empirical argument. Of the 137 full length papers in the 1990s, 80 papers made both mistakes (Question 7=0 and Question 16=0). To put it differently, of the 87 papers using only statistical significance as a criterion of importance at first use, fully 92% considered statistical significance the last word. Cross tabulations on the 1980s data reveal a similar though slightly better record (Table 5).

Table 5
If Only Statistical Significance Is Said To Be Of Importance At Its
First Use (Questions 7), Then Statistical Significance Tends To
Decide The Entire Argument

In the 1990s...						
Does not consider the test decisive (Question 16)						
Considers more than the test at the first use (Question 7)				0	1	Total
	0	80	7	87		
	1	32	18	50		
	Total	112	25	137		

Notes: `0' means "no, did the wrong thing;" `1' means "yes, did the right thing." In the 1980s data, when Question 7=0 and Question 16=0 the first row becomes 86-10-96 [McCloskey and Ziliak 1996, Table 1 and Table 5]; in other words, practice was in this additional sense somewhat better in the 1980s.

WE ARE NOT ORIGINAL

We are not the first social scientists to make the distinction between economic and statistical significance. One of us has been making the point since 1985 (McCloskey 1985a, 1985b, 1992, 1995), but she learned it from a long, long line of distinguished if lonely protesters of the arbitrary procedures laid down in the 1920s by the blessed Fisher. We have pointed out before that in the 1930s Neyman and Pearson and then especially Abraham Wald had distinguished sharply between practical and statistical significance (McCloskey and Ziliak 1996, 97-98; McCloskey 1985a). But Wald died young, and Neyman and Pearson carried the day only at the level of high-brow statistical theory (and Fisher we have just noted failed to measure or mention the matters of substantive significance that occupied Wald and Neyman and Pearson [Fisher 1925 (1941), 42, 123-124, 138-140, 326-329]). Statistical practice on the ground stayed with a predetermined level of 5% significance (mainly), regardless of the loss function, misleading even the Supreme Court of the United States.

Yet some simple souls got it right. Educators have written about the difference between substantive and statistical significance early and late (Tyler 1931; Shulman 1970; Carver 1978). Psychologists have known about the difference for nearly a century, though most of them continue like economists to ignore it (a committee of the American Psychological

Association was recently charged to re-open the question). In 1919 an eminent experimental psychologist, the alarmingly named Edwin Boring, published an article unmasking the confusion between arbitrarily-judged-statistical significance and practical significance (Boring 1919). And empirical sociology would be less easy for economists to sneer at if more realized that a good many sociologists grasped the elementary statistical point decades before even a handful of the economists did (Morrison and Henkel 1970).

Of late the protest has grown a little louder, but is still scattered (we detailed in the 1996 paper the evidence that almost all econometrics textbooks teach the students to ignore substantive significance in favor of testing without a loss function and without substantive judgments of the size of coefficients). James Berger and Robert Wolpert in 1988, though making a slightly different point (the Bayesian one that Jeffreys and Zellner emphasize), noted the large number of theoretical statisticians engaging in “discussions of important practical issues such as ‘real world’ versus ‘statistical’ significance”: Edwards, Lindman, and Savage (1963), I. J. Good (1981), and the like. What we find bizarre is that in the mainstream statistical literature this “important” point is hardly mentioned (we found in our 1996 article, though, some honorable exceptions, such as the first edition of the elementary text by Freedman, Pisani, and Purvis [1978; we note with alarm that later editions have soft-peddled the issue]). Among economists the roll of honor is likewise short but distinguished. J. M. Keynes (virtually), Oskar Lange, Arnold Zellner, Arthur Goldberger, A. C. Darnell, Clive Granger, Edward Leamer, Milton Friedman, Robert Solow, Kenneth Arrow, Zvi Griliches, Glen Cain, Gordon Tullock, Gary Solon, Daniel Hamermesh, Thomas Mayer, David Colander, Jeffrey Wooldridge, Jan Magnus, and Hugo Keuzenkamp are not dunces and they haven’t minced words (Lange [1959], 13-15, 133-157 [on page 151 Lange speaks of “practical significance,” his main concern]; Cain and Watts 1970, 229, 231-232; Keuzenkamp and Magnus 1995; McCloskey and Ziliak 1996, 99 and numerous other references on 112-114; McCloskey’s citations in her works cited; Darnell’s comprehensive review of 1997; Hamermesh 1999; Colander 2000; Wooldridge 2000, 131-134; Keuzenkamp 2000, 266; and so forth). Recently, to pick one among the small, bright stream of revisions of standard practice that appear in our mailboxes, Clinton Greene (2003) has applied the argument to time-series econometrics, showing that tests of cointegration based on arbitrary levels of significance miss the economic point: they are neither necessary nor sufficient.

We are sometimes told that “You’re rehashing issues decided in the 1950s” or “Sure, sure: but the hot *new* issue is [such and such new form of specification error, say]” or “I have a metaphysical argument for why a universe should be viewed like a sample.” When we are able to get such people-in-a-hurry to slow down and listen to what we are saying (which is not often), we discover that in fact they do *not* grasp our main point, and their own practice shows why. It is dangerous, for example, to mention Bayes in this connection, because the reflexive reply of most econometrically minded folk is to say “1950s” and have done with it. Our point is not Bayesian (although we honor the Bayesians such as Leamer and Zellner who have made similar--and also some different—criticisms of econometric practice). Our (idiotically simple) point has nothing to do with Bayes’ Theorem: it applies to the most virginal classical regressions.

Our experience is that in the rare cases when people *do* grasp our point—that fit and importance are not the same---they are appalled. They realize that almost everything that has been done in econometrics since the beginning needs to be redone. The wrong variables have been included, for example (which is to say errors in specification have vitiated the conclusions); tiny coefficients have inflated reputations; mistaken policies have been recommended; science has stopped.

We believe we have shown from our evidence in the *American Economic Review* over the two last decades what scientists from Edgeworth to Goldberger have been saying: science is about magnitudes. Seldom is the magnitude of the sampling error the chief scientific issue. (A sympathetic reader might reply it's not the size that counts; it's what you do with it. But that too is mistaken. As Friedman's alloy regression and hundreds of other statistical experiments reveal, what matters is size *and* what you do with it. Scientific judgment, like any judgment, is about loss functions—what R. A. Fisher was most persistent in denying.)

WHAT SHOULD ECONOMISTS DO?

We should act more like the Gary Solons and the Claudia Goldins. We should be economic scientists, not machines of walking dead recording 5% levels of significance. In his acceptance speech for the Nobel Prize, Bob Solow put it this way:

[Economists] should try very hard to be scientific with a small *s*. By that I mean only that we should think logically and respect fact. . . . Now I want to say something about fact. The austere view is that “facts” are just time series of prices and quantities. The rest is all hypothesis testing. I have seen a lot of those tests. They are almost never convincing, primarily because one senses that they have very low power against lots of alternatives. There are too many ways to explain a bunch of time series. . . . My hunch is that we can make progress only by enlarging the class of eligible facts to include, say, the opinions and casual generalizations of experts and market participants, attitudinal surveys, institutional regularities, even our own judgments of plausibility. (Solow 1988)

Solow recommends we “try very hard to be scientific with a small *s*”; the authors we have surveyed in the *AER*, by contrast, are trying very hard to be scientific with a small *t*. As Solow says, it’s almost never convincing.

What to do? One of us was advised to remove the 1996 article from his CV while job hunting—it wasn’t “serious” research. Shut up and follow R. A. Fisher. The other served fleetingly on the editorial board of the *AER*. Each time she saw the emperor had no clothes of oomph she said so (by the way, in the original Danish of the story the child is *not* identified as to gender: we think it was probably a little *girl*). The behavior did not endear her to the editors. After a while she and they decided amiably to part company.

The situation was strange: economic scientists, for example those who submit and publish papers in the *AER*, or serve on hiring committees, routinely violate elementary standards of statistical cogency. And yet it is the messengers who are to be taken out and shot. This should stop. We should revise publication standards, and cease shooting messengers who bring the old news that fit is not the same thing as importance. If the *AER* were to test papers for cogency, and refused to publish papers that used fit irrelevantly as a standard of oomph, economics would in a few years be transformed into a field with empirical standards. At present (we can say until someone starts claiming that *in the 2000s* practice has improved), we have shown, it has none. Ask: “Is the paper mainly about showing and measuring *economic* significance?” If not, the editor and referees should reject it. It will not reach correct scientific results. Its findings will be biased by misspecification and mistaken as to oomph. (Requiring referees to complete

a 19-item questionnaire would probably go against the libertarian grain of the field; a short form would do: “Does the paper focus on the *size* of the effect it is trying to measure, or does it instead recur to irrelevant tests of the coefficient’s *statistical* significance?”) To do otherwise—continuing to decorate our papers with stars and signs while failing to interpret size—is to discard our best unbiased estimators, and to renege on the promise of empirical economics: measurement. No size, we should say, no significance.

REFERENCES

- American Economic Review*. January 1980 to December 1999. The 319 full-length papers using tests of statistical significance. May Supplement excluded.
- Angrist, Joseph.** 1995. The Economic Returns to Schooling in the West Bank and Gaza Strip. *American Economic Review* 85(5): 1065-1086.
- Baird, Davis.** 1988. Significance tests, history and logic. In *Encyclopedia of Statistical Sciences* 8, ed. S. Kotz and N.L. Johnson. New York: John Wiley, 466-471.
- Becker, Gary S., Michael Grossman, and Kevin M. Murphy.** 1994. An Empirical Analysis of Cigarette Addiction. *American Economic Review* 84(3): 396-418.
- Berger, James O. and Robert L. Wolpert.** 1988. *The Likelihood Principle*, 2nd ed. Hayward, CA: Institute of Mathematical Statistics.
- Bernanke, Ben S. and Alan S. Blinder.** 1992. The Federal Funds Rate and the Channels of Monetary Transmission. *American Economic Review* 82(4): 901-921.
- Bernheim, B. Douglas and Adam Wantz.** 1995. A Tax-Based Test of the Dividend Signaling Hypothesis. *American Economic Review* 85(3): 532-551.
- Boring, Edwin G.** 1919. Mathematical versus Scientific Significance. *Psychological Bulletin* 16(10): 335-38.
- Cain, Glen G. and Harold W. Watts.** 1970. Problems in Making Policy Inferences from the Coleman Report. *American Sociological Review* 35(2): 228-242.

- Card, David and Alan B. Krueger.** 1994a. Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. *American Economic Review* 48(2): 772-793.
- Card, David and Alan B. Krueger.** 1994b. *Myth and Measurement: the New Economics of the Minimum Wage*. Princeton: Princeton University Press.
- Carter, Ronald P.** 1978. The Case Against Statistical Significance Testing. *Harvard Educational Review* 48(3): 378-398.
- Colander, David.** 2000. New Millenium Economics: How Did It Get This Way, and What Way Is It? *Journal of Economic Perspectives* 14(1): 121-132.
- Darnell, A. C.** 1997. Imprecise Tests and Imprecise Hypotheses. *Scottish Journal of Political Economy* 44 (3): 247-268.
- Edgeworth, Francis Y.** 1885. Methods of Statistics. *Jubilee Volume of the Statistical Society*, June 22-24. Royal Statistical Society of Britain: 181-217
- Edwards, W., H. Lindman, and L. J. Savage.** 1963. Bayesian Statistical Inference for Psychological Research. *Psychological Review* 70(3): 193-242.
- Fisher, Ronald A.** [1925] 1941. *Statistical Methods for Research Workers* 8th ed. New York: G. E. Stechart and Co. 8th edition.
- Friedman, Milton.** 1985. In *Lives of the Laureates*, ed. W. Breit and R.W. Spencer. Cambridge: MIT Press, 77-92. Selection reprinted: M. Friedman and A. J. Schwartz. 1991. Alternative Approaches to Analyzing Data. *American Economic Review* 81(1): 39-49.
- Good, I. J.** 1981. Some Logic and History of Hypothesis Testing. In *Philosophy in Economics*, ed. J.C. Pitt. Dordrecht, The Netherlands: Reidel.
- Greene, Clinton A.** 2003. Towards Economic Measures of Cointegration and Non-Cointegration. Unpublished paper, Department of Economics, University of Missouri, St. Louis. April. Email: clinton_greene@umsl.edu
- Hamermesh, Daniel S.** 1999. *The Art of Labormetrics*. Cambridge, MA: National Bureau of Economic Research, Inc.
- Hendricks, Kenneth and Robert H. Porter.** 1996. The Timing and Incidence of Exploratory Drilling on Offshore Wildcat Tracts. *American Economic Review* 86(3): 388-407.
- Kachelmeier, Steven J. and Mohamed Shehata.** 1992. Examining Risk Preferences Under High Monetary Incentives: Experimental Evidence

from the People's Republic of China. *American Economic Review* 82(5): 1120-1141.

Keuzenkamp, Hugo A. 2000. *Probability, Econometrics and Truth*. Cambridge: Cambridge University Press.

Keuzenkamp, Hugo A. and Jan Magnus. 1995. On Tests and Significance in Econometrics. *Journal of Econometrics* 67(1): 103-28.

Kruskal, William S. 2002. Personal interview. August 16, 2002, University of Chicago.

Kruskal, William S. 1968a. Tests of Statistical Significance. In *International Encyclopedia of the Social Sciences* 14, ed. David Sills. New York: MacMillan, 238-250.

Kruskal, William S. 1968b. Statistics: The Field. In *International Encyclopedia of the Social Sciences* 15, ed. David Sills. New York: MacMillan. 206-224

Lange, Oskar. 1959 [1978]. *Introduction to Econometrics*. Warsaw: PWN—Polish Scientific Publishers, and Oxford, England: Pergamon Press.

McCloskey, Deirdre, and Stephen Ziliak. 1996. The Standard Error of Regressions. *Journal of Economic Literature* 34(March): 97-114.

McCloskey, Deirdre. 1985. The Loss Function Has Been Mislaid: The Rhetoric of Significance Tests. *American Economic Review*, Supplement 75 (2, May): 201-205.

McCloskey, Deirdre. 1995. The Insignificance of Statistical Significance, *Scientific American* April: 32-33.

McCloskey, Deirdre. 1992. The Bankruptcy of Statistical Significance, *Eastern Economic Journal* 18(Summer): 359-361.

McCloskey, Deirdre. 1985b. *The Rhetoric of Economics*. Especially chapters 8 and 9. Madison: University of Wisconsin Press.

Morrison, Denton E. and Ramon E. Henkel. 1970. *The Significance Test Controversy: A Reader*. Chicago: Aldine.

Shulman, L.S. 1970. Reconstruction of Educational Research. *Review of Educational Research* 40: 371-393.

Solon, Gary. 1992. Intergenerational Income Mobility in the United States. *American Economic Review* 82(3): 393-408.

- Solow, Robert.** 1988. In *Lives of the Laureates*, ed. W. Breit and R.W. Spencer. Cambridge: MIT Press.
- Twain, Mark.** [1883] 1946. *Life on the Mississippi*. New York: Bantam.
- Tyler, R.W.** 1931. What is Statistical Significance? *Educational Research Bulletin* 10: 115-118, 142.
- Wooldridge, Jeffrey M.** 2000. *Introductory Econometrics*. U.S.A.: South-Western College Publishing (Thomson Learning).
- Zellner, Arnold.** 1984. *Basic Issues in Econometrics*. Chicago: University of Chicago Press.
- Zimmerman, David J.** 1992. Regression Toward Mediocrity in Economic Stature. *American Economic Review* 82 (3): 409-429.

ABOUT THE AUTHORS



Steve Ziliak is Associate Professor Economics in the School of Policy Studies at Roosevelt University. Professor Ziliak has previously taught at Emory University, Bowling Green State University, and the Georgia Institute of Technology. In 1996 he earned the Ph.D. in Economics at the University of Iowa and at the same time the Ph.D. Certificate in the Rhetorics of the Human Sciences, also at the University of Iowa. Ziliak's current research concerns are in economic history and the rhetoric of economics. He is especially interested in 19th and 20th century histories of welfare and poverty in the United States and in the philosophy and rhetoric of economics in public affairs. With Deirdre McCloskey he is presently writing a book on the misuse of significance testing in the life and human sciences.



Deirdre McCloskey is the Distinguished University Professor of Economics, English, and History at the University of Illinois-Chicago and the Distinguished Professor of Economics, Philosophy, and Art and Cultural Studies at Erasmus University-Rotterdam. Professor McCloskey has published many books and articles but she is perhaps best known for her book *The Rhetoric of Economics* (1985 [1998]), a classic in the field. She is currently at work on a four-volume book entitled *Bourgeois Virtues*.



*Econ Journal Watch,
Volume 1, Number 2,
August 2004, pp 359-368.*

CHARACTER ISSUES

Preference Falsification in the Economics Profession

WILLIAM L. DAVIS*

[Abstract, Keywords, JEL Codes](#)

Mathematical and statistical rigor is a prized aspect of academic economic research. The economics profession officially claims that such rigor advances our understanding of the economy and public policy issues, that there is, in effect, a public good created by research published in economics journals. Yet, a majority of AEA members who responded to a survey I conducted admit, at least privately, that academic research mainly benefits academic researchers who use it to advance their own careers and that journal articles have very little impact on our understanding of the real world and the practice of public policy. Understanding why this contradiction between official position and private belief exists and persists is important if we hope to fulfill the original promise of economics to serve society.

* Department of Economics, University of Tennessee at Martin.
I am indebted to the anonymous *EJW* referee of this paper for his or her fruitful comments.

THE CHARACTER AND MISSION OF ECONOMICS: TWO ORIENTATIONS

Economists must fashion their work along many dimensions, including the kind of subjects to address, the audiences to address, the methods to use, the style to adopt, and so on. Following in the footsteps of Deirdre McCloskey, Thomas Mayer, David Colander, Arjo Klammer, and many others, Dan Klein (2001) has argued that the rich texture of dimensions can be usefully simplified as a tension between two orientations: namely, the scholastic orientation and the public discourse orientation. The scholastic affirms academic norms about the various questions, for their own sake, with the belief that academic institutions, like an invisible hand, channel intellectual effort toward long-run enlightenment and better decision making. The public discourse orientation emphasizes communicating with lay people by addressing issues in the policy discourse, being more relevant and outspoken, and hence not adhering to paradigmatic modes of discourse, such as model building and statistical significance.

Currently, the scholastic orientation reigns among the editors and reviewers of most academic journals—and, apparently, among their economist-contributors. But economists who want to publish in such journals must toe the line, even if they are not enthusiastic about the scholastic orientation. The economist who publicly voices discontent risks his or her professional reputation among so-called mainstream economists who support the scholastic orientation, some of whom may have influence over that person's career.

This claim is widely recognized, although difficult to confirm. The scholastic influence appears to begin in graduate school where future economists are taught about the profession's priorities, including its distinction between important and unimportant topics, and which methods are acceptable for studying those topics. In their book *The Making of an Economist*, Klammer and Colander (1990) conducted numerous interviews with graduate students in economics. They summarized the interviews with the following quotes:

The interviews suggested a definite tension, frustration, and cynicism that, in our view, went beyond the normal graduate school blues. There was a strong sense that

economics was a game and that hard work in devising relevant models that demonstrated a deep understanding of institutions would have a lower payoff than devising models that were analytically neat; the façade, not the depth of knowledge, was important. This cynicism is not limited to the graduate school experience, but is applied also to the state of the art as they perceive it. (Klamer and Colander 1990, 18)

Furthermore, disillusionment and cynicism are not conditions limited to graduate students. Many members of the profession appear to lack faith in what they do. They will confess, usually at unguarded moments, that their highly sophisticated research produces ultimately meaningless results—but they will demand their students follow their lead anyway. (Klamer and Colander 1990, 184)

The intergenerational influence on economists continues long after the completion of graduate school via different means. The work in the “top” journals is typically held in high esteem among the economics profession, even if it has little or no relevance to the public discourse of society’s problems. Through their academic journals, professional associations control the modes of professional discourse, largely restricting work to mathematical model-building and statistical significance. As a result, the dominant modes of discourse—and all the associated limitations and blind spots—are preserved.

PREFERENCE FALSIFICATION

The University of Southern California economist Timur Kuran developed a theory of *preference falsification* to explain why some government policies and social practices go on for such a long time and then can suddenly, and dramatically, change. In his book *Private Truths, Public Lies: The Social Consequences of Preference Falsification* (Harvard University Press, 1995), Kuran suggests that when individuals have private preferences about a public matter they often falsify their preferences about the matter out of a desire to maintain acceptance and respect. Preference falsification leads to

inefficiencies and propagates ignorance. Kuran's work has attracted great attention throughout the social sciences, and the book has been translated into German, Swedish, Turkish, and Chinese.

My hypothesis is that many members of the economics profession exhibit preference falsification. By falsifying his preferences, an economist can publicly support the scholastic orientation, thereby not jeopardizing his reputation, while actually holding beliefs to the contrary. We do not know the extent of such behavior. Preference falsification is hard to show, since only the individual—and sometimes not even the individual—can know whether his publicly expressed views correspond with his true, privately held beliefs. One possible method to detect preference falsification is the use of a survey that elicits private preferences without compromising an individual's public reputation.

Recently I conducted an anonymous survey of professional economists that explored their attitudes about the economics profession. The paper is forthcoming in the *American Journal of Economics and Sociology* (Davis, 2004). Although the survey did not employ the “scholastic versus public discourse” formulation, it addressed that tension. Here I summarize some of the results and suggest that they reveal a significant degree of disenchantment with the scholastic orientation of the profession. After summarizing the results, I will follow Kuran's theory and discuss possible consequences of preference falsification in the economics profession.

SURVEY RESULTS INDICATE DISENCHANTMENT WITH SCHOLASTICISM

The survey, conducted among a randomly selected sample of 1,000 AEA members, attempted to ascertain economists' perceptions of the progress of economic research, its usefulness for society, some of the factors that determine research publication, and the influence of gender and race on economic research. The summary here, based on the 373 responses received, is made with a broad brush, in fairness to the journal that is publishing the full paper. The reader can consult the paper for more precise reporting.

Contrary to the official image of economics journals serving as forums to advance understanding by presenting quality, meaningful research, a majority of responding economists agree that research that lacks

a mathematical component is less likely to be published in an economics journal and that school affiliation and author recognition are factors used by editors and reviewers to determine what gets published. Further, most economists believe a “good-old-boy” network in the profession influences which articles are published in nationally recognized economics journals, and that most articles in nationally recognized journal are published, at least partly, for reasons other their contribution to economic science.

More to the point of “scholastic” vs. “public discourse” orientation, the majority say that economic research in nationally recognized journals is not useful for individuals in business or industry or for teachers of college-level principles courses. It is mostly useful for academic economists engaged in research and for economics graduate students. And most economists agree that the profession and its members are ineffective at communicating with the lay public. Finally, a majority of responding economists either agreed with, or were neutral about, the statement that economic research in nationally recognized journals is not useful to government policy makers and does not provide spillover benefits for society.

All these findings support the view that, despite the publicly held view of economics, less than half of economists really positively believe in the official view of the profession as one that advances science, human understanding, and makes society better. Many economists, maybe around 50 percent, are at least somewhat disgruntled with the scholastic orientation of the profession.

The survey asked how many articles per year the respondent publishes in national journals, and indeed, the higher-output respondents generally had stronger support for the scholastic orientation. This may suggest that the disaffection of the respondents stemmed from low productivity. But even if the population is to some extent segmented in such a fashion, it remains that a huge portion of the profession does not believe in what the “leaders” say is professionally important. Outside the ivory tower, nonacademic economists generally believe that economic science has not improved its explanatory capacity over the last several years and therefore research is published for reasons other than a legitimate scientific contribution and does not entail spillover benefits to society.

Many respondents provided written comments that complained about the scholastic emphasis of the profession.

- “There is a crisis in economics. The bifurcation of the economics profession into researchers, teachers, and policy-makers has gotten

worse and the number of individuals who are respected for contributions in all three areas gotten fewer and farther between. As a consequence, the layperson and persons in business and industry have less and less respect for the economics profession. People want to understand the economy, but we are not helping them.”

- “The economics profession has failed miserably in its efforts to explain observable events. We have too often fallen in love with our models and used too little observed empirical evidence. The issue is not what our models tell us, but what statistical evidence reveals.”
- “Economists’ main audience seems to be other economists. That would be fine if taxpayers were not paying their salaries.”
- “The economics profession is a bad joke. More and more economists are saying less and less to fewer and fewer people. And they conceal their vacuity in abstruse language and mathematical formulae.”
- “An economic expert is someone who knows more and more about less and less until he knows absolutely everything about absolutely nothing.”
- “It is a shame that so many very talented people, as it is the case with PhDs in economics, spend so much valuable time writing, quite often, useless papers.”
- “The top journals have over-emphasized math, gaining an elegance of sorts but at the cost of . . . relevance.”
- “When are economists going to have the guts to criticize their own work?”

THE CONSEQUENCES OF PREFERENCE FALSIFICATION

Kuran uses his theory of preference falsification to explain a number of social patterns. Here I discuss some of the basic ideas and ask whether we see a parallel phenomenon in the economics profession. Two prominent consequences of preference falsification are observed in a society's inhibitions to change and its distortions of knowledge.

Preference falsification can inhibit change because a generation may influence public opinion long after its days are over with an established public opinion that carries over to future generations that have no responsibility to it. . . . A collective reluctance to change a bad political regime or public policy may be observed as a result. Preference falsification can distort knowledge through the removal of facts and arguments from public discourse that imparts credibility to myths by shielding them from corrective disclosures. (Kuran 1995, 114)

The persistence of India's caste system and the long duration of communism in the Soviet Union are, to an extent, products of preference falsification within these countries. In each case, according to Kuran, individuals living in these countries formed their public preferences, in part, on the basis of the reputational incentives associated with historical public opinion, even though their private preferences were at odds with that opinion. And while it is difficult to prove, Kuran observes that subtle signs of preference falsification were obvious in each country that perpetuated a popular public opinion at the expense of differing private opinion.

In India for instance, the caste system's vocal opponents are ostracized and threatened with punishment for questioning its wisdom and fairness. As a result, public discourse on the subject is diluted with distortions, and new generations of Indians are raised in a culture where the caste system is presumably favored by the masses. High-caste Indians perpetuate this belief by creating sophisticated reasons for its existence that ultimately become an important component of conventional Indian wisdom.

In the former Soviet Union public discourse was regulated with propaganda on a regular basis, causing citizens who challenged the

prevailing economic system to use euphemisms to disguise their complaints about Communism. When citizens publicly challenged an official position or a specific policy they did so by appearing sympathetic to broader communist goals. And while protests were allowed, they were generally confined to a “Party-defined zone of acceptability,” which meant that protestors never really criticized the doctrine of Communism itself.

The effect of such behavior in each case has been the reproduction of a public opinion over generations that perpetuated an inefficient policy or institution. And this in turn affects the public discourse of events over time. Further, Kuran argues that public discourse may deteriorate to the point where unthinkable beliefs ultimately become unthinkable thoughts.

An unthinkable belief is a thought that one cannot admit having, or even characterize as worth entertaining, without raising doubts about one’s civility, morality, loyalty, practicality, or sanity. An unthought belief is an idea that is not even entertained. . . . By transferring beliefs from the realm of the thinkable to that of the unthinkable, social pressures induce the withdrawal of those beliefs from public discourse. The consequent reconstitution of public discourse distorts private knowledge. (Kuran 1995, 176)

The economics profession, having a public character and being established for a long period of time, appears to be ripe with such attributes, which, to an extent, parallel the signs of preference falsification present in the cases just mentioned. These signs are manifested in the profession much the same way they were among the respective opponents of the caste system and communism. In fact, one could argue that the scholastic orientation has become conventional wisdom in economics and is sustained implicitly through the process of preference falsification, which has distorted knowledge by deemphasizing, removing, or ignoring facts and formulation from how economists conceive their own selfhood. The notion of changing from the scholastic orientation to the public discourse orientation seems to be unthinkable—at least publicly—and, for many, unthought. As a result, a reproduction of opinion within the profession perpetuated by a previous generation of economists has, and continues to, occur.

Preference falsification may be endemic in various spheres of all societies. Kuran suggests, however, that as more individuals become aware of the possibility of it, they will increase their efforts to encourage candid,

truthful expression. And the development of new media, new surreptitious criticism, and new ways for dissidents to develop and voice criticism, can help enliven fundamental debate in those societies. It can help dissidents to connect with others and suffer less from an existential sense of alienation.

Kuran's studies of political change show that old orthodoxies can be reformed and even deposed. Academic economics, however, is based on neither mass democracy nor consumer choice. Rather, the academic structure is self-referential, self-validating, and therefore self-perpetuating, and as such, the lay person—whether as tax-payer or as tuition payer—has little ability to change the situation. If the current economic scholasticism is to be challenged, it will not likely be by a revolution of lay people, but instead by growing cynicism within the profession. The results of my survey indicate precisely such developments.

Economists should read Kuran's work and take stock of its relevance to their own profession and selfhood.

REFERENCES

- Davis, William L.** 2004. Economists' Opinions of Economists' Work. Forthcoming. *American Journal of Economics and Sociology*.
- Klamer, Arjo and David Colander.** 1990. *The Making of an Economist*. Boulder, CO: Westview Press.
- Klein, Daniel B.** 2001. A Plea to the Economists Who Favor Liberty: Assist the Everyman. *Eastern Economic Journal* 27(2): 185-202.
- Kuran, Timur.** 1995. *Private Truths, Public Lies: The Social Consequences of Preference Falsification*. Cambridge, MA: Harvard University Press.

ABOUT THE AUTHOR



William L. Davis is Professor of Economics at the University of Tennessee at Martin where he has taught undergraduate and graduate courses for the last fifteen years and currently serves as the associate editor for the *Journal of Business and Economic Perspectives*. During his tenure at UT he has been a member of the Editorial Review Board for the University of Tennessee Press for many years and was past Director of Graduate Programs in Business Administration. He has been a visiting research fellow at the American Institute for Economic Research in Great Barrington, Massachusetts, and has received research grants from the Tennessee Valley Authority, Goodyear Tire and Rubber Company, and Kerr McGee Oil Company and other organizations. His articles have appeared in various social science and economics journals on topics ranging from drug prohibition to education reform. He earned his Ph.D in economics at Oklahoma State University.



*Econ Journal Watch,
Volume 1, Number 2,
August 2004, pp 369-379.*

CHARACTER ISSUES

Learning to Lose a Leg: Casualties of PhD Economics Training in Stockholm

ANNE D. BOSCHINI, MATTHEW J. LINDQUIST,
JAN PETTERSSON,* AND JESPER ROINE**

[Abstract, Keywords, JEL Codes](#)

THE SWEDISH ECONOMIST ASSAR LINDBECK HAS RECENTLY expressed concern that PhD programs are not educating enough “two-legged” economists. In the late 1980s Robert Eisner, president of the American Economic Association, appointed the Commission on Graduate Education in Economics (COGEE), to evaluate the state of economics graduate education in the United States. The commission delivered its recommendations in 1991 in a report that include an oft-quoted concern about graduate programs generating “too many *idiots savants*, skilled in technique but innocent of real economic issues” (Krueger et al. 1991, 1044-1045).¹ Around the same time many European universities were introducing

* Boschini, Lindquist, and Pettersson: Department of Economics, Stockholm University.

** Corresponding author: Department of Economics, Stockholm School of Economics.

¹ The members of the commission were Professors Anne Krueger (chair), Kenneth J. Arrow, Olivier Jean Blanchard, Alan S. Blinder, Claudia Goldin, Edward E. Leamer, Robert

"US-style" graduate programs. This adoption of US practices has in many ways lead to improvements but also to some problems.² Today, protests and concerns about the state of graduate education in economics can be heard all over Europe.³

In both the previous American debate, and the current European one, a central issue is the balance between the teaching of standard modes of research and the teaching of how to really understand and address public policy issues. When Assar Lindbeck said that PhD programs are not educating enough "two-legged economists" (Lindbeck 2001, 32), he meant economists "who both master analytical techniques and have a feel for real-world problems." The risk, according to Lindbeck, is that many young economists go on to apply the techniques in a mechanical way—lacking economic insight and failing to cultivate economic intuition. Another concern is that the "one-legged" economists are retreating from the public debate, allowing lobbyists to advance socially harmful positions. For example, Skedinger and Johansson (2004) demonstrate the lack of participation by economics professors in the public debate over globalization.

If too many economists are "one-legged," a central question must be whether that condition is a result of self-sorting or of reshaping. Do economics PhD programs attract students interested in the social sciences and public debate, or do they attract students with a prime interest in technique?

We surveyed all graduate students enrolled in the Stockholm Doctoral Program in Economics (SDPE) in October 2001. The program includes all of the PhD candidates in economics at the Stockholm School of Economics and Stockholm University.⁴ The SDPE is a good representative of

Lucas, John Panzar, Rudolph G. Penner, T. Paul Schultz, Joseph E. Stiglitz, and Lawrence H. Summers.

² See Frey and Eichenberger (1993) for a discussion of traditional differences between European and American economics.

³ Under the banner "autisme-économique" a protest movement began among French economics students in June 2000. This has been followed by other petitions, like the so-called "Cambridge Proposal," and has resulted in the Post-Autistic Economics network. See <http://www.paecon.net> for further information. Also, throughout Europe, there is a growing network of Hayekian students, organized around policy think tanks, research institutes, and certain university departments.

⁴ In terms of admission there are two separate programs; one at Stockholm University and one at the Stockholm School of Economics. However, the compulsory first-year courses are taken jointly by students in both programs, and students can also choose courses given at both schools. There are no other PhD programs in economics in Stockholm.

a strong European program, which in the early 1990s implemented an American-style research character.⁵ Though minor in scope, our survey is in the same spirit as the one conducted by David Colander and Arjo Klamer in the mid 1980s (Colander and Klamer 1987) which allows us to compare our results with some of theirs.

This article came about as a response to a debate among Swedish economists regarding the lack of policy and "real-world" interest among young economists. At the time, all four of us were enrolled in the PhD program at Stockholm University and we found the discussion rather simplistic. While established economists criticized graduate students for not participating in the policy debate, little attention was paid to the incentives in the PhD program for such engagement. We, therefore, formulated a short questionnaire to investigate whether the PhD students were "one-legged" at the time of admission to the program, or if they risked "losing a leg" in the process of becoming PhDs.

THE SURVEY

The questionnaire is very brief and lacks depth. Nonetheless, the high response rate (73 percent) gives us a high level of confidence in the veracity of the limited results obtained.

During the month of October 2001, we sent the questionnaire by e-mail to the 95 graduate students (57 male and 38 female) active in the Stockholm Doctoral Program in Economics (SDPE). Of these, 69 (44 male and 25 female) responded to the questionnaire, i.e., 73 percent. This can be compared to a response rate of about 25 percent, considered normal, for the much more detailed survey used by Colander and Klamer (1987). Answers were treated confidentially. The survey and responses follow.

⁵ The SDPE recruits students internationally and in this sense it is indeed a European, rather than Swedish program. Out of those enrolled in the program, one in four is not from Sweden. By "American-style" we mean a rather standardized two-year course-program and a thesis consisting of separate articles (rather than a monograph).

Stockholm Doctoral Program in Economics—Survey

1. How many credits/points did you have in mathematics, statistics, and economics before you began at the SDPE? [Note: The Swedish academic year has two semesters, fall and spring. Studying full time for one semester gives the student 20 credits and hence, one academic year of study gives 40 points. This was explained in the e-mail sent out with the questionnaire.]

Answer:	< 80	80 - 120	> 120
Responses (percentage):	13 (19)	26 (38)	30 (43)

2. How many credits/points did you have from courses in the social science and the humanities before you began studying at the SDPE? Please exclude those credits/points in economics that you reported in question (1) above.

Answer:	< 20	20 - 60	> 60
Responses (percentage):	11 (16)	25 (36)	31 (45)

[Note: 2 non-respondents]

3. Have you ever written or co-authored a debate/policy article in economics since you enrolled in the SPDE?

Answer:	Yes	No
Responses (percentage):	12 (17)	57 (83)

4. If not, why?

<u>Answer</u>	<u>Responses (percentage, of 57)</u>
Not interested:	2 (3.5)
Lack of time/incentives:	45 (79)
Other:	8 (14)
Other + Lack of time:	2 (3.5)

If Other, please write a short explanation here:

Responses of 8 who answered 'Other':

- One answered: "Lack of contacts . . . Who would publish a debate article by a Ph.D. student?"
- One answered that they did not want to be "stamped" with a particular political label.
- Two wrote nothing, while four actually wrote some form of lack of time + just starting the program.

5. Are you an active member of a volunteer organization and/or political party?

Answer:	Yes	No
Responses (percentage):	11 (16)	58 (84)

6. Why did you choose to begin a Ph.D. in economics? Please rank the following alternatives from 1 to 6 by putting a number in the box next to each statement, e.g., [2].

<u>Answer</u>	<u>Responses:</u>	
	<u>Ranked 1st by</u>	<u>Ranked 6th by</u>
[] My interest in math and/or statistics	0	10
[] My interest in the social sciences	45	3
[] Future career possibilities	10	0
[] My desire to serve in the community	8	10
[] To increase my earning potential	2	3
[] Other	6	13

[Notes: The ranked 1st numbers add up to 71 because two respondents marked two alternatives a first place ranking. Also a lot of people did not need and/or where reluctant to use all 6 numbers.]

Highlights of the results follow:

- 81 percent of the respondents have studied more than one full-time semester in other subjects than the so-called core subjects.
- 45 percent of the respondents have studied more than three full-time semesters, which is more than 1.5 years, in other fields within the social sciences and the humanities. This result should be judged

in light of the fact that there are no formal requirements for such courses in order to be accepted to the SDPE.⁶

- 57 percent of the students had less than three years of full-time study in economics, mathematics, and statistics upon starting the graduate program. It is also noteworthy that there is no correlation—either positive or negative—between having a strong background in quantitative subjects and having studied a large number of courses in other subjects within the social sciences or humanities.
- 16 percent of the respondents were active members of a volunteer organization and/or political party, 17 percent of the respondents had during their time in the doctoral program published a debate article. Interestingly, there was only a one percent overlap between these groups meaning that at least 32 percent of the doctoral students exhibit some form of civic engagement. In terms of characteristics, this group of "active" students seems representative, with a slight overrepresentation of women, and a slight underrepresentation of foreign students.⁷ It does not seem to be the case that students with stronger mathematics backgrounds are less active. If anything, the students with strong backgrounds in both mathematics/statistics and social science/humanities are overrepresented in the group that has written debate articles.

⁶ The National Economy department website at Stockholm University describes the requirements as follows: "Qualified applicants meet one of the following two requirements: 1) Undergraduate degree from a Nordic university with 60 credits (three semesters of full-time study) in economics (including an undergraduate thesis) and 60 credits of undergraduate studies in other subjects. A qualified applicant has written a senior thesis in economics. Applicants are encouraged to take the Graduate Record Examination (GRE) General Test. 2) An undergraduate degree, BA, from a non-Nordic university with a major in economics and a senior thesis in economics. Applicants are strongly encouraged to take the Graduate Record Examination (GRE) General Test. Applicants who do not have Swedish, or English, as their native language must show proficiency in English to be considered for admission. We recommend such applicants to take the Test of English as a Foreign Language (TOEFL). Although not required a good background in mathematics or/and statistics is often helpful for successful completion of a PhD in economics." (Online: <http://www.ne.su.se/education/graduate/qual.html>)

⁷ The under-representation of foreign students is hardly surprising since it is clearly more difficult to engage in these activities during the first years in a new environment.

- Students who had not written debate articles were asked why. 80 percent said lack of time and/or incentives, while only 3 percent answered that they were not interested. Thus, it seems as if it is a lack of time and incentives, which lies behind the observed non-participation, not a lack of interest.
- As for the motivation to pursue a PhD in economics in the first place, 65 percent answered that they were primarily motivated by their interest in the social sciences. Second most important were career concerns (14 percent) and third was the "community interest." No one ranked their interest in mathematics and/or statistics as the primary reason for entering the graduate program in economics. Furthermore, the two alternatives which were ranked last by most of the respondents were "my interest in mathematics and/or statistics" and "other". Thus, it seems as if the typical doctoral student is primarily motivated by an interest in the social science, not so much an interest in mathematics or statistics per se.⁸

In terms of the questions raised in the introduction, this survey suggests that economics as a subject is not attracting "one-legged" students—quite the opposite.

PARALLELS TO THE U.S. DEBATE

The survey results parallel the results from the 1985 survey of American graduate students done by Colander and Klamer (1987). As in their study, we find that students enter the program because they have an interest in understanding and contributing to society. As Colander and Klamer put it: "If graduate schools are graduating *idiots savants* who have no interest in policy, it is not because students enter graduate school with no

⁸ As a point of speculation one could, based on our survey results, argue that students who choose economics rather than any other field in the social sciences do so because of the relatively greater emphasis on formal and quantitative work. After all, most have in their undergraduate studies encountered other approaches and have still chosen economics.

interest" (97).⁹ This statement is equally true for the Stockholm situation sixteen years later.

Our results on participation in policy debates, or in other issue-oriented activities, are also similar to those reported by Colander and Klamer (1987). They report that the lack of time is the main reason for not pursuing their other (work related) interests. In our study 80 percent answer that it is lack of time and/or incentives, which keeps them from engaging in policy relevant side activities, while only 3 percent of the non-active students say that it is a lack of interest.

So how can it be that the perceptions of graduate students in a typical European PhD program in economics were so alike those of their American counterparts from 1985? One possible answer is of course that this is what graduate economics inevitably is like. Proper graduate education is about forgetting the questions you want to study so that you can focus on learning techniques and methods. This is certainly not the view of anyone concerned about economics being one-legged. Given that the findings of Colander and Klamer (1987) and others led to the American Economic Association appointing a commission on graduate education in economics, it also seems unlikely that this view was held by the "profession" in the U.S. at the time.¹⁰ Another possibility is that European schools forgot about the possible problems with the American programs they copied and did not pay much attention to the U.S. debate about graduate education.

It could also be that the "learning to lose a leg" syndrome will be short-lived. Maybe in the Stockholm case in October 2001, the process that generates expectations in those entering was yet to catch up to the character changes in the PhD programs. In the process of reviewing our paper for this journal, David Colander, who refereed the paper (and waived anonymity), wrote: "I have recently redone the survey that Arjo and I did . . . and in the new survey [of current PhD students] I find less cynicism and concern about mathematics" (Colander 2004; quoted with permission). One possible explanation for the difference from 1985 is that, now U.S. programs are less focused on technique, another is that the expectations-generation process has caught up, and the would-be two-legged economists are either being sorted out of the population or are entering PhD programs with a clearer understanding of what they are in for.

⁹ The *idiots savants* reference is from Robert Kuttner (1985), who in summarising the views of Wassily Leontief and John Kenneth Galbraith writes: "Departments of economics are graduating a generation of idiots savants, brilliant at esoteric mathematics yet innocent of actual economic life" (Kuttner 1987, 77).

¹⁰ See Krueger et al. (1991) for references to why the commission was appointed.

CONCLUDING REMARKS

Whether the above findings are a concern or not depends on one's view of what graduate education in economics should be about. If one is of the opinion that the primary goal of economics education is to teach students mathematics and econometrics for their own sake, then the fact that this is what students think is being done is nothing but encouraging. We do not, however, think that too many economists hold this view.

A much more common view is that, even though real-world problems and policy issues are considered important, one must first learn the tools, then apply them. We believe that such a view is dubious for at least three reasons. First, it assumes that the interest in real problems and policy can be put aside for years and then picked-up as if there was no interruption. Second, it implicitly assumes that there are no specific skills besides the model-building and statistical techniques that dominate the training program. Once you know those techniques you are automatically able to formulate and explore the meaningful and important questions. Third, and perhaps most important, is that graduate school is a socialization process through which students learn what is valued by the profession. If students, especially during their first years, are told to focus exclusively on technical crafts, their ability, interest, and regard for real-world issues and projects are bound to wane.

There have been no visible initiatives from the economics profession in Europe to evaluate European graduate programs and European graduate students in economics. Indeed, if there is a major difference between the debate in the US in the late 1980's and the recent discussions in Europe, it lies in the fact that, in Europe, the initiatives to discuss the contents of graduate education have come mainly from students. We think it is time for a European Commission on Graduate Education in Economics.

It is important to emphasize that the views expressed in this article should not be interpreted as if we believe that one must choose between rigorous mathematical or statistical techniques and social engagement. On the contrary, we are strong proponents of a rigorous analytical approach to important social questions. We would even go so far as to say that formal modelling tools are one of the strengths of economics relative to other approaches. This does not, however, imply that relevance can be neglected.

REFERENCES

- Colander, David.** 2004. Referee report: Learning to Lose a Leg. Written for *Econ Journal Watch*. Unpublished.
- Colander, David and Arjo Klamer.** 1987. The Making of an Economist. *Journal of Economic Perspectives* 1(2): 95-111.
- Frey, Bruno S. and Reiner Eichenberger.** 1993. American and European Economics and Economists. *Journal of Economic Perspectives*. 7(4): 185-193.
- Krueger, Anne et al.** 1991. Report of the Commission on Graduate Education in Economics. *Journal of Economic Literature* 29(September): 1035-1053.
- Kuttner, Robert.** 1985. The Poverty of Economics. *The Atlantic Monthly*. February, 74-84.
- Lindbeck, Assar.** 2001. Economics in Europe. *CESifo Forum* 2(1): 31-32.
- Skedinger, Per and Dan Johansson.** 2004. In Sweden, Anti-Globalizationists Dominate Public Discourse, Econ Profs Do Little. *Econ Journal Watch* 1(1): 175-184.

ABOUT THE AUTHORS



Anne Boschini is a researcher at the Department of Economics, Stockholm University. In 2002 she completed her PhD on political economics and economic growth at Stockholm University and spent the following academic year as a post-doc at the Centre for Economic Performance, London School of Economics. She has taught Advanced Macroeconomics and Engendering Macroeconomics at the undergraduate level at Stockholm University. She has participated in two governmental commissions and written a book (in Swedish) on the topic of the lack of female managers in the private sector. Her research interests range from economic growth, institutions to gender. At present, Anne Boschini investigates the role of natural resources and

higher education for economic development. Her e-mail address is: anne.boschini@ne.su.se.



Matthew Lindquist received his Ph.D. from Stockholm University in 2002. He is currently working as a research associate at the Swedish Institute for Social Research, Stockholm University. His main research interests lie within macroeconomics and labor economics, both theoretical and empirical. His current research projects deal with relative wage determination in Sweden and the US. He has published several articles in international journals and has taught mathematics, time series analysis, and macroeconomics at Stockholm University. His email address is: Matthew.Lindquist@sofi.su.se.



Jan Pettersson is a Ph.D. candidate at the Department of Economics, Stockholm University. His research deals with the efficiency of foreign development assistance, the relation between democracy and economic growth, and the impact of institutions and natural resource endowments on economic development. His email address is: jrp@ne.su.se.



Jesper Roine received his PhD in Economics from Stockholm University in 2002 and is currently research associate at the Department of Economics at Stockholm School of Economics. He also serves as secretary of the Economic Council of Sweden. He has previously been visiting researcher at Northwestern University (1999-2000) and post-doctoral fellow at the London School of Economics (2002-2003). His main fields of research are political economics, economic growth and economic sociology. His email address is: jesper.roine@hhs.se.



*Econ Journal Watch,
Volume 1, Number 2,
August 2004, p 380.*

CORRESPONDENCE

Editors,

Congratulations on an excellent new journal. I found your article on the Social Science Citation Index [[Klein with Chaing 2004](#)] particularly enlightening.

But one minor point: Your survey actually missed one SSCI-registered journal that has taken an explicitly conservative/libertarian position (even while publishing high-quality articles and not all from one perspective), namely the *Harvard Journal of Law and Public Policy*.

Sincerely,

Peter Kurrild-Klitgaard
Associate Professor
Dept. of Political Science & Public Management
University of Southern Denmark and
President, Danish Political Science Association
kurrild@sam.sdu.dk

Klein responds: Yes, *HJLPP* should have been included as indicated. Our thanks to Professor Kurrild-Klitgaard for identifying and correcting our error.

Econ Journal Watch welcomes letters commenting on the journal or articles therein. Please send correspondence to editor@econjournalwatch.org, Subject line: EJW Correspondence.