

# Macrocomparative Research

Lane Kenworthy

January 2013

## *My ten points*

---

1. Ask a question
2. Go macro
3. The social scientist as Sherlock Holmes
4. Describe
5. Graph
6. Know your cases
7. Step away from the pool
8. Beware multivariate overload
9. Address multiple sources of uncertainty
10. What should your goal be?

# ***1. Ask a question***

---

Too often social scientists aim to "add to literature A" or "contribute to our understanding of topic B"

A paper or book should be driven by a question (one per paper)

Literatures and topics are too broad, too vague

# ***1. Ask a question***

---

We often choose a research question because

- data are available

- it allows you to use an analytical technique you've mastered

Better to pick a question that

- is substantively important

- interests you

# ***1. Ask a question***

---

Some questions I've tried to answer:

Do government social programs reduce poverty?

Why have the incomes and living standards of the poor improved more in some affluent nations than in others?

Do universal social programs yield more redistribution than targeted ones?

How do taxes contribute to redistribution?

Do public expenditures or private expenditures do more to boost the living standards of the poor?

Are nations with a sizeable low-wage sector doomed to high poverty?

# ***1. Ask a question***

---

Some questions I've tried to answer:

How much do presidents influence income inequality?

Does a rise in inequality lead more support for government redistribution?

Does public opinion determine welfare state generosity?

Is income inequality bad for the poor?

Have generous social programs, heavy taxation, and labor market regulations impeded employment growth?

Are institutions a key contributor to economic growth?

# ***1. Ask a question***

---

Questions and hypotheses need not be original

Social scientists do *far* too little replication and reanalysis

## ***2. Go macro***

---

Countries (or regions) as the unit of analysis

This is what I do in almost all of my research

But is it the right way to go?



## 2. *Go macro*

---

Objection #1: There's no longer much variation among the rich nations, due to globalization, the EU, etc.

Apart from monetary policy and inflation and perhaps a few others, this is empirically false

## 2. *Go macro*

---

Objection #2: There are insurmountable identification problems — not enough cases, too many unmeasured sources of variation

I think prioritizing the identification strategy over the research question is wrongheaded

Some research questions require macro analysis, even if the data, measures, and estimation techniques are less than ideal

### ***3. The social scientist as Sherlock Holmes***

---

The research process as learned in methodology courses tends to focus on doing the analysis correctly

I think of macrocomparative analysis as more like detective work

Seldom do you have the evidence you want

So you piece together a conclusion from multiple imperfect and incomplete bits of evidence

Various types of data

New measures

A mix of methods: medium-N quantitative, small-N most-similar-cases comparison, single-case process tracing, natural experiments

### ***3. The social scientist as Sherlock Holmes***

---

A useful device: Ask yourself what sorts of things you would expect to observe if the hypothesis were true (or false)

### 3. The social scientist as Sherlock Holmes

---

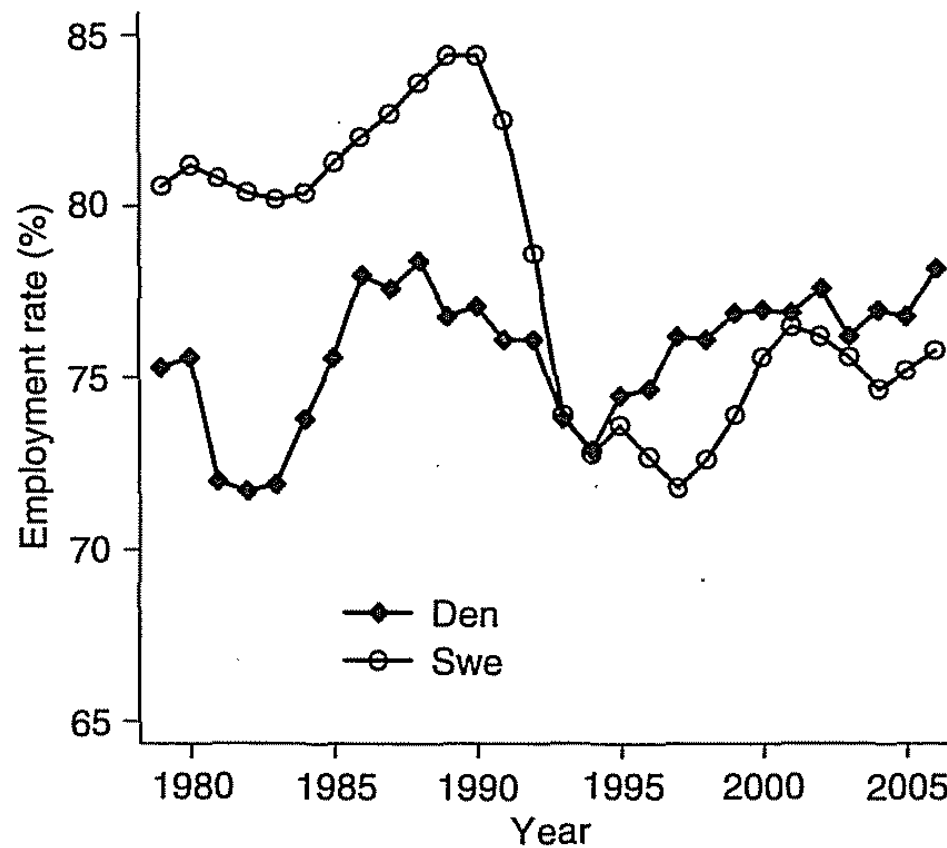


Figure 6.10. Employment trends in Denmark and Sweden, 1979ff.

### 3. The social scientist as Sherlock Holmes

---

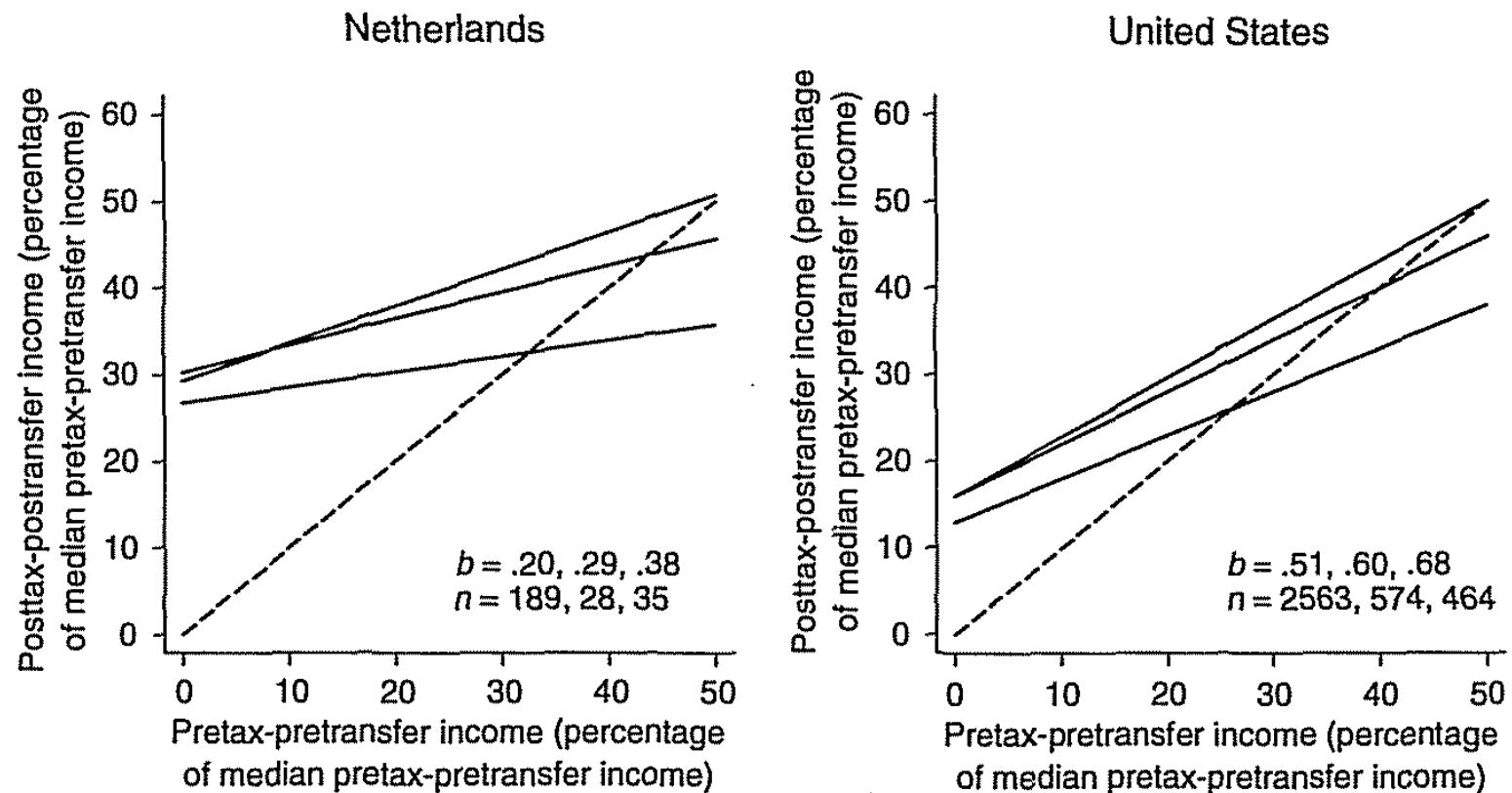


Figure 7.1. Government benefit generosity for three types of low-income households, 2000.

## ***4. Describe***

---

Social scientists (or at least journal editors and reviewers) prize explanation

We tend to undervalue description

## ***4. Describe***

---

Much of the disagreement about determinants of welfare state generosity, poverty, inequality, globalization, and other issues in comparative political economy hinges on how to describe (define and measure) the dependent variable



## 4. *Describe*

---

For my *Jobs with Equality* book, some simple descriptive charts helped me understand what my research question should be and how to proceed analytically

In analyzing employment growth, it's crucial to disaggregate by sector

## 4. Describe

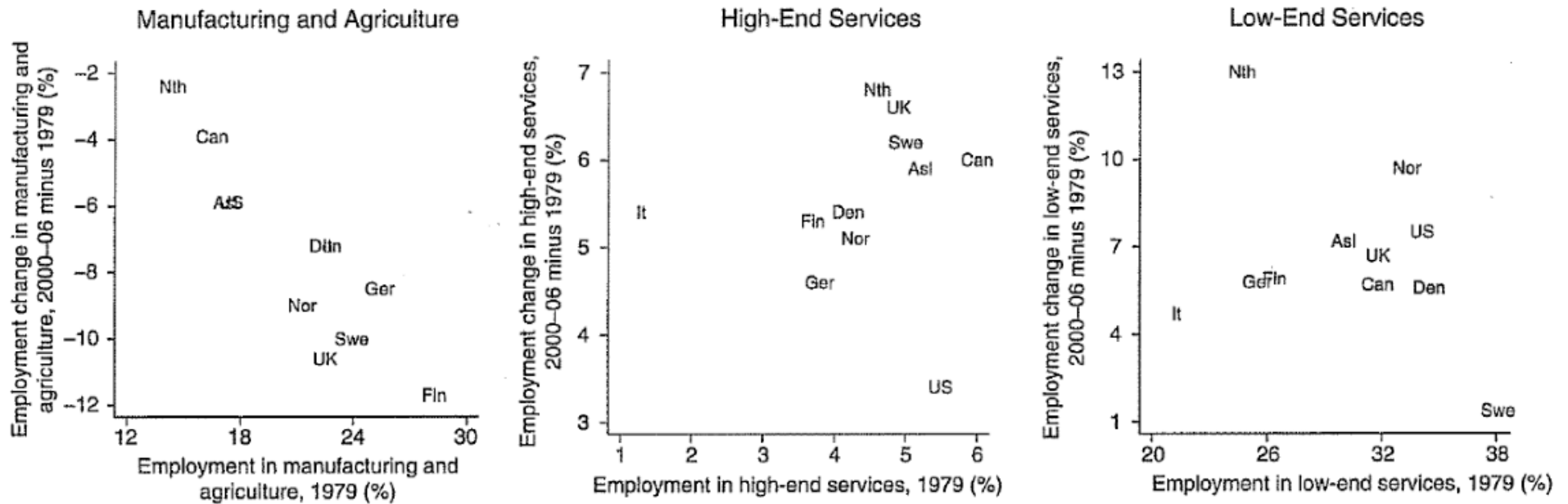


Figure 4.7. Employment change in manufacturing and agriculture, high-end services, and low-end services, 1979 to 2000-06, by initial employment level.

## ***5. Graph***

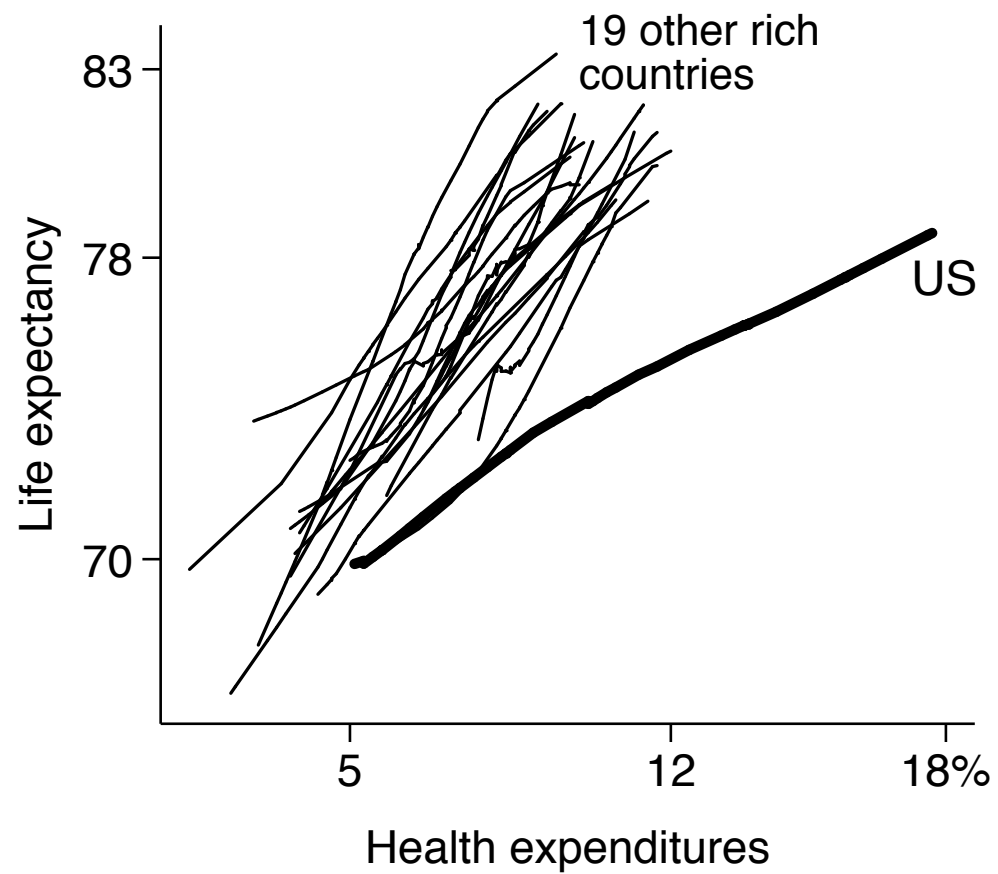
---

Analysts and readers tend to spot and understand patterns in data better when they're displayed graphically

Simple dot plots (bar graphs), line plots, and scatterplots can take you a long way

## 5. Graph

---



## ***6. Know your cases***

---

The more you know about the details of the cases, the less likely you are to draw mistaken inferences from statistical analyses

With 20 or so cases, it's possible to know something about them (or at least some of them)

This knowledge also allows you to use over-time developments in individual countries as an additional source of analytical leverage

## *7. Step away from the pool*

---

Most quantitative macrocomparative analyses use data that are pooled across countries and over time, usually annually

This is unfortunate

## *7. Step away from the pool*

---

The assumption that the causes of cross-sectional variation are the same as the causes of over-time variation is frequently wrong

Our default strategy should be to examine these types of variation separately, at least at the outset

## *7. Step away from the pool*

---

Differences-in-differences analysis often is better

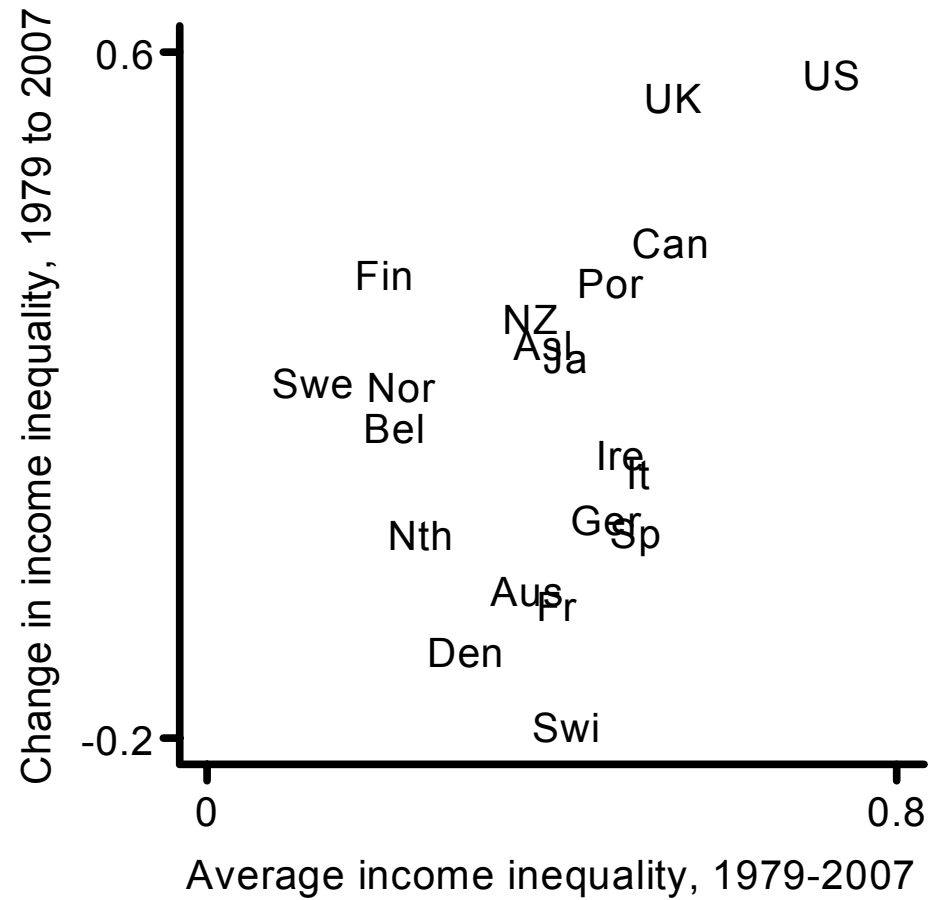
if you have a reasonably lengthy time series

and if there is a decent amount of cross-country variation in change in the hypothesized cause(s)



## 7. Step away from the pool

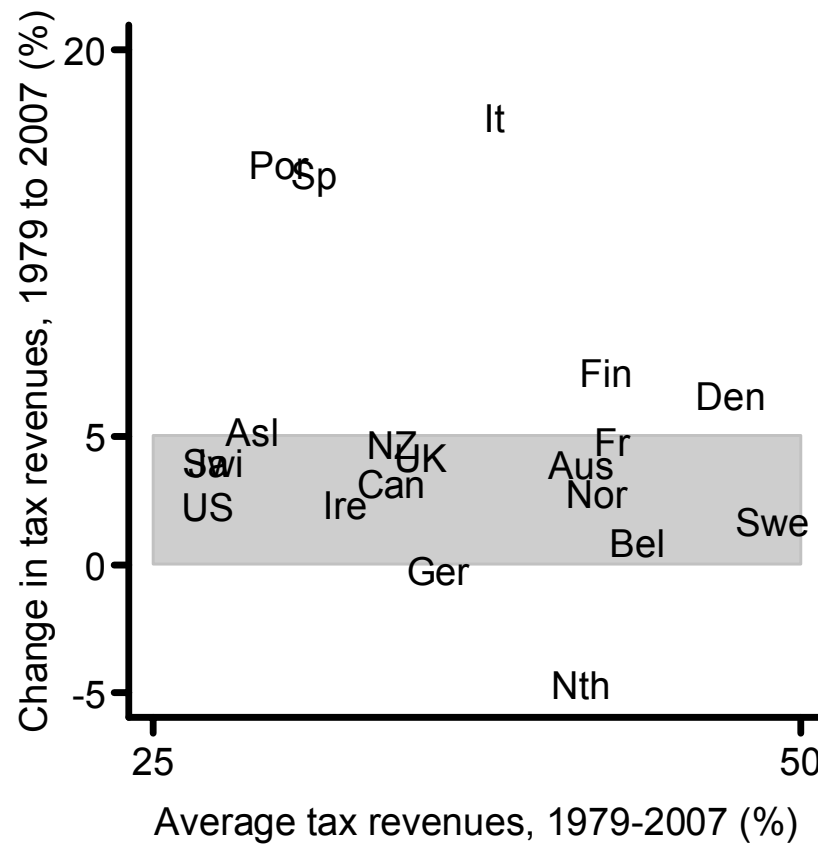
---



## 7. Step away from the pool

---

These conditions aren't always met



## *7. Step away from the pool*

---

I sometimes used pooled regression as a descriptive device, to get a better feel for what's there in the data — akin to playing around with scatterplots and crosstabs

But very rarely is it my main analytical tool

## 8. Beware multivariate overload

TABLE 5.10. Prais-Winsten estimates of determinants of income inequality

Variables	Model 1	Model 2	Model 3	Model 4
Debt crisis	.479	.554	.650	.446
1990s	.704	.831	.677	.830
2000s	.876	.934	.328	1.005
Gross income	3.234***	2.950***	3.010***	3.000***
No household adjustment	-3.001***	-3.101***	-2.721***	-4.125***
GDP per capita	.298 <sup>a</sup>	.207	.313 <sup>a</sup>	.300 <sup>a</sup>
Sector dualism	.063	.121**	.166***	.093*
Inflation	.001**	.001**	.001**	.001
Youth population	-.301 <sup>a</sup>	-.270 <sup>a</sup>	-.117	-.444 <sup>a</sup>
Stock of FDI	.030	.018	.043	.047*
FDI flow	.321**	.276**	.278**	.313**
Ethnic heterogeneity	5.576***	4.342***	4.860***	3.889***
Employment in industry	-.473***	-.508***	-.399***	-.403***
Female labor force participation	-.045	-.041	-.010	.019
IMF agreements	-.058 <sup>a</sup>	-.067 <sup>a</sup>	-.033	-.015
External debt (% GDP)	.003	.002	.004	.005
Trade	.023*	.023*	.027*	.022
Politics and policy				
Democracy	-.191***			-.148***
Democracy (20+ years)		-.238***		
Left political strength			-.182***	
Democracy*social security welfare				-.012***
Average years of education				-1.274***
Social security and welfare				.154
Health (cumulative ave)				.172
Education (cumulative ave)				.017
Common p	.24	.26	.34	.20
Constant	70.503***	68.285***	57.386***	76.942***
R <sup>2</sup>	.81***	.82***	.86***	.81***
N	271	271	271	259

<sup>a</sup> Significant but sign of coefficient opposite of directional hypothesis.

\*  $p \leq .05$ ; \*\*  $p \leq .01$ ; \*\*\*  $p \leq .001$ .

## ***8. Beware multivariate overload***

---

Sometimes a dozen or more independent variables is the right way to go, but I prefer Christopher Achen's "rule of three"

Achen, Christopher H. 2002. "Toward a New Political Methodology: Microfoundations and ART." *Annual Review of Political Science* 5: 423-450.

Achen, Christopher H. 2005. "Let's Put Garbage-Can Regressions and Garbage-Can Probits Where They Belong." *Conflict Management and Peace Science* 22: 327-339.

## ***8. Beware multivariate overload***

---

I think it's usually best to start with the bivariate relationship and then move to greater complexity

And when possible, look at graphical depictions of the partial associations, not just coefficients

## ***9. Address multiple sources of uncertainty***

---

Tests for statistical significance tell us about only one source of uncertainty: sample-to-population

Others may be more worrisome

- Flawed measures of key concepts

- Missing variables

- Vague theory to guide model specification

- Influence of particular cases

## ***10. What should your goal be?***

---

Show off your knowledge of the methodological tool du jour?

Flawlessly execute the scientific method?

Emulate the most recent *ASR* or *APSR* article in your field?

NO



## ***10. What should your goal be?***

---

Ask a good question and get us closer to the answer

# Further reading

## TOWARD IMPROVED USE OF REGRESSION IN MACRO-COMPARATIVE ANALYSIS

Lane Kenworthy

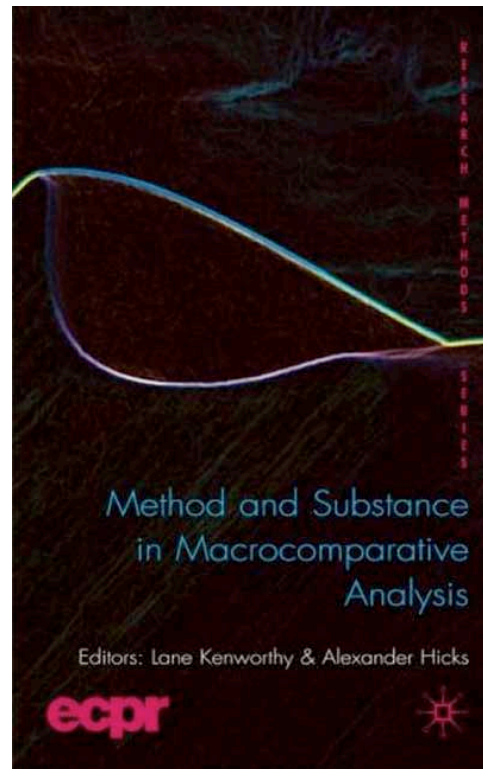
I agree with much of what Michael Shalev (2007) says in his paper, both about the limits of multiple regression and about how to improve quantitative analysis in macro-comparative research. With respect to the latter, Shalev suggests three avenues for advance: (1) improve regression through technical refinement; (2) combine regression with case studies (triangulation); (3) turn to alternative methods of quantitative analysis such as multivariate tables and graphs or factor analysis (substitution). I want to suggest some additional ways in which the use of regression in macro-comparative analysis could be improved. None involves technical refinement. Instead, most have to do with relatively basic aspects of quantitative analysis that seem, in my view, to be commonly ignored or overlooked.

### LOOK AT THE DATA

Shalev's third suggested path for progress consists of using tables, graphs, and tree diagrams to examine causal hierarchy and complexity and to identify cases meriting more in-depth scrutiny. This should be viewed not as (or at least not solely as) a substitute for regression but rather as a critical component of regression analysis. All of us were (I hope) taught in our first

*Capitalisms Compared*  
Comparative Social Research, Volume 24, 343–350  
Copyright © 2007 by Elsevier Ltd.  
All rights of reproduction in any form reserved  
ISSN: 0195-6310/doi:10.1016/S0195-6310(06)24010-9

343



*Qualitative & Multi-Method Research, Fall 2011*

### *Step Away from the Pool*

**Lane Kenworthy**  
University of Arizona  
[lane.kenworthy@gmail.com](mailto:lane.kenworthy@gmail.com)

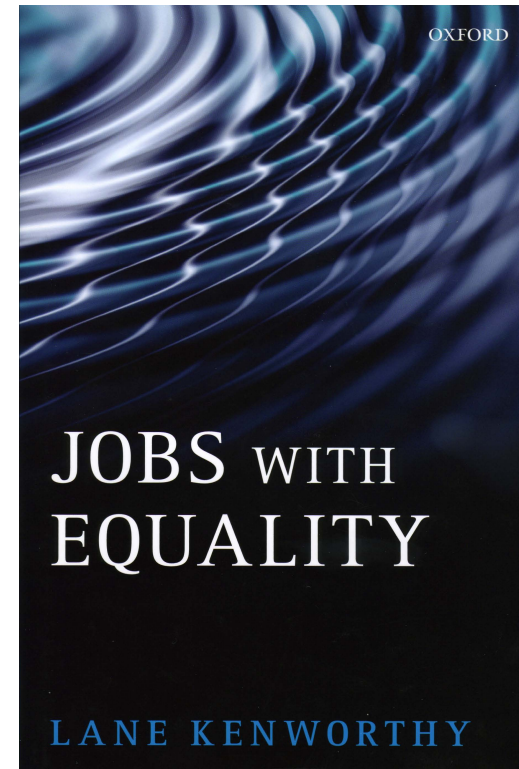
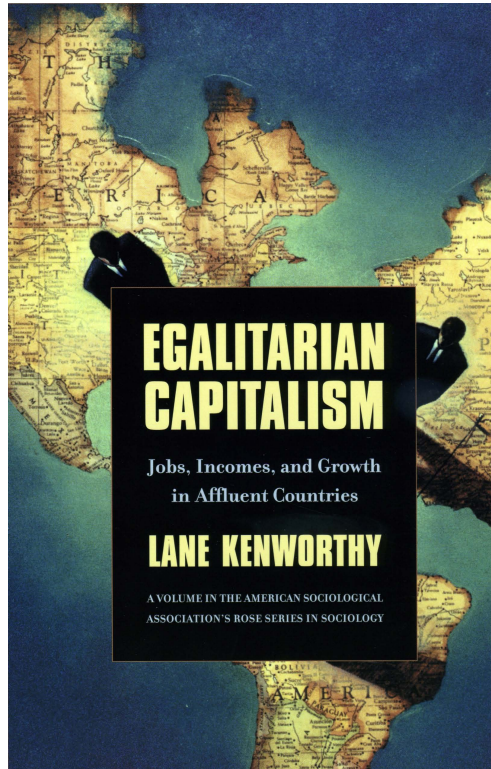
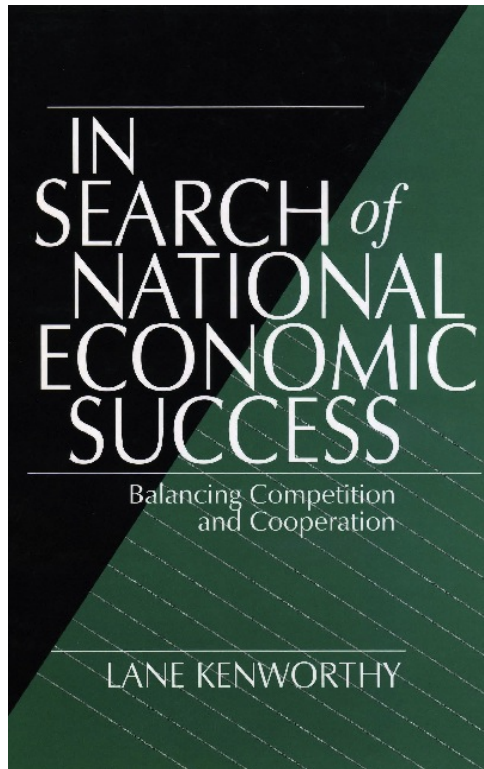
Quantitative macrocomparative analysis of the rich long-standing-democratic nations—"medium-N analysis"—is dominated by pooled time-series cross-section regression. I guess-timate that more than 90% of the medium-N papers I read in journals and as journal submissions use pooled regression.<sup>1</sup>

Quantitative data on many of the institutions, policies, and socioeconomic outcomes studied by comparativists first became available for more than a handful of countries in the 1970s. For a while quantitative macrocomparative research consisted mainly of cross-sectional analysis of single-point-in-time data or period averages. By the 1990s reasonably lengthy time series existed and analysts began examining over-time patterns. Pooling over time and across countries helped alleviate what had long been considered the achilles heel of cross-sectional comparative analysis: the small-N problem, which limits the number of control variables that can be included in a regression model. Pooled regression became the tool of choice. It has remained so for two decades.

In my view, that's unfortunate.

## *Further reading*

---



## *Further reading*

---

