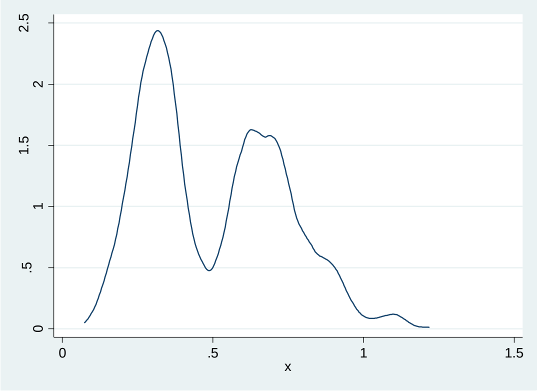
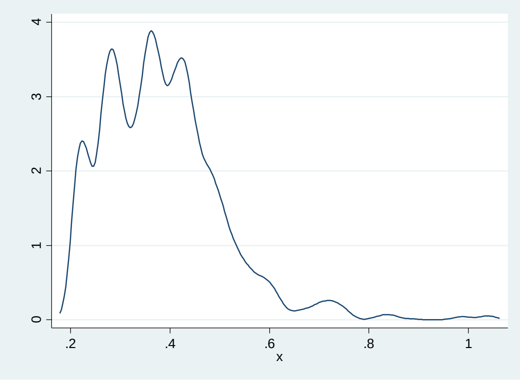




Model Uncertainty and Robustness: a computational approach

Cristobal Young

Department of Sociology
Stanford University

October, 2017

Sociological Methods & Research
2017, Vol. 46(1) 3-40

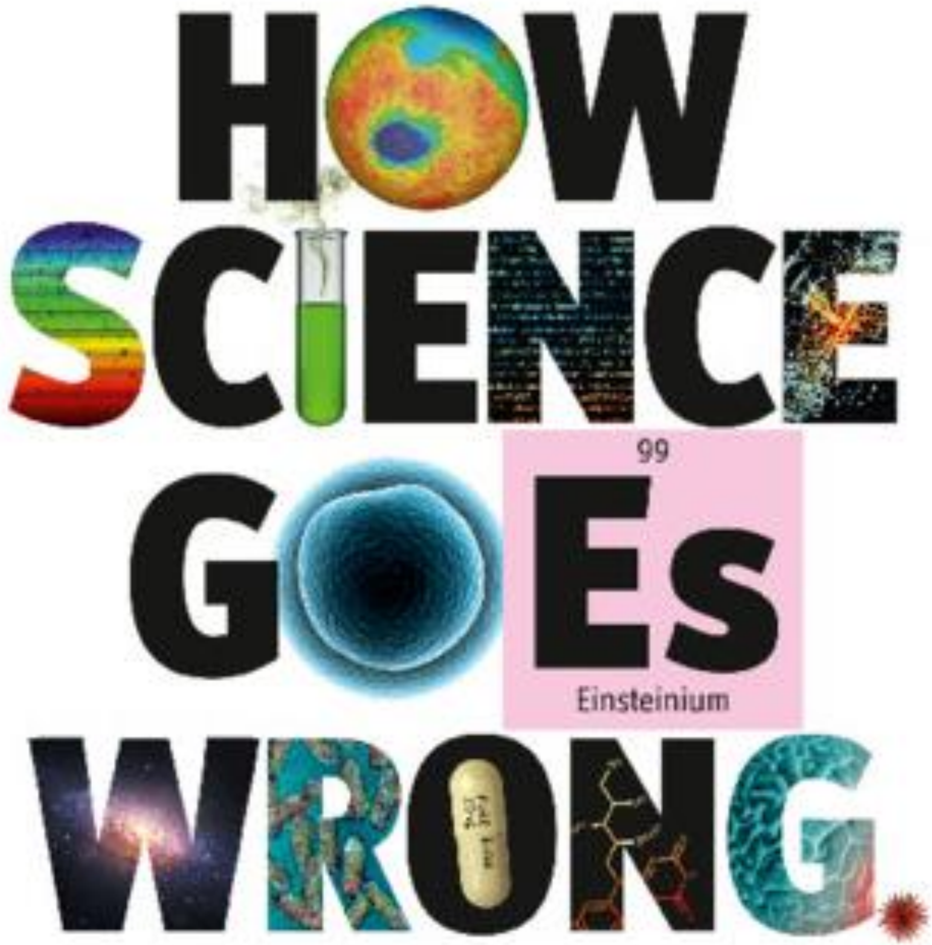
DOI: 10.1177/0049124115610347
journals.sagepub.com/home/smr
SAGE

Model Uncertainty and Robustness: A Computational Framework for Multimodel Analysis

Cristobal Young[1] and Katherine Holsteen[2]

Replication package in Stata:

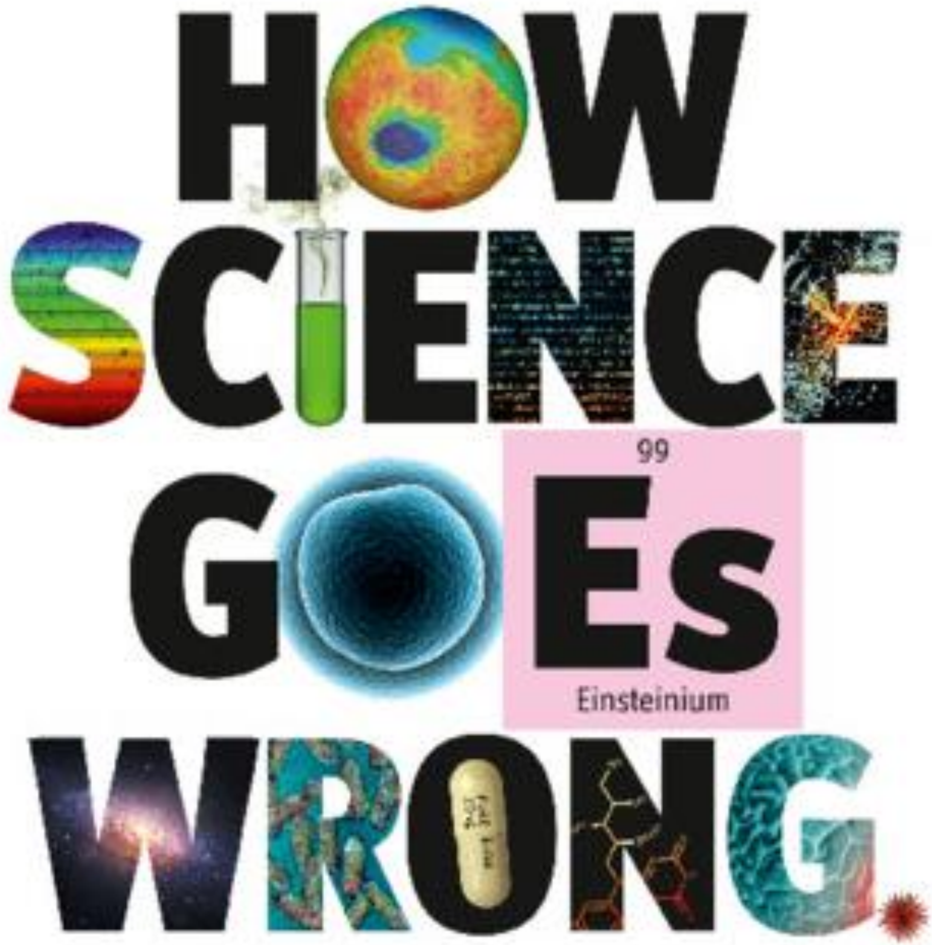
do http://web.stanford.edu/~cy10/public/mrobust/install_mrobust.do



Crisis in Science

Non-reproducible research

Researchers don’t believe the work of other researchers



Crisis in Science

Non-reproducible research

Researchers don’t believe the work of other researchers

Result:
Public doesn’t believe our research

Crisis of confidence

Problems of Non-Robust Research

1. **Bio-medical research on cancer**

   - most research on potential treatments is not replicable by industry labs

(Bio-tech giant Amgen had 100 scientists spend 10 years on replicating 53 landmark studies – only 11 percent were replicable.)

Problems of Non-Robust Research

1. **Bio-medical research on cancer**
   - most research on potential treatments is not replicable by industry labs

(Bio-tech giant Amgen had 100 scientists spend 10 years on replicating 53 landmark studies – only 11 percent were replicable.)

2. **Genetic research on intelligence**
   - vast research on *specific genes* linked to IQ appears to be all false positives (Jeremy Freese)

Problems of Non-Robust Research

1. **Bio-medical research on cancer**

- most research on potential treatments is not replicable by industry labs

(Bio-tech giant Amgen had 100 scientists spend 10 years on replicating 53 landmark studies – only 11 percent were replicable.)

2. **Genetic research on intelligence**

- vast research on *specific genes* linked to IQ appears to be all false positives (Jeremy Freese et al)

3. Determinants of Economic Growth

- Why do some countries have higher economic growth?

- of the 67 growth factors identified in the literature, three-quarters of them are non-robust

Solutions:

1. Transparency & Replication

“Nullius in Verba” as the motto of the Royal Society

“take no one's word for it” – “see for yourself”

Solutions:

1. Transparency & Replication

"Nullius in Verba" as the motto of the Royal Society
"take no one's word for it" – "see for yourself"

Production of research has run far ahead of the replication of research.

Researchers act as if their work will never be replicated – bad incentive

Solutions:

1. Transparency & Replication

“Nullius in Verba” as the motto of the Royal Society

“take no one's word for it” – “see for yourself”

2. Comprehensive Model Robustness Analysis

Build more transparency and skepticism into original articles

Incorporate the spirit of replication into our research.

Model Uncertainty in Applied Research

Model Uncertainty in Applied Research

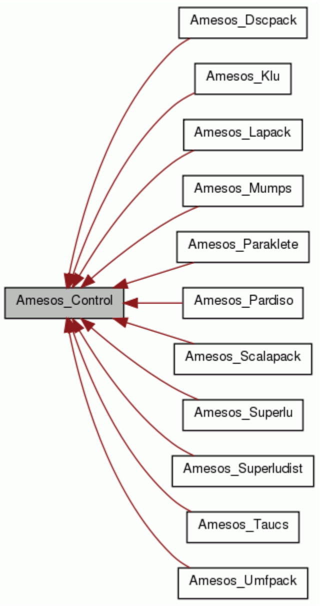
Edward Leamer describes the process of model specification:

> Sometimes I take the error terms to be correlated, sometimes uncorrelated; … sometimes I include observations from the decade of the fifties, sometimes I exclude them; sometimes the equation is linear and sometimes nonlinear; sometimes I control for variable *z*, sometimes I don't.
>
> (Leamer 1983: 37-38)

Three Elements of Model Specification:

Three Elements of Model Specification:

1) Choice of Control Variables

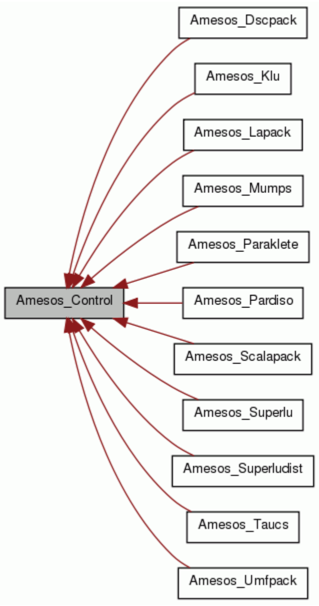


Three Elements of Model Specification:

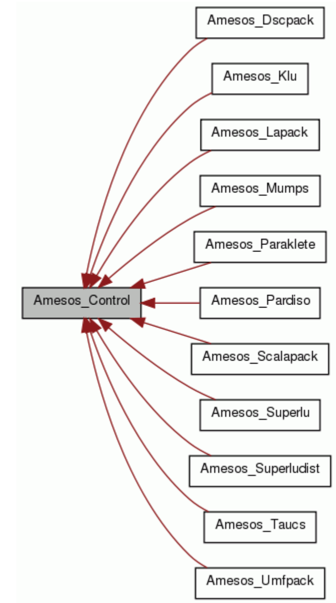
1) Choice of Control Variables

2) Choice of Key Variable Definitions

Social capital: voter turnout or volunteer activity?
Religiosity: church attendance or belief in god?
Inequality: gini coefficient or top 1% share of income?



Three Elements of Model Specification:

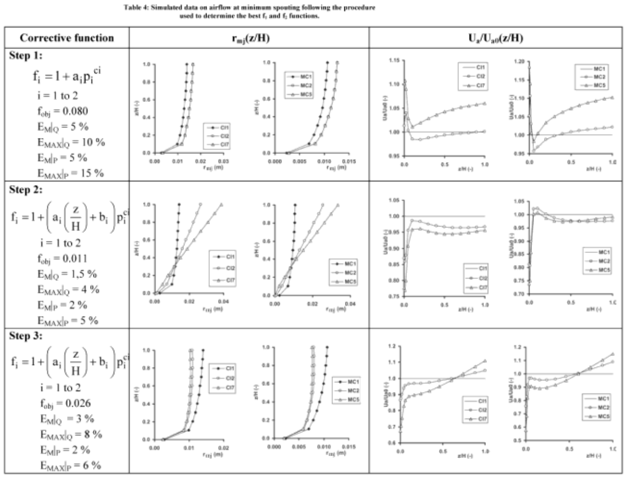


1) Choice of Control Variables

2) Choice of Key Variable Definitions

Social capital: voter turnout or volunteer activity?
Religiosity: church attendance or belief in god?
Inequality: gini coefficient or top 1% share of income?

3) Choice of Functional Form



OLS, logit, probit, fractional regression, non-linear least squares, instrumental variables, matching, poisson, negative binomial, etc… link function

Model assumptions are necessary

Point estimates *cannot* be calculated until you've made some modeling choices / assumptions.

Model assumptions are necessary

Point estimates *cannot* be calculated until you've made some modeling choices / assumptions.

Model assumptions can be relaxed

Define the *model space* as all possible combinations of these (reasonable) elements.

Model assumptions are necessary

Point estimates *cannot* be calculated until you've made some modeling choices / assumptions.

Model assumptions can be relaxed

Define the *model space* as all possible combinations of these elements.

Example:
12 plausible controls,
3 plausible definitions of the outcome variable,
4 plausible functional forms

Model assumptions are necessary

Point estimates *cannot* be calculated until you've made some modeling choices / assumptions.

Model assumptions can be relaxed

Define the *model space* as all possible combinations of these elements.

Example:
12 plausible controls,
3 plausible definitions of the outcome variable,
4 plausible function forms

yields 32,768 possible models.

Empirical Example:

Union Wage Premium

Do union members earn higher wages?

Table 2: Determinants of Log Hourly Wage

	Model: OLS	
Union member	11.1***	(2.2)
Usual hours worked	0.3**	(0.1)
Age	-0.6	(0.3)
Education (grade completed)	6.3***	(0.6)
College graduate	4.6	(3.6)
Married	1.1	(2.0)
Lives in south	-12.2***	(2.0)
Lives in metro area	22.4***	(2.3)
Lives in central city	-3.7	(2.3)
Total work experience	3.2***	(0.3)
Job tenure (years)	0.9***	(0.2)
Constant	56.5***	(15.0)
Observations	1865	
Adjusted R-squared	0.408	

* p<0.05   ** p<0.01   *** p<0.001. Standard errors in parentheses. The outcome variable, log of hourly wages, has been scaled by 100 so that coefficients can be interpreted as percent changes in wages for a unit change in the predictor.

Table 2: Determinants of Log Hourly Wage

	Model: OLS	
Union member	11.1***	(2.2)
Usual hours worked	0.3**	(0.1)
Age	-0.6	(0.3)
Education (grade completed)	6.3***	(0.6)
College graduate	4.6	(3.6)
Married	1.1	(2.0)
Lives in south	-12.2***	(2.0)
Lives in metro area	22.4***	(2.3)
Lives in central city	-3.7	(2.3)
Total work experience	3.2***	(0.3)
Job tenure (years)	0.9***	(0.2)
Constant	56.5***	(15.0)
Observations	1865	
Adjusted R-squared	0.408	

* p<0.05 ** p<0.01 *** p<0.001. Standard errors in parentheses. The outcome variable, log of hourly wages, has been scaled by 100 so that coefficients can be interpreted as percent changes in wages for a unit change in the predictor.

Is it Robust?

How robust is this finding to the choice of control variables in the model?

Is it Robust?

How robust is this finding to the choice of control variables in the model?

10 possible controls = 2^{10} or 1,024 unique combinations of the control variables.

Is it Robust?

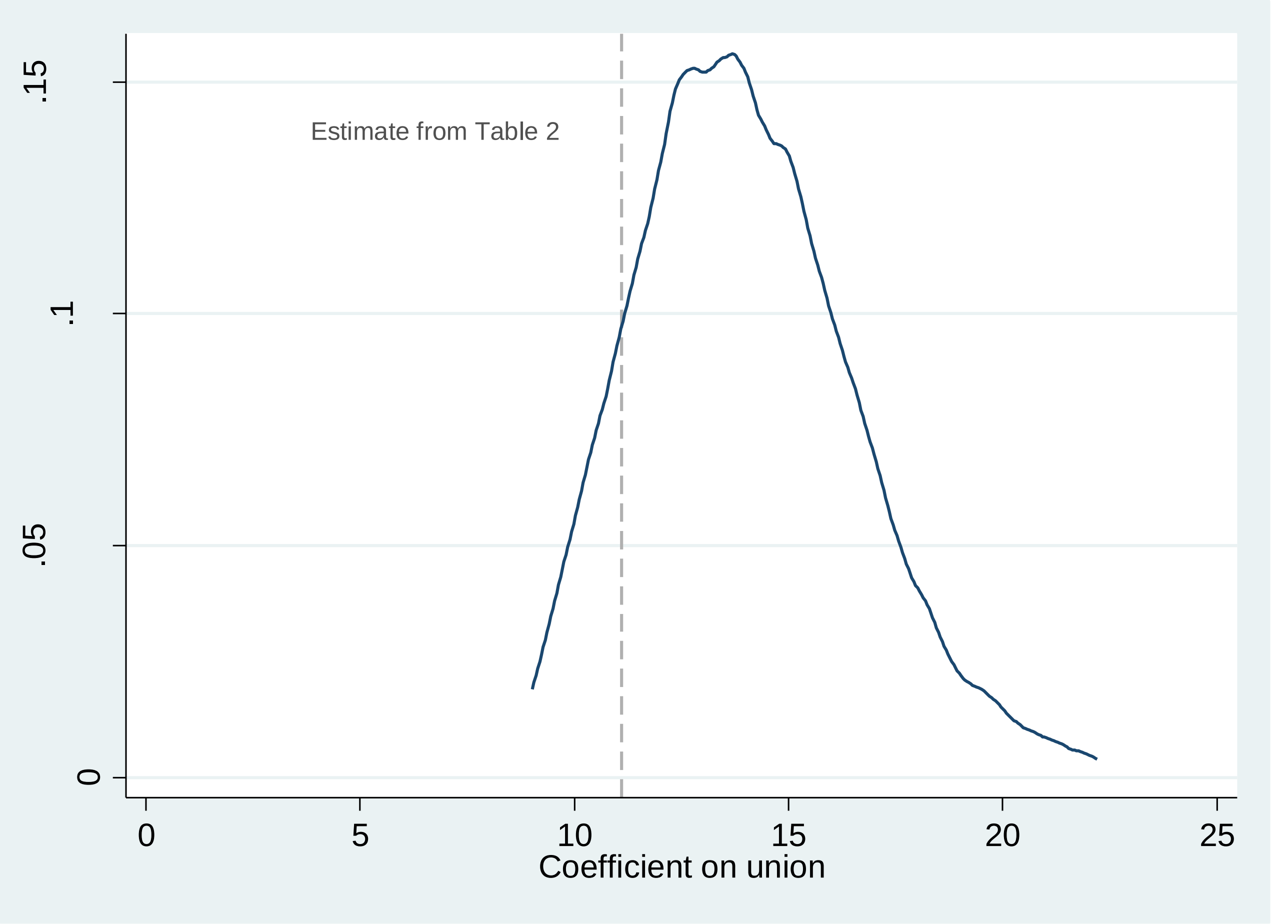
How robust is this finding to the choice of control variables in the model?

10 possible controls = 2^{10} or 1,024 unique combinations of the control variables.

Run each of these models and store all of the estimates for the effect of union on log wage

then check how much the estimate varies.

Figure 1: Modeling Distribution of Union Wage Premium



Note: Vertical line indicates the "preferred estimate" of an 11 percent union wage premium as reported in Table 2.

Compare back with table

Table 3: Model Robustness

Linear regression;			
Variable of interest	union		
Outcome variable	wage	Number of observations	1865
Possible control terms	10	Mean R-squared	0.26
Number of models	1,024	Multicollinearity	0.06
Model Robustness Statistics:		**Significance Testing:**	
Mean(b)	14.00	Sign Stability	100%
Sampling SE	2.37	Significance rate	100%
Modeling SE	2.51		
Total SE	3.46	Positive	100%
		Positive and Sig	100%
Robustness Ratio:	4.05	Negative	0%
		Negative and Sig	0%

Table 3: Model Robustness

Linear regression;			
Variable of interest	union		
Outcome variable	wage	Number of observations	1865
Possible control terms	10	Mean R-squared	0.26
Number of models	1,024	Multicollinearity	0.06

Model Robustness Statistics:		**Significance Testing:**	
Mean(b)	14.00	Sign Stability	100%
Sampling SE	2.37	Significance rate	100%
Modeling SE	2.51		
Total SE	3.46	Positive	100%
		Positive and Sig	100%
Robustness Ratio:	4.05	Negative	0%
		Negative and Sig	0%

2nd Applied Example

Mortgage Lending in Boston: Interpreting HMDA Data

By Alicia H. Munnell, Geoffrey M. B. Tootell, Lynn E. Browne, and James McEneaney *

THE AMERICAN ECONOMIC REVIEW *MARCH 1996*

A data set on the factors influencing banks willingness to lend to mortgage applicants. N = 2,355 mortgage applications collected by the Boston Federal Reserve.

2^{nd} Applied Example

Mortgage Lending in Boston: Interpreting HMDA Data

By ALICIA H. MUNNELL, GEOFFREY M. B. TOOTELL, LYNN E. BROWNE, AND JAMES MCENEANEY *

THE AMERICAN ECONOMIC REVIEW *MARCH 1996*

A data set on the factors influencing banks willingness to lend to mortgage applicants. N = 2,355 mortgage applications collected by the Boston Federal Reserve.

Are banks less likely to approve mortgages from female applicants?

Effect of Gender on Mortgage Lending Decisions of Banks

Table 4: Determinants of Mortgage Application Acceptance

	Model: OLS	
Female	3.7*	(1.6)
Black	-11.4***	(1.8)
Housing Expense Ratio	5.8	(10.5)
Self Employed	5.6**	(1.8)
Married	4.6***	(1.3)
Bad Credit History	-25.2***	(2.3)
Payment-Income Ratio	-50.2***	(9.3)
Loan-to-Value Ratio	11.9***	(3.4)
Denied Mortgage Insurance	-71.2***	(4.2)
Constant	113.8***	(3.4)
N	2355	
adj. R-sq	0.226	

* p<0.05   ** p<0.01   *** p<0.001. Standard errors in parentheses. The outcome variable, mortgage acceptance (1 = accepted, 0 = denied), has been scaled by 100 so that coefficients can be interpreted as percent changes in the acceptance rate for a unit change in the predictor.

Effect of Gender on Mortgage Lending Decisions of Banks

Table 4: Determinants of Mortgage Application Acceptance

	Model: OLS	
Female	3.7*	(1.6)
Black	-11.4***	(1.8)
Housing Expense Ratio	5.8	(10.5)
Self Employed	5.6**	(1.8)
Married	4.6***	(1.3)
Bad Credit History	-25.2***	(2.3)
Payment-Income Ratio	-50.2***	(9.3)
Loan-to-Value Ratio	11.9***	(3.4)
Denied Mortgage Insurance	-71.2***	(4.2)
Constant	113.8***	(3.4)
N	2355	
adj. R-sq	0.226	

* p<0.05 ** p<0.01 *** p<0.001. Standard errors in parentheses. The outcome variable, mortgage acceptance (1 = accepted, 0 = denied), has been scaled by 100 so that coefficients can be interpreted as percent changes in the acceptance rate for a unit change in the predictor.

Model Robustness Testing

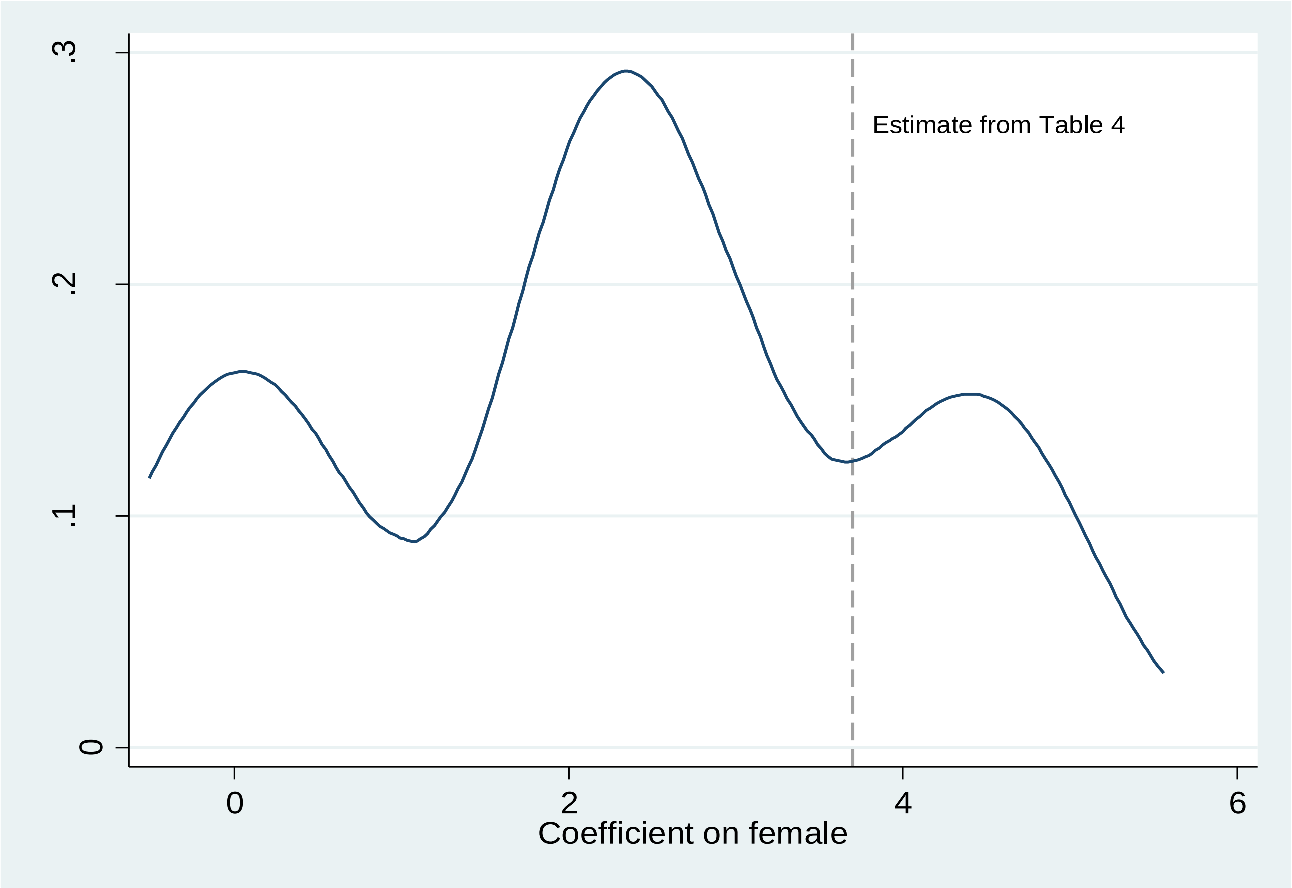
Table 5: Model Robustness of the Gender Effect on Mortgage Lending

Linear regression			
Variable of interest	female		
Outcome variable	acceptance	Number of observations	2355
Possible control terms	8	Mean R-squared	0.13
Number of models	256	Multicollinearity	0.19

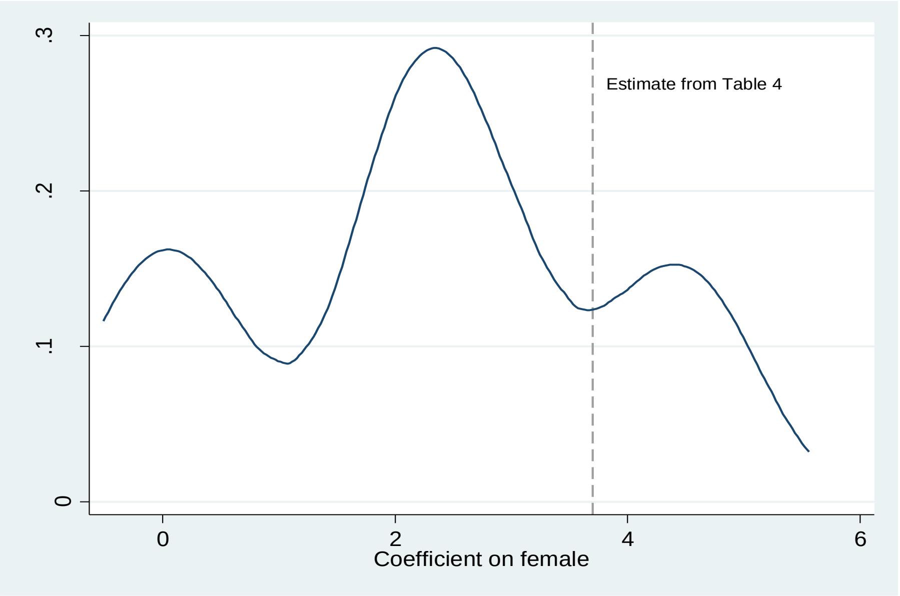
Model Robustness Statistics:		**Significance Testing:**	
Mean Estimate	2.29	Sign Stability	88%
Sampling SE	1.61	Significance rate	25%
Modeling SE	1.60		
Total SE	2.27	Positive	88%
		Positive and Sig	25%
		Negative	12%
Robustness Ratio:	1.01	Negative and Sig	0%

Modelling Distribution

Figure 2. Modeling Distribution of the Gender Effect on Mortgage Lending



Note: Estimates from 256 models. See table 7 for more information about the distribution. The vertical line shows the "preferred estimate" from Table 4 (3.7 percent higher acceptance rate for women).



Hard to draw conclusions from the evidence

without knowing more about the modeling distribution.

Why do these estimates vary so much?

Why is the distribution so non-normal?

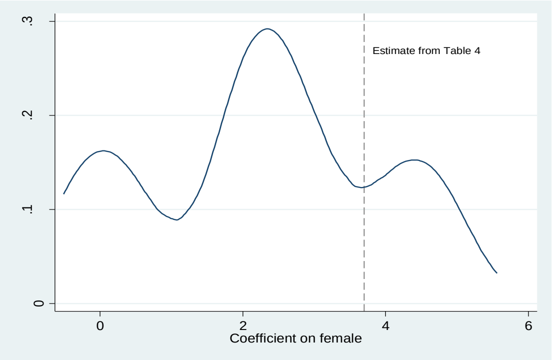
What control variables are critical to finding a positive and significant result?

These questions lead us to the **model influence analysis**

Model Influence: $\Delta\beta$ as the Effect of Interest

After estimating the full model space, decompose the modeling distribution.

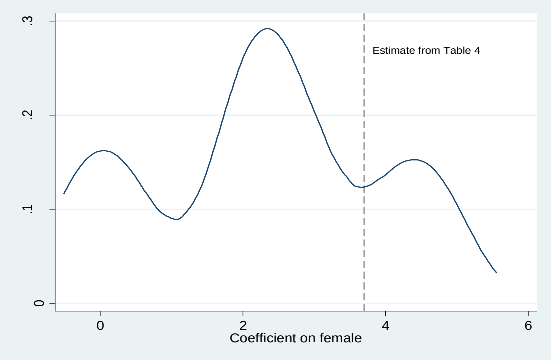
A meta-regression analysis of:



Model Influence: $\Delta\boldsymbol{\beta}$ as the Effect of Interest

After estimating the full model space, decompose the modeling distribution.

A meta-regression analysis of:



Influence regression:

$$b_j = \alpha + \theta_1 D_{1j} + \theta_2 D_{2j} + \cdots + \theta_P D_{Pj} + \varepsilon_j$$

b_j represents all the "coefficients of interest" estimated in the model space models. If 1,024 models, then 1,024 b_j estimates.

What model ingredients explain why you get different results?

Model Robustness Testing

Table 6: Model Influence Results for Gender Effect on Mortgage Lending

	Effect of Variable Inclusion	Percent Change From Mean Estimate
Married	2.47	107.8%
Black	1.91	83.3%
Self Employed	-0.30	-13.3%
Loan-to-Value Ratio	-0.25	-10.7%
Bad Credit History	-0.23	-10.1%
Housing Expense Ratio	0.19	8.4%
Payment-Income Ratio	-0.18	-8.1%
Denied Mortgage Insurance	-0.03	-1.1%
Constant	0.50	
R-squared	0.98	

Note: Based on 256 estimates reported in table 5.

Model Robustness Testing

Table 6: Model Influence Results for Gender Effect on Mortgage Lending

	Effect of Variable Inclusion	Percent Change From Mean Estimate	
Married	2.47	107.8%	Most influential
Black	1.91	83.3%	
Self Employed	-0.30	-13.3%	
Loan-to-Value Ratio	-0.25	-10.7%	
Bad Credit History	-0.23	-10.1%	
Housing Expense Ratio	0.19	8.4%	
Payment-Income Ratio	-0.18	-8.1%	
Denied Mortgage Insurance	-0.03	-1.1%	Least influential
Constant	0.50		
R-squared	0.98		

Note: Based on 256 estimates reported in table 5.

Model Robustness Testing

Table 6: Model Influence Results for Gender Effect on Mortgage Lending

	Effect of Variable Inclusion	Percent Change From Mean Estimate	
Married	2.47	107.8%	Most influential
Black	1.91	83.3%	
Self Employed	-0.30	-13.3%	
Loan-to-Value Ratio	-0.25	-10.7%	
Bad Credit History	-0.23	-10.1%	
Housing Expense Ratio	0.19	8.4%	
Payment-Income Ratio	-0.18	-8.1%	
Denied Mortgage Insurance	-0.03	-1.1%	Least influential
Constant	0.50		
R-squared	0.98		

Note: Based on 256 estimates reported in table 5.

NOTE: Influence ≠ statistical significance

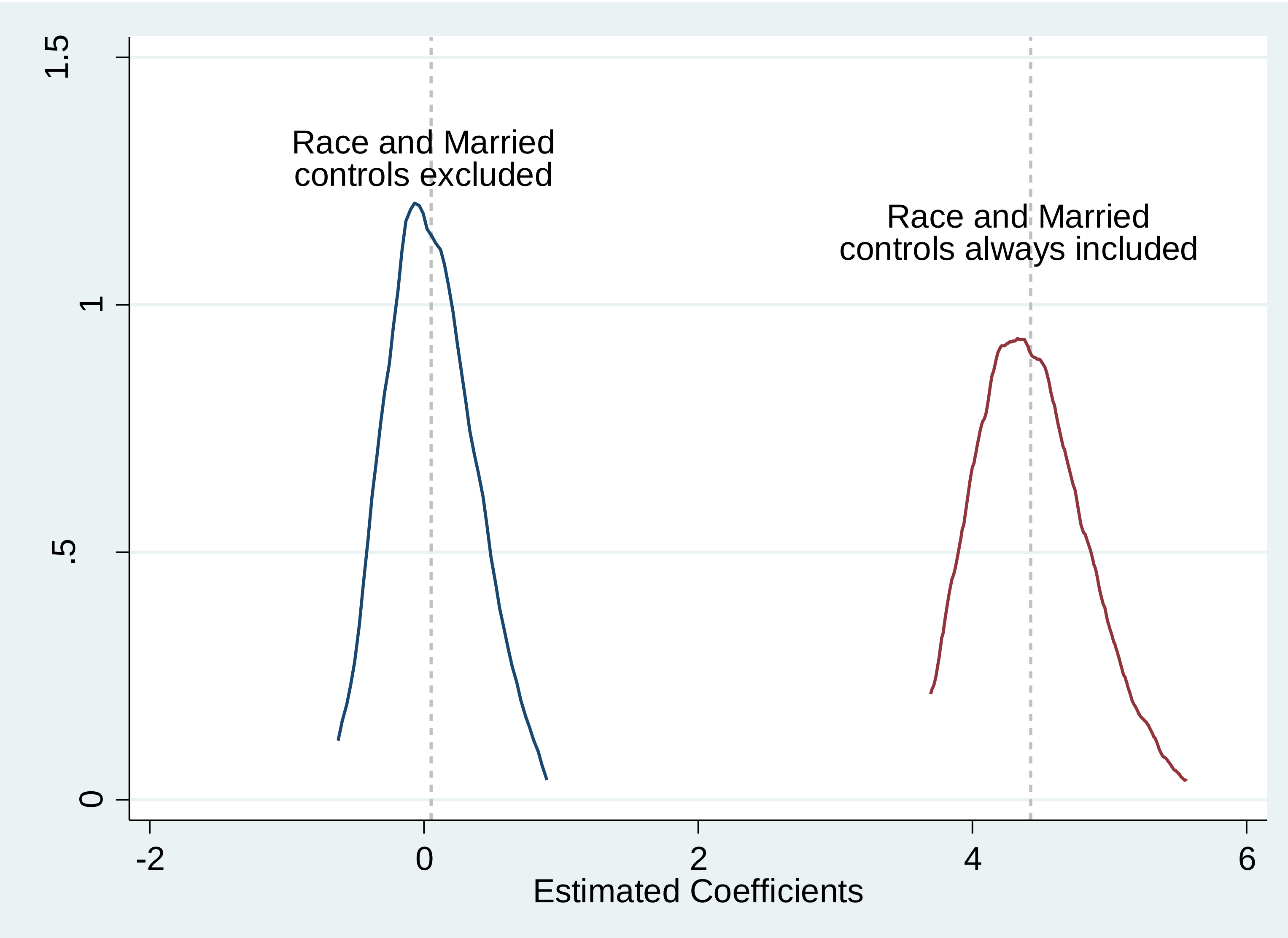
Effect of Gender on Mortgage Lending Decisions of Banks

Table 4: Determinants of Mortgage Application Acceptance

	Model: OLS	
Female	3.7*	(1.6)
Black	-11.4***	(1.8)
Housing Expense Ratio	5.8	(10.5)
Self Employed	5.6**	(1.8)
Married	4.6***	(1.3)
Bad Credit History	-25.2***	(2.3)
Payment-Income Ratio	-50.2***	(9.3)
Loan-to-Value Ratio	11.9***	(3.4)
Denied Mortgage Insurance	-71.2***	(4.2)
Constant	113.8***	(3.4)
N	2355	
adj. R-sq	0.226	

* p<0.05 ** p<0.01 *** p<0.001. Standard errors in parentheses. The outcome variable, mortgage acceptance (1 = accepted, 0 = denied), has been scaled by 100 so that coefficients can be interpreted as percent changes in the acceptance rate for a unit change in the predictor.

Modeling Distributions for the Gender Effect under Different Assumptions



Research articles often seek to tell a ‘perfect story’ with an unblemished set of supportive evidence.

Yet, acknowledging ambiguity in empirical results can lead to deeper thinking and greater insight into the social process at work.

Model influence analysis takes us well beyond the robustness results or a simple model averaging approach:

we can see which model assumptions matter, evaluate their merits, and explore their implications.

Last empirical Example:

Tax-Induced Migration

Do people tend to move from high-tax to low-tax states?

Table 7: Determinants of Cross State Migration

Poisson Models

	Model 1	Model 2
	ACS	ACS
Income tax difference	1.38	2.42*
	(1.53)	(1.23)
Population - origin	0.79***	0.83***
	(0.04)	(0.03)
Population - destination	0.73***	0.81***
	(0.03)	(0.03)
Log distance		-0.32***
		(0.04)
Contiguity		1.09***
		(0.07)
Sales tax difference		0.02
		(0.01)
Property tax difference		0.02
		(0.05)
Avg income		0.01**
		(0.00)
Natural amenities (landscape)		-0.00
		(0.00)
Constant	-16.22***	-18.64***
	(0.86)	(0.85)
N	2015	2015
pseudo R-sq	0.525	0.788

* p<0.05 ** p<0.01 *** p<0.001 Robust standard errors in parentheses

Table 7: Determinants of Cross State Migration

Poisson Models

	Model 1	Model 2	Model 3
	ACS	ACS	IRS
Income tax difference	1.38	2.42*	3.00*
	(1.53)	(1.23)	(1.33)
Population - origin	0.79***	0.83***	0.82***
	(0.04)	(0.03)	(0.03)
Population - destination	0.73***	0.81***	0.80***
	(0.03)	(0.03)	(0.03)
Log distance		-0.32***	-0.30***
		(0.04)	(0.03)
Contiguity		1.09***	1.10***
		(0.07)	(0.07)
Sales tax difference		0.02	0.02
		(0.01)	(0.01)
Property tax difference		0.02	0.07
		(0.05)	(0.05)
Avg income		0.01**	0.01
		(0.00)	(0.00)
Natural amenities (landscape)		-0.00	-0.00
		(0.00)	(0.00)
Constant	-16.22***	-18.64***	-18.30***
	(0.86)	(0.85)	(0.87)
N	2015	2015	2015
pseudo R-sq	0.525	0.788	0.780

* $p<0.05$ ** $p<0.01$ *** $p<0.001$ Robust standard errors in parentheses

Table 7: Determinants of Cross State Migration

Poisson Models

	Model 1	Model 2	Model 3
	ACS	ACS	IRS
Income tax difference	1.38	2.42*	3.00*
	(1.53)	(1.23)	(1.33)
Population - origin	0.79***	0.83***	0.82***
	(0.04)	(0.03)	(0.03)
Population - destination	0.73***	0.81***	0.80***
	(0.03)	(0.03)	(0.03)
Log distance		-0.32***	-0.30***
		(0.04)	(0.03)
Contiguity		1.09***	1.10***
		(0.07)	(0.07)
Sales tax difference		0.02	0.02
		(0.01)	(0.01)
Property tax difference		0.02	0.07
		(0.05)	(0.05)
Avg income		0.01**	0.01
		(0.00)	(0.00)
Natural amenities (landscape)		-0.00	-0.00
		(0.00)	(0.00)
Constant	-16.22***	-18.64***	-18.30***
	(0.86)	(0.85)	(0.87)
N	2015	2015	2015
pseudo R-sq	0.525	0.788	0.780

* $p<0.05$ ** $p<0.01$ *** $p<0.001$ Robust standard errors in parentheses

Tax-Induced Migration

Do people tend to move from high-tax to low-tax states?

12 possible controls,

2 possible data sets,

3 possible estimation commands

24,576 possible models

We've seen 3 models. What happens when you look at the other 24,573?

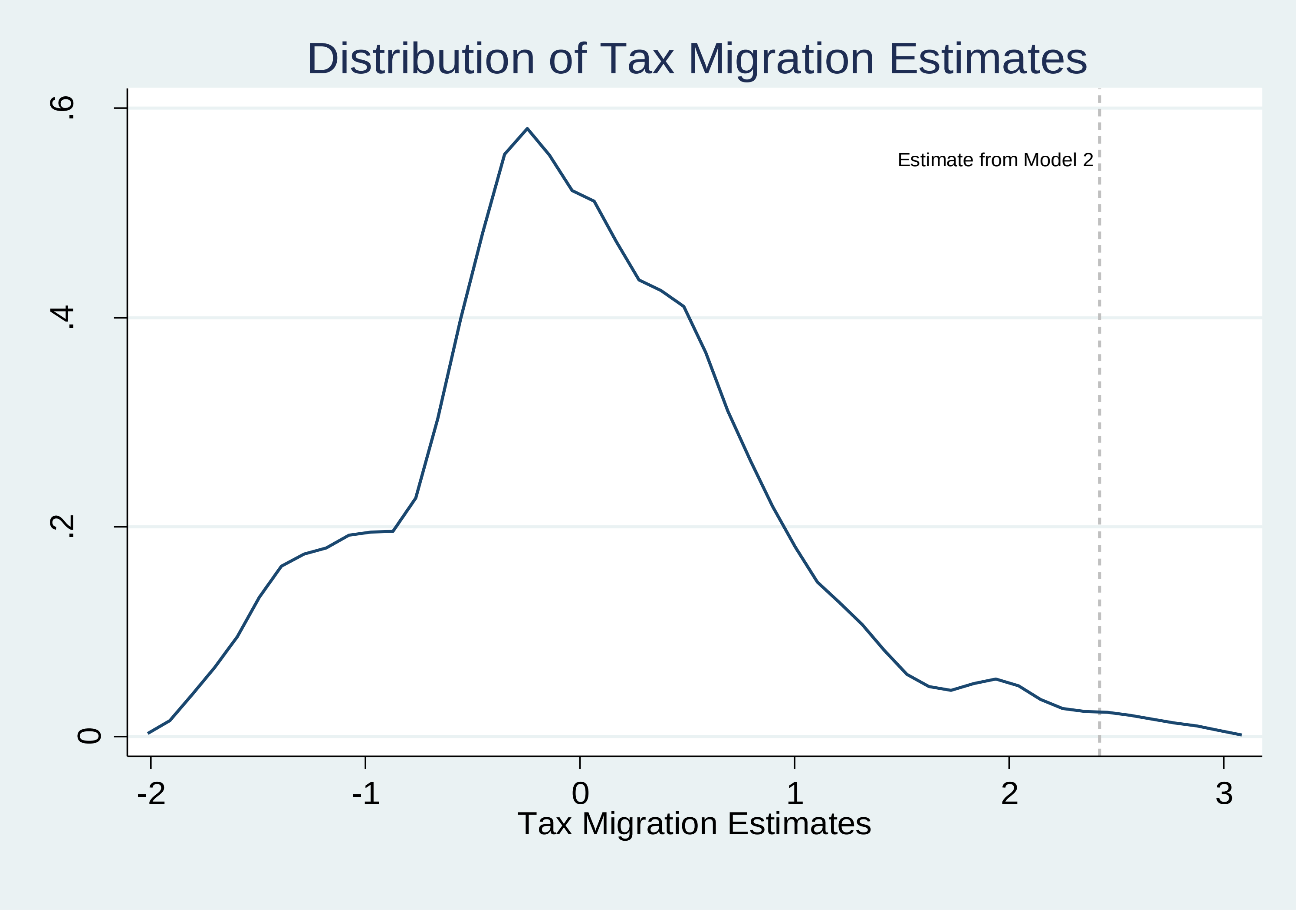
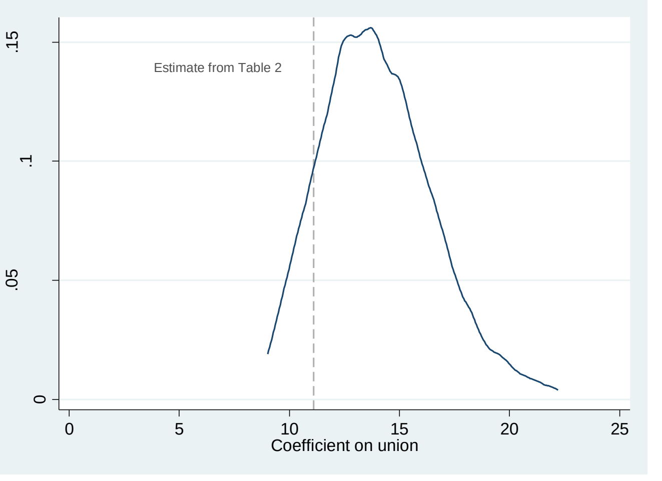

Distribution of Tax Migration Estimates
Estimate from Model 2
0
.2
.4
.6
-2
-1
0
1
2
3
Tax Migration Estimates

Table 8: Model Robustness of Tax Migration

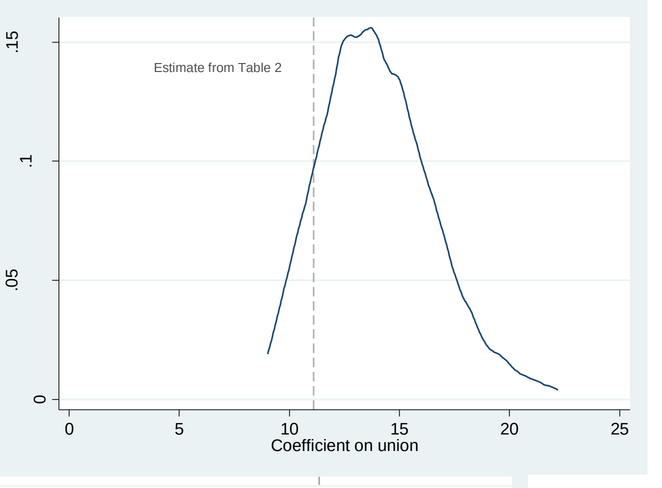
Variable of interest	income tax rate	Number of models	24,576
Outcome variable	migration	Number of observations	2,015
Possible control terms	17	Mean R-squared	.479
Model Robustness Statistics:		**Significance Testing:**	
Mean(b)	0.01	Sign Stability	51.9%
Sampling SE	1.10	Significance rate	1.5%
Modeling SE	0.83		
Total SE	1.38	Positive	48.9%
		Positive and Sig	0.2%
Robustness Ratio:	0.01	Negative	51.1%
		Negative and Sig	1.3%

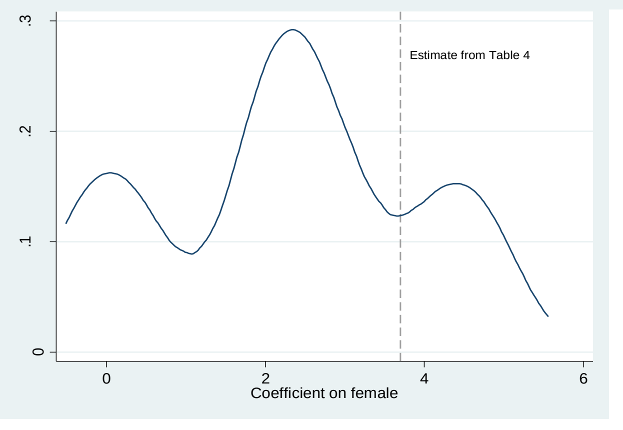
What happens when we are transparent about the model?

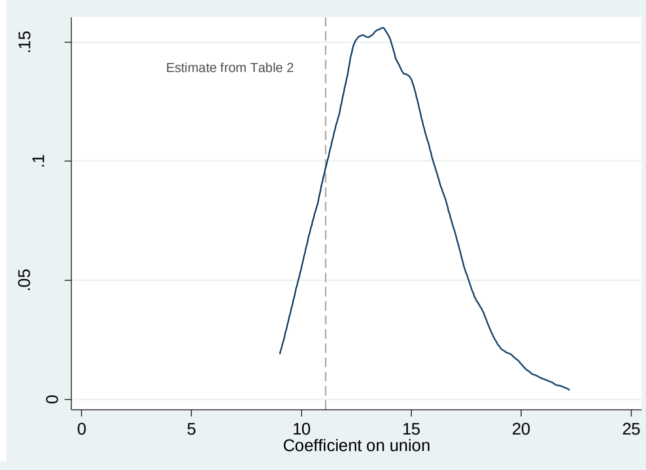
Relax model assumptions, and consider plausible alternative specifications…

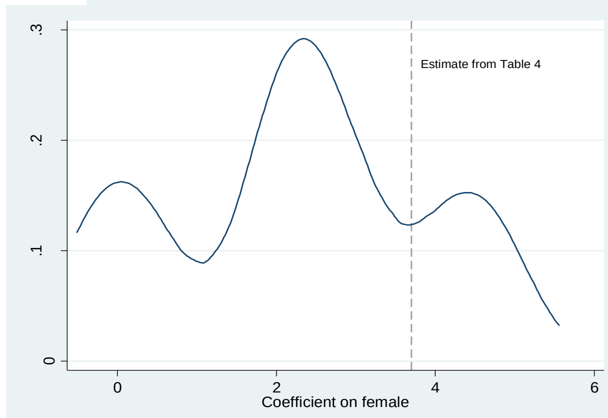


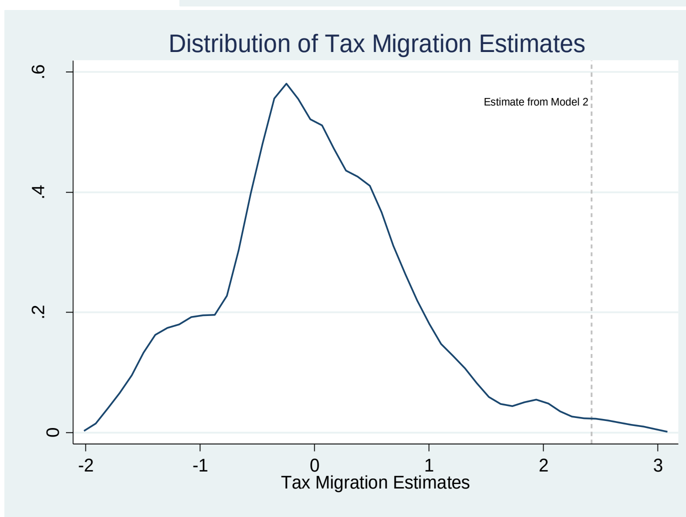
Estimate from Table 2
0
.05
.1
.15
0
5
10
15
20
25
Coefficient on union


Estimate from Table 2
0
.05
.1
.15
0
5
10
15
20
25
Coefficient on union


Estimate from Table 4
0
.1
.2
.3
0
2
4
6
Coefficient on female







HOW SCIENCE GOES WRONG.

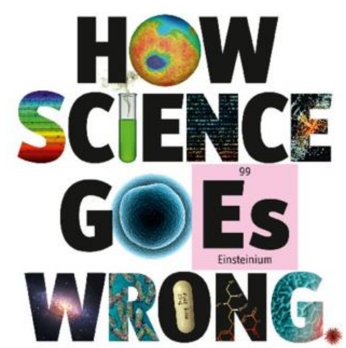
Transparency in Applied Science

Crisis in Science:
There are good reasons why scientists do not always believe other people's research:

There are many ways to do the analysis, but the "true model" is unknown – model uncertainty is *inherent* in applied research

Results should be evaluated, not just by their significance levels, but also by model robustness

(and transparency levels!)



Thank you!

Cristobal Young
cristobal.young@stanford.edu

Software & one-click replication package:
www.cristobalyoung.com

See for yourself!