

ORIGINAL ARTICLE

Comparative politics and causal evaluation of structural reforms: the case of the UK national minimum wage introduction

Thomas König^{1*}, Guido Ropers² and Anika Buchmann³

¹Department of Political Science, University of Mannheim, PO Box 103462, D-68131 Mannheim, Germany, ²Graduate School of Economic and Social Sciences, University of Mannheim, 68131 Mannheim, Germany and ³Chair of Political Science II, University of Mannheim, 68131 Mannheim, Germany

*Corresponding author. Email: tkoenig@uni-mannheim.de

(Received 12 October 2017; revised 22 March 2019; accepted 28 March 2019; first published online 12 November 2019)

Abstract

In comparative studies, causal evaluations attempt to improve our understanding of the effectiveness of structural reforms by counterfactually inspecting post-treatment effects. Yet, even if comparative scholars find similar treatment and comparison units, the interpretation of the post-treatment trajectory is difficult as short-term estimates can be subject to strategic timing of reform implementation, while long-term effects are likely affected by further interventions. To illustrate these difficulties we apply the generalized synthetic control method to evaluate the introduction of a British national minimum wage. We find a short-term decreasing effect on youth unemployment that turns into an increasing effect over time. This suggests the presence of an upward biased selection effect from strategic timing. We also inspect two post-treatment interventions and find that they differ in their general and country-specific implications for the long-term trajectory.

Keywords: Comparative politics, structural reforms, causal inference

Following the “credibility revolution” (cf. Angrist and Pischke, 2010), comparative scholars increasingly engage in the causal evaluation of policy outcomes from structural reforms. Examples include the analysis of tobacco rules for smoking (Abadie *et al.*, 2010), fiscal rules for debt (Koehler and König, 2015), and federal institutions for welfare spending (Arnold and Stadelmann-Steffen, 2017). Compared to the analysis of regime types (Duverger, 1951; Sartori, 1976; Lijphart, 1984) and their implications for policy outcomes (Tsebelis, 2002; Strom *et al.*, 2008; Martin and Vanberg, 2011), these studies follow Mill’s Method of Difference and interpret the post-treatment reform effects using the synthetic control method (SCM) (Abadie *et al.*, 2010, 2015). The SCM’s matching-based extension of the difference-in-differences approach allows to conduct causal inference analyses even if the treatment is not randomly assigned and the number of treatment and comparison units consists of only a “few aggregate entities” (Abadie *et al.*, 2015, 496). This is an important advantage for the comparative study of structural reforms, as they rarely occur and typically do not provide for variation at the subnational level.

Statistically, the SCM follows the idea of the potential outcomes model (Rubin, 1974; Holland, 1986). It is based on a counterfactual approach that assumes “as-if” randomization of the reform treatment controlling for observed and unobserved factors in the pre-treatment period. The credibility of this approach rests on the ability to reconstruct the pre-intervention trajectories of characteristics and outcomes of the treated unit with a synthetic control unit. For this, the SCM

provides a systematic way to choose a weighted combination of comparison units that better fits the treated unit than any single unit itself (Abadie *et al.*, 2015). The approach of the SCM has been praised for bridging the quantitative/qualitative divide in comparative politics and described as “the most important innovation in the policy evaluation literature in the last 15 years” (Athey and Imbens, 2017, 9).¹

In this article, we aim to draw attention to two distinct but highly connected challenges for comparative studies, which interpret causal reform estimates of structural reforms.² The first challenge concerns a form of “time selection bias” that originates from a non-random implementation of reforms in time. Especially in the case of important reforms, policy-makers will avoid policy outcomes that risk their electoral survival in office (Franzese, 2002; Ashworth, 2012). Instead, important reforms are implemented in times when their effects are unlikely to harm the reelection chances of policy-makers. In the short-term post-treatment period, this effect may cause an upward bias of otherwise potentially negative estimates on policy outcomes.

The second challenge concerns the internal validity of the estimates over time due to omitted interactions and differential trends in treatment and comparison units. In comparative studies, they may result from further (structural) post-treatment interventions that affect treatment and comparison units differently. Taking a very restrained view, Abadie *et al.* (2015) propose that none of the units used for the SCM should be “subject to structural shocks to the outcome variable during the sample period of the study” (2015, 497). Yet, although post-treatment interventions are likely to occur over time, they may not affect the insight into the reform effects if treatment and comparison units are similarly affected by them. Accordingly, our recommendation is that comparative scholars assess the degree to which their estimates are subject to and affected by both general and country-specific post-treatment interventions.

Although the implications of these challenges are well known (see, e.g., the discussion by Meyer, 1995), they can—independently from each other—result in an interaction with the post-treatment effects that is often neglected in causal comparative analyses. Together, they challenge the external validity of findings in causal comparative evaluations: On the one hand, when a time selection effect is likely to exist, a logical recommendation would be to draw more attention to the long-term trajectory because policy-makers tend to discount long-term implications in their policy-making (Jacobs, 2016). In fact, Abadie *et al.* advocate a “sizable number of post-intervention periods (...) in cases when the effect of the intervention emerges gradually after the intervention or changes over time” (2015, 497). On the other hand, the probability of further (structural) interventions is increasing over time. This makes it difficult to attribute differences in policy outcomes between treatment and comparison units to either the reform, additional interventions, or some interaction of it.

To illustrate the presence and consequences of these two interrelated challenges, we apply the generalized synthetic control method (GSC) developed by Xu (2017). One advantage of this method is that it provides frequentist uncertainty estimates which are easier to interpret than the “placebo studies” used for inference in the SCM. Furthermore, the GSC allows to additionally condition the counterfactual estimation on controls associated with (long-term) interventions in the post-treatment periods. If the reform estimates remain robust to such controls, comparative scholars can more confidently attribute the post-treatment effects to the reform implementation. We exemplify the interpretation of the post-treatment trajectory with our comparative analysis of the implications of a national minimum wage (NMW) for youth unemployment in the United Kingdom (UK) in 1999, which experts considered as the most successful structural reform of the past 30 years in British history (BBC, 2010). As expected, we find strong support for a non-

¹Moreover, the application of the SCM to a diverse set of reform topics is stimulating the methodological discussion on causal inference such as the design of credible control groups (Allegritto *et al.*, 2017) or the formalization of the inference procedure (Firpo and Possebom, 2017).

²See also Duso and Röller (2003), Allcott (2015), and Kleis and Moessinger (2016).

linear effect on the long-term trajectory that indicates the presence of an upward bias. We further assess potential bias in the longer-term from two additional post-treatment interventions, the 2004 Eastern Enlargement and the 2007/8 economic and financial crisis, but find the trajectory of the reform estimates to be robust.

The contribution of this article is twofold. First, we contribute to the growing literature, which attempts to bridge the divide between qualitative and quantitative methods in comparative politics, by raising attention to an important issue: When comparative scholars identify non-linear trends, the validity of short-term estimates has to be taken with caution due to possible time selection bias. At the same time, the assessment of the long-term trajectory needs to carefully consider the presence of further interventions that may affect treatment and comparison units differently. Second, we explore an important substantive question on the implications of NMW for youth unemployment. We confirm previous findings that minimum wages do not increase (youth) unemployment in the short-run (e.g., Card and Krueger, 1995; Schmitt, 2013). However, the longer post-treatment trajectory suggests negative implications for youth unemployment over time. In addition to providing support for the presence of time selection bias in the short-term post-treatment period, it confirms NMW studies exploiting subnational variation (Dolton *et al.*, 2015).

1. Electoral considerations and time selection of policy reforms

Do policy-makers randomly implement structural reforms in time, if these reforms are expected to have implications on relevant policy outcome variables? To the contrary, political economists stress the strong incentive of policy-makers to engage in targeted policy-making as a result of electoral and partisan competition; that is, they are not independent of macro-economic policy outcomes and political or institutional factors that influence their stay in office (Nordhaus, 1975; Tufte, 1978; Persson and Tabellini, 1999; Besley and Case, 2000). In representative democracies, policy-makers have an incentive to engage in (macro-economic) policy-making that ensures their re-election. This argument finds support in the retrospective voting literature. Accordingly, voters evaluate governmental parties based on macro-economic performance. Not only does it generate democratic accountability (Ferejohn, 1986; Ashworth, 2012), but, more importantly in our case, it also increases the incentive for policy-makers to implement reforms only if the policy output does not create short-term electoral costs. As a consequence, policy-makers are unlikely to implement structural reforms that may generate negative macro-economic effects until election day.

The strategic timing of reform implementation may produce non-linear effects that call the validity of causal estimates in the immediate period shortly after implementation into question. For example, Hellman (1998) discusses such a time selection effect as “J-curve” with regard to the transformation of Eastern European countries after 1990. Moreover, economists have used narrative analyses of policy documents to identify ideological rather than electoral considerations behind policy reform-making to better account for the selection and reverse causality bias in reform estimates (Romer and Romer, 2004, 2010; Kleis and Moessinger, 2016). The challenges from strategic timing are also highlighted by Hotz *et al.* (2005) in their discussion of the external validity of causal estimates in other contexts. Even if the treatment is randomly assigned in space, treatment effects may still be dependent on invariant (economic) context characteristics of a particular implementation setting. In such cases, it is very difficult for researchers to generalize findings from randomized controlled trials to other settings or contexts. Similarly, Allcott (2015) emphasizes “site selection bias” in randomized controlled trials as another threat to the external validity of treatment effect estimates.

In comparative studies, strategic timing may impose a further hurdle for choosing comparison units, even if the comparison units are unaffected by the reform, as demanded by Mill’s Method of Difference. Observed reform estimates subject to strategic timing will be positively biased in comparison to the population or target effect, if one considers the actual implementation time

of a structural reform as a draw from all possible time points. In this context and in analogy to Allcott (2015, 1130), the term “bias” underscores that this difference is due to a systematic rather than a random time selection mechanism. Therefore, comparative scholars need to consider the policy-makers’ (expected) responsiveness to favorable economic and political context in their interpretation of reform estimates.

The question arises how to assess and address this time selection bias of structural reforms. Because important structural reforms occur rarely, it is difficult, if not impossible, to assess time selection bias in a comparative manner. However, one way to address this problem is to consider that policy-makers operate in a sphere of incomplete information (e.g., Hall, 1993; Blyth, 2002). They face uncertainty about the potential reform effects, in particular over the long-run (Jacobs, 2016). Because structural reforms attempt to change existing conditions, which may not be beneficial for all persons concerned, the challenge for policy-makers is to implement reforms that overcome the anticipated short-term opposition of the reform losers (Przeworski, 1991; Pierson, 1994; Weaver, 2003, 138). Accordingly, the incentive for policy-makers is to ensure positive short-term effects while they are similarly likely to discount long-term effects in their decision-making. Furthermore, changing economic and political context conditions the incentives and abilities to manipulate the policy outcomes over time as well as the effectiveness of such manipulation (Franzese, 2002). Thus, the evaluation of the long-term trajectory may constitute a possible way to evaluate the presence and degree of time selection bias. If time selection bias is indeed present, comparative analyses are likely to find a non-linear reform effect evolving over time.

2. Post-treatment interventions

The more the outcomes of a structural reform matter for the political survival of a policy-maker, the more likely it becomes that strategic timing impacts the short-term observations in comparative studies. As outlined, the evaluation of the long-term trajectory may help to assess the presence of such time selection bias. Nevertheless, long-term estimates also run danger of being affected by structural changes or interventions in the post-treatment period. Yet, motivated by the experimental ideal, causal analyses primarily draw attention to confounders in the pre-treatment assignment mechanism rather than the post-treatment period (Rubin, 1974; Holland, 1986). As a consequence, the role of post-treatment interventions for the validity of causal estimates are rarely discussed in practice. At the same time, estimates, in particular in small-*N* comparative studies, are very sensitive to both general and country-specific interventions in the post-treatment period.

In case of country-specific post-treatment interventions, the assumption that they are equally distributed across treatment and comparison units is likely to be violated. If interventions hit treatment and comparison units differently, it becomes very difficult to contribute differences in policy outcomes between treatment and comparison units to either the reform, the post-treatment interventions, or some interaction of it because the chosen control units may no longer be valid to construct the counterfactual unit of interest in the post-treatment period. This is different from interventions that affect treatment and comparison units to a similar degree. These types of general post-treatment interventions may not call into question the overall validity of the estimated reform effects because the weighted combination of comparison units used for imputing the counterfactual of interest remains valid. Nonetheless, even such interventions, like a worldwide economic recession, may still impact countries differently, which is similar to a missing variable in the data generating process of a counterfactual construction. Therefore, it is recommendable to assess the type of post-treatment interventions on the long-term trajectory in comparative case studies.

Overcoming the divide between qualitative and quantitative approaches that focus on causal evaluation, the SCM was the first to provide a systematic way to choose comparison units in

comparative studies by combining a qualitative case study design with a generalized difference-in-differences regression design (Abadie and Gardeazabal, 2003; Abadie *et al.*, 2010, 2015). A major premise of this approach is that “when the units of analysis are a few aggregate entities, a combination of comparison units often does a better job of reproducing the characteristics of the unit or units representing the case of interest than any single comparison unit alone” (Abadie *et al.*, 2015, 496). When the number of pre-treatment periods is large enough, the matching algorithm facilitates to control for unobserved factors as well as for heterogeneous effects of observable factors on the outcome of interest (Abadie *et al.*, 2010)—at least in the pre-intervention period.

To better assess the role of further interventions in the post-treatment period, we propose using a recent extension of the SCM, the GSC developed by Xu (2017). Similar to the SCM, the GSC method tackles the problem of the missing counterfactual by imputing treated counterfactuals from a donor pool of untreated comparison units. It is based on a semiparametrical linear interactive fixed effects model that includes time-varying coefficients interacted with unit-specific intercepts.³ Like the SCM, the GSC relies on a reweighting scheme that builds on predicting pre-treatment outcomes as benchmark (Xu, 2017, 58). Yet, while the GSC is not immune against bias in the post-treatment periods, it offers a more robust and transparent estimation of the long-term trajectory than the SCM for two reasons. In comparison to the placebo tests of the SCM, the GSC provides easily interpretable frequentist uncertainty estimates. Moreover, the GSC allows incorporating control variables (assuming constant effects over time) to estimate the counterfactual of interest. Thus, the out-of-sample predictions for the post-treatment periods no longer rely on an imputation structure that is solely grounded on a weighted average of the observed post-treatment outcomes of interest from the comparison units.

Following the logic of “proxy controls” (Angrist and Pischke, 2008, 66), controlling for post-treatment variables in the GSC approach can be helpful to assess the type of post-treatment interventions. They may help to control for relevant omitted factors, even if they might themselves be affected by the treatment because the estimates might better capture the reform effect than without control. Nevertheless, the inclusion of post-treatment variables can also result in a form of selection bias if they are outcome variables of the treatment themselves (cf. Acharya *et al.*, 2016). Thus, we only control for the effect of post-treatment interventions in order to find out whether the inclusion of the intervention proxies change the reform estimates. If the (non-linear) long-term reform estimates remain robust to these controls, we can be more confident to attribute the observed differences in the long-term trajectory between treatment and comparison units to the implications of the reform implementation.

3. Case: UK national minimum wage

The introduction of a NMW, as put in place in the UK in 1999, constitutes an example of an important structural reform where electoral considerations of policy-makers are likely to have played a role at the time of implementation. The Labour Party has always been in favor of a minimum wage, while the Conservative Party was opposed to the introduction of a NMW (Pyper, 2014). After the 1992 defeat of the Conservative Party, Labour Prime Minister Tony Blair decided to build the political case for a NMW (Coats, 2007, 24). Given ideological controversy between Labour and Conservatives, it is likely that the timing of implementation of the NMW has been responsive to the prevailing economic conditions (in order to avoid the support of the Conservatives’ argument on negative employment effects). Speaking in favor of a time selection bias in the short-term, the UK government implemented a particularly low minimum wage when it first introduced the NMW with the explicit goal to avoid any damaging employment effects (Low Pay Commission, 2016, 1). Thus, we can expect that the short-term effect of the

³See Appendix A for a detailed explanation of the estimation strategy which follows Xu (2017).

NMW on youth employment was negligible or positive in economically favorable times of implementation.

The reform effect is, however, likely to have changed in the longer-term. The simple reason is that electoral considerations might have been limited, for example, due to the end of the governmental term. Moreover, far more than being a single-shot treatment, the implementation of NMW was adjusted over time through subsequent upgradings (e.g., Blanchflower *et al.*, 2017). It resulted in the first “real-valued” increases of the minimum wage. These are more likely to result in negative labor market effects, at least among particularly vulnerable subgroups such as the youth. A subnational level study by Dolton *et al.* (2015) suggests that this might have resulted in an overall reform effect that only unfolded over time. In addition to the NMW implementation, the development of UK youth unemployment after 1999 was likely influenced by other exogenous changes. If they were unaccounted for before the NMW introduction, they possibly affected the post-treatment outcomes directly or through some interaction. While not being exclusive, we consider two labor market interventions in the context of the UK labor market in our analysis, the Eastern enlargement of the EU in 2004 and the worldwide economic and financial downturn of 2007/08.

In preparation for Eastern enlargement, the UK government was guided by macro-economic considerations when it decided to open the labor market for more immigrants already in 2004, followed by Sweden and Ireland. As hoped for by the UK government (Wright, 2010), it resulted in initially overall positive economic effects—such as the so-called Polish plumper effect—of the inflowing workforce from Eastern Europe. This confirms studies in labor economics, which find generally small or no effects of immigration on average native employment rates (National Academies of Sciences, and Medicine, Engineering, 2017, 204). While immigration from Eastern Europe was welcome in the UK in the beginning, this situation started to change in the aftermath of the economic and financial crisis in 2007/8. In the 4th quarter of 2008, UK GDP fell by 1.5 per cent and the country officially entered a period of recession. Compared to the dominating macro-economic considerations regarding Eastern enlargement, the unexpected crisis hit the UK and its labor market in particular. Unemployment rose, especially among the 18–24 years old (Bell and Blanchflower, 2011).

At first sight, British (youth) unemployment was more affected at the time of the economic and financial crisis than in the years before when the UK government upgraded minimum wage levels and Eastern enlargement opened the doors for immigration. Furthermore, as argued, the impact of strategic timing decreases over time, which suggests a higher credibility of the effects in the post-treatment periods. By using the GSC method, we are able to illustrate the sensitivity of the reform effect to observable implications of post-treatment interventions. If the reform estimate is unaffected or if both the treatment and comparison units are affected similarly, the long-term reform estimates should not change substantially. Yet, if the estimates become smaller or even change direction, we can interpret it as an indicator that the long-term reform estimate is affected by an intervention. In this case, it becomes very difficult to assign differences in policy outcomes between treatment and comparison units to either the reform, the interventions, or some interaction of it, which calls the validity of the estimates into question.

4. Data and sample

Even though most of the credibility literature on causal identification recommends exploiting subnational variation, a limited focus on institutionally similar labor markets may also challenge the external validity of those results (e.g., Sturn, 2017, 2). Furthermore, the UK government approved a structural reform and introduced a minimum wage that sets the standards nationwide. For comparative research, this requires to assess the reform implications of a nationally set minimum wage with regard to the structural rather than subnational labor market characteristics. Our data consists of annual country-level panel data for the period from 1984 to 2013. Limitations of

data availability for our main variable of interest, youth unemployment rate in the UK, prevented us from extending the time frame to periods before 1984.

The selection of the donor pool of potential comparison countries is guided by four principles (see Abadie *et al.*, 2015, 500):

- First, countries which have been affected by a similar intervention of interest are excluded from the donor pool, i.e. countries which introduced a nationally determined minimum wage during the study period.
- Second, the donor pool is restricted to structurally similar countries in order to avoid over-fitting and bias from interpolation across countries with very different characteristics. Therefore, we restrict our pool of control units to OECD countries as they are of similar economic structure.
- Third, Abadie *et al.* (2015, 500) recommend excluding units that experienced large idiosyncratic shocks to the outcome of interest as a result of the reform implementation.
- Fourth, it requires a sufficiently high number of both pre-treatment and post-treatment observations of the outcome of interest and potential control variables.

In the end, it leaves us with a final donor pool consisting of Germany, Finland, Sweden,⁴ Norway, Denmark, Italy, and Austria.⁵

The policy outcome variable of interest Y_{jt} is the youth unemployment rate measured as the percentage of youth labor force aged 15–24 years and taken from the OECD database on employment and labor market statistics (OECD, 2014). Following the minimum wage literature, young workers should be the more impacted by a NMW implementation, as they have, on average, fewer skills, experience, and training on the job than their adult counterparts. Moreover, the youth is weakly organized in the labor market which reduces other endogeneity concerns about the implementation of NMW by policy-makers and their responses to exogenous interventions in the post-treatment period. We rely on the standard OECD definition and focus on the 15–24 age group.

A crucial assumption of causal inference studies is that the structural reform has no effect during the pre-treatment periods in order to estimate the post-treatment effects (Abadie *et al.*, 2015, 3). Consequently, prior reform effects to the selected treatment period would bias the estimation as the estimator is based on a comparison between the treatment and comparison units for the post-treatment period. To examine potential bias, Abadie *et al.* (2010, 494) recommend redefining the treatment period, T_0 , to the first period in which the outcome of interest might have already reacted to the reform treatment. With regard to NMW in the UK, there are multiple points which could be considered in addition to the official year of the NMW implementation in 1999, such as the election of Labour into government in 1997 or the establishment of the LPC in 1998. Although we use the year of the official implementation of the NMW as the first treatment period when the final regulations regarding the NMW suggested by the LPC became public and official, we will run several robustness checks to test whether the reform effect changes when adjusting the pre-treatment period.

To account for the discussed strategic timing of reform implementation we need to include a sizable number of post-treatment periods since the effects are expected to emerge gradually and change over time. Therefore, our post-treatment time ranges from 1999 until 2013. We also add two control variables as proxies for the post-treatment labor market interventions taken from the OECD database to our model specification (OECD, 2014). To account for the inflow of (low-skill) labor as result of the Eastern enlargement in 2004, we construct a logged variable measuring

⁴We deliberately include Sweden in the control group even though it did not establish a transitory period for labor migration after the Eastern enlargement in 2004. We will control for the resulting bias by explicitly modeling the labor market consequences of the Eastern enlargement in our application of the GSC.

⁵South Korea and Switzerland had to be excluded due to non-availability of data.

the number of immigrants coming into the UK from countries which joined the EU in 2004⁶ as share of the active labor force population. To account for the effects of the downturn following the 2007/8 economic and financial crisis, we add the percentage change in GDP as a second control variable.

5. Results

We start our investigation with estimating the average treatment effect without conditioning the counterfactual on additional observables.⁷ Figure 1 shows the counterfactual trajectory of youth unemployment in the untreated UK from the GSC estimation as well as the actual development. Compared to conventional DID estimates shown in Figure A1 in the appendix, the GSC estimation fits the development of youth unemployment in the pre-treatment period well. This result increases the confidence to associate differences in the post-treatment period to the reform implementation.

With this confidence we can turn to the post-treatment trajectory in Figure 1, which suggests to distinguish three distinct post-treatment periods how youth unemployment in the “treated” UK with a NMW developed over time (solid line). First, we observe an unemployment reducing effect. The contrasting youth unemployment trends between the treated UK and the comparison units (dashed line) indicate a selection effect through strategic timing of reform implementation. The estimated “counterfactual UK” would have had a higher youth unemployment rate without the NMW immediately after its introduction. In contrast, the actual UK youth unemployment rate continues to follow the declining trend of the years prior to reform. Afterwards, the actual youth unemployment rate in the UK remains below the pre-treatment level until 2004.⁸ This short-term effect suggests that the implementation of the NMW has been a targeted action of policy-makers. The Labour Party under Tony Blair would have been unlikely to implement a minimum wage if they were expecting negative (short-term) consequences that would have lowered their chances of re-election in 2001.

In 2004, the point estimate of youth unemployment in the counterfactual UK is for the first time lower than the actual unemployment rate. From this point on, we observe a clearly opposing trend of youth unemployment compared to the control group. It characterizes a distinct second period of reform implementation toward the end of the second governmental term of Tony Blair. It may be explained by country-specific interventions like inherent reform adjustments or the post-intervention effect of Eastern enlargement that also happened during this period. Finally, we find indication of a further post-treatment intervention in 2008. Compared to the previous period, the parallel development of the youth unemployment trend in the treated and comparison units suggests a more general intervention with similar effects. Most likely, this reflects the economic and financial crisis, which started to unfold in that same year.

Together, the diverging trends between the actual youth unemployment rate and the estimated counterfactual during the period between 2001 and 2007 hint at post-treatment interventions during this time that are responsible for the increasing effect on youth unemployment. At this point, however, it remains open whether to account these effects to country-specific adjustments or more general labor market interventions. Therefore, we subsequently include two control variables, immigration from Eastern Europe and GDP growth, to assess the robustness of the findings to two (exogenous) post-treatment interventions that might bias the NMW reform estimates but are unlikely to have been anticipated by the policy-makers at the time of the NMW

⁶Estonia, Latvia, Lithuania, Malta, Poland, Slovak Republic, Slovenia, Czech Republic, Hungary, and Cyprus.

⁷We conduct all our analysis in R using the `gsynth` package provided by Xu (2017).

⁸Note that all models as well as the actual youth unemployment rate experience a slight increase in 2003. It indicates that the GSC is able to capture unobserved common shocks. After a short economic recovery in 2002, the countries of the EU in general experienced stalling employment growths and rising rates of unemployment in 2003 (EC, 2004).

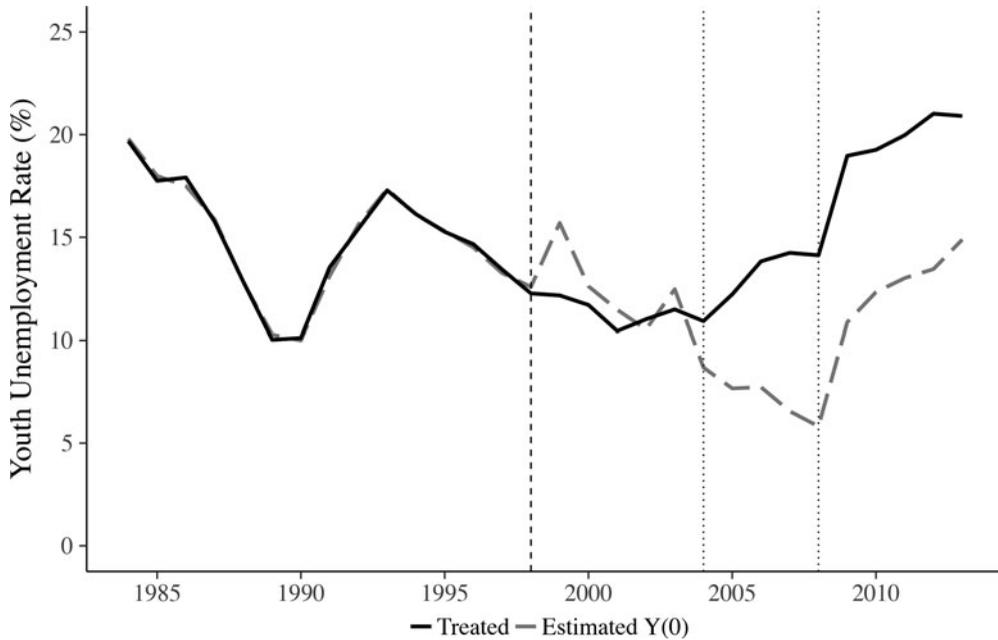


Figure 1. Comparison of estimated counterfactual, $\hat{Y}(0)$, from the main model with the actual development of youth unemployment rate in the UK after the implementation of the NMW.

implementation in 1999. Table 1 presents the numerical results. Column 1 shows the average treatment effect in the main model when no covariates are included. Columns 2 and 3 list the average treatment effect when we, additionally, condition the estimation of the counterfactual on the (constant) effect of immigration from Eastern European countries as a share of the active labor force and the effect of the economic and financial crisis for GDP growth. Column 4 presents the results when conditioning on both control variables. All models include country and year fixed effects.

The findings reveal five latent factors to be important which all models condition on.⁹ According to the main model of column 1, the estimated average treatment effect over the whole post-treatment period (1999–2013) is an increase of youth unemployment rate of about 3.9 percentage points (with a standard error of about 2.3 based on parametric bootstraps), but is statistically not distinguishable from zero. This is not surprising given our expectation of a non-linear trend over time. Accordingly, the short-term decreasing effect should cancel out the long-term increasing effect.

When adding the controls for the two labor market interventions in the post-treatment period (models 2–4), we see a small but negligible decrease in the estimates, while the standard errors remaining sizable. Without being conclusive, it suggests that at least a small part of the long-term increasing effect, which we observe in the main model, is attributable to the exogenous labor market intervention. The substantive size of the long-term NMW effect on youth unemployment remains, however, similar.

Following our discussion about the validity of reform estimates in comparative studies, we are mainly interested in the confidence to attribute changes in the dependent variable to the post-treatment reform implementation. Figure 2(a) shows the average annual treatment effect of the UK NMW on youth unemployment from 1999 to 2013 with a 95 percent confidence interval

⁹See Figure A2 in the appendix for the trajectory of the factors over time.

Table 1. The effect of the UK NMW on youth unemployment rate 1999–2013

	(1)	(2)	(3)	(4)
Average treatment effect NMW	3.9 (2.32)	3.74 (2.45)	3.37 (2.46)	2.51 (2.37)
EE immigration (% of active labor force; ln)		0.47 (0.28)		0.63 (0.30)
Δ GDP			0.02 (0.05)	0.06 (0.05)
Country FE	x	x	x	x
Year FE	x	x	x	x
Unobserved factors	5	5	5	5
N	240	240	240	240
Control countries	7	7	7	7

for our main model. First, we find that the youth unemployment rate in our main model is about 3.5 percentage points lower than without the NMW with a confidence interval (CI) clearly below zero ranging from -4.5 to -2.3 percentage points in the first year of the implementation. Second, after this short initial period, the size of the average treatment effect begins to rise and around 2007/8, we see an overall positive average treatment effect of the NMW implementation on youth unemployment rate, which is statistically different from zero. The baseline model suggests an average increase as a result of the NMW of 8.3 percentage points in 2008 (CI: [1.27 percentage points, 16.7 percentage points]).

The non-linear effect of the UK NMW on youth unemployment also holds when we include proxy controls for the two post-treatment interventions in our models. [Figure 2\(b\)](#) controls for the increase in immigration from Eastern Europe to approximate the effect of the Eastern enlargement. [Figure 2\(c\)](#) shows the average treatment effect when controlling for GDP growth rates to directly capture the consequences of the economic and financial crisis of 2007/8. [Figure 2\(d\)](#) shows the development of the average treatment effect of the UK NMW when accounting for both control variables. Overall, the dynamic development of the effect remains similar to the main model in [Figure 2\(a\)](#): The implementation of the NMW decreased the youth unemployment rate in the short-run, in particular, in the first year of the implementation. Afterwards, we observe a steady increase in the period from 2000 to 2007 compared to the predicted development without the implementation of a NMW in all models.

In sum, the analysis with the GSC approach points to reform-inherent post-treatment adjustments rather than general exogenous structural post-treatment interventions that explain the non-linear effect of the UK NMW on youth unemployment. Yet, our cross-country findings are based on a relatively small number of comparison units only. Therefore, we conduct a number of diagnostic checks including applying the classical synthetic control method and re-running our main analysis by including additional demand and supply variables of youth unemployment to ensure that the (imputed) results are not excessively extrapolated and robust to other specifications (see Appendix B). Overall, the tests support the conclusions drawn from our main analysis. Nevertheless, we emphasize that our analysis based on aggregate times-series cross-sectional data still does not allow us to fully exclude the possibility of an interaction effect with the Eastern enlargement or another unobserved post-treatment intervention from our analysis.

6. Concluding discussion

The focus on causal inference in comparative studies has increased with the popularity of the synthetic control method. One key message of our analysis is that comparative scholars should carefully consider the conditions under which policy-makers implement structural reforms. Electoral considerations can result in an upward time selection bias of otherwise negative reform effects in

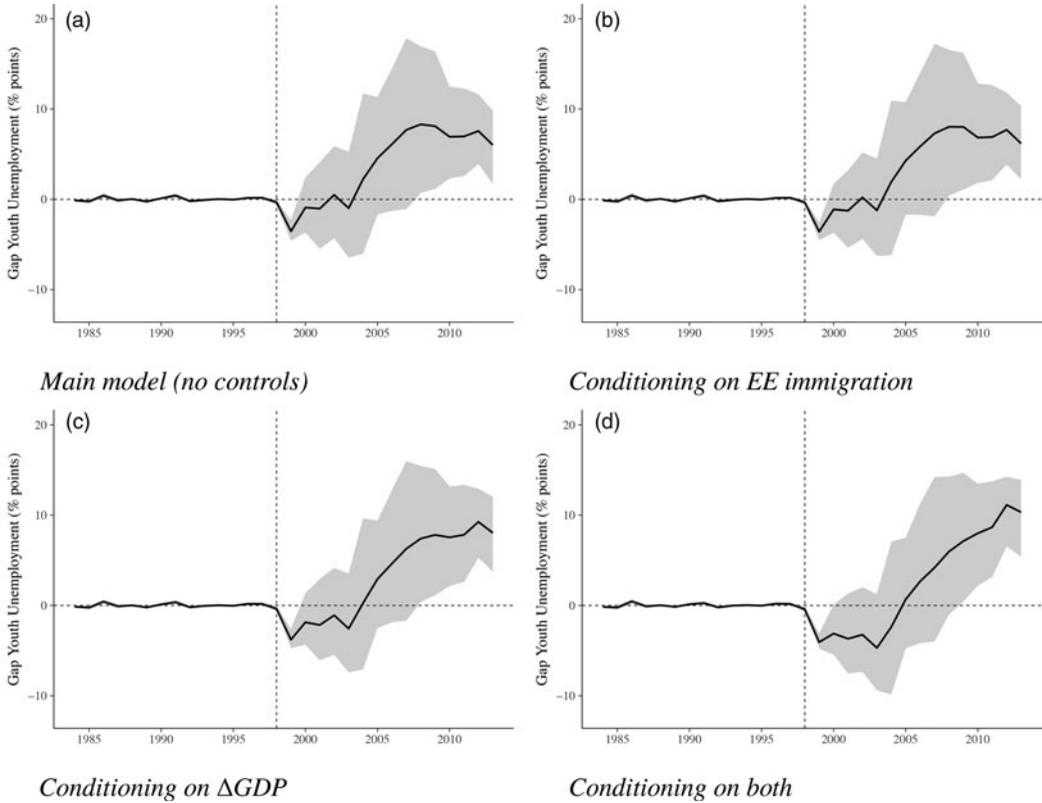


Figure 2. Average treatment effect of the UK NMW over time as the difference (gap) between the estimated counterfactual and the actual development of youth unemployment in the UK. Figure (a) shows the effect estimated without additional controls. The remaining three models, additionally, condition the estimation of the counterfactual on (a) immigration inflow from Eastern European countries as a share of active labor force, (b) growth in GDP, and (c) both control variables. (a) Main model (no controls), (b) conditioning on EE immigration, conditioning on Δ GDP, and (d) conditioning on both.

the early post-treatment period if structural reforms are not randomly implemented in time. As a consequence, the long-term reform trajectory is likely to unfold in a non-linear trend over time, as policy-makers tend to discount possible long-term implications. At the same time, further post-treatment interventions in both treatment and comparison units can make it difficult to attribute reform effects to either the reform implementation or other post-treatment interventions in the longer-term. To better understand the implications of such post-treatment interventions, we applied the generalized synthetic control method by Xu (2017) to evaluate the reform effect of UK NMW for youth unemployment.

In this study, we caution against the interpretation of reform estimates independent of the political and economic context of reform implementation. Yet, by accounting for the larger context of policy implementation based on qualitative insight or theoretical knowledge, comparative research can enhance our understanding about the effectiveness of structural reforms. Take our empirical assessment of the UK NMW effect as an example. Because NMW tend to persist and as the UK NMW has been implemented at the national level, the comparative analysis requires methods that help to carefully interpret the post-treatment trajectory. At first, we find that the NMW has reduced youth unemployment in the short-term. As argued, electoral competition inclines policy-makers to adjust their behavior to the future economic environment and makes the implementation of a NMW more likely if its employment effects are expected to

be small or even positive during their own re-election period. Yet, the estimated long-term trajectory of the NMW effect is non-linear with an increasing youth unemployment effect unfolding over time.

On closer inspection of this trend, we differentiated between three post-treatment periods. In the first years of the introduction, we observe an unemployment decreasing effect which, however, slows down and turns into an increasing trend in the second period toward the end of the second governmental term of Tony Blair. We conditioned our analysis on observable implications of another post-treatment intervention, the Eastern enlargement of the EU in 2004, but find the general trend to be robust. We also find indication of an intervention in the third period, which is likely to result from the worldwide 2007/8 economic and financial crisis. Even though we cannot examine the exact underlying mechanisms of these post-treatment interventions, a plausible explanation for the turnaround of the reform effect is that the UK government started to continuously increase the inflation-corrected minimum wage only two years after the NMW implementation in 1999 until the crisis. In addition, our analysis suggests that the NMW did not further increase youth unemployment rates during the crisis in the final period of our study. One potential explanation points to the flexibility of the policy-makers during this time period.

The enormous youth unemployment rates in many countries are one of the key problems, which representative democracies are facing today in Europe. The 2007/8 crisis led to significant falls in gross domestic product for most OECD countries. The conventional view is that, in the face of falling demand, a potential reaction of firms is to adjust labor costs by lowering or freezing wages (Bryan *et al.*, 2012, 7). With a minimum wage in place, however, firms have much smaller margins to react to any kind of demand or supply shock, making the reduction of employment a likely response. Using survey data on full- and part-time restaurant staff in Pennsylvania and New Jersey, Card and Krueger (1995) were able to call this conventional wisdom into question. However, compared to insights from US state-level analyses, empirical studies about NMW effects are characterized by very mixed findings. Our analysis highlights time selection of reform implementation as a threat to the external validity of short-term estimates. In the extreme, the non-linear development of reform effects in the post-treatment periods due to time selection may explain some of the scholarly controversies regarding the implication of reform effects.

Supplementary material. To view supplementary material for this article, please visit <https://doi.org/10.1017/psrm.2019.45>

Acknowledgments. We wish to thank Ziad Elgendy for his support when gathering the data used in this article.

References

- Abadie A and Gardeazabal J (2003) The economic costs of conflict: a case study of the basque country. *American Economic Review* **93**, 113–132.
- 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.
- Acharya A, Blackwell M and Sen M (2016) Explaining causal findings without bias: detecting and assessing direct effects. *American Political Science Review* **110**, 512–529.
- Allcott H (2015) Site selection bias in program evaluation. *The Quarterly Journal of Economics* **130**, 1117–1165.
- Allegretto S, Dube A, Reich M and Zipperer B (2017) Credible research designs for minimum wage studies: a response to Neumark, Salas, and Wascher. *ILR Review* **70**, 559–592.
- Angrist J and Pischke J-S (2008) *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, New Jersey: Princeton University Press.
- Angrist J and Pischke J-S (2010) The credibility revolution in empirical economics: how better research design is taking the con out of econometrics, Working Paper 15794. National Bureau of Economic Research, March. Accessed December 15, 2017. Available at: <https://www.nber.org/papers/w15794>.
- Arnold T and Stadelmann-Steffen I (2017) How federalism influences welfare spending: Belgium federalism reform through the perspective of the synthetic control method. *European Journal of Political Research* **56**, 680–702.

- Ashworth S** (2012) Electoral accountability: recent theoretical and empirical work. *Annual Review of Political Science* **15**, 183–201.
- Athey S and Imbens GW** (2017) The state of applied econometrics: causality and policy evaluation. *The Journal of Economic Perspectives* **31**, 3–32.
- BBC** (2010) Minimum wage ‘most successful government policy’. December 3. Accessed June 26, 2017. Available at: <https://www.bbc.com/news/uk-politics-11896971>.
- Bell DNF and Blanchflower DG** (2011) Young people and the great recession. *Oxford Review of Economic Policy* **27**, 241–267.
- Besley T and Case A** (2000) Unnatural experiments: estimating the incidence of endogenous policies. *Economic Journal* **110**, 672–694.
- Blanchflower D, Costa R and Machin S** (2017) The return of falling real wages. CEP Real Wages Updates 6.
- Blyth M** (2002) Great transformations: economic ideas and institutional change in the twentieth century. Accessed January 17, 2018.
- Bryan M, Salvatori A and Taylor M** (2012) The impact of the national minimum wage on earnings, employment and hours through the recession.
- Card D and Krueger AB** (1995) *Myth and Measurement: The New Economics of the Minimum Wage*. Princeton, New Jersey: Princeton University Press.
- Coats D** (2007) The national minimum wage. Retrospect and prospect. The Work Foundation.
- Dolton P, Bondibene CR and Stops M** (2015) Identifying the employment effect of invoking and changing the minimum wage: a spatial analysis of the UK. *Labour Economics* **37**, 54–76.
- Duso T and Röller L-H** (2003) Endogenous deregulation: evidence from OECD countries. *Economics Letters* **81**, 67–71.
- Duverger M** (1951) The influence of the electoral system on political life. *International Social Science Bulletin* **3**, 3114–352.
- (EC), European Commission** (2004) The EU economy: 2003 review. Review 6/2003. Directorate General For Economic and Financial Affairs. Accessed October 2, 2017. Available at: https://ec.europa.eu/economy_finance/publications/pages/publication7694_en.pdf.
- Ferejohn J** (1986) Incumbent performance and electoral control. *Public Choice* **50**, 5–25.
- Firpo S and Possebom V** (2017) Synthetic control method: inference, sensitivity analysis and confidence sets. Accessed April 19, 2018. Available at: <https://www.sites.google.com/site/vitorapossebom/research>.
- Franzese RJ Jr** (2002) Electoral and partisan cycles in economic policies and outcomes. *Annual Review of Political Science* **5**, 369–421.
- Hall PA** (1993) Policy paradigms, social learning, and the state: the case of economic policymaking in Britain. *Comparative Politics* **25**, 275–296.
- Hellman JS** (1998) Winners take all: the politics of partial reform in postcommunist transitions. *World Politics* **50**, 203–234.
- Holland PW** (1986) Statistics and causal inference. *Journal of the American Statistical Association* **81**, 945–960.
- Hotz VJ, Imbens GW and Mortimer JH** (2005) Predicting the efficacy of future training programs using past experiences at other locations. *Journal of Econometrics, Experimental and Non-Experimental Evaluation of Economic Policy and Models* **125**, 241–270.
- Jacobs AM** (2016) Policy making for the long term in advanced democracies. *Annual Review of Political Science* **19**, 433–454.
- Kleis M and Moessinger M-D** (2016) The long-run effect of fiscal consolidation on economic growth: evidence from quantitative case studies. SSRN Scholarly Paper ID 2801590. Rochester, NY: Social Science Research Network, June 17. Accessed December 11, 2017. Available at: <https://papers.ssrn.com/abstract=2801590>.
- Koehler S and König T** (2015) Fiscal governance in the eurozone: how effectively does the stability and growth pact limit governmental debt in the euro countries? *Political Science Research and Methods* **3**, 329–351.
- Lijphart A** (1984) *Democracies: Patterns of Majoritarian and Consensus Government in Twenty-One Countries*. New Haven: Yale University Press.
- Low Pay Commission** (2016) National minimum wage: low pay commission report: Spring 2016. Accessed April 10, 2018. Available at: <https://www.gov.uk/government/publications/national-minimum-wage-low-pay-commission-reportspring-2016>.
- Martin LW and Vanberg G** (2011) *Parliaments and Coalitions: The Tole of Legislative Institutions in Multiparty Governance*. Oxford: Oxford University Press.
- Meyer BD** (1995) Natural and quasi-experiments in economics. *Journal of Business & Economic Statistics* **13**, 151–161.
- National Academies of Sciences, and Medicine, Engineering** (2017) The economic and fiscal consequences of immigration. Washington, DC: The National Academies Press.
- Nordhaus WD** (1975) The political business cycle. *The Review of Economic Studies* **42**, 169–190.
- OECD** (2014) *Labour Force Statistics by Sex and Age: Indicators*. Paris: Organisation for Economic Co-operation and Development, July. Accessed April 7, 2017. <https://www.oecd-ilibrary.org/content/data/data-00310-en>.
- Persson T and Tabellini G** (1999) The size and scope of government: comparative politics with rational politicians. *European Economic Review* **43**, 699–735.
- Pierson P** (1994) *Dismantling the Welfare State? Reagan, Thatcher, and the Politics of Retrenchment*. Cambridge, New Jersey: Cambridge University Press.
- Przeworski A** (1991) *Democracy and the Market: Political and Economic Reforms in Eastern Europe and Latin America*. Princeton, New Jersey: Cambridge University Press.

- Pyper D** (2014) The national minimum wage: historical background. Commons Briefing Papers, no. SN06897.
- Romer CD and Romer DH** (2004) A new measure of monetary shocks: derivation and implications. *American Economic Review* **94**, 1055–1084.
- Romer CD and Romer DH** (2010) The macroeconomic effects of tax changes: estimates based on a new measure of fiscal shocks. *American Economic Review* **100**, 763–801.
- Rubin DB** (1974) Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* **66**, 688–701.
- Sartori G** (1976) *Parties and Party Systems: a Framework for Analysis*. Princeton, New Jersey: Cambridge University Press.
- Schmitt J** (2013) Why does the minimum wage have no discernible effect on employment? CEPR Reports and Issue Briefs. Center for Economic and Policy Research (CEPR). Accessed April 7, 2017. Available at: <https://ideas.repec.org/p/epo/papers/2013-04.html>.
- Strom K, Müller WC and Bergman T** (2008) *Cabinets and Coalition Bargaining: The Democratic Life Cycle in Western Europe*. Comparative Politics. Oxford; New York: Oxford University Press.
- Sturn S** (2017) Do minimum wages lead to job losses? Evidence from OECD countries on low-skilled and youth employment. ILR Review: 0019793917741259.
- Tsebelis G** (2002) *Veto Players: How Political Institutions Work*. Princeton, New Jersey: Princeton University Press.
- Tufte ER** (1978) *Political Control of the Economy*. Princeton, New Jersey: Princeton University Press.
- Weaver RK** (2003) Cutting old-age pensions. In Pal LA and Weaver RK (eds), *The Government Taketh Away: The Politics of Pain in the United States and Canada*. Washington, DC: Georgetown University Press, pp. 41–70.
- Wright C** (2010) The regulation of European labour mobility: national policy responses to the free movement of labour transition arrangements of recent EU enlargements. *SEER: Journal for Labour and Social Affairs in Eastern Europe* **13**, 157–179.
- Xu Y** (2017) Generalized synthetic control method: causal inference with interactive fixed effects models. *Political Analysis* **25**, 57–76.