Open Policy Analysis: Principles and Applications

Fernando Hoces de la Guardia

UC Berkeley:
Berkeley Initiative for Transparency in the Social Sciences

GiveWell
August 23rd, 2018

Motivation: To Producers of Policy Analysis

Cynical view:

$$\text{Policy Analysis} = \text{Research} - \text{Novelty} - \text{Rigor}$$

Motivation: To Producers of Policy Analysis

Cynical view:

$$\text{Policy Analysis} = \text{Research} - \text{Novelty} - \text{Rigor}$$

Optimists view:

$$\text{Policy Analysis} = \text{Research} - \text{Novelty} + \text{Relevance} - \text{Rigor}$$

Motivation: To Producers of Policy Analysis

Cynical view:

$$\text{Policy Analysis} = \text{Research} - \text{Novelty} - \text{Rigor}$$

Optimists view:

$$\text{Policy Analysis} = \text{Research} - \text{Novelty} + \text{Relevance} - \text{Rigor}$$

Our proposal:

$$\text{Open Policy Analysis} = \text{Research} - \text{Novelty} + \text{Relevance}$$

Motivation: To Producers of Policy Analysis

Cynical view:

$$\text{Policy Analysis} = \text{Research} - \text{Novelty} - \text{Rigor}$$

Motivation: To Producers of Policy Analysis

Cynical view:

$$\text{Policy Analysis} = \text{Research} - \text{Novelty} - \text{Rigor}$$

Optimists view:

$$\text{Policy Analysis} = \text{Research} - \text{Novelty} + \text{Relevance} - \text{Rigor}$$

Motivation: To Producers of Policy Analysis

Cynical view:

$$\text{Policy Analysis} = \text{Research} - \text{Novelty} - \text{Rigor}$$

Optimists view:

$$\text{Policy Analysis} = \text{Research} - \text{Novelty} + \text{Relevance} - \text{Rigor}$$

Our proposal:

$$\text{Open Policy Analysis} = \text{Research} - \text{Novelty} + \text{Relevance}$$

Motivation: To Producers of Research

Think for a second of the one paper that you are most proud of. Now think of the best estimate that you have in that paper.

Motivation: To Producers of Research

Think for a second of the one paper that you are most proud of. Now think of the best estimate that you have in that paper.

- Do you know if that paper had any **direct** effect on public policy? (direct: estimate used in policy report, law, testimony
  indirect: general knowledge, NYT op-ed)

Motivation: To Producers of Research

Think for a second of the one paper that you are most proud of. Now think of the best estimate that you have in that paper.

- Do you know if that paper had any **direct** effect on public policy? (direct: estimate used in policy report, law, testimony
  indirect: general knowledge, NYT op-ed)
- If that estimate where to be revise to by a factor of 2 (or 10). How should the policy analysis change?

Why We Need Open Policy Analysis

Fernando Hoces de la Guardia[1]
Sean Grant[2]
Edward Miguel[1]

[1]UC Berkeley:
Berkeley Initiative for Transparency in the Social Sciences
[2]RAND

GiveWell
August 23rd, 2018

Policy Analysis And The Evidence-Based Policy Movement

Evidence-Based movement is growing.

- "The golden age of evidence-based policy" (Haskins 2017).
- Credible causal evidence (Angrist & Pischke, 2010)
- Transparency and reproducibility of research (Miguel et al. 2014).
- Commission on Evidence-Based Policymaking (CEBP, 2017)

Policy Analysis is a fundamental link.

- As many definitions as textbooks (Dunn, 2015; Weimer & Vining, 2017; Williams, 1971)
- Common denominator: client-oriented empirical analysis meant to inform a specific policy debate
- Aspires at scientific rigor. (Wildavsky 1979),

Policy Analysis And The Evidence-Based Policy Movement

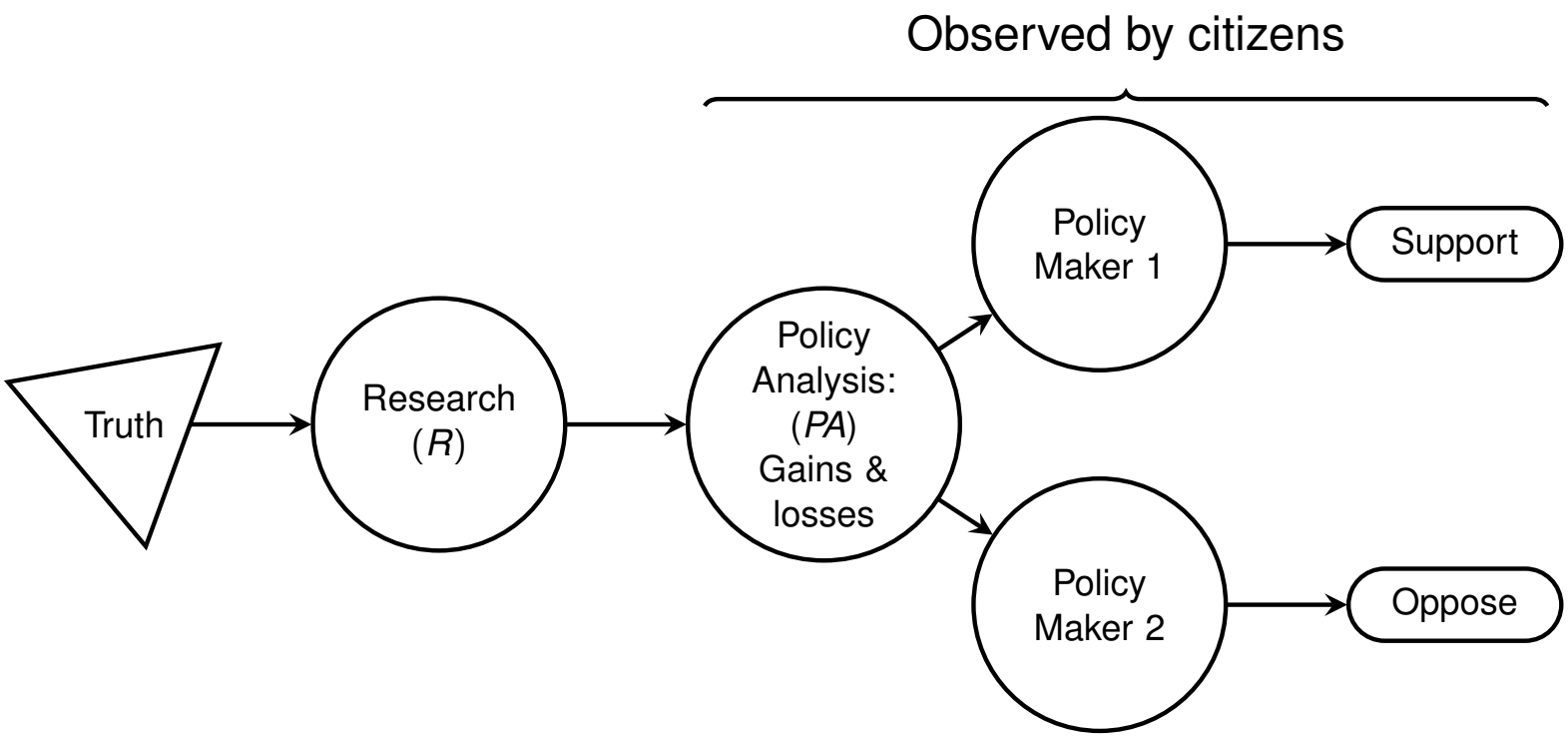
Evidence-Based movement is growing.

- "The golden age of evidence-based policy" (Haskins 2017).
- Credible causal evidence (Angrist & Pischke, 2010)
- Transparency and reproducibility of research (Miguel et al. 2014).
- Commission on Evidence-Based Policymaking (CEBP, 2017)

Policy Analysis is a fundamental link.

- As many definitions as textbooks (Dunn, 2015; Weimer & Vining, 2017; Williams, 1971)
- Common denominator: client-oriented empirical analysis meant to inform a specific policy debate
- Aspires at scientific rigor. (Wildavsky 1979),

One Ideal Evidence-Based Policy Link



Reproducibility Crisis In Empirical Research

- Large magnitude of publication bias (Franco et al 2014).
- Evidence of extensive p-hacking across social science disciplines (Gerber et al 2008, Brodeur et al 2016).
- Replication rates are low (Collaboration et al, 2015 , Camerer et al, 2016).
- Computational reproducibility is also low (Stodden et al 2016, Chang and Li 2015, Gertler et al 2018).

Credibility Crisis Of Policy Analysis

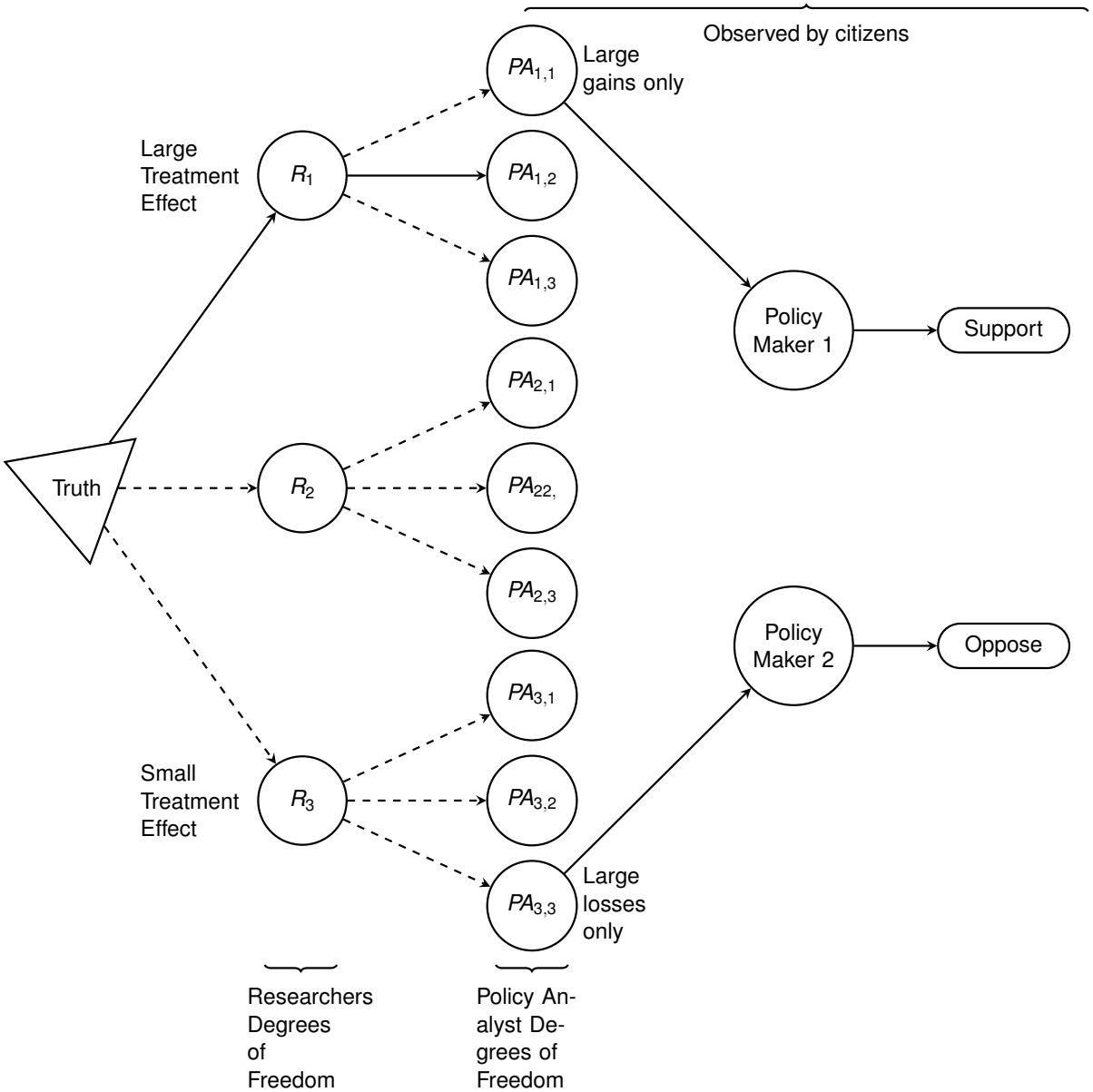
- Incredible Certitudes (Manski, 2013)
- Report wars (Wesselink et al, 2013)
- Alternative facts ("The Death of Expertise" Nichols, 2017; "The Death of Truth", Kakutani 2018; "Truth Decay", Rich & Kavanagh 2018)

Credibility Crisis Of Policy Analysis

- Incredible Certitudes (Manski, 2013)
- Report wars (Wesselink et al, 2013)
- Alternative facts (“The Death of Expertise” Nichols, 2017; “The Death of Truth”, Kakutani 2018; “Truth Decay”, Rich & Kavanagh 2018)

How This Affects The Evidence Based Policy Link?

How This Affects The Evidence Based Policy Link?



Relevance

Main consequences of policy analysis that lacks openness:

1. Cherry picking evidence.
2. Challenging to automate and Improve systematically recurring reports.
3. Difficulty understanding how research informs policy analysis.

Cherry Picking Evidence

"When I was director of the CBO, I was very frustrated when we would write a policy report [saying] a certain policy would have these two advantages and these two disadvantages, and the advocates would quote only the part about the advantages, and the opponents would quote only the part about the disadvantages. That encourages the view that there are simple answers. There aren't generally simple answers. There are trade-offs."

— Douglas Elmendorf (Director of CBO, 2009-2015)
Harvard Magazine, 2016

Cherry Picking Evidence

"When I was director of the CBO, I was very frustrated when we would write a policy report [saying] a certain policy would have these two advantages and these two disadvantages, and the advocates would quote only the part about the advantages, and the opponents would quote only the part about the disadvantages. That encourages the view that there are simple answers. There aren't generally simple answers. There are trade-offs."

— Douglas Elmendorf (Director of CBO, 2009-2015)
Harvard Magazine, 2016

Difficulty Understanding how Research Informs Policy Analysis

- What happens when new research emerges?
  - What if $\hat{\tau}$(Blattman, Fiala, and Martinez 2020) $= \frac{1}{2}\hat{\tau}$(Blattman, Fiala, and Martinez 2013)? Or $\tau_{2020} = 2\tau_{2013}$?
- Where a the largest unknowns in the policy analysis?
  - GiveWell lists at least 100 parameters in its cost-effectiveness analysis. What are the 5/10 most important ones?
- Where is the marginal piece of research most informative for this analysis?
  - Are the gaps in knowledge for this PA guiding the research agenda?

Difficulty Understanding how Research Informs Policy Analysis

- What happens when new research emerges?
  - What if $\hat{\tau}$(Blattman, Fiala, and Martinez 2020) $= \frac{1}{2}\hat{\tau}$(Blattman, Fiala, and Martinez 2013)? Or $\tau_{2020} = 2\tau_{2013}$?
- Where a the largest unknowns in the policy analysis?
  - GiveWell lists at least 100 parameters in its cost-effectiveness analysis. What are the 5/10 most important ones?
- Where is the marginal piece of research most informative for this analysis?
  - Are the gaps in knowledge for this PA guiding the research agenda?

Difficulty Understanding how Research Informs Policy Analysis

- What happens when new research emerges?
  - What if $\hat{\tau}$(Blattman, Fiala, and Martinez 2020) $= \frac{1}{2}\hat{\tau}$(Blattman, Fiala, and Martinez 2013)? Or $\tau_{2020} = 2\tau_{2013}$?
- Where a the largest unknowns in the policy analysis?
  - GiveWell lists at least 100 parameters in its cost-effectiveness analysis. What are the 5/10 most important ones?
- Where is the marginal piece of research most informative for this analysis?
  - Are the gaps in knowledge for this PA guiding the research agenda?

The Open Science Movement

- Definition of principles of Open Science/Research Transparency (Miguel et al 2014)
- Development of guidelines to operationalize principles of Open Science (Nosek et al 2015)
- Journals and funders: Journals (Science + 5k other journals), Registries (AEA), Funders (NIH, NSF and multiple donors)

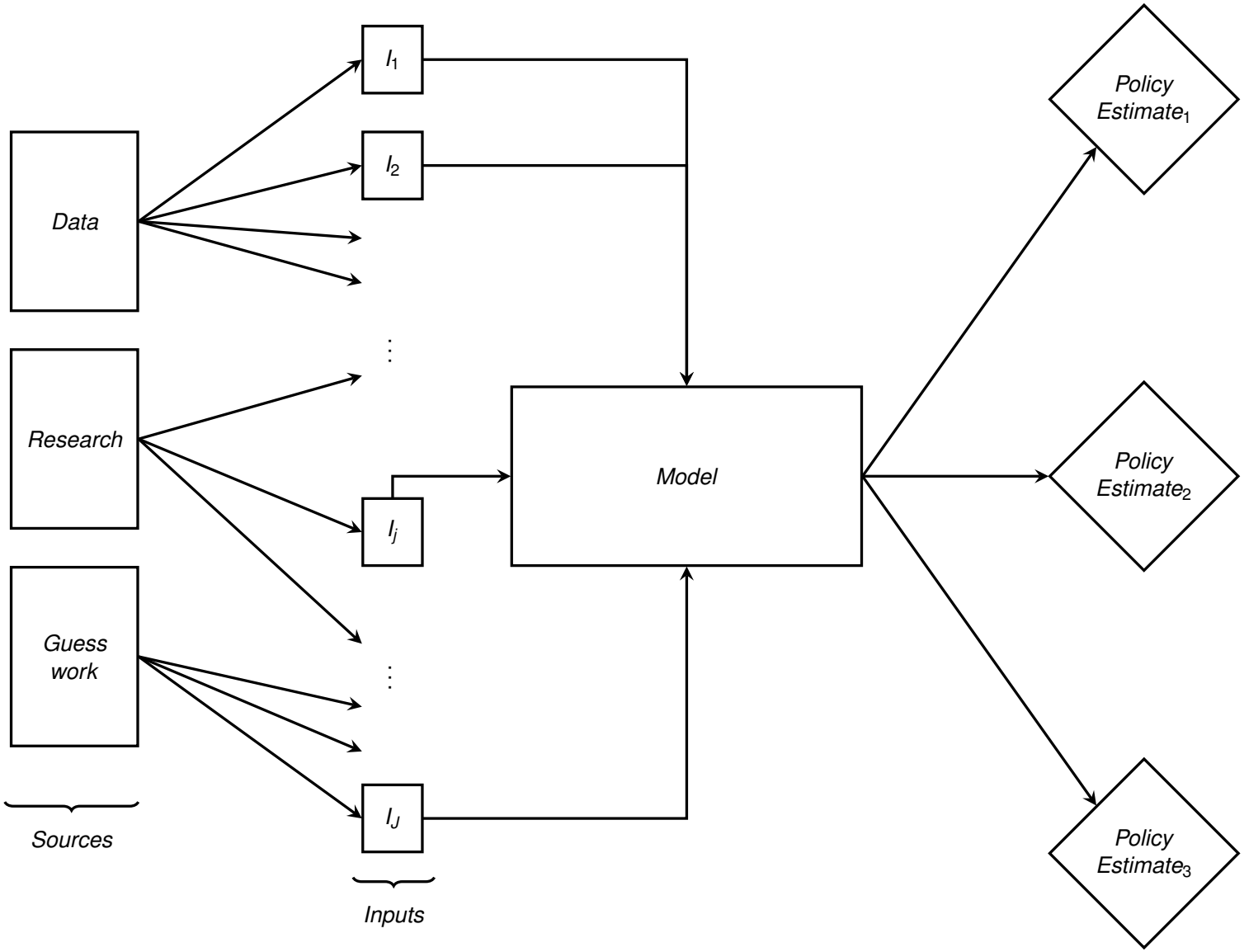
Open Science

	Empirical Research	Policy Analysis
Problems	Reproducibility Crisis	Credibility Crisis
Solutions	*Open Science* Principles, Guidelines, Applications	

Open Policy Analysis

	Empirical Research	Policy Analysis
Problems	Reproducibility Crisis	Credibility Crisis
Solutions	*Open Science* Principles, Guidelines, Applications	*Open Policy Analysis* Principles

The Process of Policy Analysis



Principles for Open Policy Analysis

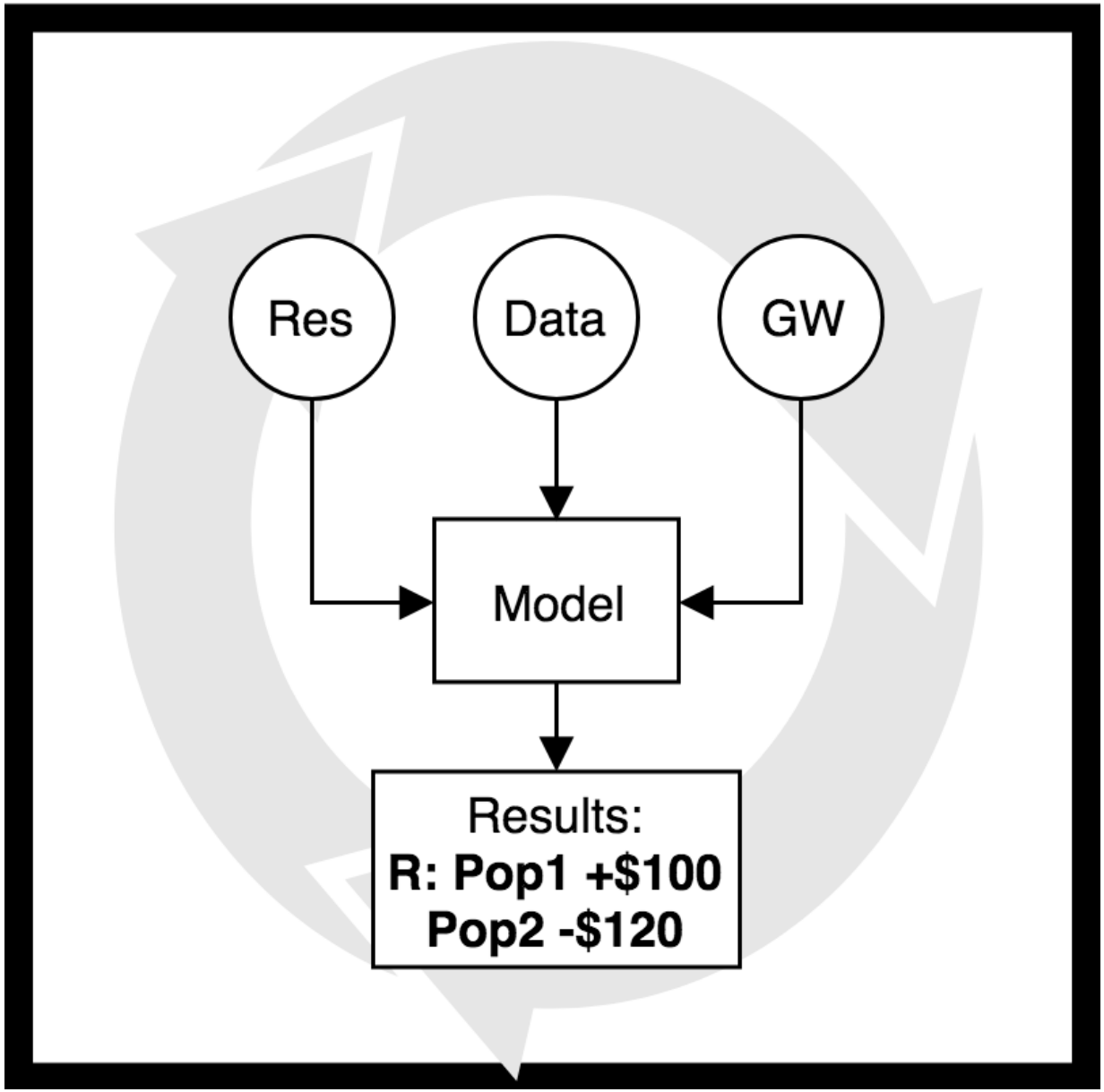
Proposed principles:

1. Computational Reproducibility
2. Analytic Transparency
3. Output Transparency

Principle 1

Computational Reproducibility

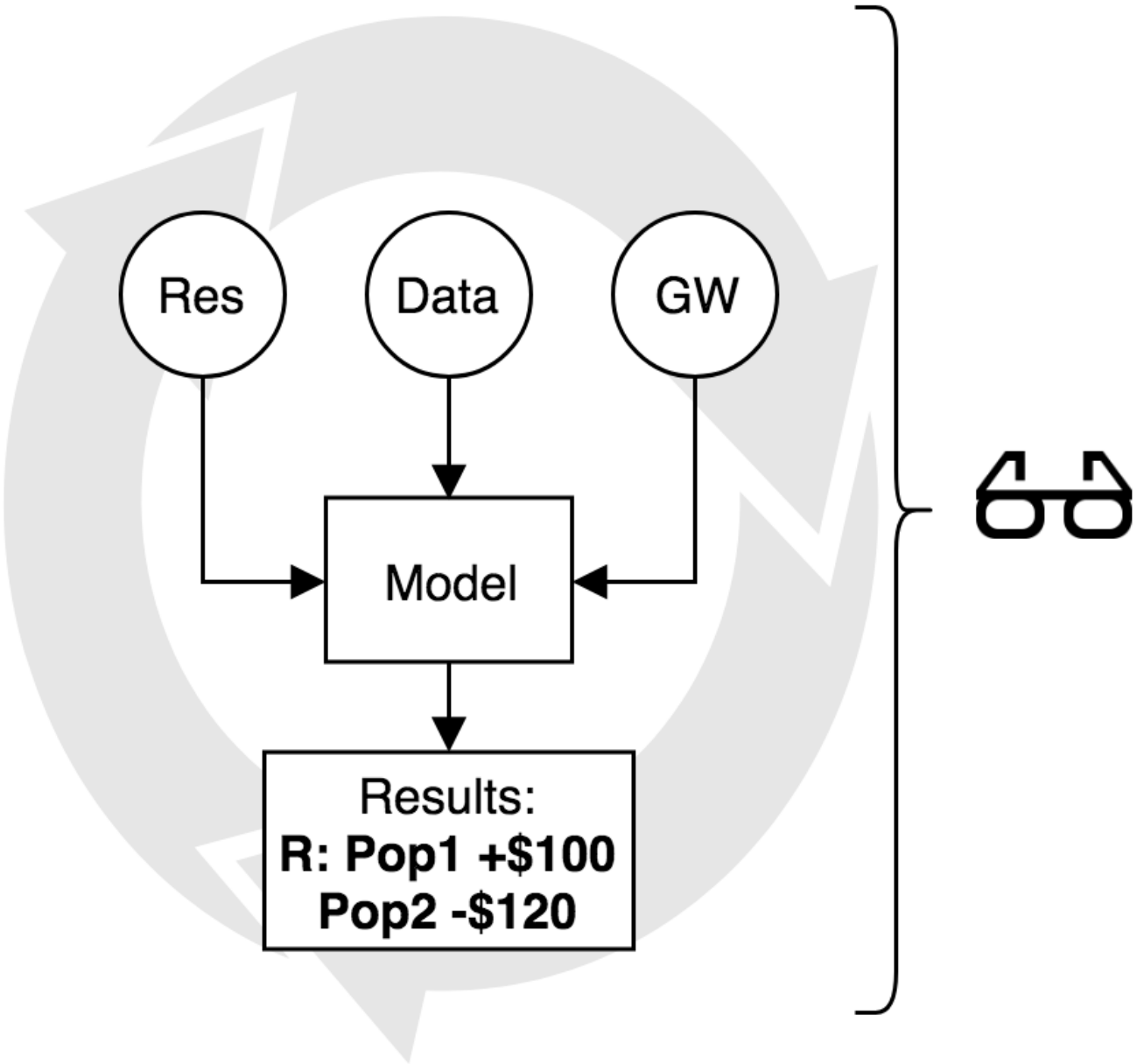
- Literate Programming
- Version control
- File structure
- Label sources



Principle 2

Analytic Transparency

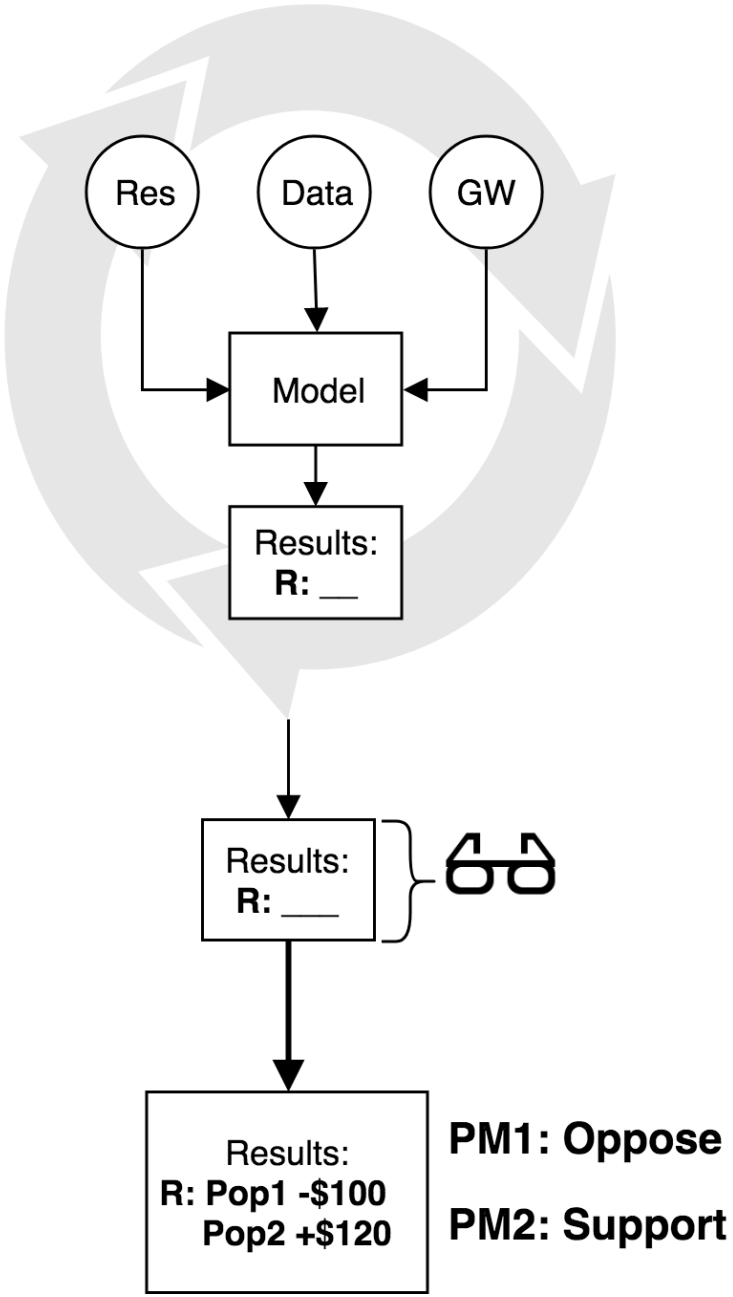
- Open code
- Open data
- Report as Dynamic Document



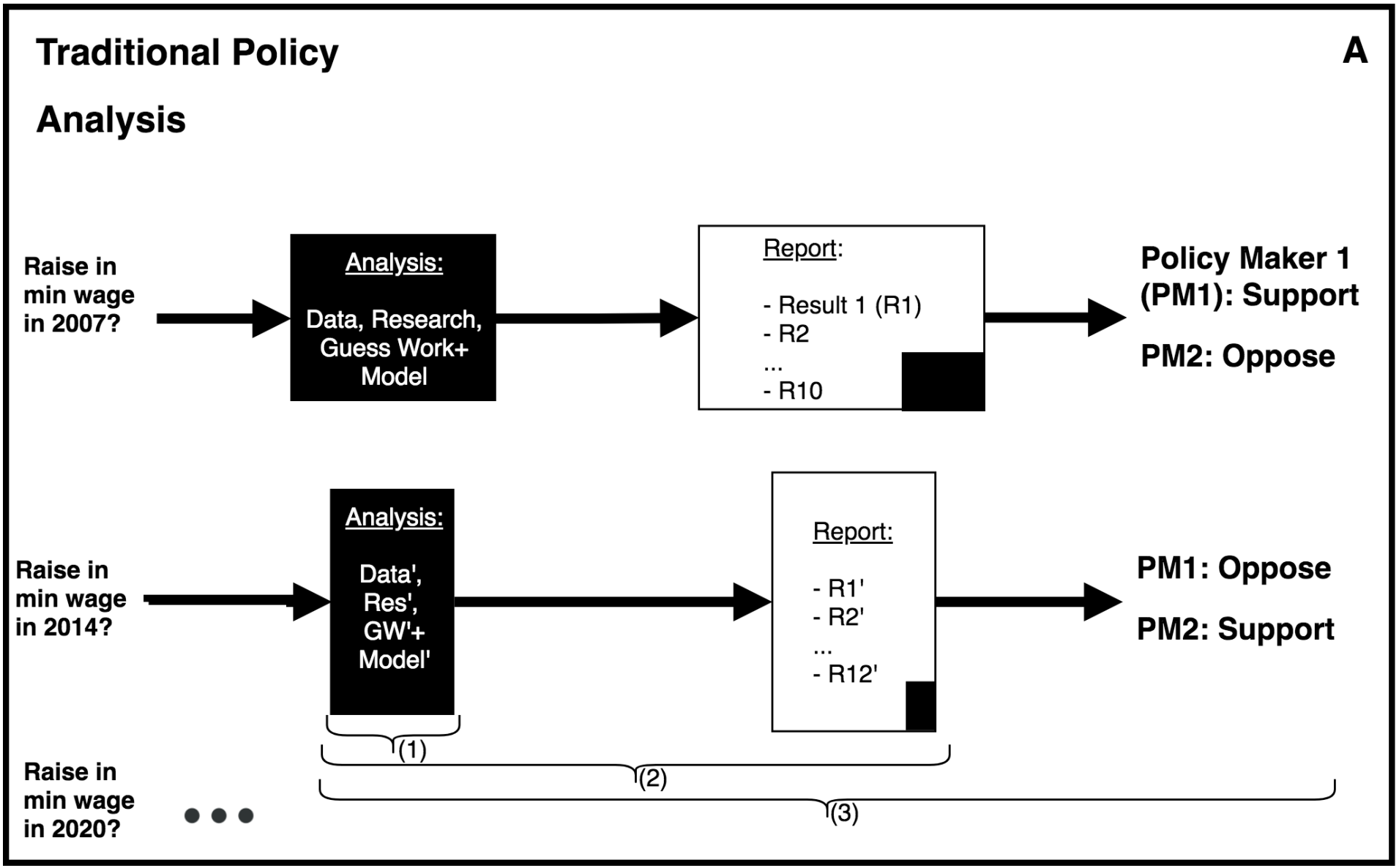
Principle 3

Output Transparency

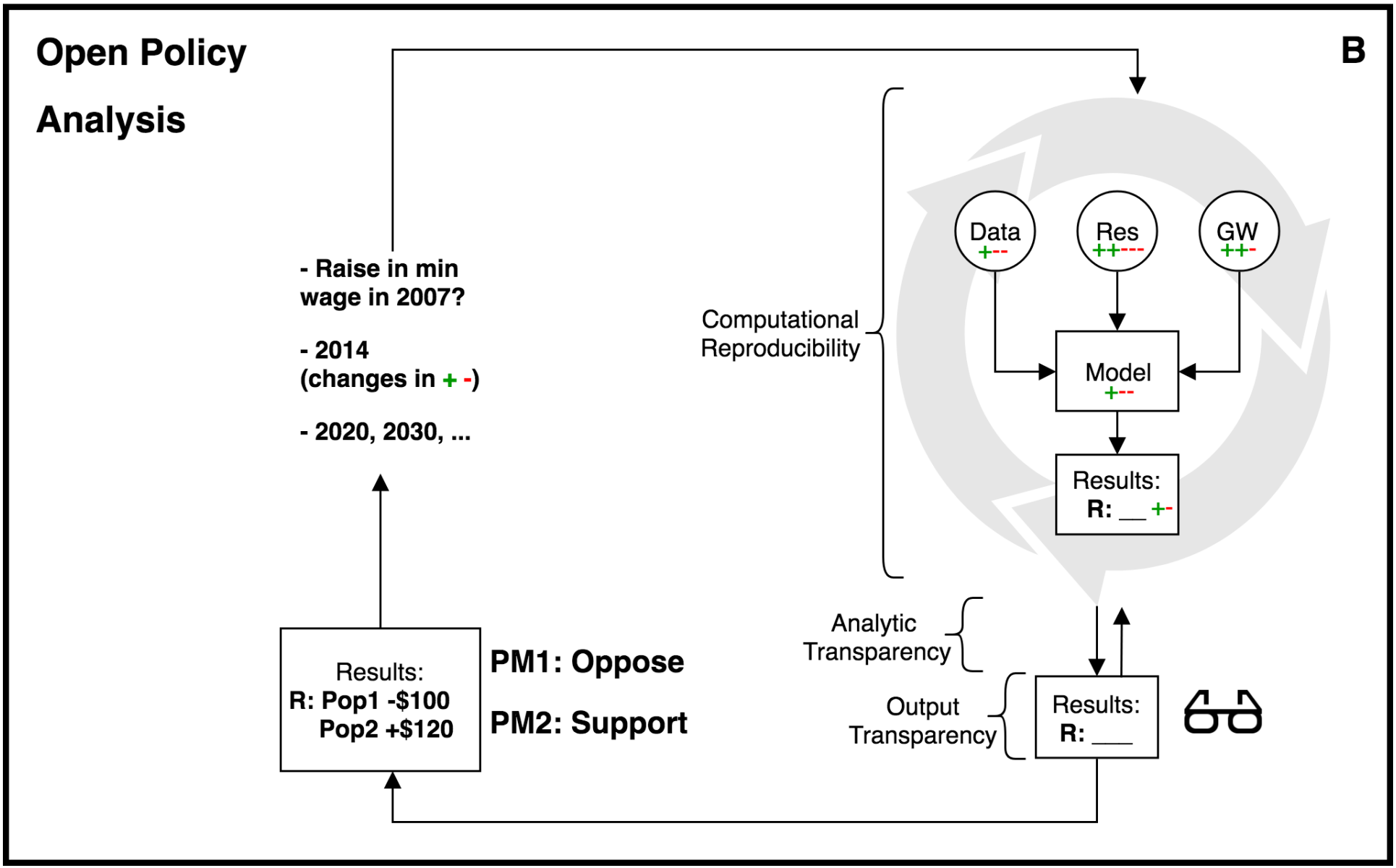
- Pre-committed output display
- Assumptions-output link



Summing Up: Where We Are



Summing Up: Where Should We Go



Open Policy Analysis: A Case Study of the Minimum Wage Policy Estimate

Fernando Hoces de la Guardia

UC Berkeley:
Berkeley Initiative for Transparency in the Social Sciences

GiveWell
August 23rd, 2018

Motivation: Gap On How to Conduct OPA

	Empirical Research	Policy Analysis
Problems	Reproducibility Crisis	Credibility Crisis
Solutions	*Open Science* Principles, Guidelines, Applications	*Open Policy Analysis* Principles

Motivation: Gap On How to Conduct OPA

	Empirical Research	Policy Analysis
Problems	Reproducibility Crisis	Credibility Crisis
Solutions	*Open Science* Principles, Guidelines, Applications	*Open Policy Analysis* Principles, Guidelines, Applications

Description of Case Study

"The Effects of a Minimum-Wage Increase on Employment and Family Income" Congressional Budget Office (2014)

Description: CBO estimated the effects of a raise in the federal minimum wage from $7.25/hr to $10.10/hr.

Main policy estimates:

- 500,000 jobs would be lost.
- 16.5 million workers would receive a salary increase.
- Distributional effects: below poverty line (PL) +$5billion; between one and three PL +$12billion; between three and six PL +$2billion; above six PL -$17billion

Key research estimate: Elasticity of labor demand for teenagers in the labor force.

Summary of Adapted Guidelines

Standard	Level 0	Level 1	Level 2	Level 3
Workflow	Policy estimates vaguely described	All the inputs, and their corresponding sources, used in the calculations are listed	Lvl 1 + Policy estimates are listed, in same unit if possible	Lvl 2 + all the components can be modified with little effort
Data	Report says nothing	Clearly stated whether all, some components, or none of the data is available, with instructions for access when possible.	Lvl 1 + report and data are in same place	Lvl 2 + Report has specific lines of code that call the data and changes in the data produce traceable changes in the report
Methods & Code	Key assumption are listed	Methods are described in prose. Large amount of work is required to reproduce qualitatively similar estimates	Methods and described in prose, with detailed formulas, and code is provided as supplementary material	Lvl 2 + All is in the same document where changes in the code affect the output automatically

From TOP guidelines (Nosek et al 2015) v1.0.1

Summary of Adapted Guidelines

Standard	Level 0	Level 1	Level 2	Level 3
Workflow	Policy estimates vaguely described	All the inputs, and their corresponding sources, used in the calculations are listed	Lvl 1 + Policy estimates are listed, in same unit if possible	Lvl 2 + all the components can be modified with little effort
Data	Report says nothing	Clearly stated whether all, some components, or none of the data is available, with instructions for access when possible.	Lvl 1 + report and data are in same place	Lvl 2 + Report has specific lines of code that call the data and changes in the data produce traceable changes in the report
Methods & Code	Key assumption are listed	Methods are described in prose. Large amount of work is required to reproduce qualitatively similar estimates	Methods and described in prose, with detailed formulas, and code is provided as supplementary material	Lvl 2 + All is in the same document where changes in the code affect the output automatically

From TOP guidelines (Nosek et al 2015) v1.0.1

Summary of Adapted Guidelines

Standard	Level 0	Level 1	Level 2	Level 3
Workflow	Policy estimates vaguely described	All the inputs, and their corresponding sources, used in the calculations are listed	Lvl 1 + Policy estimates are listed, in same unit if possible	Lvl 2 + all the components can be modified with little effort
Data	Report says nothing	Clearly stated whether all, some components, or none of the data is available, with instructions for access when possible.	Lvl 1 + report and data are in same place	Lvl 2 + Report has specific lines of code that call the data and changes in the data produce traceable changes in the report
Methods & Code	Key assumption are listed	Methods are described in prose. Large amount of work is required to reproduce qualitatively similar estimates	Methods and described in prose, with detailed formulas, and code is provided as supplementary material	Lvl 2 + All is in the same document where changes in the code affect the output automatically

From TOP guidelines (Nosek et al 2015) v1.0.1

Summary of Adapted Guidelines

Standard	Level 0	Level 1	Level 2	Level 3
Workflow	Policy estimates vaguely described	All the inputs, and their corresponding sources, used in the calculations are listed	Lvl 1 + Policy estimates are listed, in same unit if possible	Lvl 2 + all the components can be modified with little effort
Data	Report says nothing	Clearly stated whether all, some components, or none of the data is available, with instructions for access when possible.	Lvl 1 + report and data are in same place	Lvl 2 + Report has specific lines of code that call the data and changes in the data produce traceable changes in the report
Methods & Code	Key assumption are listed	Methods are described in prose. Large amount of work is required to reproduce qualitatively similar estimates	Methods and described in prose, with detailed formulas, and code is provided as supplementary material	Lvl 2 + All is in the same document where changes in the code affect the output automatically

From TOP guidelines (Nosek et al 2015) v1.0.1

Applying Guidelines to Build an Open Report

DEMO .

Demo Checklist

- One-click reproducible & machine independent.
- WYSIATI.
- Readable. Weather you know `R` or not.
- TO DO: diff between two versions.
- Sensitivity analysis.

Sensitivity Analysis: Status Quo

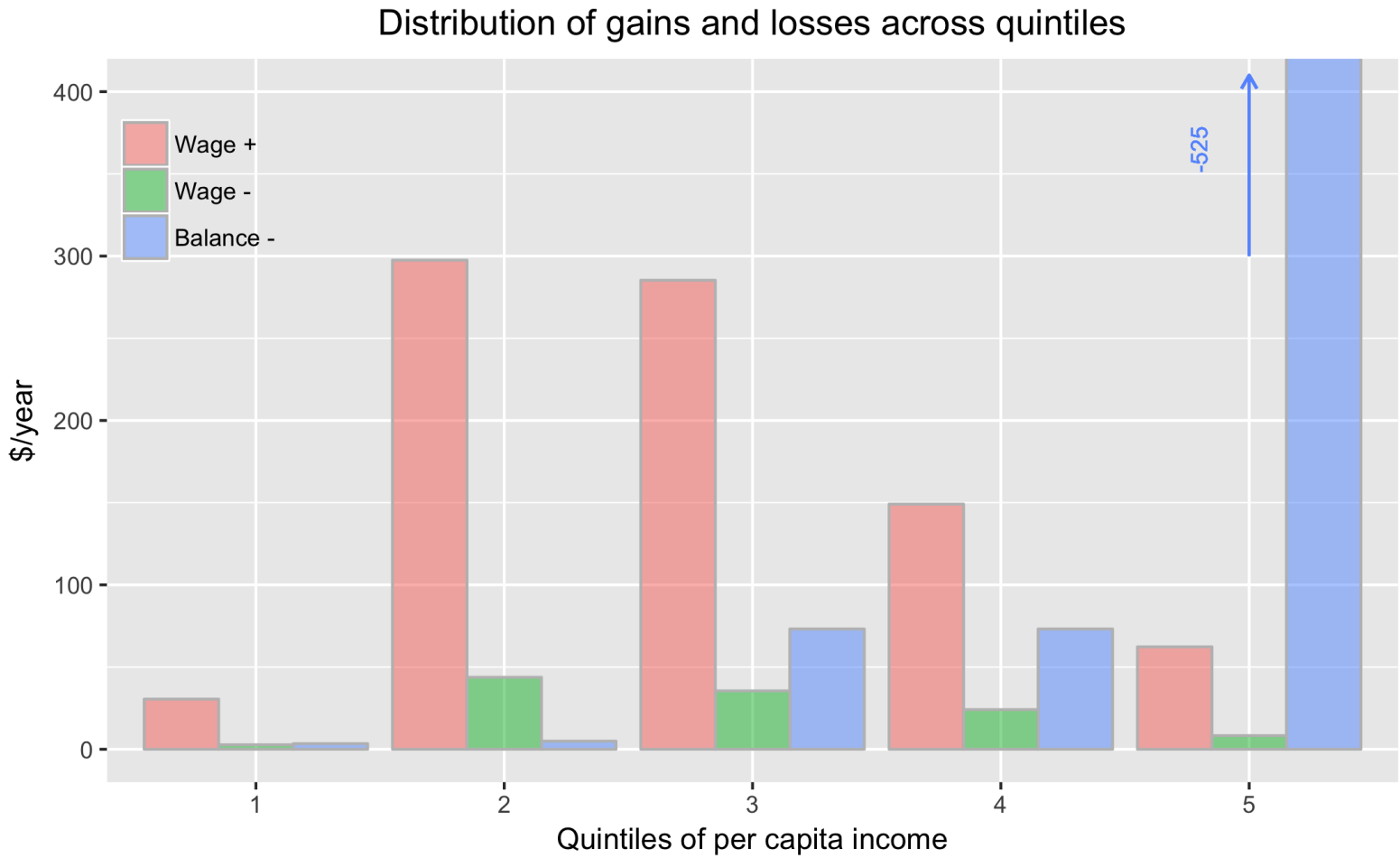


Figure: Default settings

SA: Change in Elasticity of Labor Demand

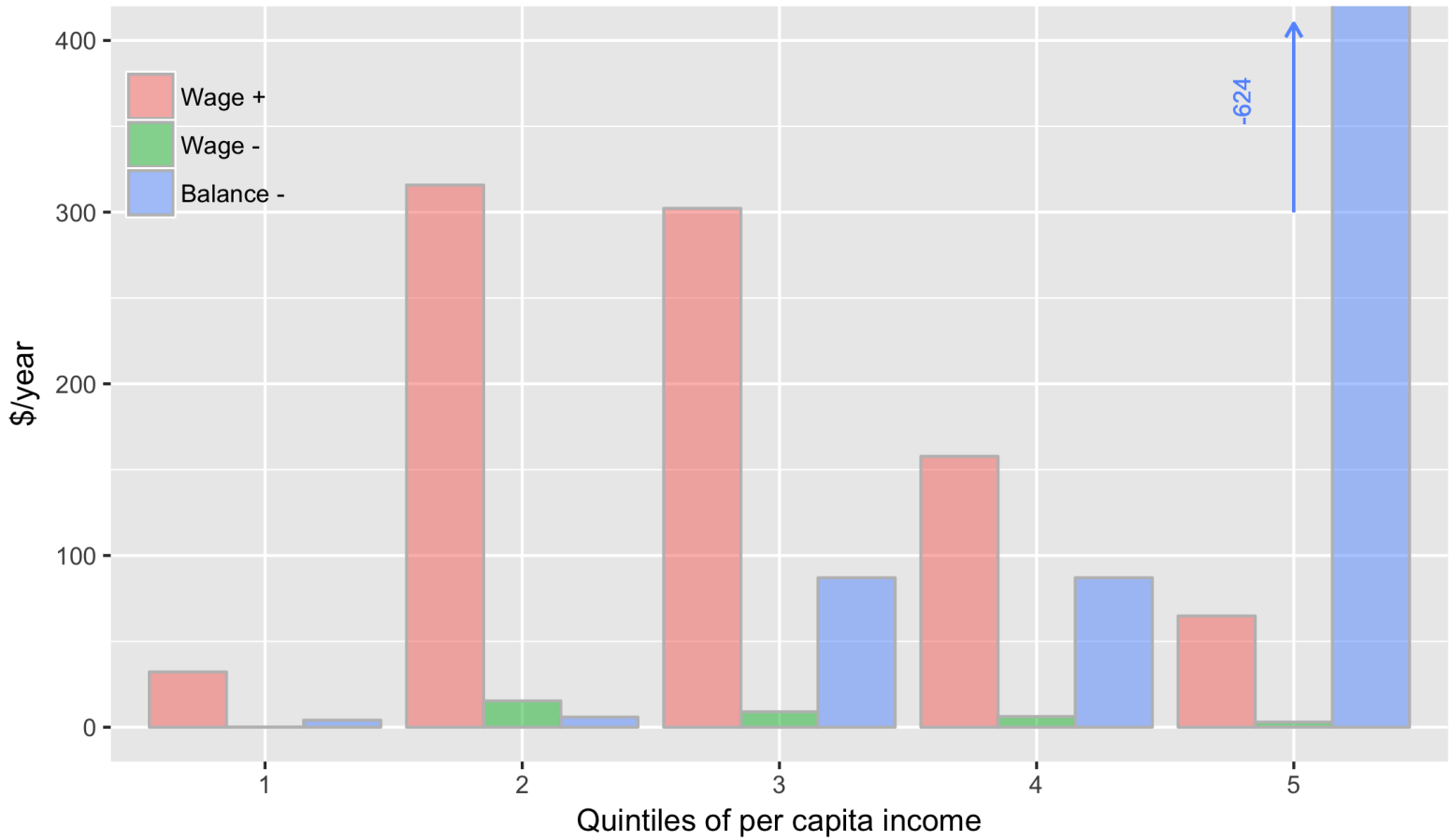


Figure: From $\eta_{lit}^{teens} = -0.1$ to $\eta_{lit}^{teens} = -0.01 (\Delta^{-} 90\%)$

Sensitivity Analysis: Status Quo

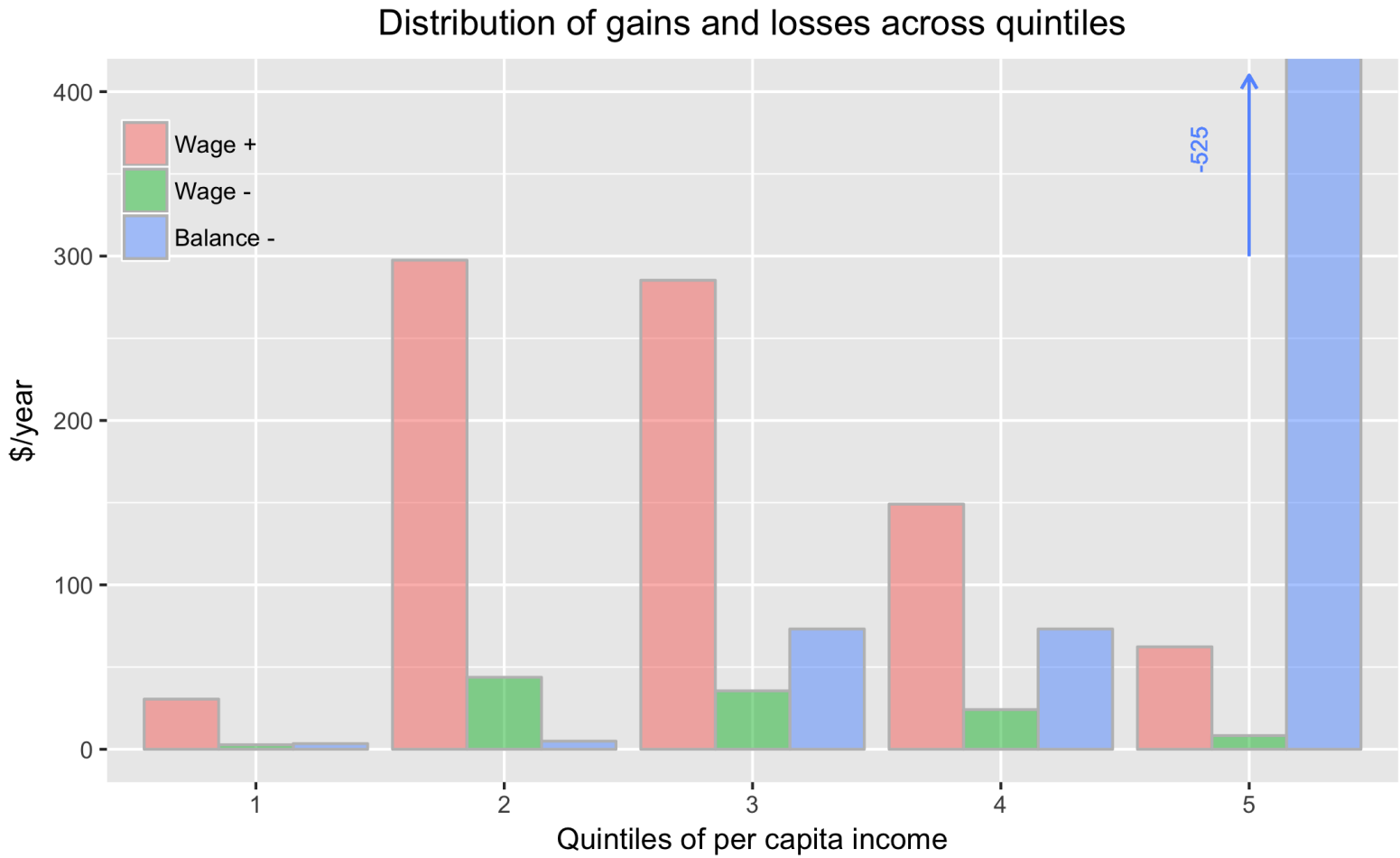


Figure: Default settings

SA: Change in Distribution of Balance Loses

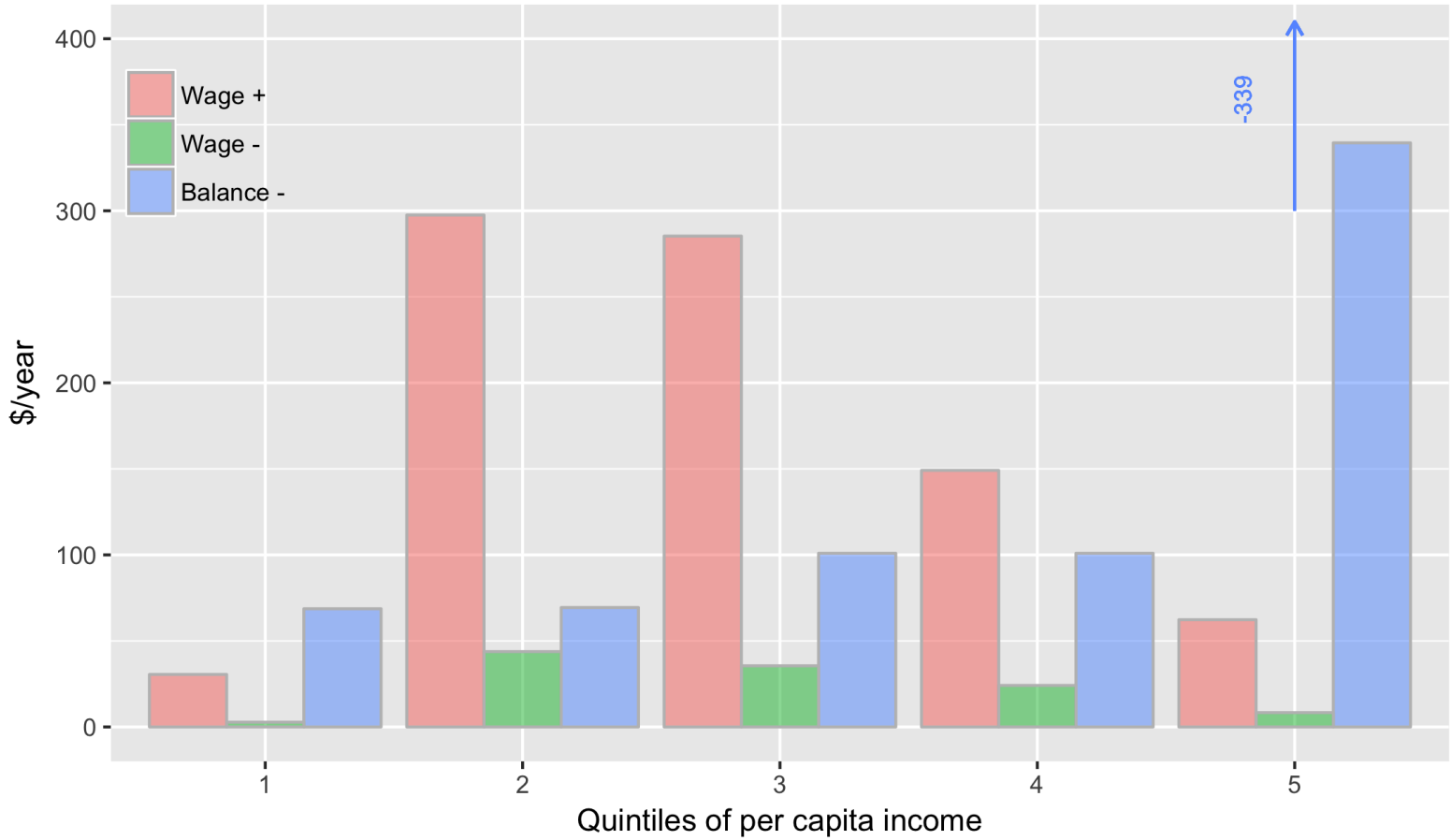


Figure: From $(1PL, 6PL) \sim (1\%, 29\%, 70\%)$ to $(20\%, 40\%, 40\%)$

Sensitivity Analysis For Multiple Parameters

Table: $\%\Delta W$ for a $\%\Delta$ in inputs. Two sample policy makers.

		Re-distributional Preferences			
		Dislikes ($\rho = -0.1$)		Likes ($\rho = 0.1$)	
Source	Input	$10\%\Delta^+$	$10\%\Delta^-$	$10\%\Delta^+$	$10\%\Delta^-$
Data					
	Annual wage growth (g_w)	-3%	2%	-2%	1%
	Annual growth in N	0.8%	-0.9%	0.5%	-0.5%
Research					
	η_{teen}	-4%	4%	-2%	2%
	Ripple Scope $(8.7, 11.5)$	37%	-24%	21%	-14%
	Ripple Intensity $(50\%\Delta w)$	5%	-5%	3%	-3%
Guess Work					
	Extrapolation factor (F_{ex})	-3%	2%	-1%	1%
	Non compliance (α_1)	-7%	7%	-4%	4%
	Substitution factor (F_{sub})		20%		-8%
	Net benefits	-5%	5%	2%	-2%
	Distribution of balance losses				
	Current: $(1\%, 29\%, 70\%)$				
	$(1\%, 4\%, 95\%)$	22%		13%	
	$(5\%, 35\%, 60\%)$	-17%		-9%	
	$1/N$	-129%		-73%	

Limitations

- There is additional scope for reproducibility.
- Complete case study requires extensive institutional knowledge.
- Guidelines need to be build based on consensus of practitioners.

Discussion

Let's assume this becomes the new status quo.

- Costs of producing the next report on effects of minimum wage will be very small.
- Every additional effort will imply improvements on the "state of the art" report (e. g. *dBL*; $\eta(MW), \alpha_1(MW)$)
- Learning about one parameter (QALYs, DWL) will update estimates *across* reports.
- Much easier to have a substantive and normative policy debate. Pilot example: `Shiny App!`.

BITSS/CEGA's next step to push OPA forward

- Partner with key policy analysts to build more case studies (CBO, Tax policy/Inequality, Chilean pension reform)
- Guidelines and trainings
- Developing a new model for collaboration among policy analysts, policy-makers and researchers, with the aim of fostering more direct impact on high-level government decisions

Your next steps to push OPA forward

- Collaborate with BITSS to open up your PA.
- Fund OPA: directly or conditionally.
- Train students/analysts in OPA.
- Present/showcase your OPA. Pioneers: GiveWell, AEI.
- Nominate a PA to be open.

An Aspiration

"Democracy Thrives In Sunlight"

The Wonk Times

Washington DC

VOL.III. . . No.14 | *MARCH 29, 2020* | THREE DOLLARS

CBO Publishes Open Report on Minimum Wage

Results Will Be Published Separately in Two Weeks

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Bipartisan Support for CBO Methodology

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum.

Thank you.

Pre-prints:
Why OPA
OPA Case Study

Slides at
github.com/fhoces/CBO2018

fhoces@berkeley.edu

Back-up slides

Easier Methodological Appraisal. Example: dis-employment effects **Before**

Steps taken to verify the analysis & employment variation ($\widehat{\Delta E} \times 1000$) at each line[1]

1. Find an elasticity: -0.1 (page 25): $\widehat{\Delta E} \approx 300$
2. What about adults? $\eta^{adults} = \frac{1}{3}\eta^{teens}$ (page 28): $\widehat{\Delta E} \approx 100$
3. What about the adjustment? $\widehat{\eta^{g}_{w \leq MW}} = \frac{\eta^{g}_{lit}}{p^{g}_{w \leq MW}} \times \frac{\%\Delta MW}{\%\Delta w^{g}}$ (page 26-28 + 2 papers): $\widehat{\Delta E} \approx 1,100$
4. The adjustment factors $\frac{1}{p^{g}_{w<MW}} \times \frac{\%\Delta MW}{\%\Delta w^{g}} = F^{g}_{adj}$ are not computed from the data (3.2 teens, 19.5 adults). Instead: $F^{teen}_{adj} = F^{adult}_{adj} = 4.5$ (page 28) $\widehat{\Delta E} \approx 500$

Steps 2-4 took several days of work!

[1] Assuming target population $\approx$ 22 million, $\overline{\Delta w_{w \leq MW'}} \approx 14\%$, and non-compliance $\approx 15\%$

Easier Methodological Appraisal. Example: dis-employment effects **Before**

Steps taken to verify the analysis & employment variation ($\widehat{\Delta E} \times 1000$) at each line[1]

1. Find an elasticity: -0.1 (page 25): $\widehat{\Delta E} \approx 300$
2. What about adults? $\eta^{adults} = \frac{1}{3}\eta^{teens}$ (page 28): $\widehat{\Delta E} \approx 100$
3. What about the adjustment? $\widehat{\eta^{g}_{w \leq MW}} = \frac{\eta^{g}_{llt}}{p^{g}_{w \leq MW}} \times \frac{\%\Delta MW}{\%\Delta w^{g}}$ (page 26-28 + 2 papers): $\widehat{\Delta E} \approx 1,100$
4. The adjustment factors $\frac{1}{p^{g}_{w<MW}} \times \frac{\%\Delta MW}{\%\Delta w^{g}} = F^{g}_{adj}$ are not computed from the data (3.2 teens, 19.5 adults). Instead: $F^{teen}_{adj} = F^{adult}_{adj} = 4.5$ (page 28) $\widehat{\Delta E} \approx 500$

Steps 2-4 took several days of work!

[1] Assuming target population $\approx$ 22 million, $\overline{\Delta w_{w \leq MW'}} \approx 14\%$, and non-compliance $\approx 15\%$

Easier Methodological Appraisal. Example: dis-employment effects **Before**

Steps taken to verify the analysis & employment variation ($\widehat{\Delta E} \times 1000$) at each line[1]

1. Find an elasticity: -0.1 (page 25): $\widehat{\Delta E} \approx 300$
2. What about adults? $\eta^{adults} = \frac{1}{3}\eta^{teens}$ (page 28): $\widehat{\Delta E} \approx 100$
3. What about the adjustment? $\widehat{\eta^{g}_{w \leq MW}} = \frac{\eta^{g}_{lit}}{p^{g}_{w \leq MW}} \times \frac{\%\Delta MW}{\%\Delta w^{g}}$ (page 26-28 + 2 papers): $\widehat{\Delta E} \approx 1,100$
4. The adjustment factors $\frac{1}{p^{g}_{w \leq MW}} \times \frac{\%\Delta MW}{\%\Delta w^{g}} = F^{g}_{adj}$ are not computed from the data (3.2 teens, 19.5 adults). Instead: $F^{teen}_{adj} = F^{adult}_{adj} = 4.5$ (page 28) $\widehat{\Delta E} \approx 500$

Steps 2-4 took several days of work!

[1] Assuming target population $\approx$ 22 million, $\overline{\Delta w_{w \leq MW'}} \approx 14\%$, and non-compliance $\approx 15\%$

Easier Methodological Appraisal. Example: dis-employment effects **Before**

Steps taken to verify the analysis & employment variation ($\widehat{\Delta E} \times 1000$) at each line[1]

1. Find an elasticity: -0.1 (page 25): $\widehat{\Delta E} \approx 300$
2. What about adults? $\eta^{adults} = \frac{1}{3}\eta^{teens}$ (page 28): $\widehat{\Delta E} \approx 100$
3. What about the adjustment? $\widetilde{\eta^{g}_{w \leq MW}} = \frac{\eta^{g}_{lit}}{p^{g}_{w \leq MW}} \times \frac{\%\Delta MW}{\%\Delta w^{g}}$ (page 26-28 + 2 papers): $\widehat{\Delta E} \approx 1,100$
4. The adjustment factors $\frac{1}{p^{g}_{w \leq MW}} \times \frac{\%\Delta MW}{\%\Delta w^{g}} = F^{g}_{adj}$ are not computed from the data (3.2 teens, 19.5 adults). Instead: $F^{teen}_{adj} = F^{adult}_{adj} = 4.5$ (page 28) $\widehat{\Delta E} \approx 500$

Steps 2-4 took several days of work!

[1] Assuming target population $\approx$ 22 million, $\overline{\Delta w_{w \leq MW'}} \approx 14\%$, and non-compliance $\approx 15\%$

Easier Methodological Appraisal. Example: dis-employment effects **Before**

Steps taken to verify the analysis & employment variation ($\widehat{\Delta E} \times 1000$) at each line[1]

1. Find an elasticity: -0.1 (page 25): $\widehat{\Delta E} \approx 300$
2. What about adults? $\eta^{adults} = \frac{1}{3}\eta^{teens}$ (page 28): $\widehat{\Delta E} \approx 100$
3. What about the adjustment? $\widetilde{\eta^{g}_{w \leq MW}} = \frac{\eta^{g}_{lit}}{p^{g}_{w \leq MW}} \times \frac{\%\Delta MW}{\%\Delta w^{g}}$ (page 26-28 + 2 papers): $\widehat{\Delta E} \approx 1,100$
4. The adjustment factors $\frac{1}{p^{g}_{w \leq MW}} \times \frac{\%\Delta MW}{\%\Delta w^{g}} = F^{g}_{adj}$ are not computed from the data (3.2 teens, 19.5 adults). Instead: $F^{teen}_{adj} = F^{adult}_{adj} = 4.5$ (page 28) $\widehat{\Delta E} \approx 500$

Steps 2-4 took several days of work!

[1] Assuming target population $\approx$ 22 million, $\overline{\Delta w_{w \leq MW'}} \approx 14\%$, and non-compliance $\approx 15\%$

Equations from Model in DD

$$\widehat{\Delta E} = N \times \eta \times \%\Delta w + \text{Other factors} \tag{1}$$

$$\widehat{N^s_{final}} = g_N(\hat{t'}|t) \times \hat{N}^s_t \times P(\hat{w'} \leq MW^{new}|s) \times (1 - \hat{\alpha}^s_1 - \hat{\alpha}^s_2) \qquad s = \{teens \tag{2}$$

The elasticity for adults from the literature is define as the one for teenagers with an extrapolation factor.

$$\eta^{adults}_{lit} = \eta^{teens}_{lit} \times F_{extrapolation} \tag{3}$$

Adjustments to the elasticity of labor demand

Following Neumark and Wascher (2008), Brown (1999). First:

$$\eta^s_{lit} = p^s_{w \leq MW} \eta^s_{w \leq MW} + (1 - p^s_{w \leq MW}) \eta^s_{w > MW} \qquad s = \{teens, adults\}$$

Second, assume $\eta^s_{w \leq MW} = 0$:

$$\eta^s_{w \leq MW} = \frac{\eta^s_{lit}}{p^s_{w \leq MW}} \qquad s = \{teens, adults\}$$

And third, adjust for the effective average wage variation for each group ($\overline{\%\Delta w^s}$):

$$\widetilde{\eta^s_{w \leq MW}} = \frac{\eta^s_{lit}}{p^s_{w \leq MW}} \times \frac{\%\Delta MW}{\overline{\%\Delta w^s}} = \eta^s_{lit} \times F^s_{adjs} \qquad s = \{teens, adults\} \tag{4}$$

Final Effect on Employment

$$\widehat{\Delta E} = \sum_{g \in \{A,T\}} \left(\widehat{N_g^{final}} \times \widetilde{\eta_{w \leq MW}^{g}} \times \overline{\%\Delta w^g} \right) - \widehat{OF} \tag{5}$$

Effect on Wages

$$w'' = \begin{cases} w' & if \quad w \in U[0,1] < \alpha_1 \\ w^{new} & o/w \end{cases} \tag{6}$$

$$w^{new} = \begin{cases} w'/2 & if \quad w \in U[0,1] < \alpha_{aux} \\ \widetilde{w^{new}} & o/w \end{cases} \tag{7}$$

Ripple Effects

$$\widetilde{w^{new}} = \begin{cases} MW' & if \quad w' < R_{lb} \\ MW' + R^l(w' - R^s_{lb}) & if \quad w' \in [R_{lb}, MW') \\ w' + R^l(R^s_{ub} - w') & if \quad w' \in [MW', R_{ub}) \\ w' & o/w \end{cases} \tag{8}$$

Computing Income

$$y'_{i,h} = \sum_{i \in N_h} \left(g_{nw}(t'|t) nw_i + w'_i \right) / N_h$$

$$y''_{i,h} = \sum_{i \in N_h} \left(g_{nw}(t'|t) nw_i + w''_i \right) / N_h \tag{9}$$

Final Policy Estimates

$$WG_i = \left(y''_i - y'_i \right) \mathbf{I} \left(y''_i > y'_i \right) \tag{10}$$

$$WL_i = \left(y'_i - y''_i \right) \mathbf{I} \left(y''_i < y'_i \right) \tag{11}$$

$$BL = \sum_i WG_i - F_{sub} \sum_i WL_i; \quad BL_i = BL \times dBL \tag{12}$$

$$\overline{WG_Q} = \frac{\sum_{i \in Q} WG_i}{N_{pop}/5} \quad \overline{WL_Q} = \frac{\sum_{i \in Q} WL_i}{N_{pop}/5}$$

$$\overline{BL_Q} = \frac{\sum_{i \in Q} BL_i}{N_{pop}/5} \tag{13}$$

Snapshots of DD

https://rpubs.com/fhoces/dd_cbo_test1

Reader Companion for CBO report on Min Wage (Preliminary Version. Do Not Circulate)

Fernando Hoces de la Guardia + (hopefully) a lot more people

Last edit: 2016-10-16

1 Introduction

The role of policy analysis is to connect research with policy. Because of heavy time constrains, policy analyses are typically ambiguous regarding the details of how the analysis was carried out. This creates three problems: (i) its hard to understand the connection between research and policy, (ii) allows policy makers to cherry pick policy reports, and (iii) hinders systematic improvement and/or automation of parts of the analysis. In this document we demonstrate the use of a reproducible workflow to reduce the ambiguity in policy analysis.

Here we attempt to contribute to the policy discussion of the minimum wage. The minimum wage is a contentious policy issue in the US. Increasing it has positive and negative effects that different policymakers value differently. We aim to add clarity on what those effects are, how much do we know about them, and how those effects vary when elements of the analysis change. We select the most up-to-date, non-partisan, policy analysis of the effects of raising the minimum wage, and build an open-source reproducible analysis on top of it.

In 2014 the Congressional Budget Office published the report titled "The Effects of a Minimum-Wage Increase on Employment and Family Income". The report receive wide attention from key stakeholders and has been used extensible as an input in the debate around the minimum wage[1]. To this date we consider the CBO report to be the best non-partisan estimation of the effects of raising the minimum wage at the federal level. Although there was disagreement among experts around some technical issues, this disagreement has been mainly circumscribed around one of the many inputs used in the analysis, and we can fit the opposing positions in to our framework.

Our purposes are twofold: First, promote the technical discussion around a recurrent policy issue (minimum wage) by making explicit and visible all the components and key assumptions of its most up-to-date official policy analysis. Second, demonstrate how new scientific practices of transparency and reproducibility (T & R) can be applied to policy analysis. We encourage the reader to collaborate in this document and help develop an ever-improving version of the important policy estimates[2] (re)produced here.

Snapshots of DD.

2 Employment effects

At a general level the effects on employment ($\widehat{\Delta E}$) will be calculated using a more detailed version of the following equation:

$$\widehat{\Delta E} = N \times \eta \times \%\Delta w + \text{Other factors}$$

Where N represents the relevant population, η the elasticity of labor demand, Δw the relevant percentual variation in wages, and the *Other factors* will encapsulate effects on employment through an increase in the aggregate demand.

To describe the methodology behind each of those four components we first describe the data used, the wage variable choose, and the procedure used to forecast the wage and population distribution of 2016 using data from 2013.

2.1 Data, wages, and forecast

To simulate the policy effects we need the distribution of wages and employment under the status quo. From the perspective of 2013, this implies forecasting to 2016 data on employment and wages.

2.1.1 Data

The Current Population Survey (CPS) was used to compute the effects on employment. From the analysis in the section on distributional effects we can deduce that the data corresponds to the Outgoing Rotation Group (ORG). CPS is a monthly cross sectional survey. The same individual is interviewed eight times over a period of 12 months. The interviews take place in the first and last 4 months of that period. By the 4th and 12th interview, individuals are asked detailed information on earnings. The CPS ORG file contains the information on this interviews for a given year. We analyze the data for 2013.

Currently three versions of these data sets can be found online: CPS raw files, ORG NBER and ORG CEPR. The analysis will be performed using the CPER ORG data base.

The weights used in our analysis will be `orgwgt/12`

2.1.1.1 Code to load the data

`R`
`Stata`

Snapshots of DD.

performed using the CPER ORG data base.

The weights used in our analysis will be `orgwgt/12`

2.1.1.1 Code to load the data

R

```
call.cps.org.data <- function(){
  data_use <- "CPER_ORG"

  # Using CEPR ORG data
  if (data_use == "CPER_ORG") {
  # Checking if working directory contains data, download if not.
    if ( !("cepr_org_2013.dta" %in% dir()) ) {
        # create name of file to store data
        tf <- "cepr_org_2013.zip"

        # download the CPS repwgts zipped file to the local computer
        download.file(url =  "http://ceprdata.org/wp-content/cps/data/cepr_org_2013.zip", tf , mode
= "wb" )

        # unzip the file's contents and store the file name within the temporary directory
        fn <- unzip( zipfile = tf , overwrite = T )
    }
    df <- read.dta("cepr_org_2013.dta")
  }

  # Using NBER ORG data
  if (data_use == "NBER_ORG") {
    # Checking if working directory contains data, download if not.
    if ( !("morg13.dta" %in% dir()) ) {
      # Downloading data 53mb
      df <- read.dta("http://www.nber.org/morg/annual/morg13.dta")
    }
    df <- read.dta("morg13.dta")
```

Snapshots of DD.

2.5 Computing effects on employment

Putting all elements together we get:

$$\widehat{\Delta E} = \sum_{g \in \{A,T\}} \left(\widehat{N_g^{final}} \times \widetilde{\eta^g_{w \leq MW}} \times \overline{\%\Delta w^g} \right) - \widehat{OF}$$

2.5.1 Code to compute each component

`R`
`Stata`

Components of Elasticities

	Adult	Teen
η_{lit}	-0.03	-0.10
$\eta_{w \leq MW'}$	-0.23	-0.13
F_{adj}	4.50	4.50
$\overline{\%\Delta w}$	13.81	16.65
$\widetilde{\eta_{w \leq MW}}$	-0.15	-0.45

Using all the components described above we get $\widehat{\Delta^- E}$ = -478 thousand jobs. The report however computes F^g_{adjs} in a different fashion and gets a value of 4.5 (when computing the values of F^g_{adjs} from the table below - as oppose to using historical values - we get $\widehat{\Delta^- E}$ = -321 thousand jobs).

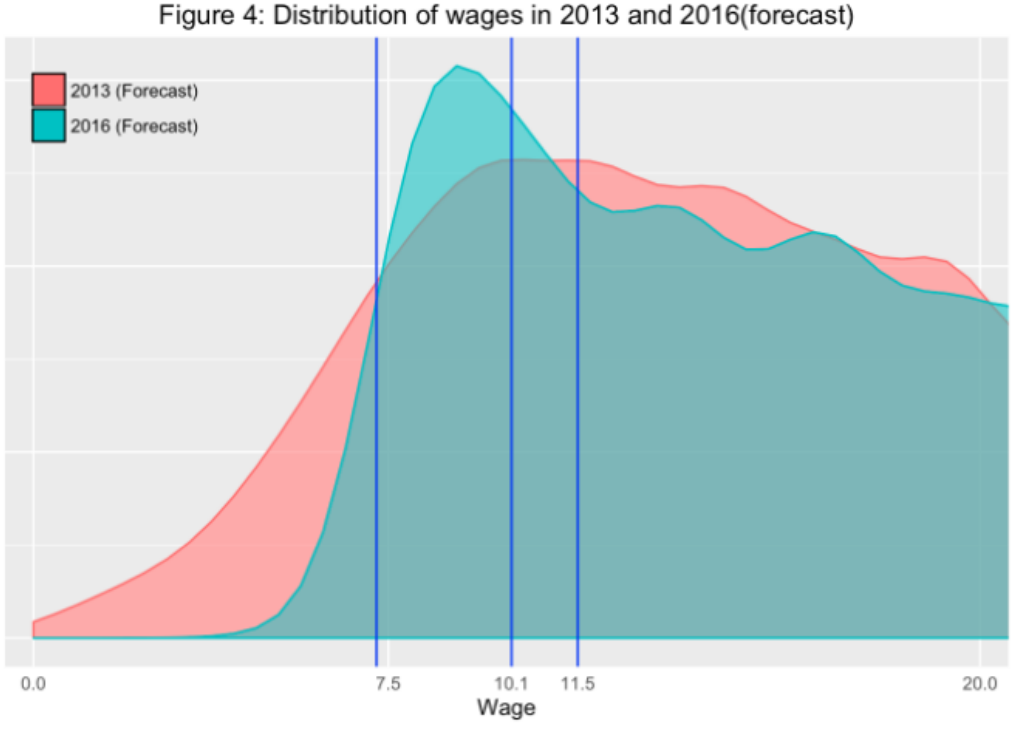
3 Distributional effects

In the first step towards obtaining the policy estimates presented in the introduction we concluded with

Snapshots of DD.

R

Stata



Comparison of 2013 and 2016 under the status quo

	2013	2016: status quo
Salary workers	122,593,557	129,545,571
Median wage	17.79	20.56

Snapshots of DD

Final replication output (nothing in the "learn more" button yet)

Policy estimates in CBO report and Replication Results

	Effects/Policy Estimates	Replication
wage gains (billions of $)	31	53.4
wage losses (bns of $)	~5	7.4
Balance losses (bns of $)	~24	43.9
Net effect (bns of $)	2	2
# of Wage gainers (millions)	16.5	23.1/16.9
#of Wage losers (millions)	0.5	0.5

	<1PL	[1PL, 3PL)	[3PL, 6PL)	>6PL
Balance losses (bns of $)	~0.3	~3.4	~3.4	~17
Net effect (bns of $)	5	12	2	-17
Replication loses	-0.4	-6.4	-6.4	-30.8
Replication NE	17.6	14.6	-0.1	-30.1

Learn more

Clear connection between sources and inputs

Source	Input
Data	
CPS ORG 2013 (CEPR version)	Number of salary workers in 2013 ($\widehat{N^g_{final}}$ $g \in \{teen, adult\}$); Fraction of workers below the new minimum wage ($P_{\hat{w} \leq MW^1 \mid g}$); Average wage variation for those below the new min wage ($\overline{\%\Delta w^g}$); Non-compliance rate (α_1^g)
CPS ASEC 2012 (CEPR version)	Wages and Non-Wage Income distribution (dF_w, dF_{nw}); Household size (N_h); Hours/weeks worked ($\hat{w}, \hat{h}$)
State level Min. Wage (DOL)	Trends in state min. wage (MW_t^s)
10-year economic forecast (CBO)	Predicted worker growth by 2016 (in 2013) ($\hat{g_N}$); Wage growth in by 2016 ($\hat{g_w}$); Non-wage growth by 2016 ($\hat{g_{nw}}$)
Research	
Elasticity of labor demand for teenagers	$\eta^{lit}_{teen} = -0.1$
Ripple effects	From $R_{lb} = \$8.7$ to $R_{ub} = \$11.5$ with a "ripple" intensity of $R_I = 50\%$
Guess Work	
Extrapolation factor from teenagers to adults	$F_{ex} = 1/3$
Net benefits	$\hat{NB} = \$2 billion$
Adjustment to account for effective wage variation and affected population	$F_{adj} = 4.5$
Aggregate consumption effects on employment	$\hat{OF} = 40,000$ *new jobs*
Distribution of balance loses	$dBL = (1\%, 29\%, 70\%)$ if income $\in [0, 1PL, 6PL, +)$
Fract. of wage loses used to pay wage gains	$F_{subs} = 1$
Job killing process: fraction of jobs	Cut wages in half for twice the number of jobs destroyed

Fully specified model

Model	Policy estimate (per quintile)
Predicted household income with and without min wage increase. **Depends on**: $\widehat{N^g_{final}}$, $P_{\hat{w} \leq MW^1 \mid g}$, $\overline{\%\Delta w^g}$, α^g_1, dF_w, dF_{nw}, N_h, $\hat{w}$, $\hat{h}$, MW^s_t, $\hat{g_N}$, $\hat{g_w}$, $\hat{g_{nw}}$, η^{lit}_{teen}, R_{lb}, R_{ub}, R_l, F_{ex}, F_{adj}, $\hat{OF}$	Average gain in per capita income due to net wage increase. ($\overline{WG_q}$)
Predicted household income with and without min wage increase. **Depends on**: $\widehat{N^g_{final}}$, $P_{\hat{w} \leq MW^1 \mid g}$, $\overline{\%\Delta w^g}$, α^g_1, dF_w, dF_{nw}, N_h, $\hat{w}$, $\hat{h}$, MW^s_t, $\hat{g_N}$, $\hat{g_w}$, $\hat{g_{nw}}$, η^{lit}_{teen}, F_{ex}, F_{adj}, $\hat{OF}$	Average loss in per capita income due to net wage decrease. ($\overline{WL_q}$)
Distribution of balance loses **Depends on**: $\overline{WG_q}(\cdot)$, $\overline{WL_q}(\cdot)$, $\hat{NB}$, F_{subs}, dBL	Average loss in per capita income to balance wage gains. ($\overline{BL_q}$)
Equations; Back	

Comparing the Trade-offs: A Toy Example

Model for the normative comparison made by a policy maker (welfare function):

$$W(\rho) = \sum_{i \in N} \left(\omega_{wg} wg_i + \omega_{wl} wl_i + \omega_{bl} bl_i\right) \omega_i^d(Q_i, \rho)$$

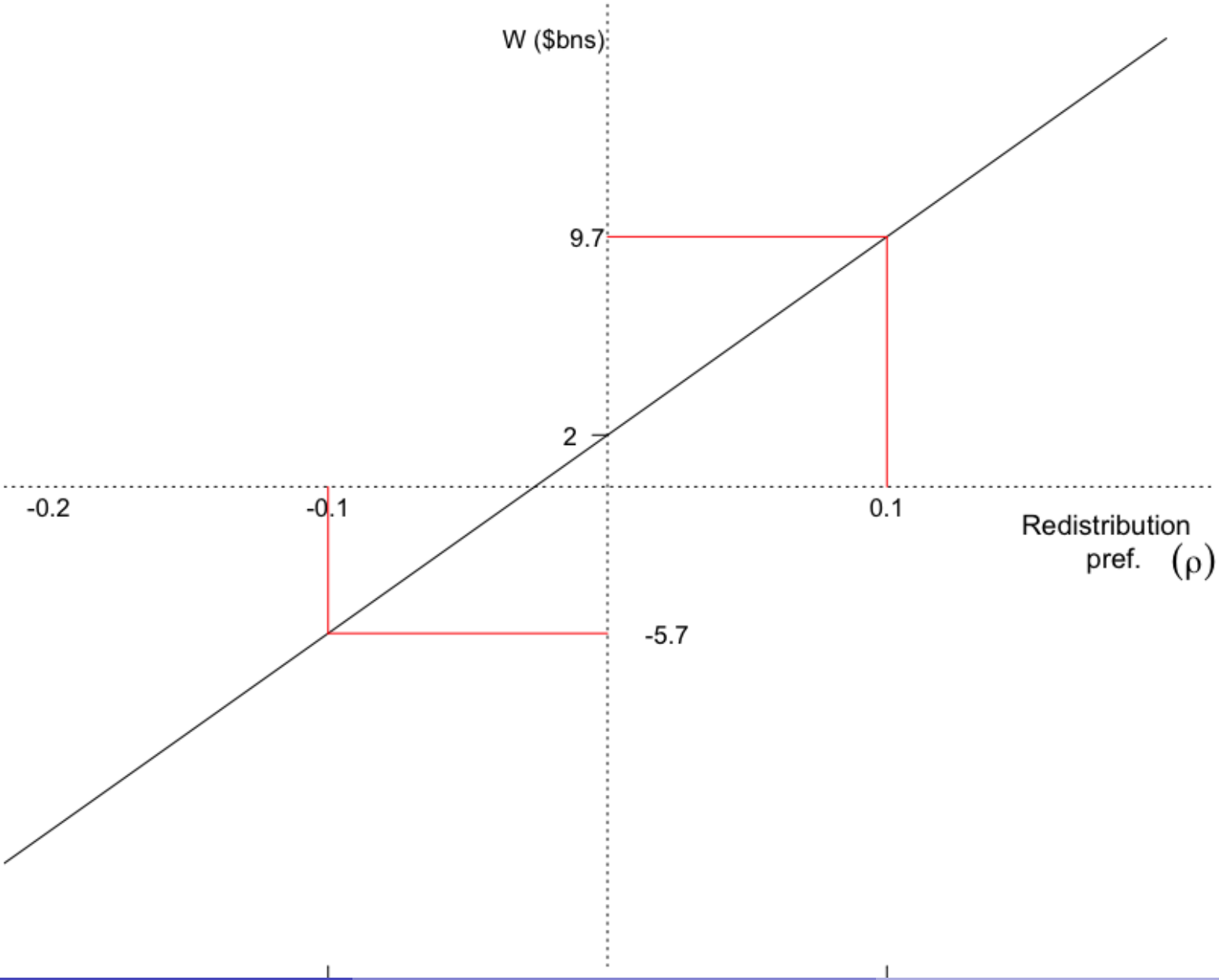
with:

$$\omega_i^d(Q_i, \rho) = \frac{(1 - \rho(Q_i - Q_{median}))}{\sum_i \omega_i^d(Q_i)} Q_{max} \quad \text{for } \rho \in \left(-\frac{1}{2}, \frac{1}{2}\right)$$

$\rho > 0$ represents positive valuation of progressive redistribution. $\rho < 0$ represents positive valuation of regressive redistribution.

Redistribiutional Preferences

Toy

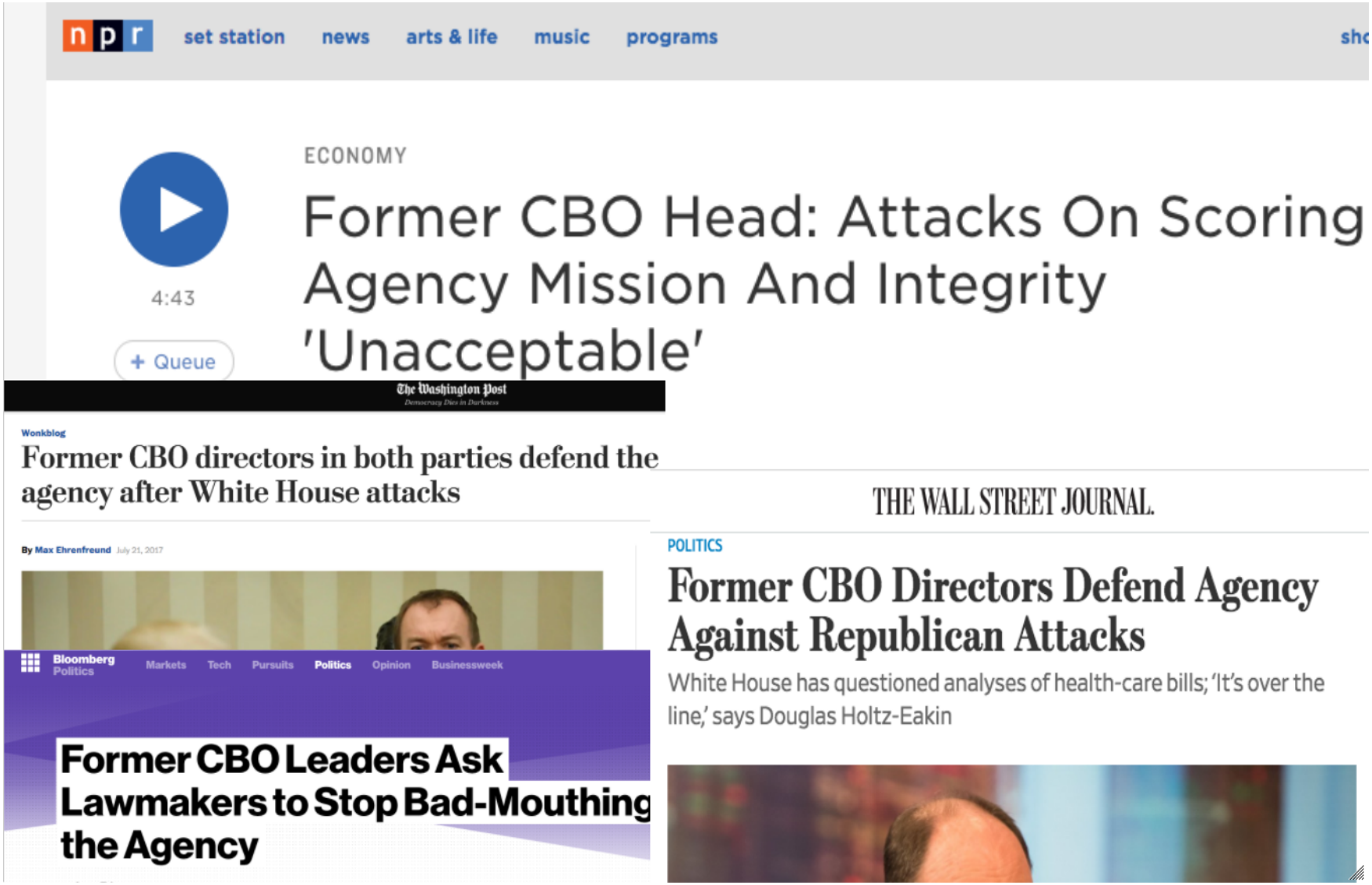


Motivation 2: An Academic Concern in 2013

"I worry that someday sooner or later the existing social contract to take CBO scores at face value will break down. Conventional Certitudes that lack foundation cannot last indefinitely."

— Charles Manski
Public Policy in an Uncertain World, 2013

Motivation 2: A Reality In 2017



Challenges And Suggestions

Challenges:

- Policymakers may not want analyses to be open.
- Analysts may wish to keep policy analyses “closed”.
- For policy analysis contracted out to third parties: Opening methods will prevent them form reselling extensions.
- Initially reproducibility represents an additional layer of work.
- Limits to sharing sensitivity of information, requires resources for adequate de-identification if open data is expected