sociological science

Supplement to:

Enns, K. Peter. 2026. "How a Seemingly Innocuous and Intuitive Methodological Choice Confused a Generation of Research on Policy Responsiveness" Sociological Science 13: 528-564.

Online Supplement to:

How a Seemingly Innocuous and Intuitive Methodological Choice Confused a Generation of Research on Policy Responsiveness

Peter K. Enns
peterenns@cornell.edu

Contents

Online Supplement 1 Additional Evidence that the Preferences of the Affluent (90th Income Percentile) *Cannot* Proxy for the Superrich

Page and Gilens (2017, 92) explain, “we suspect that much of the influence we have detected is being wielded by a tiny group within the affluent: the ‘truly wealthy’– that is, multimillionaires and billionaires.” For this to be the case, the preferences of the affluent (the 90th income percentile in Gilens’ data) must represent or reflect the preferences of the superrich. It *cannot* be the case, however, that the preferences of the affluent in Gilens’ data serve as a proxy for the preferences of the superrich. This distinction matters because Gilens and Page’s (2014, 573) conclusion that the 90th income percentile dominates politically is even more surprising when we recognize their preferences cannot serve as a proxy for the preferences of the superrich.

In 2022, the household income for 90th percentile was $155,400 (Piketty and Saez 2003, and Table A4: `https://eml.berkeley.edu/~saez/TabFig2022.xlsx`).[A-1] While households earning this amount were certainly well-off financially, $155,400 is much closer to the median income than to multimillionaires and billionaires. Professionally, these households are also likely to be much closer to the median. As the text noted, a family with two public school teachers in Bakersfield, California can reach the 90th income percentile in earnings. Figure A-1 shows the 2023-24 basic salary (182 day) in the Bakersfield City School District was $78,859 for teachers with a Bachelor’s Degree plus 36 additional semester units or a Master’s Degree (Step 8).[A-2] A family with two teachers at this level would have earned $157,718. Total household income is the appropriate metric, because the Piketty and Saez (2003) data represent gross income, excluding capital gains, for tax units, “defined as a married couple living together (with dependents) or a single adult (with dependents)” (4-5). With a high salary of $118,012 in 2023-24, a single teacher in the Bakersfield City School District could not reach the 90th income percentile, but if a teacher in the highest pay category (Step 20) had a partner earning an additional income of $40,000, that would have been sufficient to be “affluent” in Gilens’ analysis. Households earning around $155,000 are certainly well-off, but financially and professionally they have much more in common with the median than with multimillionaires and billionaires.

Indeed, the policy preferences of the affluent and the median income group in Gilens’ data correlate at a near perfect r=0.94. Theoretically and empirically, it seems inconceivable that the preferences of the affluent could serve as an indicator for the superrich. However, to empirically assess the potential feasibility for the affluent to serve as an indicator of the truly wealthy, I utilize research by Page, Bartels and Seawright (2013) and their collaborators who conducted the one-of-a-kind Survey of Economically Successful Americans (SESA), which randomly sampled and surveyed the superrich (mean wealth=$14,006,338, median=$7,500,000) in the Chicago-area. Although not a national survey, the authors note, “The Chicago-area wealthy, if not perfectly typical [of the wealthy nationally], may at least tend to occupy a middle ground—not only geographically but in

[A-1] This income amount for the 90th percentile is slightly higher than the household income threshold Gilens (2012, 301, fn 2) reports, $138,000 in 2008. The 90th percentile in 2008 in the Piketty and Saez data is $146,244. The original source of Gilens’ income estimate can be accessed here `https://web.archive.org/web/20150330155448/https://www.census.gov/newsroom/releases/archives/news_conferences/2009-09-10_remarks_johnson.html`.

[A-2] This salary data was originally accessed from `https://www.bcsd.com/Page/257`, but that link is longer active. An archived version of that webpage can be viewed via the Internet Archive Wayback Machine (`https://web.archive.org/web/20240910005812/https://www.bcsd.com/Page/257`).

BAKERSFIELD CITY SCHOOL DISTRICT
Human Resources

TEACHERS ANNUAL BASIC SALARY SCHEDULE - 182 DAY
2023-24
4.39% INCREASE
Effective July 1, 2023

BASED ON 7 HRS. 15 MIN. PER DAY INCLUDING LUNCH PERIOD

STEP	1 BACHELOR'S DEGREE CALIFORNIA CREDENTIAL	DAILY RATE	2 BACHELOR'S DEGREE PLUS 12 SEMESTER UNITS OR 18 QUARTER UNITS	DAILY RATE	3 BACHELOR'S DEGREE PLUS 24 SEMESTER UNITS OR 36 QUARTER UNITS	DAILY RATE	4 BACHELOR'S DEGREE PLUS 36 SEMESTER UNITS OR 54 QUARTER UNITS OR MASTER'S DEGREE	DAILY RATE	5 BACHELOR'S DEGREE PLUS 48 SEMESTER UNITS OR 72 QUARTER UNITS OR MASTER'S DEGREE + 12 SEM/18 QTR UNITS	DAILY RATE	6 BACHELOR'S DEGREE PLUS 60 SEMESTER UNITS OR 90 QUARTER UNITS OR MASTER'S DEGREE + 24 SEM/36 QTR UNITS	DAILY RATE
1	59,232	325.45	61,116	335.80	63,089	346.64	65,057	357.46	67,026	368.27	69,002	379.13
2	61,116	335.80	63,089	346.64	65,057	357.46	67,026	368.27	69,002	379.13	70,978	389.99
3	63,089	346.64	65,057	357.46	67,026	368.27	69,002	379.13	70,978	389.99	72,944	400.79
4	65,057	357.46	67,026	368.27	69,002	379.13	70,978	389.99	72,944	400.79	74,917	411.63
5	67,026	368.27	69,002	379.13	70,978	389.99	72,944	400.79	74,917	411.63	76,888	422.46
6	68,672	377.32	70,978	389.99	72,944	400.79	74,917	411.63	76,888	422.46	78,859	433.29
7	70,978	389.99	72,944	400.79	74,917	411.63	76,888	422.46	78,859	433.29	80,944	444.75
8	72,944	400.79	74,917	411.63	76,888	422.46	78,859	433.29	80,830	444.12	83,737	460.09
9	74,917	411.63	76,888	422.46	78,859	433.29	80,830	444.12	83,078	456.47	86,508	475.32
10			78,859	433.29	80,830	444.12	82,804	454.97	85,462	469.57	89,295	490.63
11					82,804	454.97	84,776	465.80	87,650	481.59	92,078	505.92
12							86,745	476.62	89,889	493.90	94,856	521.19
13									91,888	504.88	97,654	536.56
14									93,885	515.85	100,410	551.70
15											104,105	572.01
17											107,561	590.99
18											111,839	614.50
19											115,133	632.60
20											118,012	648.42

BONUS - ADD TO BASIC
SALARY SCHEDULE

MASTER'S	2,000
DOCTORATE	2,000
BILINGUAL	2,000
SPEC. ED.	2,000
ACADEMY PROGRAM SHIFT DIFFERENTIAL:	3,600

Mar-23

Figure A-1: Bakersfield, CA 2023-24 Teacher Salary Schedule

terms of political culture" (Page, Bartels and Seawright 2013, 53) and Page and Gilens (2017, 114) conclude, "SESA allows us to compare multimillionaires' political opinions and actions to those of ordinary Americans."

I calculate the correlation of the policy preferences of the superrich and the general public across the 66 policy-related survey questions Page, Bartels and Seawright (2013) report.[A-3] While the general public encompasses more than just the median, the correlation between the preferences of the median and general public in Gilens' data is r=0.99, indicating that the general public offer an excellent proxy for the preferences of the median. The correlation between the preferences of the superrich and general public based on the SESA data is r=0.67. The positive association between the preferences of the general public and the superrich is notable, but real differences in preferences exist (also see Page and Gilens (2017, 114-118) and Broockman and Malhotra (2020)). The results Page, Bartels and Seawright (2013) report do not allow a direct assessment of the correlation between the preferences of the 90th income percentile and the superrich, but this correlation would need to be greater than the correlation between the preferences of the affluent and the median (r=0.94) in order for the affluent's preferences to be considered a proxy for the superrich. If the strength of the relationship between the affluent's preferences and the superrich was equal to or less than the strength of the relationship between the affluent and the median, we would necessarily conclude that the affluent's preferences represented the median as much or more than they represented the superrich.

To evaluate the potential relationship between the preferences of the affluent and superrich, I attempted to simulate data which would represent the preferences of the median, the 90th income percentile, and the superrich. The mean and standard deviation of the median and 90th income percentile preferences in the simulated data match the observed mean and standard deviation in Gilens' data and the mean and standard deviation of the simulated superrich preferences match the

[A-3]The data come from Tables 4, 5, 6, 7, 8, 9, and 10 of Page, Bartels and Seawright (2013).

SESA data. Then, I specified correlations of r=0.94 and 0.67 to reflect the observed relationships in the data between preferences of the median and 90th income percentile and the median and superrich, respectively. I set the correlation between the preferences of the 90th income percentile and the superrich to equal 0.95, just above the correlation between the preferences of the affluent and the median. This is the lowest value that would allow the affluent to be a better proxy of the superrich than of the median.[A-4]

It turns out to be *impossible* to simulate data with these parameters. The correlation matrix is not positive definite. This exercise shows that it is not mathematically possible for the preferences of the affluent to match the superrich more closely than they match the median given the other observed patterns in the data. This result aligns with Page, Bartels and Seawright (2013, 52) who write, "For systematic evidence on the policy preferences of really wealthy Americans—such as the top 1 percent or the top one-tenth of 1 percent of wealth-holders—it is necessary to design special surveys that explicitly target those groups." We should not treat the preferences of the affluent in Gilens' data as a proxy the preferences of the truly wealthy.

Online Supplement 2 Confirming Simulations Mimic Observed Data

The simulated policy preferences for income groups used to estimate the values in Table 1 have the same mean, minimum and maximum values, covariance, and sample size as Gilens' data. The relationships between these variables and the simulated policy change variable also follow the observed relationships in Gilens' data. As a result, the simulated data mimic Gilens' data in every way. Thus, it is thus not surprising that the predicted values based on the simulated data in Figures A-2 and A-3, below, mirror the actual predicted values from Gilens' data in Figure 2, but these figures offer further confirmation that the simulations accurately reproduce the characteristics of the data.

[A-4]To evaluate this scenario even more precisely, I re-ran the simulation setting the correlation between the affluent and superrich just 0.001 greater than the correlation between the affluent and the median, with the exact same result. That combination of associations is not possible.

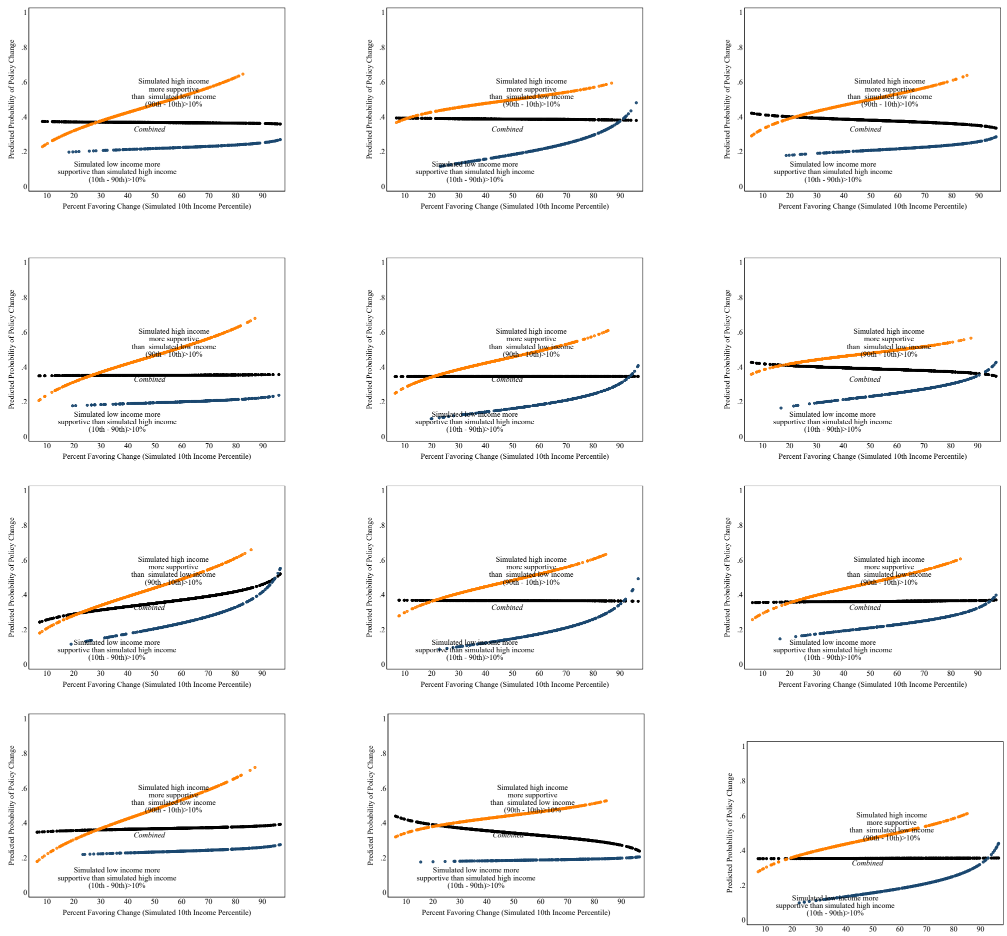


Figure A-2: The final 12 simulations (out of 1,000) showing the estimated relationship between the simulated preferences of the **10th income percentile** and the probability of policy change. As expected, the simulated data follow the patterns of actual data in Figure 2.

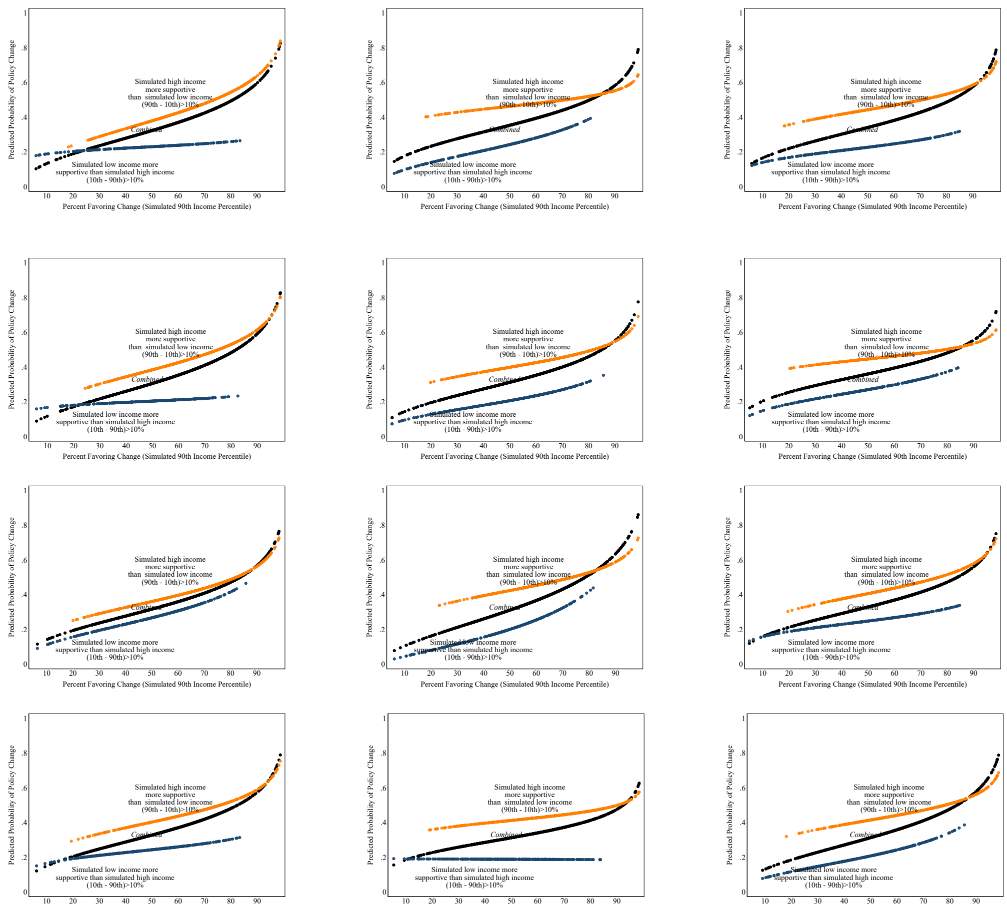


Figure A-3: The final 12 simulations (out of 1,000) showing the estimated relationship between the simulated preferences of the **90th income percentile** and the probability of policy change. As expected, the simulated data follow the patterns of actual data in Figure 2.

Online Supplement 3 Numerical Results for CEM Analysis (Figure 3) and Corresponding Results for the 50th and 90th Income Percentiles

Figure 3 in the text showed that when Coarsened Exact Matching (CEM) is used to ensure identically overlapping distributions, patterns of responsiveness to all income groups look similar. With CEM, if only a subset of observations need to be dropped to obtain balance, the algorithm randomly selects which of these observations to drop. To ensure that the specific randomly dropped observations do not account for the patterns in Figure 3, I generated 1,000 CEM-balanced data sets and estimated the relationship between policy preferences of the 10th and 90th income percentiles and the probability of policy change for all 1,000 iterations. Table A-1 reports an exact replication of Gilens' (2012) results (Table 3.2 "Greater than 10 points") and the median estimate and associated 95% uncertainty estimates (based on the 25^{th} smallest and 975^{st} largest estimates) from the 1,000 regressions.

In the first row (10th income percentile), in contrast to Gilens' estimate of near zero responsiveness (0.02) to the lowest income group, after using CEM to ensure fully balanced distributions, a positive and significant association emerges (0.44). The magnitude of this responsiveness is similar to what Gilens estimates for the 90th income percentile (0.46). As the text notes, the different estimate of responsiveness does *not* result because of a change in correlation between income groups. The correlation between the 10th and 90th income percentiles when a preference gap exists (r=0.60) is nearly identical after CEM (median correlation =0.59). After CEM, the estimate of policy responsiveness to the 90th income percentile (0.31) is slightly lower than Gilens' estimate, but the confidence interval overlaps Gilens' estimate.

The bottom half of Table A-1 reports results for the policies which the 50th and 90th income percentiles differ in support by at least 10 percentage points. The analysis reproduces Gilens' near-zero (-0.01) estimated relationship between the preferences of the median income and the probability of policy change. When CEM is used to ensure balance, we observe a positive and statistically significant relationship (0.66). The CEM analyses demonstrate that the non-overlapping portions of the data (which result because limiting the analysis to policies where a preference gap exists introduces left and right censoring and exacerbates the extent to which the affluent are more supportive of proposed policy changes than the lower income groups) account for the lack of responsiveness to the 10th and 50th income percentiles observed by Gilens.

Table A-1: Policy Responsiveness to the Preferences of the 10th, 90th, and 50th Income Percentiles When a Preference Gap Exists: Gilens (2012) Replication and after Coarsened Exact Matching (CEM)

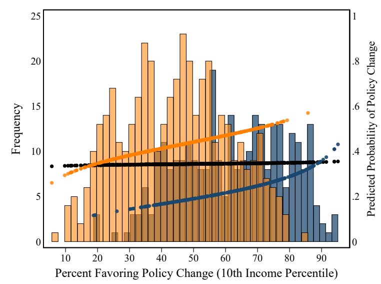
	Gilens Replication	CEM Balanced	Gilens Replication	CEM Balanced
	10th and 90th Income Percentiles			
10th Income Percentile	0.02	0.44*		
	[-0.16, 0.19]	[0.24, 0.64]		
90th Income Percentile			0.46*	0.31*
			[0.27, 0.65]	[0.14, 0.45]
Constant	-0.65*	-0.81*	-0.77*	-0.76*
	[-0.80,-0.49]	[-0.95, -0.70]	[-0.93,-0.60]	[-0.86, -0.68]
N	723	432	723	460
	50th and 90th Income Percentiles			
50th Income Percentile	-0.01	0.66*		
	[-0.28, 0.26]	[0.30, 1.06]		
90th Income Percentile			0.47*	0.30*
			[0.11, 0.82]	[0.12, 0.48]
Constant	-0.80*	-0.78*	-0.86*	-0.85*
	[-1.03,-0.56]	[-0.96, -0.62]	[-1.11,-0.62]	[-0.95, -0.76]
N	322	166	322	266

Dependent variable is coded 1 if the proposed policy change took place within four years of the survey date, 0 otherwise. Gilens Replication reproduces Table 3.2 of Gilens (2012) "Greater than 10 points." CEM Balanced reports the median estimate from 1,000 analyses where Coarsened Exact Matching creates fully balanced distributions. *=p<0.05; 95% uncertainty estimates in brackets.

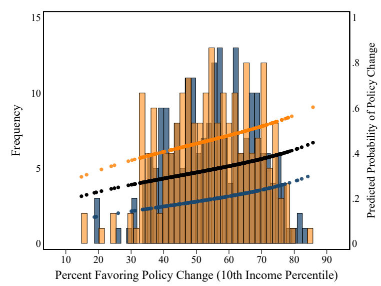
Online Supplement 4 Comparison of CEM Distribution and Distribution with All the Data

Figure A-4 (a and b) reproduce Figure 3 (c and d) from the text. Figure A-4 (a) shows that limiting the analysis to policies where more than a 10 percentage point preference gap exists shifts the distribution of data analyzed, changing the overall estimate of responsiveness (black line). Figure A-4 (b) confirms this conclusion, showing that after Coarsened Exact Matching (CEM), even though the data still contain a preference gap, the overall estimate of responsiveness (black) matches the estimate of responsiveness to both groups of data (orange and blue lines).

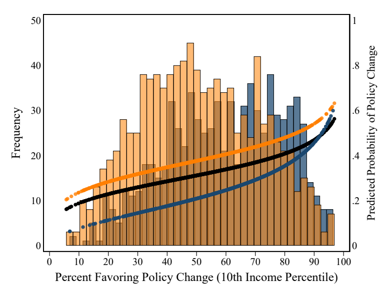
Figure A-4 (c) plots *all* data—not just data limited to policies where a preference gap exists—and the corresponding predicted probabilities of policy responsiveness. Orange bars indicate policies the 90th percentile prefer more than the 10th income percentile; i.e., (90th-10th)>0 instead of (90th-10th)>10, and blue bars indicate policies the 90th percentile prefer less than the 10th percentile; i.e., (90th-10th)<0 instead of (90th-10th)<–10. Two patterns stand out. First, the distributions of orange and blue bars when all data are analyzed look much closer to the CEM analysis than to the preference gap analysis Gilens conducted. The overlap is not identical, because the data are grouped by policies the affluent support more and less than the 10th income percentile, but Figure A-4 (c) illustrates how restricting the data to policies where a preference gap exists (a) alters the patterns in the data. Second, the predicted probabilities when analyzing all data look much closer to the CEM analysis than to the preference gap analysis. In addition to further illustrating how only analyzing policies with a preference gap can alter the data, the similarities between Figure A-4 (c) and (b) confirm that CEM does not introduce unexpected patterns into the data.



(a) Partially Overlapping Distributions Resulting from Analyzing Preference Gaps



(b) Fully Overlapping Distributions After CEM



(c) Distributions when Including *All* Data

Figure A-4: Distribution of Policy Preferences for the 10th Income Percentile when the 90th Income Percentile Supports Policies More (Orange) or Less (Blue) and the Corresponding Predicted Policy Responsiveness: When a Preference Gap Exists as Analyzed by Gilens (a), When a Preference Gap Exists with Coarsened Exact Matching (b), and with All Data (c)

Online Supplement 5 Extending Table 2 (Group Intercept Model) to the 50th and 90th Income Percentiles

Table 2 in the text showed that the addition of the group intercept variable resolves Simpson's paradox and changes the results so we no longer observe zero responsiveness to the 10th income percentile. Table A-2 reports corresponding results for the policies which the 50th and 90th income percentiles differ in support by at least 10 percentage points. The analysis reproduces Gilens' near-zero estimated relationship between the preferences of the median income and the probability of policy change. The estimate of responsiveness to the 50th income percentile (0.25) in the Group Intercept model is positive as expected, but not statistically significant (p=0.138, two-tailed), though the confidence interval overlaps the estimate for the 90th income percentile.

Table A-2: Policy Responsiveness to the Preferences of the 50th and 90th Income Percentiles When a Preference Gap Exists: Replication of Gilens (2012) and Accounting for the Group Intercept

	Gilens Replication	Group Intercept	Gilens Replication	Group Intercept
	50th		*90th*	
50th Income Percentile	-0.01	0.25		
	[-0.28, 0.26]	[-0.08, 0.59]		
90th Income Percentile			0.47*	0.41*
			[0.11, 0.82]	[0.05, 0.77]
Group Intercept		0.82*		0.47
		[0.22, 1.41]		[-0.03, 0.96]
Constant	-0.80*	-1.27*	-0.86*	-1.11*
	[-1.03,-0.56]	[-1.70,-0.84]	[-1.11,-0.62]	[-1.48,-0.74]
N	322	322	322	322

Dependent variable is coded 1 if the proposed policy change took place within four years of the survey date, 0 otherwise. Replication reproduces Table 3.2 of Gilens (2012). Group Intercept is a dichotomous variables coded as 1 when the 90th income percentile supports the policy more than other income groups, 0 otherwise, to account for the higher probability of policy adoption when the affluent support a policy more. 95% confidence intervals in brackets; *=p<0.05

Online Supplement 6 Numerical Results for Figure 4 and Additional Gender Gap Analysis

Figure 4 in the text showed evidence that Kopkin and Roberts' (2023) finding of no policy responsiveness to women and Mathisen's (2024) finding of negative policy responsiveness to women result because the data were restricted to policies for which a preference gap exists. The predicted values in Figure 4 are based on the estimates in Table A-3. The top half of Table A-3 reproduces Kopkin and Roberts' (2023, Table 3) estimates of policy responsiveness to men in Column 1 (0.56) and women in column 4 (0.02) when a preference gap exists. The bottom half of the table reproduces results from Mathisen (2024, Figure 3 and Appendix Table B1), including the statistically significant estimate of *negative* responsiveness to women (-0.22).

If the analysis of preference gaps introduced Simpson's paradox, estimating the models with the group intercept or after CEM should yield different results. The Group Intercept model (columns 2 and 5) adds a dichotomous indicator to account for the higher probability of policy adoption when men prefer the policy more than women. As expected, the biggest change occurs in the estimate of responsiveness to women (Column 5). Using the same data as Kopkin and Roberts (2023) (top half of Table A-3) and Mathisen (2024) (bottom half), we see a positive and significant relationship between the policy preferences of women and the probability of policy change (0.52 and 0.38, respectively). Columns 3 and 6 report the results after CEM. Again, the results change. Estimated responsiveness to women is significant using the Kopkin and Roberts (2023) approach (0.68). Using the data Mathisen analyzed, the coefficient associated with women in the CEM analysis is positive, but not statistically significant (0.21), perhaps due to the uncertainty that results from the small sample size after CEM (n=138) (e.g., Pimentel et al. 2018). Nevertheless, the estimated responsiveness to women is close to Mathisen's (2024) original estimate of responsiveness to men (0.293), and the uncertainty around these two estimates overlap substantially. Consistent with the patterns in Figure 4, the estimate of negative responsiveness to women appears to be driven by Simpson's paradox when the data are restricted to policies where a preference gap exists.

Table A-3: Policy Responsiveness to the Preferences of Men and Women When a Preference Gap Exists: Replication of Kopkin and Roberts (2023) and Mathisen (2024), Accounting for the Group Intercept and After CEM

	(1)	(2)	(3)	(4)	(5)	(6)
	Responsiveness to Men			*Responsiveness to Women*		
	Kopkin and Roberts (2023) Data					
	K & R Replication	Group Intercept	CEM	K & R Replication	Group Intercept	CEM
Men	0.56*	0.74*	0.70*			
	(0.14)	(0.15)	[0.55, 0.85]			
Women				0.02	0.52*	0.68*
				(0.09)	(0.13)	[0.22, 1.17]
Group Intercept		0.92*			1.46*	
		(0.20)			(0.27)	
Constant	-0.81*	-1.32*	-0.84*	-0.74*	-1.58*	-0.88*
	(0.10)	(0.15)	[-0.93, -0.76]	(0.09)	(0.19)	[-1.07,-0.71]
N	538	538	426	538	538	182
	Mathisen (2024) Data					
	Mathisen Replication	Group Intercept	CEM	Mathisen Replication	Group Intercept	CEM
Men	0.293*	0.44*	0.53*			
	(0.147)	(0.14)	[0.37, 0.69]			
Women				-0.22*	0.38*	0.21
				(0.10)	(0.15)	[-0.26, 0.72]
Group Intercept		0.24*			0.35*	
		(0.04)			(0.06)	
Constant	0.160*	-0.03	0.04	0.16*	-0.06	0.20
	(0.078)	(0.08)	[-0.03, 0.12]	(0.05)	(0.10)	[-0.07,0.46]
N	466	466	370	466	466	138

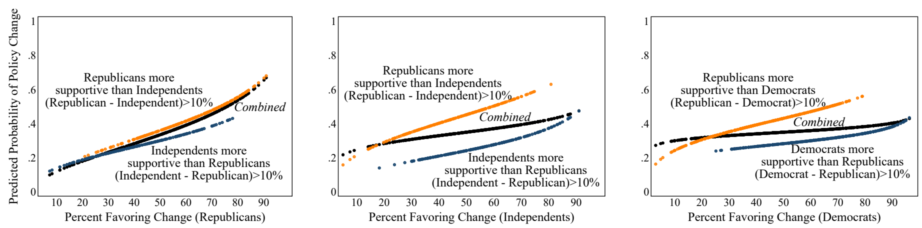
Dependent variable is coded 1 if the proposed policy change took place within four years of the survey date, 0 otherwise. Coefficients with standard errors in parentheses, 95% uncertainty estimates in brackets (*=p<0.05). Kopkin and Roberts' (K & R) preference gap cutoff is 8 percentage points and Mathisen's cutoff is 10 percentage points.

Online Supplement 7 Full Partisan Group Analysis Referenced in the Text

As the text notes, partisan groups offer a third example of how analyzing preference gaps can introduce Simpson's paradox, leading to misleading conclusions about policy responsiveness. Figure A-5 follows the previous figures, and presents the predicted probability of policy change for Republicans, independents, and Democrats when Republicans prefer the policies by at least 10 percentage points more than independents or Democrats (orange), 10 percentage points less than these groups (blue), or combining the data (black). All data come from Gilens. The left panel (a) shows compelling evidence of responsiveness to Republicans. This result makes sense given the Republican advantage in the federal government during the years covered by Gilens' data (1981 to 2005).[A-5]

[A-5]Gilens survey data stop in 2002, but the final policy change in his data occurred in 2005. During this period, Republicans controlled the house, senate, and presidency in five years and two of these institutions in another 12 years. Democrats only controlled the house, senate, and presidency in two years, and two institutions in another six.

In the middle panel (b), if we consider the preferences of independents when Republicans support the policy more (orange) or when independents support the policy more (blue), we see evidence of responsiveness. This responsiveness is consistent with the electoral importance of the median and of independents (Downs 1957; Dalton 2013; Fiorina 2016; Kang and Powell 2010; Wright 1989) and evidence that independents pay attention to and respond to legislator and candidate positioning (Jessee 2009, 2012; Shor and Rogowski 2018). However, when the data are analyzed together (black), the relationship is attenuated and not statistically different from zero (see Table A-4). The same patterns emerge for Democrats in the right panel (c). When we analyze all policies together (black), the evidence of responsiveness in the orange and blue predicted values largely disappears. Although we might expect more responsiveness to Republicans due to the partisan composition of the federal government during the period of analysis, the lack of evidence of responsiveness (black lines) to independents and Democrats is stunning and appears to result from Simpson's paradox.



(a) Republicans (b) Independents (c) Democrats

Figure A-5: Predicted Policy Responsiveness to Republicans (left panel), Independents (middle panel), and Democrats (right panel) When Republicans Prefer the Policy More Than Independents or Democrats (Orange), Independents or Democrats Prefer the Policy More Than Republicans (Blue), and Combined (Black).

Table A-4 reports the numerical results associated with the predicted probabilities (black lines) in Figure A-5 as well as additional evidence that Simpson's paradox influences the conclusions. Results for Republicans, Independents, and Democrats reflect separate analyses. The first column of Table A-4 reports the estimated relationship between each partisan group's preferences and the probability of policy change when following Gilens' preference gap approach. Consistent with the analyses based on all data in Figure A-5, we cannot conclude that the relationships between the preferences of independents or Democrats and policy change is different from zero.

The second column accounts of the different intercepts of the blue and orange lines in Figure A-5 by adding a group intercept binary indicator for when Republicans support the policy more than independents or Democrats. As with the earlier income group and gender analyses, this indicator resolves Simpson's paradox, and the estimated relationships align with the slopes of the orange and blue lines. These results also align with theoretical expectations. We continue to see evidence of responsiveness to Republicans' preferences, but we also see evidence of responsiveness to independents' preferences, to roughly the same magnitude as Republicans. For Democrats, we see a statistically significant relationship between preferences and policy outcomes, but to a smaller degree. This smaller point estimate also aligns with theoretical expectations, since Republicans controlled more of the federal government during the period of analysis.

The far right shows the results after CEM. A significant relationship between preferences and policy outcomes emerges for all three groups, again suggesting that the absence of evidence of responsiveness to independents and Democrats in Figure A-5 and column 1 of Table A-4 results from Simpson's paradox. The smaller coefficient for Republicans is unexpected, perhaps a function of the small sample size that results after CEM drops unmatched observations (e.g., Pimentel et al. 2018). The positive and significant associations for all partisan groups in the CEM analysis offers further evidence that the null results for responsiveness to independents and Democrats when restricting the data to policies which a preference gap exists results from Simpson's paradox.

Table A-4: Policy Responsiveness to Republicans, Independents, and Democrats When at Least a 10 Percentage Point Preference Gap Exists

	Gilens' Approach	Group Intercept	CEM
Republicans	0.60*	0.52*	0.31*
	(0.13)	(0.15)	[0.03,0.57]
Indicator for Republican support > Independent		0.28	
		(0.23)	
Constant	-0.69*	-0.84*	-0.68*
	(0.10)	(0.16)	[-0.78,-0.57]
N	450	450	270
Independents	0.23	0.48*	0.59*
	(0.13)	(0.15)	[0.34,0.84]
Group Intercept		0.96*	
		(0.23)	
Constant	-0.62*	-1.16*	-0.63*
	(0.10)	(0.17)	[-0.72,-0.54]
N	450	450	322
Democrats	0.10	0.30*	0.60*
	(0.08)	(0.09)	[0.40,0.84]
Group Intercept		0.68*	
		(0.19)	
Constant	-0.63*	-1.00*	-0.62*
	(0.08)	(0.13)	[-0.70,-0.52]
N	771	771	416

Dependent variable is coded 1 if the proposed policy change took place within four years of the survey date, 0 otherwise. Group Intercept is a dichotomous variables coded as 1 when Republicans support the policy more than independents or Democrats, 0 otherwise. *=p<0.05; Standard errors in parentheses. 95% uncertainty estimates, based on 1,000 CEMs, in brackets.

Online Supplement 8 Conclusions about Type I and Type II Errors from Table 3 Hold When Including the Group Intercept Variable Instead of CEM

Table 3 in the text reported results of simulations that showed CEM resolves Simpson's paradox and associated Type II error by eliminating incorrect evidence of differential responsiveness, but when group preferences are highly correlated and only one group influences policy, CEM can produce Type I error. Table A-5 shows the same conclusions apply when including the group intercept in the model.

Columns 1 and 2 reproduce the results from Table 1, showing that if both income groups influence policy outcomes in the DGP, restricting the analysis to policies where a preference gap exists can produce *incorrect* evidence of differential responsiveness. In columns 3 and 4, we see that if both income groups influence policy change, including the group intercept resolves Simpson's paradox and we see evidence of the true relationship. However, due to omitted variable bias from only analyzing one income group at a time, the estimated relationship is inflated. By contrast, if only one group influences policy change (columns 5–8), we incorrectly conclude both groups influence policy change. This false positive results because only one group's preferences are included in the model and income group preferences are highly correlated. In sum, the simulations demonstrate that while the group intercept can resolve Simpson's paradox, since we do not know the true DGP, even when only analyzing policies for which a preference gap exists, income group preferences are too highly correlated in Gilens' data to evaluate differential responsiveness.

Table A-5: Simulations Show that Even when Including a Group Intercept Variable, Income Group Preferences are Too Highly Correlated to Assess Responsiveness to Income Groups in Gilens' Data

	Gilens' Approach		*DGP (β):*					
	1	2	3	4	5	6	7	8
	10th=0.164	90th=0.164	10th=0.164	90th=0.164	10th=0	90th=0.33	10=0.33	90th=0
10th Income Percentile	-0.01		0.32*		0.31*		0.33*	
	[-0.19, 0.17]		[0.13, 0.53]		[0.11, 0.52]		[0.14, 0.54]	
90th Income Percentile		0.49*		0.31*		0.33*		0.29*
		[0.32, 0.67]		[0.12, 0.51]		[0.14, 0.54]		[0.10, 0.48]
Group Intercept			1.44*	0.92*	1.69*	1.18*	1.18*	0.69*
			[1.07, 1.86]	[0.59, 1.31]	[1.31, 2.10]	[0.83, 1.56]	[0.82, 1.59]	[0.38, 1.05]
Constant	-0.56*	-0.75*	1.44*	0.92*	1.69*	1.18*	1.18*	0.69*
	[-0.71,-0.40]	[-0.92, -0.57]	[1.07, 1.86]	[0.59, 1.31]	[1.31, 2.10]	[0.83, 1.56]	[0.82, 1.59]	[0.38, 1.05]
N	726	726	726	726	726	726	726	726

Dependent variable is a simulated dichotomous indicator based on observed patterns in Gilens' data of whether policy change occurred. Median estimates from 1,000 simulated datasets. *=p<0.05; 95% uncertainty estimates in brackets.

Online Supplement 9 Correcting for Correlated Measurement Error Does Not Clarify Conclusions About Responsiveness

As the text notes, the similar results between Gilens' multivariable analyses with measurement error correction using all the data (not just data where a preference gap exists) (Gilens 2012, 85-87; Gilens and Page 2014; Page and Gilens 2017) and the results when limiting the analyses to policies where a preference gap exists were viewed as reassuring. It turns out, however, that the multivariable analysis with measurement error correction is also unable to provide conclusive evidence of differential responsiveness. This section shows that the estimates of correlated measurement error are necessarily inflated and that the conclusions from the multivariable analysis are extremely sensitive to the estimate of correlated measurement error used.

To understand Gilens' measurement error correction, recall that the analysis of preference gaps was designed to solve the challenge of estimating responsiveness to income groups whose preferences are extremely highly correlated (the preferences of the affluent and median correlate at r=0.94). If the issue was imprecise estimates due to larger standard errors from extreme multicollinearity, this would not be a problem (Bartels 2023, 89-90). But Gilens and Page report a *precisely* estimated negative coefficient—which they refer to as "implausible" (2014, Appendix 1; also see Gilens 2012, 253)—for the relationship between the preferences of average citizens and the probability of policy change ($\beta = -0.93$, s.e.=0.24, $p<.001$). Their analysis also produces what they refer to as an "implausibly large" positive coefficient (2014, Appendix 1; also see Gilens 2012, 253) that suggests more than a one-to-one relationship between affluent preferences and policy outcomes (β=1.66, s.e.=0.24). Thus, according to Gilens (2012, 253) and Gilens and Page (2014), the extreme collinearity in the data leads to incorrect conclusions from multivariable analysis (also see Achen (1985)). To address this issue, Gilens' multivariable analysis attempts to correct for the correlated measurement error across the measures of income group preferences.

To make this correction, Gilens (2012) identifies 161 sets of questions in his data that were asked in similar ways two or more times within the same year.[A-6] Gilens explains, "If there were no correlated error in these preference measures, the covariance of preferences across different income groups on the same version of a proposed policy change would (on average) equal the covariance of preferences across those income groups on alternative versions" (254). Gilens estimates that correlated error accounts for 19, 17, and 14 percent of the observed covariance between the 10th and 90th, 50th and 90th, and 10th and 50th income percentiles, respectively (255-256). He then deflates the covariance matrix in the multivariable regression by these amounts. The intuition of this approach is sound, but it turns out that the results are sensitive to the estimate of correlated measurement error and Gilens' approach inflates these estimates. These two patterns suggest that if we had a more accurate estimate of correlated measurement error, the results would likely be different. Unfortunately, there is no way to assess by how much the estimate of correlated measurement error is inflated. I detail these considerations below.

Online Supplement 9.1 An Inflated Estimate of Correlated Measurement Error

The estimate of how much correlated measurement error accounts for the covariance in observed preferences is inflated because Gilens' approach assumes correlated error accounts for *all* of the difference in covariances between "same-version" and "alternate-version" questions. Three considerations indicate that this cannot be the case.

First, since there are only 161 sets of questions and many of the alternate question versions come from different surveys, random sampling error would contribute to different covariances across question types. To evaluate how much of the covariance difference random sampling could account for, I generated two sets of simulated data representing the paired preferences of the low, middle, and high income groups in Gilens' analysis.

The simulation starts by generating data with same means, minimum values, maximum values, and covariances as Gilens' data for each income group. The total number of questions is set to match the number of paired questions Gilens' analyzes, n=161. From each of these simulated income group preferences, I generate two variables which equal the original simulated variable plus a normally distributed random error with a mean of 0 and a standard deviation of 0.05, so approximately 95 percent of values fall within +/-9.8 percentage points, which is the estimated sampling margin of error for 100 respondents. I select this sampling margin of error because a typical survey

[A-6] There are 116 sets with two alternative versions, 25 sets with three alternative versions, and 20 sets with four alternative versions, for 387 total questions (Gilens 2012, p.254).

has about 1,000 respondents and we are interested in the preferences of the 10th, 50th, and 90th income percentiles. The result is two variables for each income group that simulate the two sets of 161 paired questions Gilens' analyzes. The *only* difference across the two sets of simulated data is simulated random sampling error.

I repeat the above DGP 1,000 times. The paired questions are *identical* except for simulated random sampling error. I then calculate the difference in covariance for the 10th and 90th, 50th and 90th, and 10th and 50th income percentiles for the two simulated sets of data and compare these values to the corresponding covariance differences in Gilens' data that he attributes entirely to correlated measurement error. The simulations show that random sampling error would likely account for (median simulation) 20.0%, 17.7%, and 23.6% of Gilens' error covariance for the 10th and 90th, 50th and 90th, and 10th and 50th income percentiles, respectively. Ninety five percent of simulations fall between 0.9%—66.2%, 0.7%—65.3%, and 0.7%—78.9%, suggesting that random sampling error alone could account for as little as 0.7% and as much as 78.9% of the covariance that Gilens attributes to correlated measurement error.

Some of the question pairs in Gilens data come from the same survey, so random sampling error would not affect the covariance estimates across these questions. However, these simulations do not account for additional error that results because Gilens imputes the preferences of income groups based on a regression of policy preferences on income and income squared (Gilens 2005, 783). Based on Gilens' data, it is not possible to determine how many of the 161 question pairs came from the same polls or how much additional error imputing income group preferences introduces, but these factors are likely offsetting. More questions from the same surveys would reduce the previous estimate of random sampling error, but error from the imputation of preferences offers another explanation for differences in covariance across paired questions that is distinct from correlated measurement error. Because these unobserved factors have offsetting directional effects, the above simulations offer a reasonable approximation of how much sampling error accounts for Gilens' overestimate of correlated measurement error. Importantly, sampling error is *just one* of the ways Gilens' analysis overestimates how much of the difference in covariances is due to correlated measurement error.

Second, true opinion change will account for some of what Gilens classifies as correlated measurement error. Gilens and Page (2014, Appendix 1) indicate that "topics in the news at the time the survey was in the field" can produce correlated measurement error. News and other events can also produce *true* shifts in preferences and different income groups consistently update their preferences in the same direction (e.g., Page and Shapiro 1992; Enns and Wlezien 2011). Other groups also tend to update their preferences in tandem and this "parallel" opinion updating consistently reflects groups responding systematically to the information environment (Coppock 2022; Enns and Kellstedt 2008; Page and Shapiro 1992). Based on this research, it seems implausible to assume that when different income groups shift their preferences in tandem, *no* actual opinion updating occurs and, instead, *all* of this shift reflects correlated measurement error.

Third, different question wording across question pairs adds another source of error that inflates the estimate of correlated measurement error. None of the paired questions that Gilens uses to estimate correlated measurement error are worded the same (Gilens 2012, 58). While it is possible that different question wording could produce correlated measurement error (Gilens 1012, 86), it is likely that different question wording also elicits *true* preference differences and these true preference differences could be similar across income groups. For example, consider the question pair Gilens highlights about privacy violations (Gilens 2012, 64). One question asks respondents whether they favor or oppose federal laws that would make privacy violations "a criminal offense" and the other asks whether they favor or oppose federal laws that "could put companies out of business" for privacy violations. It would not be surprising if support for putting companies out of business was lower across all income groups, since this is a more specific and harsher consequence than designating a privacy violation a criminal offense.[A-7] Consistent with this expectation, in Gilens' data, support for putting companies out of business for privacy violations among the 10th, 50th, and 90th income percentiles is 16, 15, and 13 percentage points *lower* than making privacy violations a criminal offense. If any of this lower support reflects true preference differences based on different question wording, Gilens' estimate of correlated measurement error would be further inflated. Gilens (2012, 63 italics mine) seems to acknowledge this possibility when he writes, "such uncertainties [due to different question wording] contribute to the unreliability of my measures of public preferences, *even if they do not reflect measurement error*, per se."

In sum, random sampling error, parallel opinion updating (Coppock 2022; Enns and Kellstedt 2008; Page and Shapiro 1992), and different true preferences in response to different question wording (i.e., support for making privacy violations a criminal offense but not for putting companies out of business for privacy violations) would all inflate the estimate of correlated measurement error used by Gilens.

Online Supplement 9.2 Multivariable Estimates Are Sensitive to the Measurement Error Correction

The previous section explained that at least some of what Gilens attributes to correlated measurement error is due to random sampling error and true response differences due to different question wording and/or different topics in the news (or other relevant factors) at the time of survey. The overestimate of correlated measurement error would not be a problem if the results were not sensitive to these estimated values. But this is not the case. Because we cannot know the true amount of correlated measurement error, I replicate Gilens' measurement error analysis, allowing the proportion of error attributed to correlated measurement error to range from a lower bound of 1% to an upper bound of Gilens' estimate minus 1% (i.e., 18%, 16%, 13% for the 10th and 90th percentiles, 50th and 90th percentiles, and 10th and 50th percentiles, respectively). These ranges produce 3744 combinations of the proportion of the covariance across income group preferences that correlated measurement error might account for.

In 12.0% of these combinations, a positive and significant relationship between the preferences of the 10th or 50th income percentile and policy change emerges. More than half (56.1%) yield either a positive and significant relationship or what Gilens' (2012, 253) calls an "implausible" precisely estimated ($p<0.05$) negative relationship. In other words, more than half of the potential measurement error possibilities change the results or produce relationships that—according to Gilens—are implausible. The data do not allow us to draw conclusions with confidence or certainty. Branham, Soroka and Wlezien (2017) have provided additional evidence of the sensitivity of results with measurement error correction. They replicate Gilens and Page's (2014) measurement error correction and find the evidence of differential responsiveness to the middle and upper income groups disappears when they limit the analysis to policies which the majority of the middle and the

[A-7] Gilens (2012) provides three other examples of question sets in his Table 2.2.

affluent prefer different outcomes—where we would expect the greatest differential responsiveness (Branham, Soroka and Wlezien 2017, Table A3).[A-8]

I also conducted analyses using the reliability coefficients Gilens reports in conjunction with Gilens' proposed measurement error correction. These analyses find further evidence that multivariable conclusions are highly sensitive to the analytic approach. Gilens provides reliability estimates (2012, 88), which account for the random measurement error in the variables. While conceptually distinct from the correlated measurement error that Gilens tries to account for, we might wonder if incorporating these reliability estimates into the analysis to account for random measurement error affects the results. I assess this possibility here. The top half of Table A-6 (Column 1) presents an exact replication of Gilens's (2012, Table A3.3, Column 3) measurement error correction model and the bottom half of Table A-6 presents a nearly exact replication of Gilens and Page's (2014) Table 3, Model 4, which incorporates interest group support for each of the proposed policies.[A-9]

Column 2 assesses how the results would change if we incorporated Gilens' reliability estimates to correct for random measurement error instead of correlated measurement error. Comparing column 1 and column 2 is a bit of an "apples to oranges" comparison as the two approaches deal with two different types of error. However, Column 2 is not an implausible approach. This is the estimation strategy used by Bhatti and Erikson (2011) in their analysis of policy responsiveness in the U.S. senate to income group preferences and we have seen that as much as 78.9 of what Gilens attributes to correlated measurement error could be due to random measurement error.[A-10] Despite the potential plausibility of the analysis in Column 2, the results essentially reverse those in Column 1. As noted above, we would not necessarily expect the two different measurement approaches to produce the same results, but the opposite patterns of responsiveness to income groups highlight the sensitivity of the conclusions to the measurement error correction used.

Column 3 combines the approaches in columns 1 and 2, adjusting the covariance and utilizing the reliabilities Gilens' generated. We might expect incorporating both types of measurement error that Gilens provides to be an especially reasonable approach, but again, the results change dramatically, producing what Gilens has called "implausible" significant negative responsiveness (e.g., Gilens 2012, 253; Gilens & Page 2014, Appendix 1). The estimated relationship is negative and significant for the 10th income percentile in the top half of the table and for the 50th percentile in the bottom half of the table. The estimate for the 90th income percentile in the bottom half of the table seems implausibly large at nearly 25 times the estimate of interest groups. Because this model uses Gilens' reliability estimates and Gilens' estimates of correlated measurement error, it might seem like an especially reasonable estimation strategy. But the results highlight a broader theme. The analyses with multiple income group preference variables are extremely sensitive to the specific measurement error correction and even seemingly sensible measurement error strategies can produce what Gilens refers to as implausible results. The exception appears to be the esti-

[A-8] Bashir (2015) uses simulations based on Gilens and Page's (2014) analysis and similarly concludes, "the test on which the original study is based is prone to underestimating the impact of citizens at the 50th income percentile by a wide margin" (1); but also see Gilens (2016).

[A-9] The only replication difference is the coefficient and standard error on the interest group variable from Gilens and Page (2014, Table 3, Model 4) differ by 0.01 and 0.01, respectively. These tiny differences may reflect my use of Stata instead of AMOS to estimate the models.

[A-10] Bhatti and Erikson (2011) use errors-in-variables regression. I use the Stata sem command, which Lockwood and McCaffrey (2020) show produces the same coefficient estimates but avoids negatively biased standard errors reported by eivreg in Stata (also see Petrescu (2013)).

Table A-6: Multivariable Analysis with Different Measurement Error Corrections Produce Different Conclusions about Policy Responsiveness

	(1) Gilens Replication (Covariance)	(2) Gilens' Reliability Estimates	(3) Gilens' Reliability Estimates + Covariance
		Gilens 2012	
10th Income Percentile	-0.10	0.30*	-0.22*
	(0.09)	(0.06)	(0.04)
50th Income Percentile	0.08	0.05*	0.10
	(0.10)	(0.01)	(0.05)
90th Income Percentile	0.51*	-0.25*	0.23*
	(0.09)	(0.05)	(0.05)
Constant	0.33*	0.33*	0.33*
	(0.01)	(0.01)	(0.01)
N	1779	1779	1779
		Gilens & Page 2014	
50th Income Percentile	0.03	0.98*	-10.02*
	(0.08)	(0.11)	(2.97)
90th Income Percentile	0.76*	-0.18	11.14*
	(0.08)	(0.11)	(3.05)
Interest Groups	0.55*	0.55*	0.45*
	(0.08)	(0.08)	(0.11)
Constant	0.07	0.06	0.11*
	(0.04)	(0.04)	(0.05)
N	1779	1779	1779

Dependent variable is coded 1 if the proposed policy change took place within four years of the survey date, 0 otherwise. Gilens (2012) Replication reproduces Column 3, Table A3.3 and (nearly) reproduces Gilens & Page (2014) Model 4, Table 3. Gilens Reliability Estimates use the reliabilities reported in Gilens 2012, 88. Gilens' covariance estimates reported in Gilens 2012, 254-256. Standard errors in parentheses; *=p<0.05. SEM estimates using log-odds of preferences (except results in top half of column 1).

mated relationship between interest group preferences and the probability of policy change, which is consistently positive and significant (columns 1, 2, and 3).

The results show that conclusions about responsiveness to different groups vary substantially depending on the measurement error correction (also see Branham, Soroka and Wlezien (2017)). This variation is especially concerning because Gilens' estimates of correlated measurement error are necessarily inflated, meaning we do not know—and thus our models cannot include—the true amount of correlated measurement error. The sensitivity of results reinforces this article's main conclusions based on the analyses of preference gaps. The data are too highly correlated to assess differential responsiveness.

Online Supplement 10 The Problems with Preference Gaps as Independent Variables

As the main text indicates, instead of restricting the analysis to data where preference gaps exist, researchers are increasingly including the preference gap (the percent of one group who support a policy minus the percent of another group that support the policy) as an independent variable and interpreting the coefficient as the estimate of differential responsiveness (Flavin and Franko 2017; Lupu and Castro 2023; Mathisen et al. 2023; Mathisen 2024; Peters and Ensink 2014; Schakel, Burgoon and Hakhverdian 2020). Bartels (2023, 89) seems to endorse this approach, writing, "it may be feasible to recast our analyses (by redefining our explanatory variables) to focus directly on the impact of differences in subgroup preferences, which are less likely to be highly correlated." This approach has produced results that seem to align with Gilens' analyses of preference gaps, but these results should be reconsidered. Adding a preference gap to the right-hand-side of the equation can produce misleading results.

To see why, suppose we calculate the difference in preferences between the 90th income percentile and 10th income percentile, $(Preferences_{90th} - Preferences_{10th}) = Preference\,Gap_{90th-10th}$. The estimated relationship between this variable and the probability of policy change,

$$Policy\ Change = \alpha + \beta_1 Preference\ Gap_{90th-10th} + \epsilon, \tag{A-1}$$

can be rewritten as,

$$Policy\ Change = \alpha + \beta_1 (Preferences_{90th} - Preferences_{10th}) + \epsilon, \tag{A-2}$$

which equals,

$$Policy\ Change = \alpha + \beta_1 Preferences_{90th} - \beta_1 Preferences_{10th} + \epsilon. \tag{A-3}$$

When we include a preference gap on the right-hand side of the equation, we must recognize that the coefficient on the preference gap variable (β_1) is constrained to be equal and opposite the estimated relationship between the preferences of the other group and the probability of policy change ($-\beta_1$). This approach assumes that responsiveness to different groups diverges, which is problematic because that is what the analysis is supposed to test.

Once we recognize that including a preference gap as a variable forces two variables—which are typically highly correlated—to have equal and opposite signed relationships, it should not be surprising that this approach does not recover the DGP. However, given the growing number of publications that include a preference gap variable to estimate differential responsiveness, below I demonstrate this result with Gilens' data and use simulations to illustrate misleading conclusions that can emerge with this approach.

Importantly, if both the preference gap and the preferences of one group that is part of the preference gap are included as independent variables (e.g., $Policy\,Change = \alpha + \beta_1 Preference\,Gap_{90th-10th} + \beta_2 Preferences_{90th} + \epsilon$), the preference gap variable does *not* force equal and opposite signed relationships. This is Bartels's (2023) specific recommendation, but it is also problematic. This specification is an exact reparameterization of the original multivariable model and would thus reproduce the results Gilens and Page (2014, Appendix 1; also see Gilens 2012, 253) call "implausible," meaning this cannot be a viable estimation strategy with Gilens' data. The final subsection of this section demonstrates this result and corresponding concerns with Gilens' data.

10.1 Preference Gaps Produce *Opposite Signed* Coefficients: Gilens' Data

Equations A-1, A-2, and A-3, above, demonstrate that it *has* to be the case that the estimated relationships for the 90th and 10th income percentiles are constrained to be equal and opposite signed when a preference gap is used to estimate differential responsiveness to income groups (as noted above, the final subsection in this section considers the one exception to this conclusion). The results in Table A-7 illustrate this outcome with Gilens' data. Column 1 reports the estimated relationship between a variable capturing the preference gap between the preferences of the 90th and 10the income percentiles (i.e., 90th income percentile preferences minus 10th income percentile preferences) and the probability of policy change. The coefficient is positive and significant (2.70).

Although many scholars have interpreted this type of result as evidence of unequal responsiveness, the results in Column 2 illustrate why this is not the case. The model includes the preferences of the 10th and 90th income percentiles, but before estimating the model, I constrain the coefficients to be equal and opposite signed.[A-11] With this approach, the coefficient and standard error associated with the 90th income percentile, 2.70 (0.45), are *identical* to the coefficient and standard error associated with the preference gap variable in Column 1. The estimated values for the constant and the model log likelihood are also identical in Column 1 and 2. However, in Column 2, we see that the coefficient associated with the 10th income percentile is identical but opposite signed as the 90th income percentile, -2.70 (0.45). Even if policy responds equally to two groups, including a preference gap as an independent variable estimates responsiveness to the 90th income percentile when responsiveness to the 90th and 10th income percentiles are forced to be equal and opposite.

[A-11] In Stata, `constraint define 1 _b[pred90_sw] = - _b[pred10_sw]`.

Table A-7: Preference Gaps as Independent Variables Force Coefficients to be *Opposite Signed*

	(1)	(2)
Preference Gap (90th-10th)	2.70*	
	(0.45)	
90th Income Percentile		2.70*
		(0.45)
10th Income Percentile		-2.70*
		(0.45)
Constant	-0.76*	-0.76*
	(0.05)	(0.05)
N	1779	1779
Log Likelihood	-1110.668	-1110.668

Dependent variable is coded 1 if the proposed policy change took place within four years of the survey date, 0 otherwise. Column 2 constrains the coefficients associated with the 90th and 10th income percentiles to be equal and oppositely signed, demonstrating that including a preference gap as a right-hand-side variable estimates the relationship between the preferences of the 90th income percentile and the probability of policy change *when the estimated relationship for the 90th and 10th income percentiles are constrained to be equal and opposite signed.* Coefficients with standard errors in parentheses. *=p<0.05

10.2 Opposite Signed Coefficients and Increased Type I Error Rate: DGP with Equal Responsiveness

To further illustrate the challenges associated with including a preference gap on the right-hand-side of the equation, I conduct simulations with a DGP that ensures *equal* responsiveness to both income groups. As with previous simulations, variables representing the preferences of the 90th and 10th income percentiles are generated to have the same mean, minimum and maximum values, and same covariance as the preference variables in Gilens' data. The probability of policy change is generated to be equally influenced by the preferences of the 90th and 10th income percentiles. The exact value of this relationship is based on the observed value of the relationship between the 90th income percentile preferences and the probability of policy change.[A-12]

Although I conduct 1,000 simulations, Table A-8 reports the results from the 97^{th}, 98^{th}, 99^{th}, and 100^{th} simulation. The table confirms the conclusions from Equations A-1, A-2, and A-3 and the results in Table A-7. Although the DGP specifies *equal* responsiveness to both income groups, the coefficients associated with the preference gap in the top half of Table A-8 are equal to the coefficients for the 90th income percentile in the bottom half of the table when responsiveness to the 90th and 10th income percentiles is constrained to be identical in magnitude and opposite signed. Furthermore, while the true relationship between the simulated preferences of the 90th and 10th income percentiles and the probability of policy change is 2.43, for all 1,000 simulations,

[A-12] Unlike prior simulations, for simplicity of presentation, I do not take the log odds of simulated preferences.

the mean estimated coefficient on the preference gap variable is -0.94, meaning the coefficient on the preference gap variable does not recover the true DGP. Additionally, despite simulating equal responsiveness to both income groups, the preference gap coefficient is statistically significant in 31% of simulations, indicating Type I errors where we would incorrectly conclude that a significant difference in responsiveness exists.

Table A-8: 97^{th}, 98^{th}, 99^{th}, and 100^{th} Simulations: Preference Gaps as Independent Variables Force Coefficients to be *Opposite Signed* Even When Responsiveness Is *Equal* to Both Groups

	(1)	(2)	(3)	(4)
Preference Gap (90th-10th)	0.10	-0.01	-0.73	-0.59
	(0.63)	0.65	0.64	0.62
Constant	0.49	0.53	0.53	0.47
	(0.05)	(0.05)	(0.05)	(0.05)
N	1918	1892	1922	1935
90th Income Percentile	0.10	-0.01	-0.73	-0.59
	(0.63)	(0.65)	(0.64)	(0.62)
10th Income Percentile	-0.10	0.01	0.73	0.59
	(0.63)	(0.65)	(0.64)	(0.62)
Constant	0.49	0.53	0.53	0.47
	(0.05)	(0.05)	(0.05)	(0.05)
N	1918	1892	1922	1935

Dependent variable is coded 1 if the proposed policy change took place within four years of the survey date, 0 otherwise. The bottom half of the table constrains the coefficients associated with the 90th and 10th income percentiles to be equal and oppositely signed. DGP specifies *equal* responsiveness to the 90th and 10th income percentiles. Coefficients with standard errors in parentheses. *=p<0.05

10.3 Even When Responsiveness Differs Across Groups in the DGP, the Preference Gap Variable Can Produce Misleading Conclusions

We have seen that including a preference gap on the right-hand-side does *not* estimate the difference in responsiveness to groups. As an additional test, I follow the above simulation procedure, but this time the DGP ensures that simulated policy is more responsive to the 90th income percentile than the 10th income percentile. Specifically, the gap in responsiveness is 0.61, which represents 25% less responsiveness to the 10th income percentile than the 90th income percentile.

Figure A-6 reports the frequency distribution of coefficients associated with the preference gap variable. If including a preference gap as in independent variable estimated the difference in responsiveness, this parameter would, on average, equal 0.61 (the vertical dashed line). This is not the case. Since including a preference gap as a variable forces two highly correlated variables to have equal and opposite signed relationships with the outcome variable, it is not surprising that this approach does not recover the DGP, but the results in Figure A-6 further illustrate potential incorrect conclusions when using a preference gap to assess differential responsiveness.

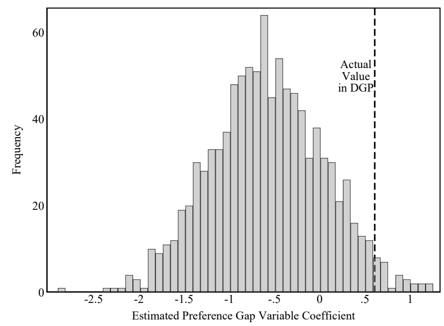


Figure A-6: Distribution of Simulated Coefficients associated with the Preference Gap Variable when the True Parameter = 0.61

10.4 The One Exception Reproduces the Original Multivariable Model of Income Group Preferences (and Associated Concerns)

To avoid forcing estimates of responsiveness to be equal and opposite signed when including a preference gap variable in the model, researchers can also include one of the variables that was used to generate the preference gap variable. Mathisen et al. (2023, Table 2.6) use this approach, including a variable that captures the preference gap between the 90th and 50th income percentiles and a variable for the 50th income percentile preferences.[A-13] Bartels (2023) endorses this approach, writing, "Relating policy outcomes to P50, (P90-P50), and (P50-P10) rather than to P50, P90, and P10 captures the same information about preferences, but isolates the differential impact of affluent and poor citizens' preferences relative to those of middle-income citizens" (89).

Bartels' summary also illustrates the challenge with this approach. Because this specification "captures the same information" as including all income group preferences (Bartels 2023, 89), all multicollinearity concerns and what Gilens (and Gilens and Page) refer to as "implausible" results (e.g., Gilens 2012, 253; Gilens & Page 2014, Appendix 1) that existed with the original model still exist. At first glance, this may seem like a surprising claim, as the results in Table A-9 suggest including preference gaps solves the multicollinearity problem.

Table A-9: Correlation between the Preferences of the 50th Income Percentile and Preference Gaps (i.e., 90th Minus the 50th Income Percentile and 50th Minus the 10th Income Percentile)

	50th	(90th - 50th)	(50th - 10th)
50th	1.00		
(90th - 50th)	-0.29	1.00	
(50th - 10th)	0.27	0.11	1.00

Unfortunately, despite the seeming promise of this approach, adding the preference gap to the

[A-13] Their analysis also includes variables for Left government, an interaction between the preference gap variable and Left government, and country dummy variables.

model does *not* actually change the model or the resulting conclusions. To see why, suppose we estimate the following model,

$$Policy\ Change = \alpha + \beta_1 50th + \beta_2 90th + \epsilon. \tag{A-4}$$

We would estimate *the exact same quantities* if we replaced the preferences of the 90th income percentile with the preference gap subtracting the preferences of the 50th percentile from the 90th,

$$Policy\ Change = \alpha + \gamma_1 50th + \beta_2 (90th - 50th) + \epsilon. \tag{A-5}$$

To demonstrate the equivalence of Equations A-4 and A-5, we simply multiply β_2 in Equation A-5 by each term within parentheses and then add like terms,

$$Policy\ Change = \alpha + \gamma_1 50th + \beta_2 90th - \beta_2 50th + \epsilon = \tag{A-6}$$
$$Policy\ Change = \alpha + (\gamma_1 - \beta_2) 50th + \beta_2 90th + \epsilon. \tag{A-7}$$

In Equations A-4 and A-5, β_2 is identical, and $(\gamma_1 - \beta_2)$ in Equation A-5 equals β_1 in Equation A-4. To illustrate these equivalencies, I estimate two models using Gilens' data. The first model estimates the probability of policy change as a function of the preferences of the 50th and 90th income percentiles. The second estimates the probability of policy change as a function of the preferences of the 50th income percentile and the preference gap between the preferences of the 90th and 50th income percentiles.[A-14] Table A-10 reports the results of these two analyses, which according to the above equations, are simply alternate parameterizations of the same model.

Table A-10: Estimating Policy Responsiveness to Preference Gaps Can Produce the *Exact Same* Quantities as Estimating Responsiveness to Preferences

Preferences		Preference Gap	
50th (β_1)	-2.765*	50th (γ_1)	2.401*
	(0.711)		(0.271)
90th (β_2)	5.166*	(90th-50th) (β_2)	5.166*
	(0.759)		(0.759)
Constant (α)	-2.114*	Constant (α)	-2.114*
	(0.170)		(0.170)
Log likelihood	-1077.565		-1077.565
Psuedo R2	0.045		0.045

From Equation A-7, $\beta_1 = \gamma_1 - \beta_2 = 2.401 - 5.166 = -2.765$.

In Table A-10, the results in the Preferences column on the left portion of the table illustrate the multicollinearity concerns with Gilens' data. The estimated relationship between the preferences of the median income group and policy change is "implausibly" (Gilens 2012, 253) negative and statistically significant (also see Achen (1985)). Looking at the Preference Gap results on the right portion of the table, it could be tempting to conclude that we see some evidence of responsiveness

[A-14] Bartels (2023, 89) recommends including low, median, and high income groups, but the mathematical point is the same. Including preference gaps estimates the exact same model as preferences in levels.

to the 50th percentile (γ=2.401) and even more responsiveness to the 90th (β_2=5.166). This would be the wrong conclusion. Equation A-7 shows that $(\gamma_1 - \beta_2)$ = (2.401-5.166), which equals β_1 (-2.765). The numeric quantities are identical across Table A-10. The only difference is the parameterization. Gilens (2012, 253) has concluded that negatively signed coefficients implying negative responsiveness like we see in the Preferences column of Table A-10 are implausible (also see Gilens and Page (2014, Appendix 1)). Replacing preferences with preference gaps only changed how the coefficients are grouped. The models are identical. If, as Gilens has argued (e.g., Gilens 2012, 253; Gilens & Page 2014, Appendix 1), the data are too highly correlated to estimate meaningful conclusions, including an indicator for preference gaps in the statistical model does not provide a solution.

Online Supplement 11 Including Gilens' Measurement Reliabilities with the General Public Anaysis

The main text showed that, consistent with theoretical expectations and prior literature, even when conditioning on the preferences of interest groups, policy responds to the preferences of the general public. Columns 1 and 2 in Table A-11, below, replicate the findings in the text. Columns 3 and 4 incorporate the measurement reliability coefficients that Gilens reports.[A-15] We saw above that results can be sensitive to measurement error correction, so it is important to evaluate the robustness of these results to such corrections. While this approach addresses measurement error, their is no *correlated* measurement error to address. As Gilens and Page (2014, Appendix 2) explain, the "measures of interest group alignments are entirely distinct from the preference measures in origin, there is no reason to expect that their measurement errors would be correlated across policies."

As with Table 4, we see evidence of responsiveness to the public and to interest groups, though the magnitude of responsiveness to the public (2.59) is more than double that of responsiveness to interest groups (1.07) when we take classical measurement error into account. The results are similar in column 4, with the estimate of responsiveness to the general public (2.73) nearly double responsiveness to business-based interest groups (1.41).[A-16] It is reassuring to see that the conclusions in Table A-11 are consistent with and without measurement error adjustment.

[A-15] The reliability coefficient is 0.81 for the policy outcome variable (Gilens 2012, 63), 0.82 of the public's policy preferences (p.65), and 0.87 for the interest group alignment variable (p.292).

[A-16] Gilens (2012) provides one reliability estimate for all interest groups, so I use that estimate for both interest group measures in column 4.

Table A-11: Policy Responsiveness to the General Public and to Interest Groups

	(1)	(2)	(3)	(4)
			w/ Reliabilities	
General Public	2.95*	3.11*	2.59*	2.73*
	(0.40)	(0.40)	(0.10)	(0.11)
Interest Groups (all)	2.66*		1.07*	
	(0.39)		(0.07)	
Mass-based Interest Groups		1.11*		0.45*
		(0.35)		(0.08)
Business Interest Groups		2.37*		1.41*
		(0.36)		(0.07)
Constant	-3.37*	-3.82*	0.33*	0.33*
	(0.28)	(0.34)	(0.01)	(0.01)
N	1779	1779	1779	1779

Dependent variable is coded 1 if the proposed policy change took place within four years of the survey date, 0 otherwise. All predictors scaled to range from 0 to 1. Columns 1 and 2: Logistic regression estimates; columns 3 and 4: Structural equation models with measurement reliability coefficients (Gilens 2012; Gilens and Page 2014). Standard errors in parentheses. *=p<0.05

Online Supplement References

Achen, Christopher H. 1985. "Proxy Variables and Incorrect Signs on Regression Coefficients." *Political Methodology* 11(4):29–317.

Bartels, Larry M. 2023. Measuring Political Inequality. In *Unequal Democracies: Public Policy, Responsiveness, and Redistribution in an Era of Rising Economic Inequality*, ed. Noam Lupu and Jonas Pontusson. Cambridge: Cambridge University Press pp. 77–97.

Bashir, Omar S. 2015. "Testing Inferences about American Politics: A Review of the "Oligarchy" Result." *Research & Politics* 2(4):1–7.

Bhatti, Yosef and Robert S. Erikson. 2011. How Poorly are the Poor Represented in the US Senate? In *Who Gets Represented?*, ed. Peter K. Enns and Christopher Wlezien. New York: Russell Sage Foundation.

Branham, J. Alexander, Stuart Soroka and Christopher Wlezien. 2017. "When Do the Rich Win?" *Political Science Quarterly* 132(1):43–62.

Broockman, David and Neil Malhotra. 2020. "What Do Partisan Donors Want?" *Public Opinion Quarterly* 84(1).

Coppock, Alexander. 2022. *Persuasion in Parallel*. Chicago: University of Chicago Press.

Dalton, Russell J. 2013. *The Apartisan American*. Thousand Oaks, California: CQ Press.

Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York: Harper and Row.

Enns, Peter K. and Christopher Wlezien. 2011. Group Opinion and the Study of Representation. In *Who Gets Represented?*, ed. Peter K. Enns and Christopher Wlezien. New York: Russell Sage Foundation chapter 1, pp. 1–25.

Enns, Peter K. and Paul M. Kellstedt. 2008. "Policy Mood and Political Sophistication: Why Everybody Moves Mood." *British Journal of Political Science* 38(3):433–454.

Fiorina, Morris P. 2016. "Independents: The Marginal Members Of An Electoral Coalition." *Hoover Institition Essay On Contemporary American Politics* 6:1–16.

Flavin, Patrick and William W. Franko. 2017. "Government's Unequal Attentiveness to Citizens' Political Priorities." *Policy Studies Journal* 45(4):659–687.

Gilens, Martin. 2005. "Inequality and Democratic Responsiveness." *Public Opinion Quarterly* 69(5):778–796.

Gilens, Martin. 2012. *Affluence and Influence: Economic Inequality and Political Power in America*. Princeton, NJ: Princeton University Press and Russell Sage Foundation.

Gilens, Martin. 2016. "Simulating representation: The devil's in the detail." *Research & Politics* 3(2):1–7.

Gilens, Martin and Benjamin I. Page. 2014. "Testing Theories of American Politics: Elites, Interest Groups, and Average Citizens." *Perspectives Politics* 12(3):564–581.

Jessee, Stephen A. 2009. "Spatial Voting in the 2004 Presidential Election." *American Political Science Review* 103(1):59–81.

Jessee, Stephen A. 2012. *Ideology and Spatial Voting in American Elections*. New York: Cambridge University Press.

Kang, Shin-Goo and G. Bingham Powell, Jr. 2010. "Representation and Policy Responsiveness: The Median Voter, Eleciton Rules, and Redistributive Welfare Spending." *Journal of Politics* 72(4):1014–1028.

Kopkin, Nolan and Andrew Roberts. 2023. "Gender Gaps in Policy Responsiveness." *Political Research Quarterly* 76(3):1340–1356.

Lockwood, J.R. and Daniel F. McCaffrey. 2020. "Recommendations about estimating errors-in-variables regression in Stata." *The Stata Journal* 20(1):116–130.

Lupu, Noam and Alejandro Tirado Castro. 2023. "Unequal policy responsiveness in Spain." *Socio-Economic Review* 21(3):1697–1720.

Mathisen, Ruben B. 2024. "The Influence Gap: Unequal Policy Responsiveness to Men and Women." *Journal of Politics* 86(4).

Mathisen, Ruben, Wouter Schakel, Svenja Hense, Lea Elässer, Mikael Persson and Jonas Pontusson. 2023. Unequal Responsiveness and Government Partisanship in Northwest Europe. In *Unequal Democracies: Public Policy, Responsiveness, and Redistribution in an Era of Rising Economic Inequality*, ed. Noam Lupu and Jonas Pontusson. New York: Cambridge University Press chapter 2, pp. 29–53.

Page, Benjamin I., Larry M. Bartels and Jason Seawright. 2013. "Democracy and the Policy Preferences of Wealthy Americans." *Perspectives on Politics* 11(1):51–73.

Page, Benjamin I. and Martin Gilens. 2017. *Democracy in America?* Chicago: University of Chicago Press.

Page, Benjamin I. and Robert Y. Shapiro. 1992. *The Rational Public: Fifty Years of Trends in Americans' Policy Preferences*. Chicago: University of Chicago Press.

Peters, Yvette and Sander J. Ensink. 2014. "Differential Responsiveness in Europe: The Effects of Preference Difference and Electoral Participation." *West European Politics* 38(3):577–600.

Petrescu, Maria. 2013. "Marketing research using single-item indicators in structural equation models." *Journal of Marketing Analytics* 1(2):99–117.

Piketty, Thomas and Emmanuel Saez. 2003. "Income Inequality in the United States, 1913-1998." *Quarterly Journal of Economics* 118(1):1–39.

Pimentel, Samuel D., Lindsay C. Page, Matthew Lenard and Luke Keele. 2018. "Optimal Multilevel Matching Using Network Flows." *The Annals of Applied Statistics* 12(3):1479–1505.

Schakel, Wouter, Brian Burgoon and Armen Hakhverdian. 2020. "Real but Unequal Representation in Welfare State Reform." *Politics & Society* 48(1):131–163.

Shor, Boris and Jon C. Rogowski. 2018. "Ideology and the US Congressional Vote." *Political Science Research Methods* 6(2).

Wright, Gerald C. 1989. "Policy voting in the U.S. Senate: Who is Represented?" *Legislative Studies Quarterly* 14(4):465–486.