Political Science Research and Methods (2020), **8**, 589–596
doi:10.1017/psrm.2019.36

RESEARCH NOTE

Social democratic representation and welfare spending: a quantitative case study

Henning Finseraas

Institute for Social Research, P.box 3233 Elisenberg, 0208 Oslo, Norway
Corresponding author. E-mail: henning.finseraas@samfunnsforskning.no

(Received 3 October 2017; revised 13 March 2018; accepted 22 March 2018; first published online 18 June 2019)

Abstract
The welfare state literature argues that Social Democratic party representation is of key importance for welfare state outcomes. However, few papers are able to separate the influence of parties from voter preferences, which implies that the partisan effects will be overstated. I study a natural experiment to identify a partisan effect. In 1995, the Labour Party (Ap) in the Norwegian municipality of Flå filed their candidate list too late and could not participate in the local election. Ap was the largest party in Flå in the entire post-World War period, but have not regained this position. I use the synthetic control method to study the effects on welfare spending priorities. I find small and insignificant partisan effects.

Keywords: Partisanship; Welfare spending; Political economy; Synthetic Control

1. Introduction

Do political parties matter for policy outcomes? The modern literature on this classic topic goes back at least to Hibbs (1977), who argued that class differences in preferences for unemployment and inflation policies are reflected in the behavior of left and right parties in government. The view that partisanship matters has been particularly influential in the welfare state literature, where the role of Social Democratic parties in welfare state development has been emphasized (Esping-Andersen, 1990; Korpi and Palme, 2003). Recently, the literature has taken a new direction to the study of partisan effects at the local level (Fiva *et al.*, 2018).

The partisanship literature is often unclear in its definition of partisan effects. Most studies follow in the spirit of Hibbs (1977) in that they point to class differences in economic preferences and parties' electoral constituencies as the basis for partisan effects on public policy. This understanding of partisan effects risks conflating voter effects and partisan effects (see Lee *et al.*, 2004, and the extended discussion in the online appendix). Following the political economy literature (Besley and Coate, 1997), partisan effects refer to effects of who governs, controlling for voter preferences. To separate voter and partisan effects is an empirical challenge. In essence, the challenge is to estimate the effect of partisanship while controlling for voter preferences. Unfortunately, we do not have good measures of voter preferences across different policy outcomes. Thus, one needs exogenous variation in partisanship to estimate a credible partisan effect (Lee *et al.*, 2004).

I leverage a natural experiment to get exogenous variation in partisanship. In 1995, the Labour party (Ap) in the Norwegian municipality Flå failed to file their candidate list in time to participate in the local election, implying that they had no representation in the following election period. I study the effects of this shock on four spending outcomes using the synthetic control method (Abadie *et al.*, 2010). This method is particularly suited for quantitative case studies and has been labeled "the most important innovation in the policy evaluation literature in the last 15 years" (Athey and Imbens, 2017, 9). I follow a number of recent recommendations of how to apply the synthetic control method to avoid potential pitfalls. The paper should therefore be of interest for researchers who want to apply the synthetic control method.

The overall conclusion is that the welfare spending consequences of the decline of Ap are small and insignificant. In the conclusion I relate the findings to the previous literature on partisan effects.

2. The decline of Flå Ap

Norwegian municipalities are governed by municipal councils elected in a proportional electoral system with a single electoral district. The municipal council elects an executive municipal board with proportional representation of the elected parties. One member of the executive board is elected as the major. The mayor chairs the council meetings and has representational obligations, but is granted limited legal and formal power by law. Local elections are important, as municipalities are responsible for local infrastructure and the provision of the key welfare services, including childcare, primary education, elderlycare, and social assistance.

Flå is located in Buskerud county, 2–3 hours drive north-west of Oslo. It is a rural municipality with 1081 inhabitants (2017). Ap dominated local politics in Flå in the post World War period: They had more than 50 percent vote share in all elections until 1983 (see Figure 1), had the major until 1988, and had the highest vote share in all elections prior to 1995. It was therefore a major event when Ap were not allowed to run for election in 1995. The election law demanded that parties had to register their list of candidates by June 1, 1995 in order to participate in the September election. Unfortunately, the party leader mistakenly believed that the deadline was June 25, so he failed to deliver the list on time. I argue that this personal mistake could have happened in any other small municipality where the party organization is run by voluntary work.[1]

Since Ap did not run, they had no representation in the 1995–1999 election period. The majority of Ap sympathizers stayed at home in 1995, as turnout fell from 79 percent in 1991 to 43 percent in 1995. The votes were split between Bygdelista (B) and Borgerlig Fellesliste (BF). B, Ap's main competitor, is a local list with no official ties to any of the national parties. BF is a center-right list of candidates from the Conservative Party, the Christian Democratic Party, and the Center Party. No parties to the left of Ap had any representation, thus the absence of Ap implied a strong rightward shift of the municipal council.[2] At the time when Ap was disqualified, B and BF had filed their candidate lists and decided their party programs, thus they could not credibly change their policies in response to Ap's exit from the election.

In the 1999 election Ap received 31 percent of the vote share, which (except for 1995) was the worst election result in the post-war period. B manifested its position as the major party. Ap's vote share fell to 25 percent in 2003, while in 2007 they failed to get enough candidates to run its own list and filed a joint list with the Center Party and the Christian People's Party. Thus, the 1995 error had long-run repercussions. Figure 1 displays the vote shares for Ap in Flå and the rest of Norway in local and national elections, 1972–2011. As evident, Ap's vote share in local elections was stable in the rest of Norway from 1995 and onwards. Importantly, we see that nothing dramatic happened to Ap's vote share in Flå in the national elections in 1993 and 1997, which is consistent with stable voter preferences in Flå around 1995. Thus, we have a shock in representation which is unrelated to voter preferences.

A key premise for the existence of partisan effects is ideological polarization. It is not obvious that ideological polarization exists in small municipalities, since citizens' legislated rights to welfare services have been strengthened over the last decades (Østerud and Selle, 2006). Moreover,

[1]One possibility is that the mistake is a signal of a local party organization in disarray. I have not come across any evidence indicating that this was the case. See the online Appendix B for further discussion.

[2]See the online Appendix B for an approach to quantify the size of the shift. I find that the left-right shift in Flå from 1991 to 1995 is the most extreme shift compared to the shifts in the control municipalities. The shift is also large in comparison with the shifts from 1987 to 1991.

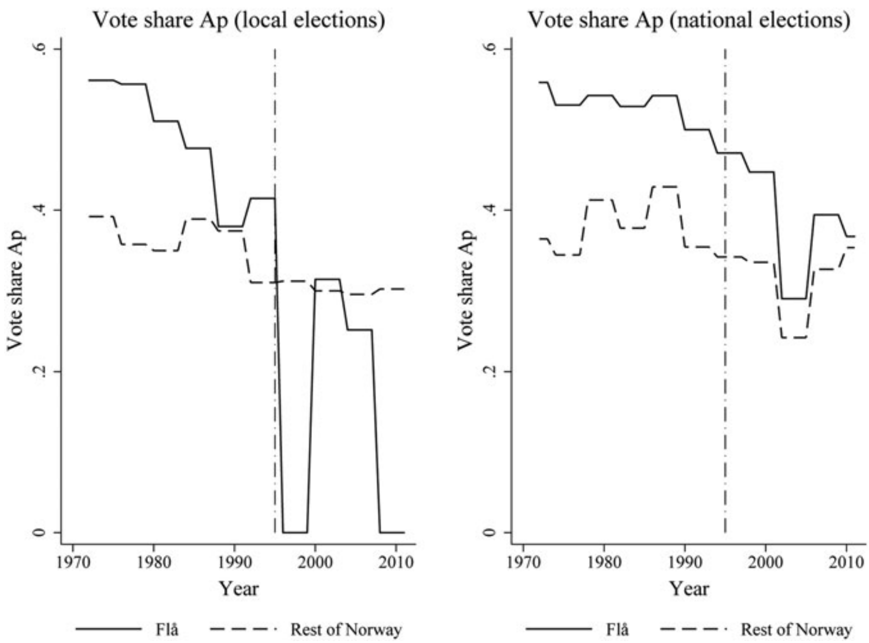

Figure 1. Ap's vote share in Flå and in the rest of Norway 1972–2011.

the importance of individual characteristics of local politicians has been emphasized in the literature (Munkerud, 2007; Hyytinen *et al.*, 2018). Still, Fiva *et al.* (2018) present results from surveys of local council members 1999–2007 which show left-right-divides on tax and spending preferences. In particular, left parties prioritize spending on children, while right parties prioritize the elderly. Although these results are averages across municipalities and might not be perfectly accurate for Flå, Fiva et al. (2018: 13) find that politicians' spending preferences vary "only moderately" across time and space. Moreover, survey data indicate that voters in small municipalities believe that local politics matter (see online Appendix B).

3. Empirical strategy

The empirical challenge is to estimate the counterfactual development of welfare spending in Flå. Two strategies dominate in the literature; the difference-in-difference (DD) approach and the synthetic control (SC) method. In this note I apply the SC method.

The main benefits of SC compared to DD are a data-driven selection of comparison cases and a less restrictive assumption regarding time-varying confounders. While the DD-approach gives equal weight to each unit in the control group, the SC method allows the units to have different weights in order to construct a synthetic control group that matches the pre-treatment outcomes of the treated unit. The idea is that similar outcomes in the pre-treatment period make it more plausible that the post-treatment outcomes of the synthetic control is a good counterfactual to the development of the treated unit. The SC modeling of the outcomes can be expressed by the following equation (Abadie *et al.*, 2010):

$$Y_{it} = \alpha_{it}\text{FLA}_{it} + Y_{it}^{C} = \alpha_{it}\text{FLA}_{it} + \delta_t + \boldsymbol{\theta}_t\boldsymbol{Z}_i + \boldsymbol{\lambda}_t\boldsymbol{\mu}_i + \epsilon_{it}$$

where Y_{it}^{C} is the counterfactual outcome and α_{it} is the treatment effect of Flå Ap's decline in period t. The counterfactual outcome is constructed from time fixed effects δ_t, a set of variables

https://doi.org/10.1017/psrm.2019.36 Published online by Cambridge University Press

$\boldsymbol{Z_i}$ which are not affected by the treatment, with $\boldsymbol{\theta_t}$ as a set of unknown associated parameters. Finally, $\boldsymbol{\lambda_t}$ is a set of time-varying unobserved variables with associated unknown and unit-varying factor loadings $\boldsymbol{\mu_i}$. By constructing a set of weights across the control units, the SC method matches the pre-treatment development of the treated unit with a weighted combination of the control units. Following Abadie *et al.* (2010, 495), if the weighted control fits the pre-treatment development of the treated unit, then it is plausible that it will approximately fit the unobserved confounders as well.

A number of issues arises in empirical applications of the SC method. The first regards the restriction of potential contributors to the synthetic control. Abadie *et al.* (2010) and Abadie *et al.* (2015) recommend that the donor pool of potential control units is restricted to units with similar pre-period characteristics as the treated unit. This is a useful restriction because the linear combination of the control units to match Flå might involve a lot of interpolation, and hence bias, if donor pool units are very different (Abadie *et al.*, 2010, 495). Since the administrative error is unlikely to have happened in a large municipality, I restrict the donor pool to the 164 municipalities who, like Flå, are classified as "least central" in the 1994 version of Statistics Norway's index of centrality.[3]

The second issue regards what variables to use as inputs. There are two decisions to be made; (i) how many pre-treatment outcomes to include and (ii) what covariates to include. The first is the most important issue since the pre-treatment outcomes have the strongest predictive power (Doudchenko and Imbens, 2016). Abadie *et al.* (2010) do not offer much guidance on this issue, and the lack of guidance can result in considerable researcher degrees of freedom to pick the specification that yields the preferred conclusions (Ferman *et al.*, 2017). To reduce this worry I estimate a set of different specifications and use decision rules to pick specifications for further analyses (see Dube and Zipperer, 2015; Ferman *et al.*, 2017).

Specifically, I estimate ten different SC specifications, following Ferman *et al.* (2017).[4] The specifications are listed in the note to Table 1. For each of the ten specifications I first run the model with Flå as the treated unit and then with each of the control units in the donor pool as the (placebo) treated unit. To assess how well the synthetic control fit the trend in Flå, I calculate (Ferman *et al.*, 2017) normalized mean squared error index $\tilde{R}^2$:

$$\tilde{R}^2 = 1 - \frac{\sum_{t=1}^{T_0} (Y_{\mathrm{FLA},t} - \hat{Y}^N_{\mathrm{FLA},t})^2}{\sum_{t=1}^{T_0} (Y_{\mathrm{FLA},t} - \bar{Y}_{\mathrm{FLA}})^2} \tag{1}$$

where T_0 are the pre-treatment years, $Y_{\mathrm{FLA},t}$ are the outcomes in Flå, $\hat{Y}^N_{\mathrm{FLA},t}$ the outcomes for the synthetic control, and $\overline{Y}_{\mathrm{FLA}}$ is the average for Flå over the pre-treatment period. $\tilde{R}^2 = 1$ implies perfect fit. Next I derive a p-value for the treatment effect using the permutation test in Abadie *et al.* (2010). The ranking of Flå's ratio in the distribution of all ratios is used to derive the p-value of the average treatment effect.

Finally, to choose among the ten treatment estimates and associated p-values, I exclude models with $\tilde{R}^2$ below 0.8 (Ferman *et al.*, 2017) since the SC should only be used when the pre-fit is good (Abadie *et al.*, 2010). Next I calculate the mean post-period MSPE for the placebo estimates, and treat the specification with the smallest MSPE as the "best" specification (Dube and Zipperer, 2015). This rule follows a cross-validation logic where the pre-period estimates are the "training" sample and the post-period estimates the "validation" sample. Thus, I choose the specification with the best out-of-sample prediction properties. As an alternative, I average over the models with good pre-fit (Imbens and Rubin 2015; Ferman *et al.*, 2017).

[3]See https://www.ssb.no/klass/#!/klassifikasjoner/128/versjon/469. I restrict the sample to municipalities with unchanged boundaries over the analysis period.

[4]I use the fully nested optimization procedure.

Table 1. Results from ten different synthetic control specifications. Short run effects

	Childcare			Education			Elderlycare			Health care		
Model	Pre $\tilde{R}^2$	Post MSPE	p	Pre $\tilde{R}^2$	Post MSPE	p	Pre $\tilde{R}^2$	Post MSPE	p	Pre $\tilde{R}^2$	Post MSPE	p
1	0.91	1.90	0.57	0.93	15.24	0.16	0.88	24.10	0.82	0.94	6.50	0.50
2	0.91	1.90	0.57	0.93	15.24	0.16	0.88	24.11	0.82	0.94	6.51	0.50
3	0.85	1.90	0.55	0.93	15.45	0.02	0.71	24.69	0.91	**0.80**	**6.02**	**0.73**
4	0.88	1.93	0.37	0.93	15.26	0.01	0.63	24.89	0.93	0.52	6.38	0.77
5	0.65	1.88	0.71	0.77	15.21	0.32	0.68	24.27	0.73	0.84	6.29	0.58
6	0.62	1.87	0.71	0.08	15.42	0.93	0.50	23.78	0.85	0.81	6.90	0.46
7	0.74	2.07	0.37	0.93	15.95	0.04	0.80	24.81	0.80	0.93	7.74	0.30
8	0.67	2.28	0.57	0.92	15.76	0.01	0.81	30.01	0.76	0.93	8.54	0.21
9	**0.88**	**1.87**	**0.49**	**0.93**	**15.02**	**0.09**	**0.87**	**24.02**	**0.83**	0.93	6.56	0.45
10	0.86	2.12	0.48	0.93	15.56	0.09	0.87	26.48	0.81	0.93	6.93	0.42

Note: The ten models are: (1) all pre-treatment outcomes + log of population size, share of the population above 65 years of age, share of population in school age, share of population below school age, (2) all pre-treatment outcomes, (3) pre-treatment outcomes for even years + the covariates (4) pre-treatment outcomes for even years, (5) pre-treatment outcomes for odd years + the covariates, (6) pre-treatment outcomes for odd years, (7) the first half of the pre-treatment outcomes + the covariates, (8) the first half of the pre-treatment outcomes, (9) the first three fourths of the pre-treatment outcomes + the covariates, (10) the first three fourths of the pre-treatment outcomes.

The outcomes are the share of total spending on child care, education, elderly care, and health care. Together these areas constitute the majority of local government spending. The data are from Fiva *et al.* (2015).

4. Empirical results

Table 1 presents the results from the ten specifications. The post-treatment period is the four budget years following the 1995 election. We see that the pre-treatment fit varies across the specifications. The models with poor fit tend to be the even/odd outcome specifications, which is because these models often fail to capture a number of spikes in the spending levels caused by large investments in particular years. The model does not fit well when these investment years are not among the pre-treatment outcomes.

The post-treatment MSPE also varies across models. In line with arguments against maximizing pre-treatment fit by using all pre-treatment outcomes (Kaul *et al.*, 2017), Models 1 and 2 never produce the smallest post-treatment MSPE. The difference in post-treatment MSPE across models with good pre-treatment fit is not always large, however, implying that model averaging across specifications with good pre-treatment fit is a necessary robustness check.

The results reveal that there is not much room for specification mining to find significant treatment effects for child care, elderly care, and health care spending. The Flå treatment effect has to be ranked number eight or better among the 164 estimates for the effect to be significant at the 5 percent level (8/164 = 0.05). This is never the case for these three outcomes. For education we see that models with equally good pre-treatment fit produce different conclusions regarding the significance of the estimates, which gives the researcher the freedom to pick a model specification with a low p-value. This result illustrates the importance of presenting results from different specifications when using the SC method.

The specifications with the lowest post-MSPE (conditional on good pre-treatment fit according to $\tilde{R}^2$) are highlighted in bold text. Figure 2 shows the trends in the four outcomes in Flå and the synthetic control in these selected specifications. The stippled line distinguishes between the pre- and the post-period. In line with the $\tilde{R}^2$ results, the synthetic control fits the trend in Flå quite well for all outcomes, which make the trends for the synthetic controls plausible counterfactuals in the treatment period.

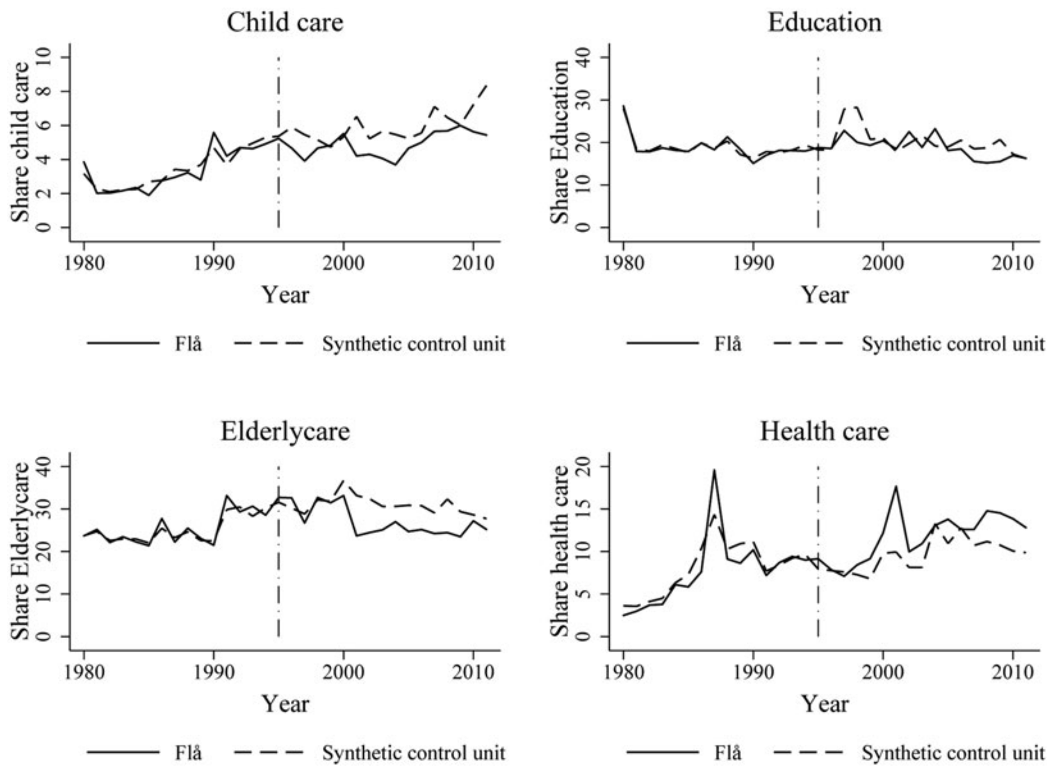

Figure 2. Trends in the outcomes. Flå versus synthetic control.

Table 2. Average spending in the 1996–2000 period

	Flå	Synthetic control	Difference	p-value
A: Lowest post-MSPE				
Child care	4.72	5.23	−0.51	0.49
Education	20.23	23.22	−2.99	0.09
Elderlycare	31.34	31.94	−0.60	0.83
Health care	8.95	7.85	1.10	0.73
B: Model averaging				
Child care	4.72	5.46	−0.74	0.53
Education	20.23	22.83	−2.60	0.11
Elderlycare	31.34	31.91	−0.57	0.85
Health care	8.95	8.14	0.81	0.45

The average outcomes for the 1996–2000 period for the two groups are presented in Table 2. Panel A presents the results when picking the specifications with the lowest post-MSPE, while Panel B presents the results when averaging over models with good pre-treatment fit. The average for Flå is about 0.5 percentage points lower for child care and elderly care spending, about 1 percentage points higher for health care spending, and about 2.5–3 percentage points lower for education spending. The education spending estimate is significant at the 10 percent level.

The long-run estimates are presented in online Appendix B. As Figure 2 visualizes, the short run effect on education diminishes over time as it is caused by a two-year up-thick in education spending in the synthetic control group. An inspection of the spending patterns in the

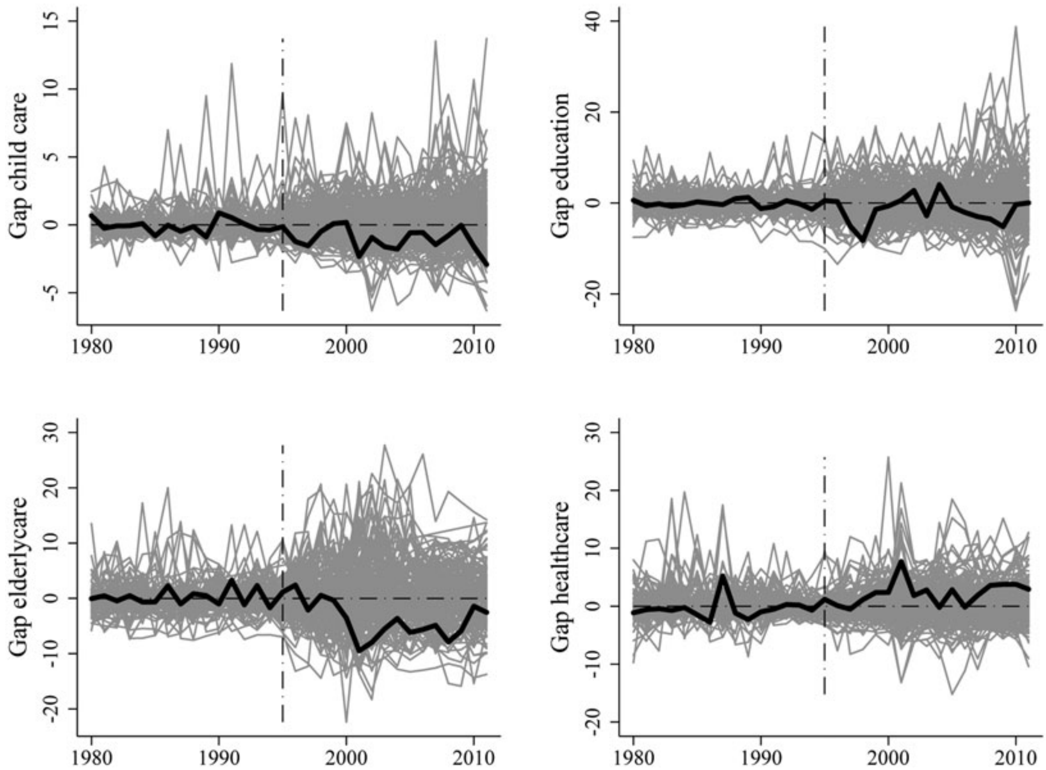

Figure 3. Differences in the outcomes. Flå versus synthetic control in black, placebos in gray.

municipalities which constitute the synthetic control—see online Appendix B for the unit weights—reveal that one of the control municipalities made large investments in education in the late 1990s. The impact of this investment diminishes over time so that the long-term effect is small and insignificant. The other treatment effects are larger in size in the long term, however, none of them are statistically significant.

Figure 3 clearly illustrates why the treatment effects are not statistically significant. For all outcomes, a relatively large number of placebos produce treatment effects of the same size. Thus, the trajectories of welfare spending in Flå do not stand out as being particularly unusual, i.e. we cannot rule out that the difference between Flå and the synthetic control would have been the same without the decline of Flå Ap.

5. Concluding remarks

The analysis provides no clear evidence of partisan effects on welfare spending. This conclusion apparently contrasts with a number of recent well-identified studies in similar institutional contexts (Pettersson-Lidbom, 2008; Folke, 2014; 2018). I rely on the sudden decline in Ap representation to identify partisan effects, which is more dramatic than the variation from close election that the previous literature relies on. The stronger treatment intensity should, in contrast to what I find, imply stronger partisan effects. That said, Fiva *et al.* (2018) report that the partisan effects are stronger for far-left and far-right parties, so my results can be read as supportive of their finding of heterogeneity on partisanship. Moreover, in my case the external validity is limited to small municipalities. Although parties appear to be polarized also in small municipalities (Fiva *et al.*, 2018), small partisan effects are in line with qualitative evaluations of the room for partisanship in small municipalities (Østerud and Selle, 2006). Given polarization, small partisan effects can be viewed as a democratic challenge since voters cannot elect policy changes. Future research might want to examine this issue further by studying partisan effects on outcomes where state regulation is weaker, such as e.g. local taxation and user fees, zoning, and local transport.

Supplementary Material. The supplementary material for this article can be found at https://doi.org/10.1017/psrm.2019.36

Acknowledgments. I would like to thank Stian Bekken and Gunnar Johansen for information about local politics in Flå and Jon Fiva, Øystein Hernæs and seminar participants at the annual meeting of Norwegian political scientists and at the Institute for Social Research for useful comments and suggestions. Grant number 255595 (Research Council of Norway) is acknowledged.

References

Abadie A, Diamond A and Hainmueller J (2010) Synthetic control methods for comparative case studies: estimating the effect of california's tobacco control program. *Journal of the American Statistical Association* **105**, 493–505.

Abadie A, Diamond A and Hainmueller J (2015) Comparative politics and the synthetic control method. *American Journal of Political Science* **59**, 495–510.

Athey S and Imbens GW (2017) The state of applied econometrics: causality and policy evaluation. *The Journal of Economic Perspectives* **31**, 3–32.

Besley T and Coate S (1997) An economic model of representative democracy. *Quarterly Journal of Economics* **112**, 85–114.

Doudchenko N and Imbens G (2016) Balancing, regression, difference-in-differences and synthetic control methods: a synthesis. NBER Working Paper No. 22791.

Dube A and Zipperer B (2015) Pooling multiple case studies using synthetic controls: an application to minimum wage policies. IZA DP No. 8944.

Esping-Andersen G (1990) The Three Worlds of Welfare Capitalism. Cambridge: Polity Press.

Ferman B, Pinto C and Possebom V (2017) Cherry picking with synthetic controls. MPRA Paper No. 78213.

Fiva JH, Halse A and Natvik GJ (2015) Local Government Dataset. Department of Economics, University of Oslo.

Fiva JH, Folke O and Sørensen RJ (2018) The power of parties: evidence from close municipal elections in Norway. *The Scandinavian Journal of Economics* **120**, 3–30.

Folke O (2014) Shades of brown and green: party effects in proportional election systems. *Journal of the European Economic Association* **12**, 1361–1395.

Hibbs D (1977) Political parties and macroeconomic policy. *American Political Science Review* **71**, 1467–1487.

Hyytinen A, Meriläinen J, Saarimaa T, Toivanen O and Tukiainen J (2018) Public employees as politicians: evidence from close elections. *American Political Science Review* **112**, 68–81.

Imbens G and Rubin D (2015) *Causal Inference for Statistics, Social, and Biological Sciences: An Introduction*. New York: Cambridge University Press.

Kaul A, Klößner S, Pfeifer G and Schieler M (2017) Synthetic Control Methods: Never use all pre-intervention outcomes together with covariates. Working Paper.

Korpi W and Palme J (2003) New politics and class politics in the context of austerity and globalization: welfare state regress in 18 countries, 1975–95. *American Political Science Review* **97**, 425–446.

Lee DS, Moretti E and Butler MJ (2004) Do voters affect or elect policies? Evidence from the US house. *Quarterly Journal of Economics* **119**, 807–859.

Munkerud LC (2007) Undersøkelse av lokalpolitikere 2006/2007. En redegjørelse for undersøkelsens gjennomføring og datakvalitet—samt noen hovedtendenser. BI Handelshøyskolen Discussion Paper 1/2007.

Østerud Ø and Selle P (2006) Power and democracy in norway: the transformation of norwegian politics. *Scandinavian Political Studies* **29**, 25–46.

Pettersson-Lidbom P (2008) Do parties matter for economic outcomes? A regression-discontinuity approach. *Journal of the European Economic Association* **6**, 1037–1056.

Cite this article: Finseraas H (2020). Social democratic representation and welfare spending: a quantitative case study. *Political Science Research and Methods* **8**, 589–596. https://doi.org/10.1017/psrm.2019.36