Comparative Case Designs

Jason Seawright

j-seawright@northwestern.edu

Jan. 22, 2024

Basics of Multi-Method Research

- Integrative multi-method designs
- Strengths and weaknesses of regression and case studies
- Detailed designs for combining case studies and regression
- Formal analysis of optimal case selection strategies

Vol. LXII.] [Part II.

JOURNAL

OF THE ROYAL STATISTICAL SOCIETY.

JUNE, 1899.

An Investigation *into the* Causes *of* Changes *in* Pauperism *in* England, *chiefly during the last* Two Intercensal Decades. (Part I.) *By* G. Udny Yule, *Assistant Professor of Applied Mathematics, University College, London.*

[Read before the Royal Statistical Society, 21st March, 1899.
Sir Robert Giffen, K.C.B., in the Chair.]

Table C.—*Table of Regression Equations of Pauperism on other Variables for all the Groups of Unions.*

1	2	3	4	5	6	7	8	9	10
							Standard Deviation.		
	Group.			Times Change in Out-Relief Ratio.	Times Change in Proportion of Old.	Times Change in Population.	(1) Round Regression Equation.	(2) Round Mean.	Per Cent. (1) on (2).
1871–81	Rural	Change per cent. in pauperism is equal to	− 27·07	+ 0·290	+ 0·271	+ 0·064	14·12	16·17	87
1871–81	Mixed	Change per cent. in pauperism is equal to	− 26·15	+ 0·282	+ 0·219	+ 0·085	15·56	17·19	91
1871–81	Urban	Change per cent. in pauperism is equal to	− 4·38	+ 0·571	− 0·094	− 0·067	20·70	25·33	82
1871–81	Metropolitan	Change per cent. in pauperism is equal to	+ 13·19	+ 0·755	− 0·022	− 0·322	10·16	16·23	63
1881–91	Rural	Change per cent. in pauperism is equal to	− 14·10	+ 0·243	+ 0·333	+ 0·178	16·38	19·46	84
1881–91	Mixed	Change per cent. in pauperism is equal to	− 11·14	+ 0·172	+ 0·470	− 0·187	16·30	18·04	90
1881–91	Urban	Change per cent. in pauperism is equal to	− 16·72	+ 0·344	+ 0·767	− 0·076	16·22	20·92	78
1881–91	Metropolitan	Change per cent. in pauperism is equal to	+ 1·36	+ 0·324	+ 1·37	− 0·369	22·86	29·16	78

Extending the Integrative Multi-Method Paradigm

- Comparative-Case Designs
- Natural experiments
- True experiments
- Case study-focused designs
- Conceptualization, measurement, and theory building

Experimental Designs

Matching

Method of Difference

$+$

Error

The Method of Difference

If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance in common save one, that one occurring only in the former; the circumstance in which alone the two instances differ is the effect, or the cause, or an indispensable part of the cause, of the phenomenon. (Mill 1843/2002)

The Method of Difference

Debunking the Method of Difference

The Potential Outcomes Framework

The Potential Outcomes Framework

Strictly speaking, for the Method of Difference to work based on a comparison between cases 1 and 2, the condition which must be met is:

$$Y_{1,t} = Y_{2,t} \tag{1}$$

$$Y_{1,c} = Y_{2,c} \tag{2}$$

Experiments

In a randomized experiment, it is true by the Law of Large Numbers that:

$$\frac{\sum_{i:D_i=t} Y_{i,t}}{\sum_{i:D_i=t} i} \approx \frac{\sum_{j:D_j=C} Y_{j,t}}{\sum_{j:D_j=c} j} \tag{3}$$

Experiments

In a randomized experiment, it is true by the Law of Large Numbers that:

$$\frac{\sum_{i:D_i=t} Y_{i,c}}{\sum_{i:D_i=t} i} \approx \frac{\sum_{j:D_j=c} Y_{j,c}}{\sum_{j:D_j=c} j} \tag{4}$$

Matching

Suppose, in an observational study, we somehow know that:

$$Y_{T,i} = f(\mathbf{X}_i) + \epsilon_i \tag{5}$$

$$Y_{C,i} = g(\mathbf{X}_i) + \delta_i \tag{6}$$

$$E(\epsilon|\mathbf{X}) = 0 \tag{7}$$

$$E(\delta|\mathbf{X}) = 0 \tag{8}$$

Matching

$$\begin{aligned}
E(Y_{T,i}|D_i = T, \mathbf{X}_i = \mathbf{W}) &= f(\mathbf{W}) + E(\epsilon_i) \\
&= f(\mathbf{W}) \\
&= E(Y_{T,i}|D_i = C, \\
&\quad \mathbf{X}_i = \mathbf{W})
\end{aligned}$$

Matching

$$
\begin{aligned}
E(Y_{T,i}|D_i = T, \mathbf{X}_i = \mathbf{W}) &= f(\mathbf{W}) + E(\epsilon_i) \\
&= f(\mathbf{W}) \\
&= E(Y_{T,i}|D_i = C, \\
&\quad \mathbf{X}_i = \mathbf{W})
\end{aligned}
$$

$$E(Y_{C,i}|D_i = T, \mathbf{X}_i = \mathbf{W}) = E(Y_{C,i}|D_i = C, \mathbf{X}_i = \mathbf{W})$$

Matching

Let:

$\tau_{\mathbf{W}} = E(Y_{T,i}|D_i = T, \mathbf{X}_i = \mathbf{W}) - E(Y_{C,i}|D_i = C, \mathbf{X}_i = \mathbf{W})$.

Matching

Let:
$\tau_{\mathbf{W}} = E(Y_{T,i}|D_i = T, \mathbf{X}_i = \mathbf{W}) - E(Y_{C,i}|D_i = C, \mathbf{X}_i = \mathbf{W})$.
Also let $\mathbf{W}$ have probability density function $w()$.

Matching

Let:
$\tau_{\mathbf{W}} = E(Y_{T,i}|D_i = T, \mathbf{X}_i = \mathbf{W}) - E(Y_{C,i}|D_i = C, \mathbf{X}_i = \mathbf{W})$.
Also let $\mathbf{W}$ have probability density function $w()$.
Then the average treatment effect of D on Y for the population is:

$$ATE = \int \tau_{\omega} w(\omega)\, d\omega \tag{9}$$

Matching as Second-Best

Matching is second-best to an experiment because:

1. We need to identify an adequate collection of matching variables, $\mathbf{X}$.

Matching as Second-Best

Matching is second-best to an experiment because:

1. We need to identify an adequate collection of matching variables, **X**.
2. We need stronger modeling assumptions, including assumptions about the existence of and conditional means of ϵ and δ.

Matching as Second-Best

Matching is second-best to an experiment because:

1. We need to identify an adequate collection of matching variables, $\mathbf{X}$.
2. We need stronger modeling assumptions, including assumptions about the existence of and conditional means of ϵ and δ.
3. For further discussion, see Freedman, Pisani, and Purves (2007), Ch. 1.

Matching as Second-Best

Matching is second-best to a fully-specified regression analysis because:

1. We lose the efficiency of using all cases to estimate a small number of parameters.

Matching as Second-Best

Matching is second-best to a fully-specified regression analysis because:

1. We lose the efficiency of using all cases to estimate a small number of parameters.
2. What is the S.E. of $\int \tau_\omega w(\omega)\, d\omega$?

Matching as Second-Best

Matching is second-best to a fully-specified regression analysis because:

1. We lose the efficiency of using all cases to estimate a small number of parameters.
2. What is the S.E. of $\int \tau_\omega w(\omega)\, d\omega$?
3. A well-specified regression sometimes comes closer to replicating experimental findings than do matching estimators (see, e.g., Peikes, Moreno, and Orzol 2008).

So, Why Match?

- Experiments aren't always available.

So, Why Match?

- Experiments aren't always available.
- Sometimes, we don't know $f()$ and $g()$.

So, Why Match?

- Experiments aren't always available.
- Sometimes, we don't know $f()$ and $g()$.
- Matching combines neatly with case studies.

Three Causal Quantities

1. ATE: $E(Y_T - Y_C)$

Three Causal Quantities

1. ATE: $E(Y_T - Y_C)$
2. ATT: $E(Y_T - Y_C | D = T)$

Three Causal Quantities

1. ATE: $E(Y_T - Y_C)$
2. ATT: $E(Y_T - Y_C | D = T)$
3. ATC: $E(Y_T - Y_C | D = C)$

Pairwise Matching

1. Take a sample of N_T treatment cases and N_C control cases.

Pairwise Matching

1. Take a sample of N_T treatment cases and N_C control cases.
2. For each treatment case, find the control case that "best" matches on $\mathbf{X}$. Save the difference on Y between those two cases

Pairwise Matching

1. Take a sample of N_T treatment cases and N_C control cases.
2. For each treatment case, find the control case that "best" matches on $\mathbf{X}$. Save the difference on Y between those two cases
3. Average the resulting paired differences. Use this as an estimate of ATT.

Propensity Score

Definition

The *propensity score* for case i is the probability that $D_i = T$ conditional on $\mathbf{X}_i$.

Propensity Score

Definition

The *propensity score* for case i is the probability that $D_i = T$ conditional on $\mathbf{X}_i$.

A well-estimated propensity score contains all the information about $\mathbf{X}_i$ that is relevant to causal inference.

Application: Female Legislators

- It is widely believed that proportional representation electoral rules produce more female legislators than do single-member district rules (e.g., Kenworthy and Malami 1999, Matland 1998, Norris 2004, Paxton and Kunovich 2003, Reynolds 1999, Rule 1987, Siaroff 2000).
- Several regression analyses have found effects in the neighborhood of 7-12%.

Application: Female Legislators

- These models have assumed that there is a single, constant effect of electoral rules on women's legislative inclusion. Is that assumption plausible?

Application: Female Legislators

- These models have assumed that there is a single, constant effect of electoral rules on women's legislative inclusion. Is that assumption plausible?
- For example, is the ATT equal to the ATC?

Application: Female Legislators

Table 5. Matching Estimates of the Effect of Electoral Institutions on Women's Legislative Representation.

Conditioning variables	Persson and Tabellini		New	
Sample	All	Quota cases excluded	All	Quota cases excluded
ATET	–0.08 (2.66)	2.00 (2.34)	–1.46 (4.77)	0.83 (4.45)
ATEC	–7.27* (3.75)	–7.02* (3.82)	–8.62* (4.35)	–6.72 (4.89)
ATE	–4.84 (3.04)	–3.88 (3.17)	–6.50 (4.19)	–4.42 (4.36)
Treatment cases	23	16	13	11
Control cases	45	30	31	25

ATE = average treatment effect; ATET = average treatment effect for the treated cases; ATEC = average treatment effect for the control cases. Standard error of effect estimate in parentheses.
*Significant at $p < .10$.

Matching for Case Selection

Matching for Case Selection

- Selecting cases by matching can be used for data preprocessing to reduce model dependence (Ho, Imai, King, and Stewart 2007).

Matching for Case Selection

- Selecting cases by matching can be used for data preprocessing to reduce model dependence (Ho, Imai, King, and Stewart 2007).
- Matching can be used to select negative cases for statistical analysis.

Application: Female Legislators

- If electoral rules cause female legislative inclusion, then a change in rules should lead to a change in the number of female legislators.

Application: Female Legislators

- If electoral rules cause female legislative inclusion, then a change in rules should lead to a change in the number of female legislators.
- The worldwide trend toward electing more female legislators is an obstacle to inference.

Application: Female Legislators

- If electoral rules cause female legislative inclusion, then a change in rules should lead to a change in the number of female legislators.
- The worldwide trend toward electing more female legislators is an obstacle to inference.
- No-change cases are needed to allow difference-in-differences analysis.

Application: Female Legislators

1. Collect data on the election before and after every change in electoral rules since 1950.

Application: Female Legislators

1. Collect data on the election before and after every change in electoral rules since 1950.
2. Collect data on every other democratic election since 1950.

Application: Female Legislators

1. Collect data on the election before and after every change in electoral rules since 1950.
2. Collect data on every other democratic election since 1950.
3. Code cases with a rule change as treatment cases, and cases without as control cases.

Application: Female Legislators

1. Collect data on the election before and after every change in electoral rules since 1950.
2. Collect data on every other democratic election since 1950.
3. Code cases with a rule change as treatment cases, and cases without as control cases.
4. Carry out propensity-score matching, and run the model on the cases in the matches.

Application: Female Legislators

Parameter	Estimate (S.E.)
Intercept	1.461** (0.587)
More Restrictive	0.041 (1.192)
Less Restrictive	0.828 (1.128)
Quota	-0.413 (1.024)
R^2	0.008
N	103

Conclusions

- Matching is rarely the ideal research design.

Conclusions

- Matching is rarely the ideal research design.
- However, for many research questions, matching allows us to proceed in spite of being unable to execute the best research design.

Comparing Cases

Table 1: Matched ATT Country Pairs, with Treatment Effects: Majoritarianism

Treatment		Control		
Country	cgexp	Country	cgexp	Effect
UK	40.4	Romania	33.4	6.98
France	45.3	Spain	36.2	9.06
Japan	20.5	Hungary	50.0	−29.44
Chile	21.0	Luxembourg	40.2	−19.15
Thailand	16.2	El Salvador	13.6	2.59
USA	21.9	Venezuela	19.3	2.53
Nepal	17.0	South Korea	17.1	−0.10
Bangladesh	12.6	South Korea	17.1	−4.50
Philippines	18.9	South Korea	17.1	1.82
Barbados	32.4	Namibia	37.3	−4.91
New Zealand	36.0	Ireland	38.1	−2.11
Canada	24.9	Ireland	38.1	−13.24
Singapore	18.5	Israel	46.3	−27.84
Trinidad&Tob	28.1	Sri Lanka	27.4	0.69
Australia	25.8	South Africa	31.3	−5.56
Bahamas	18.8	Malta	41.0	−22.17
Pakistan	23.1	Malta	41.0	−17.87
Uganda	14.7	Malta	41.0	−26.33
Gambia	24.4	Fiji	28.4	−4.00
Ghana	19.0	Fiji	28.4	−9.43
Zimbabwe	31.2	Fiji	28.4	2.81
Belize	30.1	Fiji	28.4	1.64
Averages	24.6		31.8	−7.21

Case Studies and Matching

Choosing Cases for In-Depth Study:

Case Studies and Matching

Choosing Cases for In-Depth Study:

1. Paired cases that make a very large contribution to the effect estimate

Case Studies and Matching

Choosing Cases for In-Depth Study:

1. Paired cases that make a very large contribution to the effect estimate
2. The highest propensity-score treatment case, and the lowest propensity-score control case

Case Studies and Matching

Choosing Cases for In-Depth Study:

1. Paired cases that make a very large contribution to the effect estimate
2. The highest propensity-score treatment case, and the lowest propensity-score control case
3. Extreme cases on Y

```
> ptatt.table
```

	treat.name	control.name	pairwise.att	treat.pscore	control.pscore
1	Australia	South Africa	-5.55657768	0.73547225	0.75738883
2	Bahamas	Malta	-22.17235756	0.76358562	0.75987301
3	Bangladesh	South Korea	-4.50226593	0.37124985	0.35474055
4	Belize	Fiji	1.64207077	0.81242818	0.82809120
5	Canada	Ireland	-13.24016571	0.64061184	0.61717006
6	Chile	Luxembourg	-19.15097046	0.12771933	0.12711098
7	France	Spain	9.05746841	0.09294654	0.09491685
8	Japan	Hungary	-29.44489288	0.10852921	0.10596559
9	Malawi	Fiji	-2.66178894	0.83150095	0.82809120
10	Malaysia	Fiji	-3.95250702	0.85734203	0.82809120
11	Nepal	South Korea	-0.09674072	0.33670234	0.35474055
12	Pakistan	Malta	-17.86938095	0.78094408	0.75987301
13	Philippines	South Korea	1.82348061	0.37667525	0.35474055
14	Singapore	Israel	-27.83652687	0.65895826	0.67359923
15	St. Vincent&G	Fiji	6.35684204	0.85613601	0.82809120
16	Thailand	El Salvador	2.59179211	0.26372812	0.25855176
17	Trinidad&Tob	Sri Lanka	0.68888092	0.69379038	0.67443402
18	USA	Venezuela	2.53207016	0.29672286	0.29195428
19	Uk	Italy	-8.43054199	0.07802993	0.07677351
20	Zambia	Fiji	-3.18745232	0.85634526	0.82809120
21	Zimbabwe	Fiji	2.80818176	0.80837238	0.82809120

Electoral system [edit]

In all elections where there is a single official to be elected for a given area, including the two major national elections (the election of the President of the Republic and the election of the members of the National Assembly), two-round runoff voting is used.

For elections to the European Parliament and some local elections, proportional voting is used.

e • d

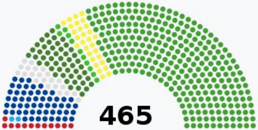
Summary of the 23 April and 7 May 2017 French presidential election results

Candidate	Party		1st round		2nd round	
			Votes	%	Votes	%
Emmanuel Macron	**En Marche!**	**EM**	**8,656,346**	**24.01**	**20,743,128**	**66.10**
Marine Le Pen	National Front	FN	**7,678,491**	**21.30**	10,638,475	33.90
François Fillon	The Republicans	LR	7,212,995	20.01		
Jean-Luc Mélenchon	La France Insoumise	FI	7,059,951	19.58		
Benoît Hamon	Socialist Party	PS	2,291,288	6.36		
Nicolas Dupont-Aignan	Debout la France	DLF	1,695,000	4.70		
Jean Lassalle	Résistons!		435,301	1.21		
Philippe Poutou	New Anticapitalist Party	NPA	394,505	1.09		
François Asselineau	Popular Republican Union	UPR	332,547	0.92		
Nathalie Arthaud	Lutte Ouvrière	LO	232,384	0.64		
Jacques Cheminade	Solidarity and Progress	S&P	65,586	0.18		
Total			**36,054,394**	**100.00**	**31,381,603**	**100.00**
Valid votes			36,054,394	97.43	31,381,603	88.48
Blank ballots			659,997	1.78	3,021,499	8.52
Null ballots			289,337	0.78	1,064,225	3.00
Turnout			37,003,728	77.77	35,467,327	74.56
Abstentions			10,578,455	22.23	12,101,366	25.44
Registered voters			47,582,183		47,568,693	

Official results published by the Constitutional Council – 1st round result · 2nd round result

The *House of Councillors* (*Sangi-in*) has 242 members, elected for a six-year term, 146 members in 47 single- and multi-seat constituencies (prefectures) by single non-transferable vote and 96 by proportional representation (by D'Hondt method) on the national level. The proportional election to the House of Councillors allows the voters to cast a preference vote for a single candidate on a party list. The preference votes exclusively determine the ranking of candidates on party lists. Half of the House of Councillors comes up for election every three years in regular/ordinary elections of members of the House of Councillors (*Sangiin giin tsūjō-senkyo*).

← Summary of the 22 October 2017 House of Representatives election results →



Parties	Constituency				PR Block				Total seats			
	Votes	%	±pp	Seats	Votes	%	±pp	Seats	Seats	±	%	±pp
Liberal Democratic Party (LDP)	26,719,032	48.21	▲0.11	218	18,555,717	33.28	▲0.17	66	284	▼6	61.08	▲0.02
Komeitō (NKP)	832,453	1.50	▲0.05	8	6,977,712	12.51	▼1.20	21	29	▼5	6.24	▼0.92
Governing coalition	**27,551,485**	**49.71**	**▲0.17**	**226**	**25,533,429**	**45.79**	**▼1.03**	**87**	**313**	**▼11**	**67.31**	**▼0.90**
Constitutional Democratic Party of Japan (CDP)	4,852,097	8.75	*New*	18	11,084,890	19.88	*New*	37	55	▲40	11.83	▲6.66
Japanese Communist Party (JCP)	4,998,932	9.02	▼4.28	1	4,404,081	7.90	▼3.47	11	12	▼9	2.58	▼1.84
Social Democratic Party (SDP)	634,719	1.15	▲0.36	1	941,324	1.69	▼0.77	1	2	–0	0.43	▲0.01
Pacifist coalition	**10,485,748**	**18.92**	**–**	**20**	**16,430,295**	**29.47**	**–**	**49**	**69**	**▲31**	**14.84**	**▲6.84**
Kibō no Tō (Party of Hope)	11,437,601	20.64	*New*	18	9,677,524	17.36	*New*	32	50	▼7	10.75	▼1.25
Nippon Ishin no Kai (JIP)	1,765,053	3.18	▼4.98	3	3,387,097	6.07	▼9.65	8	11	▼3	2.37	▼0.58
Koike coalition	**13,202,654**	**23.82**	**–**	**21**	**13,064,621**	**23.43**	**–**	**40**	**61**	**▼10**	**13.12**	**▼1.83**
Happiness Realization Party (HRP)	159,171	0.29	–	0	292,084	0.52	▲0.03	0	0	–0	0.00	–0.00
New Party Daichi	–	–	–	–	226,552	0.41	–	0	0	–0	0.00	–0.00
No Party to Support	–	–	–	–	125,019	0.22	▲0.02	0	0	–0	0.00	–0.00
Party for Japanese Kokoro (PJK)	–	–	–	–	85,552	0.15	▼2.50	0	0	–0	0.00	–0.00
Others	52,080	0.03	–	0	–	–	–	–	0	–0	0.00	–0.00
Independents	3,970,946	7.16	▲4.31	22	–	–	–	–	22	▼17	4.73	▼3.48
Total	**55,422,087**	**100.00**	**–**	**289**	**55,757,552**	**100.00**	**–**	**176**	**465**	**▼10**	**100.00**	**–**

Today the voting system for the Fijian House of Representatives is used to elect 50[1] members under a single national constituency. Fiji used the first past the post system for most of its history, but the new constitution in 1997–1998 agreed to replace it with the alternative vote (AV) system, allowing votes to be transferred from a low-polling candidate to other candidates, according to an order prescribed by the candidate, which may be customised by the elector.

AV allows voters to rank candidates in the order of their preference, with votes for low-polling candidates transferred to higher-polling candidates. Candidates who receive a minimum of 50 percent of the total vote in their respective constituencies are declared elected. If no candidate receives 50 percent, votes cast for low-polling candidates are transferred to higher-polling candidates, beginning by "eliminating" the lowest-polling candidate and continuing until one candidate has 50 percent or more of the vote.

The variant of AV chosen was taken from the Australian electoral system used for the Australian Senate where voters can opt to vote "above the line", accepting a party's prespecified preference order (as also used for the New York City Council). This system allows parties to pre-specify electoral alliances and is akin to the use of apparentment, linked party lists, in party-list proportional representation systems. Voters who disagree with the way their preferred candidate has arranged to transfer his or her votes if eliminated may opt to vote "below the line" of the ballot paper instead. Here, electors may rank all candidates in the order of their preference.

Party	Votes	%	Seats	+/–
FijiFirst	227,241	50.02	27	–5
Social Democratic Liberal Party	181,072	39.85	21	+6
National Federation Party	33,515	7.38	3	0
Unity Fiji Party	6,896	1.52	0	New
HOPE	2,811	0.62	0	New
Fiji Labour Party	2,800	0.62	0	0
Invalid/blank votes	4,179	–	–	–
Total	**458,335**	**100**	**51**	**+1**
Registered voters/turnout	637,527	71.27	–	–
Source: Fiji Village				

- Voting system: Proportional, Single-transferable-vote (STV). Each elector indicates his/her order of preference among all the candidates in his/her electoral district regardless of candidates' political affiliation. In the first count, those who satisfy the Hagenbach-Bischoff quotient are declared elected. Should any seats remain vacant, the surplus votes polled by candidates already elected are transferred proportionately to the remaining candidates on the basis of the second preferences indicated. The votes thus transferred are added to those polled by each remaining candidate. The candidate (candidates) who now possesses (possess) a number of votes equal to, or greater than the quotient is (are) elected. Candidates with the lowest number of votes are eliminated and their votes are transferred to the other remaining candidates according to the next preference shown on the ballot paper. The same operation is repeated until there are no more seats to be filled. If necessary, "bonus seats" are allocated to any party receiving the highest percentage of votes under the first-count votes to ensure that it secure a majority of seats in Parliament. The bonus seats are given to the remaining unelected candidates of the winning party irrespective of the district contested.

Party	Votes	%	Seats	+/–
Labour Party	170,976	55.04	37	–2
Forza Nazzjonali (PN–PD)[b]	135,696	43.68	30	0
Democratic Alternative	2,564	0.80	0	0
Moviment Patrijotti Maltin	1,117	0.36	0	New
Alleanza Bidla	221	0.07	0	New
Independents	91	0.03	0	0
Invalid/blank votes	4,031	–	–	–
Total	**314,696**	**100**	**67**	**–2**
Registered voters/turnout	341,856	92.06	–	–
Source: Electoral Commission				

Case Studies and Matching

Search for:

- Confounding variables

Case Studies and Matching

Search for:

- Confounding variables
- Measurement error in the treatment variable

Case Studies and Matching

Search for:

- Confounding variables
- Measurement error in the treatment variable
- Causal pathways

Manufacture Perfect Comparisons

Strictly speaking, for the Method of Difference to work based on a comparison between cases 1 and 2, the condition which must be met is:

$$Y_{1,t} = Y_{2,t} \tag{10}$$

$$Y_{1,c} = Y_{2,c} \tag{11}$$

Manufacture Perfect Comparisons

No observable condition can ever guarantee that this assumption is met, but if we were to find two cases that *exactly* match on a suitably rich set of background variables $\mathbb{X}$, perhaps, we would believe the assumption.

Manufacture Perfect Comparisons

Unfortunately, if $\mathbb{X}$ is indeed a reasonably deep list of variables, we are unlikely to find cases that in fact exactly match (or even come particularly close).

Manufacture Perfect Comparisons

Abadie and collaborators suggest that we create our own "synthetic" control cases, by averaging together existing control cases to come as close as possible to exactly matching the treatment case on $\mathbb{X}$.

Synthetic Control

The setup is one in which there are N cases, each of which is observed at multiple time periods labeled from 1 through T.

Synthetic Control

Each case has a treatment and control potential outcome for each time period. The difference between these is:

$\alpha_{i,t} = Y_{i,t}^{T} - Y_{i,t}^{C}$

Synthetic Control

Suppose that the treatment of interest happens in one case at one time period. That is, $D_{j,t} = 0$ for all $j \neq i$ and for all $t < t_{treat}$.

Synthetic Control

$$X_i - \mathbb{X}_j W$$

Synthetic Control

$$\sqrt{(X_i - \mathbb{X}_j W)^T V (X_i - \mathbb{X}_j W)}$$

Synthetic Control

Following Abadie and Gardeazabal (2003), in the empirical section of this article we choose **V** among positive definite and diagonal matrices such that the mean squared prediction error of the outcome variable is minimized for the preintervention periods (see Abadie and Gardeazabal 2003, appendix B, for details). Alternatively, if the number of available preintervention periods in the sample is large enough, researchers may divide them into an initial training period and a subsequent validation period. Given a **V**, $\mathbf{W}^*(\mathbf{V})$ can be computed using data from the training period. Then, the matrix **V** can be chosen to minimize the mean squared prediction error produced by the weights $\mathbf{W}^*(\mathbf{V})$ during the validation period.

Synthetic Control

Journal of Economic Behavior & Organization 125 (2016) 1–21

Contents lists available at ScienceDirect

Journal of Economic Behavior & Organization

journal homepage: www.elsevier.com/locate/jebo

ELSEVIER

The economic consequences of Hugo Chavez: A synthetic control analysis

CrossMark

Kevin Grier[a,*], Norman Maynard[b]

[a] Department of Economics, University of Oklahoma, United States
[b] Department of Economics, College of Charleston, United States

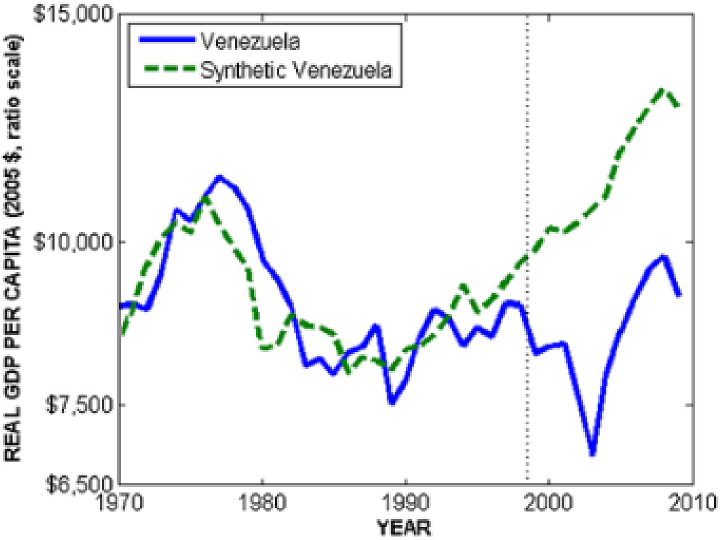
Synthetic Control

Table 2
Estimated synthetic control weights for each outcome variable.

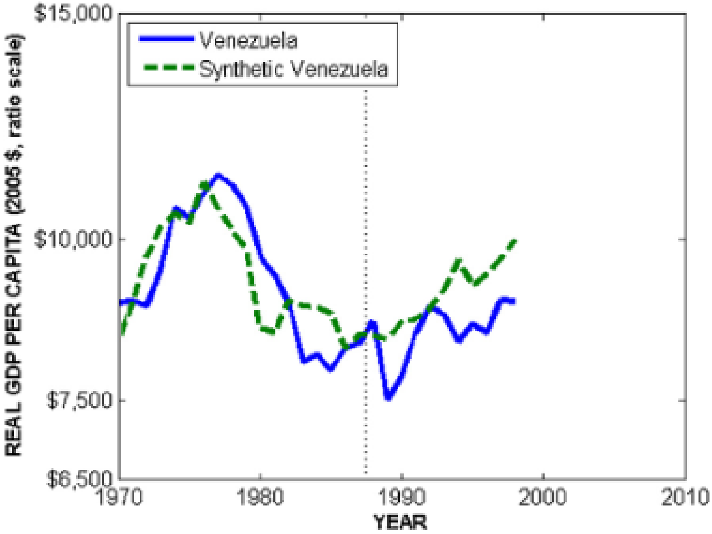
	Outcome variables				
	Income	Infant Mort	Life Exp	Poverty	Inequality
Algeria	0.00	0.00	0.00	–	–
Argentina	0.00	0.00	0.99	20.38	0.40
Brazil	7.25	0.00	0.00	0.00	0.00
Canada	20.35	0.00	0.00	–	0.00
Chile	0.00	0.00	0.00	0.00	0.00
Colombia	0.00	0.00	0.00	79.62	6.85
Costa Rica	0.00	0.00	0.00	0.00	0.00
El Salvador	0.00	10.00	6.27	0.00	0.00
Guatemala	0.00	2.25	0.00	0.00	–
Honduras	0.00	0.00	0.00	0.00	–
Indonesia	0.00	–	–	–	0.00
Iran	42.24	0.00	0.00	–	–
Iraq	0.00	–	–	–	–
Mexico	12.67	0.00	0.00	0.00	0.00
Nigeria	–	3.10	8.46	–	0.01
Norway	0.00	33.98	18.86	–	5.19
Panama	0.00	41.69	44.53	0.00	0.00
Paraguay	0.00	–	–	0.00	–
Peru	17.49	0.00	0.00	0.00	0.00
Uruguay	0.00	8.97	20.89	0.00	87.55
Model fit pre-intervention					
RMSPE	0.068	0.32	0.16	6.23	1.48
APE-to-mean ratio	0.35%	0.28%	0.002%	2.16%	0.17%
SCM inference: permutation test					
RMSPE ratio	4.27	1.36	4.76	0.80	2.44
p-value: RMSPE	0.00	0.35	0.12	0.79	0.21

Note: Columns show the weight assigned to each country in the synthetic controls for Venezuela. Each column includes a synthetic control for a different outcome variable. A dash (–) indicates that the country is not available in the dataset for the given comparison. Weights are in percentage points. Rounding errors may prevent columns from summing to 100. APE-to-mean ratio indicates the average pre-intervention prediction error divided by the average pre-intervention outcome value.

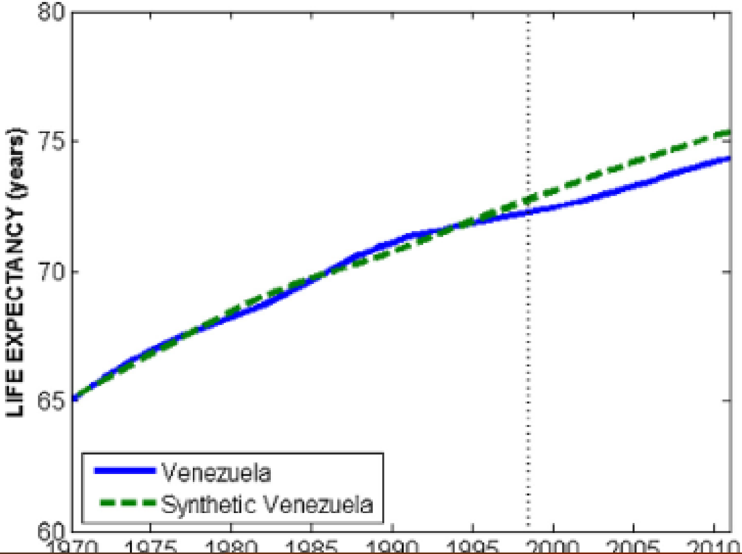
Synthetic Control



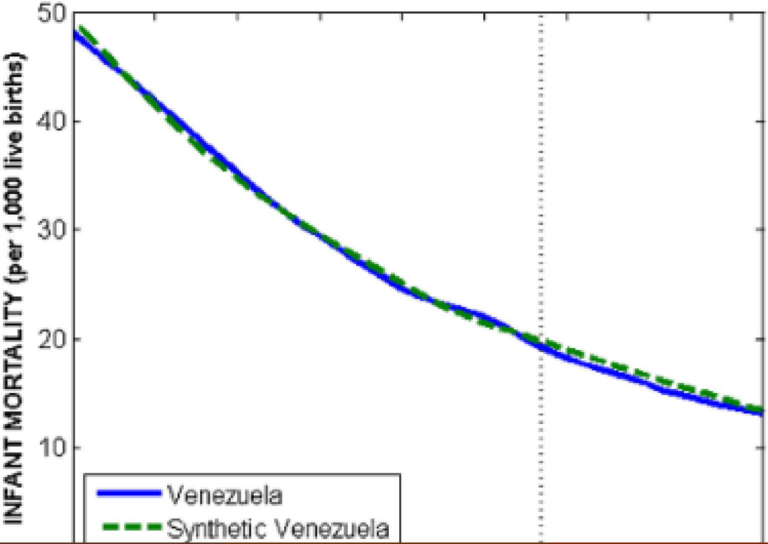
Synthetic Control



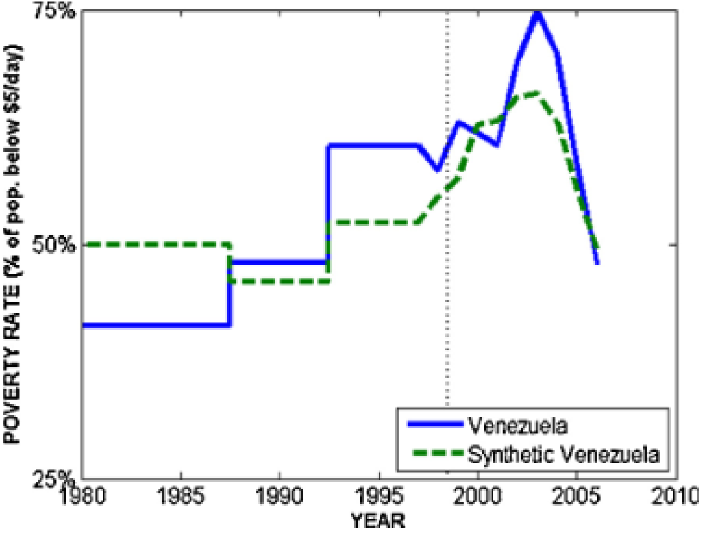
Synthetic Control



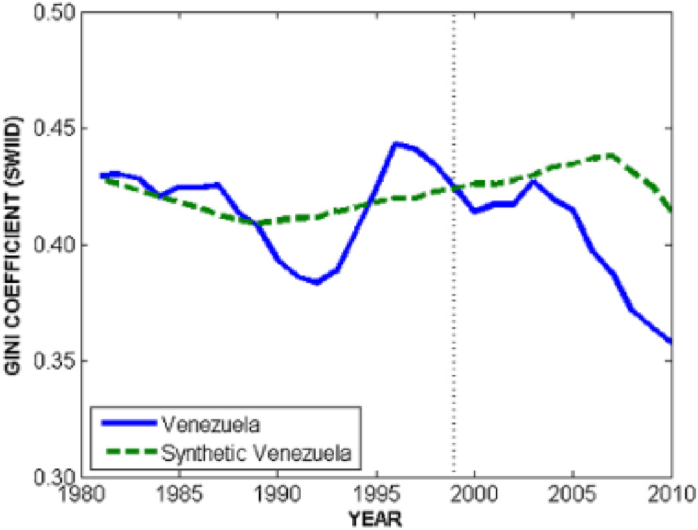
Synthetic Control



Synthetic Control



Synthetic Control



Synthetic Control

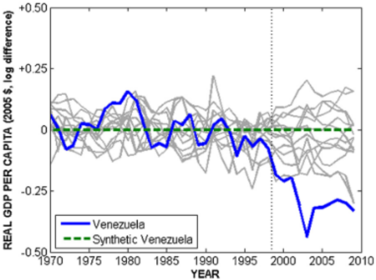


Fig. 4. Per-capita income Placebo Tests, restricted countries. *Note*: The bold line represents the difference between observed (log) income per-capita in Venezuela, 1970–2009, and the synthetic control; the synthetic control (dashed line) is normalized to zero. Gray lines represent placebo tests: deviations from synthetic control for the other countries in the dataset. This graph only include countries with pre-intervention root mean squared prediction error (RMSPE) less than 0.1025 (1.5 times that of Venezuela). It drops Chile, Indonesia, Iran, Iraq, Norway, and Panama.

Synthetic Control

Synth library