



MASTERARBEIT | MASTER'S THESIS

Titel | Title

Evaluating the Effectiveness of Austria's Kurzarbeit (Short-Time Work) Scheme during the COVID-19 Crisis

verfasst von | submitted by

Ishak Tosun

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of
Master of Science (MSc)

Wien | Vienna, 2026

Studienkennzahl lt. Studienblatt | Degree
programme code as it appears on the
student record sheet:

UA 066 913

Studienrichtung lt. Studienblatt | Degree
programme as it appears on the student
record sheet:

Masterstudium Applied Economics

Betreut von | Supervisor:

Univ.-Prof. Dipl.-Ing. Dr. Christine Zulehner Privatdoz.

Abstract

This thesis evaluates the labor-market impact of Austria's expanded short-time work scheme (Kurzarbeit) during the COVID-19 crisis. Using quarterly unemployment rates from 2017–2022, it asks how unemployment in Austria would have evolved after the first quarter of 2020 in the absence of the large-scale Kurzarbeit expansion. The analysis applies the Synthetic Control Method (SCM) to construct a "Synthetic Austria" from a weighted combination of European donor countries that matches Austria's pre-treatment unemployment dynamics and key macroeconomic predictors.

Inference and interpretation rely on model diagnostics rather than a single estimate. The thesis therefore combines pre-treatment fit assessment with placebo/permutation tests and robustness checks that vary donor pools, predictor sets and pre-treatment windows. Overall, SCM provides informative counterfactual trajectories, but estimated post-2020 gaps are sensitive to modeling choices – particularly the length of the pre-period and donor composition. The results underline both the usefulness and the limits of unemployment-based evaluation in a context where short-time work primarily stabilizes employment relationships.

Abstrakt

Diese Arbeit bewertet die Effekte der ausgeweiteten österreichischen Kurzarbeit während der COVID-19-Krise. Auf Basis quartalsweiser Arbeitslosenquoten für den Zeitraum 2017–2022 wird untersucht, wie sich die Arbeitslosigkeit in Österreich ab dem ersten Quartal 2020 entwickelt hätte, wenn die Kurzarbeit nicht in diesem Ausmaß ausgeweitet worden wäre. Methodisch wird dazu die Synthetic Control Method (SCM) verwendet: Dafür wird ein „synthetisches Österreich“ als gewichtete Kombination europäischer Donorländer konstruiert, das Österreich in der Vorperiode hinsichtlich der Entwicklung der Arbeitslosenquote sowie zentraler makroökonomischer Prädiktoren möglichst gut abbildet.

Die Interpretation stützt sich nicht auf einen einzelnen Schätzwert, sondern auf Modelldiagnostik. Dazu gehören die Bewertung der Anpassungsgüte vor dem Treatment, placebo- und permutationsbasierte Tests sowie Robustheitsanalysen, in denen Donor-Pools, Prädiktoren und Vorperioden variiert werden. Insgesamt zeichnet die SCM ein plausibles Bild davon, wie sich die Arbeitslosigkeit in Österreich ohne die Ausweitung der Kurzarbeit entwickelt hätte. Gleichzeitig zeigen die Ergebnisse, dass die geschätzten Effekte nach 2020 empfindlich darauf reagieren, wie das Modell spezifiziert wird – besonders auf die Wahl der Vorperiode und darauf, welche Länder als Vergleichsgruppe dienen. Genau darin liegt auch die zentrale Aussage der Arbeit: SCM ist ein nützliches Werkzeug für eine transparente Gegenüberstellung, stößt aber an Grenzen, wenn Arbeitslosigkeit nur indirekt auf Kurzarbeit reagiert, weil Kurzarbeit vor allem Beschäftigung hält statt Arbeitslosigkeit zu senken.

Contents

List of Tables	4
List of Figures	5
1 Introduction	7
2 Literature Review	9
2.1 Short-time work as a job retention tool: mechanisms and consensus	9
2.2 Evidence on short-time work during COVID-19	10
2.3 Unemployment as an outcome: why the evidence is less clear	12
2.4 Methodological literature: Synthetic Control Method	14
2.5 Research gap and positioning of this work	15
3 Methodology: Synthetic Control Method	16
3.1 Motivation for evaluating Austria's Kurzarbeit during COVID-19	16
3.2 Research design: Austria as treated unit and European donor pool	16
3.3 Outcome variable, treatment timing and empirical setup	17
3.4 Construction of Synthetic Austria: predictors and optimization	18
3.5 Identification assumptions in the Kurzarbeit context	19
3.6 Inference and robustness strategy	21
3.7 Advantages and limitations of SCM for short-time work evaluation	23
4 Data	24
4.1 Target variable: quarterly unemployment rate	24
4.2 Donor countries and sample	24
4.3 Data sources and data preparation	25
4.4 Data on short-time work and government-supported employment	26
4.5 Classification of countries without short-time work (STW = 0)	26
5 Results and inference	28
5.1 Comparison worlds and descriptive setup	28
5.2 Main result I: Austria vs. No-STW countries (V1)	30
5.3 Main result II: Austria vs. STW countries (V2)	32
5.4 Validation exercise: No-STW countries against STW donors (V3)	35
5.5 Robustness checks	40
5.6 Synthesis of findings	40
6 Discussion, limitations and policy implications	42
6.1 Discussion of the main findings	42
6.2 Limitations	42
6.3 Policy implications	44
7 Conclusion	45
8 List of References	47
Appendix A: Detailed robustness and additional specification checks	50
Appendix B: Extensive pre-treatment summary	66

List of Tables

Table 5.1: Classification of Countries	29
Table 5.2: Pre-treatment summary	30
Table 5.3: Main SCM specification – V1: AT vs. No-STW donor pool	30
Table 5.4: Donor weights – V1: AT vs. No-STW donor pool	32
Table 5.5: Main SCM specification – V2: AT vs. STW donor pool	33
Table 5.6: Donor weights – V2: AT vs. STW donor pool	35
Table 5.7: Main SCM specification – V3: RO vs. STW donor pool	36
Table 5.8: Donor weights – V3: RO vs. STW donor pool	36
Table 5.9: Main SCM specification – V3: NO vs. STW donor pool	38
Table 5.10: Donor weights – V3: NO vs. STW donor pool	38
Table 5.11: Placebo-in-space results – V1: AT vs. No-STW donor pool	50
Table 5.12: Placebo-in-space results – V2: AT vs. STW donor pool	51
Table 5.13: Placebo-in-space results – V3: RO vs. STW donor pool	51
Table 5.14: Placebo-in-space results – V3: NO vs. STW donor pool	52
Table 5.15: In-time placebo summary – V1: AT vs. No-STW donor pool	52
Table 5.16: In-time placebo summary – V2: AT vs. STW donor pool	53
Table 5.17: In-time placebo summary – V3: RO vs. STW donor pool	53
Table 5.18: In-time placebo summary – V3: NO vs. STW donor pool	53
Table 5.19: Leave-one-out robustness – V1: AT vs. No-STW donor pool	56
Table 5.20: Leave-one-out robustness – V2: AT vs. STW donor pool	56
Table 5.21: Leave-one-out robustness – V3: RO vs. STW donor pool	56
Table 5.22: Leave-one-out robustness – V3: NO vs. STW donor pool	56
Table 5.23: Sensitivity analysis – V1: AT vs. No-STW donor pool	57
Table 5.24: Sensitivity analysis – V2: AT vs. STW donor pool	57
Table 5.25: Sensitivity analysis – V3: RO vs. STW donor pool	58
Table 5.26: Sensitivity analysis – V3: NO vs. STW donor pool	58
Table 5.27: Lagged outcome variants V1	62
Table 5.28: Lagged outcome variants V2	62
Table 5.29: Comparison to baseline V1	65
Table 5.30: Comparison to baseline V2	65

List of Figures

Figure 5.1: Observed and synthetic unemployment rate – V1: AT vs. No-STW donor pool	31
Figure 5.2: Gap observed minus synthetic unemployment rate – V1: AT vs. No-STW donor pool.....	31
Figure 5.3: Observed and synthetic unemployment rate – V2: AT vs. STW donor pool	33
Figure 5.4: Gap observed minus synthetic unemployment rate – V2: AT vs. STW donor pool	34
Figure 5.5: Observed and synthetic unemployment rate – V3: RO vs. STW donor pool	37
Figure 5.6: Gap observed minus synthetic unemployment rate – V3: RO vs. STW donor pool.....	37
Figure 5.7: Observed and synthetic unemployment rate – V3: NO vs. STW donor pool	39
Figure 5.8: Gap observed minus synthetic unemployment rate – V3: NO vs. STW donor pool.....	39
Figure 5.9: Left: In-time placebo (fake treat. 2018Q1) – V1: AT vs. No-STW donor pool / Right: In-time placebo (fake treat. 2019Q1) – V1: AT vs. No-STW donor pool	53
Figure 5.10: Left: In-time placebo (fake treat. 2018Q1) – V2: AT vs. STW donor pool / Right: In-time placebo (fake treat. 2019Q1) – V2: AT vs. STW donor pool	54
Figure 5.11: Left: In-time placebo (fake treat. 2018Q1) – V3: RO vs. STW donor pool / Right: In-time placebo (fake treat. 2019Q1) – V3: RO vs. STW donor pool	54
Figure 5.12: Left: In-time placebo (fake treat. 2018Q1) – V3: NO vs. STW donor pool / Right: In-time placebo (fake treat. 2019Q1) – V3: NO vs. STW donor pool	55
Figure 5.13: Left: Alt. pre-window (2015–2019) – V1: AT vs. No-STW donor pool / Right: Alt. pre-window (2016–2019) – V1: AT vs. No-STW donor pool.....	59
Figure 5.14: Left: Alt. pre-window (2015–2019) – V2: AT vs. STW donor pool / Right: Alt. pre-window (2016–2019) – V2: AT vs. STW donor pool	59
Figure 5.15: Left: Alt. pre-window (2015–2019) – V3: RO vs. STW donor pool / Right: Alt. pre-window (2016–2019) – V3: RO vs. STW donor pool.....	60
Figure 5.16: Left: Alt. pre-window (2015–2019) – V3: NO vs. STW donor pool / Right: Alt. pre-window (2016–2019) – V3: NO vs. STW donor pool.....	60
Figure 5.17: Left: Gap comparison for lagged outcome variants (V1) / Right: Gap comparison for lagged outcome variants (V2)	62
Figure 5.18: Left: Gap comparison for reduced specification V1 / Right: Gap comparison for reduced specification V2	64

1 Introduction

The COVID-19 pandemic triggered an abrupt economic downturn across Europe. In response, governments introduced or massively expanded labor-market policy measures to cushion the shock. Among these instruments, short-time work (STW) played a central role. By subsidizing temporary reductions in working hours, STW is intended to stabilize employment relationships, preserve worker-firm matches and avoid layoffs during periods of sudden but potentially temporary disruption.

Austria is a particularly relevant case in this context. During the pandemic, Kurzarbeit became the central labor-market policy instrument and involved substantial public expenditure. At the same time, the increase in Austrian unemployment remained moderate by international comparison. This raises an important empirical question: to what extent can Austria's comparatively limited rise in unemployment be linked to the large-scale expansion of Kurzarbeit and to what extent does it reflect other features of the crisis environment and the relevant comparison group?

This thesis addresses that question through a deliberately narrow empirical lens: the quarterly harmonized unemployment rate. This focus is justified because unemployment is both politically salient and internationally comparable in a harmonized form. At the same time, it comes with an important limitation. Short-time work primarily stabilizes employment relationships and shifts labor-market adjustment toward reduced hours rather than open unemployment. Workers on short-time work remain statistically classified as employed, even when labor input falls sharply. For that reason, unemployment is a meaningful but incomplete indicator of the labor-market effects of Kurzarbeit.

The main question of the thesis is therefore counterfactual: how would Austria's quarterly unemployment rate have evolved between 2020 and 2022 if the large-scale expansion of COVID-era Kurzarbeit had not taken place? Answering this question is empirically difficult. Austria did not introduce Kurzarbeit randomly, but expanded it in response to an exceptional shock that affected countries differently and interacted with national institutions, sector structures and broader policy responses. Simple cross-country comparisons are therefore unlikely to provide a convincing benchmark on their own.

To address this challenge, the thesis applies the SCM to construct a transparent counterfactual "Synthetic Austria" from a weighted combination of European donor countries. The contribution of the thesis lies in three elements. First, the measurement of short-time work is based on a strict harmonized Eurostat administrative definition of government-supported employment. Second, the analysis uses SCM to make the counterfactual explicit rather than relying on pooled comparisons or an implicit parallel-trends assumption. Third and most importantly, the empirical analysis is structured across different comparison worlds. Austria is evaluated both against countries classified as $STW = 0$ in the strict Eurostat sense and against countries that themselves used STW, while an additional validation exercise examines whether similar post-2020 gaps also arise for selected non-Austrian placebo-treated cases. This design does not eliminate

every identification challenge, but it makes donor-pool choice analytically explicit and sharpens interpretation in a setting where the counterfactual is itself contestable.

The thesis does not aim to claim one definitive model-invariant causal number for Austrian Kurzarbeit. Instead, it asks how Austria's estimated post-2020 unemployment gap changes when the counterfactual is constructed under different but substantively meaningful comparison worlds. Inference therefore relies not on a single estimate alone, but on a broader set of diagnostics including pre-treatment fit, post-treatment divergence, placebo evidence, donor concentration and robustness to alternative modeling choices.

The remainder of the thesis is structured as follows. Chapter 2 situates the study in the literature on short-time work, unemployment-based evaluation and the Synthetic Control Method and derives the specific research gap addressed here. Chapter 3 presents the methodological framework, the empirical design and the identification logic underlying the SCM approach. Chapter 4 describes the data, variable construction and donor-pool classification, with particular attention to the measurement of short-time work and the distinction between recorded STW and non-recorded STW countries. Chapter 5 presents the empirical results in a staged way, beginning with the relevant comparison worlds and descriptive setup, then turning to the Austria-based SCM results, the validation exercise and the main robustness evidence. Chapter 6 discusses the findings, limitations and policy implications. Chapter 7 concludes.

2 Literature review

This chapter situates the thesis in two strands of literature: first, the literature on STW and related job-retention schemes, with a particular focus on the COVID-19 crisis and second, the methodological literature on the SCM. The review proceeds in five steps. Section 2.1 discusses the economic mechanisms through which STW is expected to affect labor-market adjustment. Section 2.2 reviews the empirical evidence on short-time work during COVID-19, with particular emphasis on Germany and Austria. Section 2.3 explains why unemployment is a relevant but conceptually difficult outcome for evaluating STW. Section 2.4 briefly situates the SCM in the methodological literature and explains why it is suitable for the present question. Section 2.5 then derives the specific research gap addressed by this thesis. Overall, the literature suggests broad agreement on the employment-preserving role of STW, but much less clarity on unemployment effects in cross-country settings and in counterfactual designs that cleanly separate countries with and without short-time work.

2.1 Short-time work as a job retention tool: mechanisms and consensus

Short-time work and related job-retention schemes are designed to stabilize employment relationships during temporary downturns by subsidizing reductions in working time rather than separations. In economic terms, STW primarily operates on the intensive margin: firms reduce hours while workers formally remain employed. This differs from policy responses that mainly address the extensive margin, such as layoffs followed by unemployment insurance. The central mechanism underlying STW is labor hoarding. When a shock is large but expected to be temporary, firms may find it efficient to retain workers and reduce hours rather than dismiss them and later incur search, hiring, training and coordination costs once demand recovers. This logic is especially relevant where employment relationships embody firm-specific human capital or where matching frictions are substantial. Earlier contributions on STW already emphasized that these schemes can reduce cyclical job losses, but also warned that the number of subsidized jobs should not be interpreted mechanically as the number of jobs saved, because deadweight effects and free-riding are possible (Boeri and Brücker, 2011; Cahuc and Carcillo, 2011).

More recent work provides a stronger theoretical and empirical micro foundation for these mechanisms. Giupponi and Landais (2023) analyze STW as a subsidy to hour reductions in the presence of temporary firm-level shocks and show that such policies can support labor hoarding and improve welfare under plausible labor-market frictions. Their framework is particularly useful because it clarifies why STW may be effective in abrupt downturns while still involving non-trivial trade-offs. If shocks are temporary and re-matching is costly, subsidizing hours reductions can protect valuable matches and prevent inefficient separations. If by contrast the downturn reflects more persistent structural change, the same policy may slow down efficient reallocation. Balleer et al. similarly argue that STW can save jobs, but distinguish between a rule-based component that may stabilize employment and a discretionary component that can be much less effective. In other words, STW is not best understood as a uniformly beneficial instrument, but as a contingent policy whose performance depends on timing, targeting, institutional design and the nature of the shock.

This more general literature already suggests two important lessons for the present thesis. First, STW should be expected to affect employment relationships more directly than open unemployment, because workers on short-time work remain formally attached to firms even when hours collapse. Second, one should not expect a simple one-to-one relationship between program size and program effectiveness. Large take-up may reflect severe crisis exposure, generous eligibility, precautionary use by firms or all three at once.

2.2 Evidence on short-time work during COVID-19

The COVID-19 crisis triggered an unprecedented expansion of job-retention support across OECD countries. OECD assessments describe job-retention schemes as the main labor-market instrument used in many OECD countries to soften the immediate labor-market impact of the pandemic. According to the OECD, take-up of job-retention support peaked at around 20% of employment on average across OECD countries in April/May 2020, supporting approximately 60 million jobs at the height of the crisis – more than ten times the scale observed during the global financial crisis. Later OECD work further suggests that these schemes may have saved up to 21 million jobs during the initial phase of the crisis, while also stressing the need to adjust, better target and eventually phase out support as the crisis evolved. The comparative review by Eichhorst et al. reaches a similar conclusion: job-retention schemes were central to the first phase of the COVID shock, but their effectiveness depended heavily on design details, the speed of implementation, the burden-sharing between firms and the state and the ability to avoid locking resources into non-viable firms.

A useful contribution for understanding the heterogeneity of program effects is Cahuc, Kramarz and Nevoux (2021). Studying the French case, they argue that short-time work may have very different effects across firms depending on the severity of the revenue shock. In severely affected firms, STW may genuinely save jobs that would otherwise be destroyed. In less affected firms, however, the same instrument may primarily subsidize hours reductions that would have occurred without layoffs anyway, thereby generating windfall effects. This is highly relevant for the present thesis because it cautions against overly mechanical interpretation of aggregate program use. Even if STW is successful on average, the measured effect on unemployment may be blurred by targeting issues, composition effects and the fact that not every subsidized job is a saved job.

Germany is the most important benchmark case in the modern literature on short-time work. This is partly because Kurzarbeit already played a major role during the Great Recession and partly because Germany again relied on the instrument during COVID-19. Earlier work on the Great Recession already treated short-time work as one component of the “German labor market miracle”, although usually not the only one. Burda and Hunt (2011) emphasize that Germany’s mild employment response to the Great Recession reflected a broader institutional and behavioral constellation, while Brenke, Rinne and Zimmermann (2013) nevertheless describe short-time work as the “German answer” to the crisis and argue that unemployment would otherwise have risen much more strongly. Boeri and Brücker (2011) also conclude that STW reduced job losses during the Great Recession, but stress that the number of jobs saved was smaller

than the full-time-equivalent volume subsidized, again pointing to the risk of deadweight effects. Together, these earlier studies are useful because they show that the main policy debate around Kurzarbeit long predates COVID-19 and already revolved around the balance between labor-market stabilization and inefficient subsidy use.

COVID-era German evidence sharpens this debate considerably. Aiyar and Dao (2021) exploit variation across German states and estimate that unemployment in 2020Q2 would have been 2.9 percentage points higher in the absence of the expansion of Kurzarbeit. Their estimate implies that the program reduced unemployment by about one third at the trough of the crisis. They also show that a 10- percentage-point increase in take-up dampened the rise in the unemployment rate by about 1.1 percentage points for the median shock. This is one of the clearest pieces of evidence linking short-time work directly to the unemployment outcome rather than to employment stability in a broader sense. At the same time, the German literature has become more nuanced in recent years. Fitzenberger and Walwei (2023) emphasize the role of short-time work as a key crisis instrument and draw attention to institutional adaptation and implementation lessons during COVID-19. Kagerl (2024), using German evidence, also finds positive employment effects but documents substantial windfall effects: his calculations suggest that up to 500,000 jobs were saved by STW in 2020, while 8.75 million jobs were supported overall, equivalent to 3.7 million employment equivalents. This large gap between supported and plausibly saved jobs strongly reinforces the idea that program scale and program impact are not identical.

Austria is even more relevant for the present thesis, yet the literature on the Austrian case is both smaller and somewhat more descriptive than the German debate. Tamesberger and Theurl (2021) are especially important because they document in detail how Austria transformed an existing short-time work instrument into a pandemic- specific emergency scheme with simplified access, broad eligibility and unusually generous support. They characterize the Austrian COVID-19 model as one of the most generous in Europe and emphasize that this design, negotiated rapidly by the government and the social partners, turned short-time work from a selective labor- market instrument into a central pillar of crisis management. Their paper is also useful because it shows that the Austrian model was not static: even within the early phase of the pandemic, the scheme was repeatedly adjusted and the authors distinguish three key reform periods corresponding to changing lockdown conditions and policy learning.

What makes Austria particularly interesting from an evaluation perspective is therefore not only that it used short-time work extensively, but that it did so in a very broad and generous form. Compared with the pre-COVID Austrian model, the Corona-Kurzarbeit offered workers a much higher replacement rate of 80–90% of net earnings, extended coverage to apprentices, reimbursed firms not only for lost-hour wage costs but also for associated social-security contributions and related payroll charges through flatrate compensation and even allowed retroactive applications during key lockdown phases. Huemer, Kogler and Mahringer (2021) show that take-up reached unprecedented levels: by the end of April 2020, 1.033 million employees, or 29.5% of all active dependent

employees, were in short-time work. Tamesberger and Theurl similarly stress the enormous scale of the early program, noting that applications in the first three months covered around 1.3 million workers, almost 40% of the Austrian labor force. They also document strong sectoral heterogeneity, with especially high take-up in accommodation and food services and in arts-related activities, which underlines how closely Austrian short-time work was tied to the specific structure of the COVID shock. Taken together, these studies show why Austria is such a compelling case: it was not a marginal or niche user of STW, but one of the countries in which the instrument became a macroeconomically important mechanism of labor hoarding and employment stabilization.

The most important recent Austrian source is the 2025 evaluation of COVID-19 Kurzarbeit commissioned by the Austrian authorities. This study moves beyond institutional description and provides a much more direct assessment of program effects over the full pandemic period from March 2020 to June 2022, during which the Austrian scheme passed through five phases with changing rules. Its central conclusion is that Kurzarbeit made a substantial contribution to employment stabilization: according to the evaluation, around 388,000 jobs were preserved and workers whose employment relationships were maintained through short-time work remained better integrated into employment afterward than comparable workers who became unemployed during the pandemic. At the same time, the study explicitly reports sizeable deadweight effects: out of 581,000 subsidized full-time-equivalent jobs, about one third would likely have continued to exist even without the scheme. For the present thesis, this is a crucial point. It suggests that Austrian Kurzarbeit was highly relevant in labor-market terms, but it also cautions against interpreting the sheer scale of program use as evidence of proportional causal effectiveness. In this respect, the Austrian literature now points toward a more balanced conclusion: COVID-19 short-time work in Austria appears to have been an important stabilizer, but its quantitative assessment depends on the counterfactual used, the level of analysis considered and the distinction between gross program size and net employment effects.

Taken together, the COVID-era literature supports a nuanced consensus rather than a simple verdict. Across OECD work, Germany, Austria and other case studies, short-time work is generally seen as an important crisis instrument that helped cushion the initial shock and stabilize employment relationships. However, the same literature repeatedly stresses heterogeneity in effects, strong dependence on program design and targeting and the presence of deadweight or windfall effects. This means that the relevant empirical question is not whether STW existed or whether it was large, but how its effect can be identified credibly for a particular country and outcome.

2.3 Unemployment as an outcome: Why the evidence is less clear

Although the literature is broadly favorable to short-time work as a job-retention instrument, the case becomes more complicated once the outcome of interest is narrowed to the harmonized unemployment rate. The first reason is conceptual. Short-time work primarily stabilizes employment relationships by reducing hours rather than by

moving workers directly from unemployment into employment. Workers on short- time work generally remain classified as employed even when their effective labor input falls sharply. As a result, labor-market slack can show up in hours worked, underemployment, reduced productivity or labor hoarding instead of appearing as open unemployment. This means that a policy can be successful in preserving jobs and still produce only a modest or noisy effect on measured unemployment. OECD discussions of job-retention schemes explicitly acknowledge this difficulty and stress that reliable effect estimation is challenging in a crisis where multiple policies operate simultaneously and employment adjustment often occurs through margins other than formal job loss.

A second reason is empirical identification. COVID-19 was a synchronized international shock, but not a uniform one. Countries differed substantially in lockdown intensity, sector composition, tourism exposure, work-from-home feasibility, fiscal support, liquidity measures and the timing of reopening. The OECD explicitly notes that the synchronized nature of the crisis and the multiplicity of policy responses severely limit the possibility of providing reliable effect estimates for job-retention schemes in simple cross-country comparisons. In other words, even if unemployment rises less in one country than in another, that difference cannot immediately be attributed to short-time work alone. The surrounding crisis environment matters and so does the composition of the comparison group.

The German evidence illustrates both the promise and the limitation of unemployment-based evaluation. On the one hand, Aiyar and Dao show that unemployment in Germany would likely have risen much more strongly in the absence of Kurzarbeit, implying that unemployment can indeed respond strongly to job-retention support under the right empirical design. On the other hand, Kagerl's evidence on substantial windfall effects shows that participant numbers and unemployment effects should not be conflated. A scheme may preserve many matches and still not reduce open unemployment in a simple one-to-one way. The Austrian evidence points in the same direction. The 2025 evaluation emphasizes substantial employment stabilization but also sizeable deadweight effects, implying that even a labor-market instrument that clearly matters may have an unemployment impact that is limited, model-dependent or difficult to isolate.

A third reason concerns the counterfactual itself. In many international comparisons, the "control" countries also use short-time work or related job-retention schemes. In such a setting, the relevant comparison is not "with STW versus without STW," but rather "one STW design versus another bundle of labor-market institutions." This is precisely where the present thesis positions itself. Instead of relying on a single undifferentiated donor pool, it explicitly distinguishes between comparison worlds based on countries with STW = 0 in the strict Eurostat administrative sense and comparison worlds consisting of STW-using countries. The aim is not to claim that this fully solves the identification problem, but to make the contamination issue visible and analytically explicit.

For these reasons, unemployment remains a relevant but deliberately narrow outcome in this thesis. It is relevant because it is politically salient, internationally comparable in

harmonized form and central to public discussion of labor-market stabilization. It is narrow because it cannot capture all the margins through which short-time work operates. A modest or unstable unemployment effect therefore does not imply that STW was ineffective. It may instead indicate that adjustment occurred mainly through hours reductions, preserved matches and hidden labor slack rather than through observable job destruction.

2.4 Methodological literature: Synthetic Control Method

The second strand of literature relevant to this thesis concerns the Synthetic Control Method. SCM was originally introduced by Abadie and Gardeazabal (2003) and further developed by Abadie, Diamond and Hainmueller (2010, 2011) as a transparent comparative case-study approach for settings with one treated aggregate unit and a donor pool of potential controls. The key idea is to construct a weighted combination of untreated units that reproduces the treated unit's pre-intervention characteristics as closely as possible. Post-treatment divergence between the treated unit and its synthetic control is then interpreted as the estimated effect of the intervention, provided that the synthetic unit offers a credible approximation to the unobserved counterfactual. Athey and Imbens (2017) describe synthetic control as one of the major innovations in modern applied econometrics for policy evaluation, precisely because it offers a transparent alternative to more implicit regression-based comparisons. Abadie (2021) further systematizes the method by emphasizing feasibility, data requirements and contextual conditions under which synthetic controls are most likely to be informative.

For the present thesis, SCM is attractive for several reasons. First, Austria is a single treated unit, which fits the comparative case-study setting for which SCM was designed. Second, the COVID-era expansion of Austrian Kurzarbeit is a large aggregate intervention that cannot plausibly be evaluated through random assignment.

Third, SCM forces the counterfactual to be “earned” through pre-treatment fit. This is helpful in a setting where labor-market adjustment is shaped by many time-invariant or slowly moving features – such as institutions, sector structure and macroeconomic exposure – that are difficult to model fully in a standard regression framework. In addition, SCM is particularly well suited to a thesis that wants to make donor-pool choice explicit. Because the synthetic control is an observable weighted combination of donor countries, one can directly examine donor dependence, weight concentration and the plausibility of different comparison worlds.

At the same time, the methodological literature also highlights important limitations. SCM is most convincing when the treated unit can be matched well in the pre-treatment period, when the donor pool is sufficiently rich and when post-treatment divergence is not plausibly driven by omitted Austria-specific shocks. Abadie (2021) stresses that the method is not a magic solution: it requires enough data and a context in which a credible synthetic counterfactual can actually be constructed. This is particularly relevant here because the pandemic involved multiple simultaneous shocks and policy responses. In such a context, inference should not rely on a single point estimate alone. Instead,

placebo exercises, sensitivity checks, donor-weight diagnostics and alternative pre-treatment windows become essential parts of the empirical argument. This is exactly the logic adopted in the present thesis.

2.5 Research gap and positioning of this work

The literature reviewed above suggests three broad conclusions. First, there is strong general and COVID-specific evidence that short-time work can stabilize employment relationships in abrupt downturns. Second, the effectiveness of STW depends heavily on program design, targeting and the broader economic environment. Third, even where STW appears successful, its impact on the unemployment rate is not straightforward to identify because unemployment is only one margin of labor-market adjustment and because international counterfactuals are hard to define cleanly. These themes appear consistently in OECD assessments, in the German literature on *Kurzarbeit*, in the Austrian case studies and in more general work on STW as a labor- hoarding instrument.

What remains less settled is how these insights translate into a transparent cross- country counterfactual for Austria when the harmonized unemployment rate is used as the main outcome. Germany has by now been studied relatively intensively, including with within-country designs that exploit regional or firm-level variation. Austria, by contrast, has more often been analyzed in institutional, descriptive or policy-evaluative terms than through a clearly structured cross-country counterfactual design. Moreover, many existing cross-country comparisons blur the distinction between countries that operated no recorded short-time work and countries that used their own job-retention schemes. As a result, the donor pool can be “contaminated” in a way that makes causal interpretation less clear.

This thesis addresses that gap in three ways. First, it evaluates Austria using the harmonized unemployment rate, a deliberately narrow but internationally comparable labor-market outcome. Second, it applies the Synthetic Control Method to construct a transparent counterfactual rather than relying on a single pooled comparison or an implicit parallel-trends assumption. Third and most importantly, it explicitly separates different donor-pool comparison worlds: one based on countries with $STW = 0$ in the strict Eurostat administrative sense and another based on countries that also used STW. This does not eliminate every identification challenge, but it sharpens interpretation by distinguishing a more credible “without-STW” benchmark from a broader peer-group comparison. In that sense, the contribution of the thesis lies less in claiming one definitive causal number and more in improving interpretability in a setting where donor-pool choice is itself a substantive modeling decision.

3 Methodology: Synthetic Control Method

This chapter sets out the empirical strategy used to evaluate Austria's COVID-19 short-time work scheme. Since the central question of the thesis is explicitly counterfactual – how Austria's unemployment rate would have evolved without the large-scale expansion of Kurzarbeit – the chapter first explains why a simple cross-country comparison is not sufficient and why the SCM is an appropriate design for this setting. It then describes the research design, defines the outcome, treatment timing and formal setup and explains how “Synthetic Austria” is constructed from donor-country weights and predictor variables. The chapter also discusses the main identification assumptions in the specific context of COVID-19 and short-time work, outlines the inference and robustness strategy and concludes with a brief assessment of the strengths and limitations of SCM for the present research question. Overall, the purpose of the chapter is to make clear not only how the method is implemented, but also under which conditions its results can be interpreted meaningfully.

3.1 Motivation for evaluating Austria's Kurzarbeit during COVID-19

Evaluating Austria's COVID-19 Kurzarbeit poses a difficult identification problem. The program was not introduced randomly, but expanded rapidly in response to an extraordinary macroeconomic shock. At the same time, labor-market outcomes during the pandemic were shaped by many overlapping influences, including lockdown intensity, sectoral exposure, fiscal support, administrative capacity and pre-existing institutional differences. In such a setting, simple cross-country comparisons are unlikely to isolate the specific role of short-time work in a convincing way.

This is the main reason why the thesis applies the Synthetic Control Method. SCM is designed for settings with one treated aggregate unit and a set of potential comparison units. Rather than assuming that Austria would have followed the same trend as an average control group, the method constructs a data-driven counterfactual from a weighted combination of donor countries that best reproduces Austria's pre-treatment unemployment dynamics and predictor profile. The identifying intuition is straightforward: if Austria is matched well before 2020, a post-2020 divergence between Austria and its synthetic counterpart can be interpreted as evidence consistent with a treatment effect, provided that no major Austria-specific confounding shock drives the gap independently of Kurzarbeit.

In the Austrian short-time work context, this feature is especially valuable because it forces the counterfactual to be earned through pre-treatment fit. This makes SCM particularly suitable for a thesis that aims to evaluate Austria transparently while also taking donor-pool choice and model stability seriously.

3.2 Research design: Austria as treated unit and European donor pool

Austria is the treated unit in this study. The donor pool consists of European countries that can plausibly serve as comparison units and for which consistent data on outcomes and predictors are available. The synthetic control (“Synthetic Austria”) is defined as a

weighted average of these donor countries, with weights chosen to best replicate Austria's pre-treatment characteristics. The principal estimand is the difference between Austria's observed post-treatment outcomes and the outcomes predicted by Synthetic Austria.

The empirical focus of this thesis differs from much of the earlier SCM literature on cross-country institutional changes such as EU membership (e.g., Campos et al., 2014, 2019; Garoupa and Spruk, 2020). Instead of aggregate income or productivity effects, the central outcome dimension here is labor market stabilization during a deep crisis. Kurzarbeit is intended to reduce separations, preserve job matches and dampen unemployment fluctuations by subsidizing hours reductions.

3.3 Outcome variables, treatment timing and empirical setup

The empirical analysis uses the quarterly harmonized unemployment rate as the main outcome variable. This measure is available at quarterly frequency, is comparably defined across countries and can therefore be used in a consistent SCM framework. The treatment is defined as the activation and more importantly, the rapid large-scale expansion of Austria's COVID-19 Kurzarbeit scheme. The intervention date is set to 2020Q1, which marks both the onset of the pandemic shock in Europe and the beginning of Austria's extraordinary labor-market policy response. Periods up to 2019Q4 are therefore treated as pre-intervention observations, while periods from 2020Q1 onward belong to the post-treatment phase.

Formally, let there be $(J+1)$ countries observed over T time periods. Austria is indexed by $j = 1$, while donor countries are indexed by $j=2, \dots, J+1$. Let Y_{jt} denote the outcome variable for country j at time t . In this thesis, Y_{jt} is the quarterly harmonized unemployment rate. The treatment effect for Austria at time t can then be written as

$$\tau_{1t} = Y_{1t}^I - Y_{1t}^N,$$

where Y_{1t}^I is the observed outcome for Austria under the COVID-19 Kurzarbeit regime and Y_{1t}^N is the counterfactual outcome that Austria would have experienced in the absence of treatment.

Since Y_{1t}^N is unobserved, SCM approximates it by constructing a synthetic control:

$$\hat{Y}_{1t}^N = \sum_{j=2}^{J+1} w_j Y_{jt},$$

where the weights satisfy $w_j \geq 0$ and $\sum_{j=2}^{J+1} w_j = 1$. These constraints imply that $j=2$

Synthetic Austria is a convex combination of real donor countries. This is useful because it avoids extrapolation and keeps the counterfactual institutionally interpretable. The main estimand is therefore the difference between Austria's observed post-2020 unemployment path and the post-2020 path predicted by Synthetic Austria.

3.4 Construction of Synthetic Austria: predictors and optimization

Synthetic Austria is constructed by choosing donor-country weights such that Austria is matched as closely as possible in the pre-treatment period. This matching is based on two elements: first, the pre-treatment path of the outcome itself and second, a set of predictor variables that capture relevant macroeconomic and labor-market characteristics. The basic idea is straightforward. A synthetic control is more credible if it does not only reproduce Austria's unemployment rate before 2020, but also resembles Austria in terms of the broader conditions under which labor-market adjustment takes place. In the context of this thesis, those conditions include cyclical activity, labor-market utilization, external exposure, hours adjustment, output dynamics and general macroeconomic conditions.

The baseline predictor set therefore includes the GDP index, the employment rate, trade openness, average hours worked, the industrial production index and inflation. These variables are not chosen arbitrarily. The GDP index serves as a broad indicator of overall economic activity and captures the general demand environment. The employment rate complements unemployment by reflecting labor-market utilization more broadly. Trade openness is included because Austria is a small open economy and external demand conditions are likely to matter for labor-market outcomes, especially during a crisis that affected countries asymmetrically. Average hours worked are particularly relevant in the short-time work context because STW is designed precisely to shift part of the adjustment from the extensive margin to the intensive margin. The industrial production index captures cyclical output movements, which were highly relevant during the pandemic shock. Inflation is included as a broad indicator of macroeconomic conditions and demand pressure.

This selection is intended to capture both cyclical and structural components of Austria's labor-market environment without overloading the model. In practical terms, SCM selects the donor-weight vector

$$W = (w_2, \dots, w_{J+1})$$

so as to minimize the distance between Austria's predictor vector and the synthetic predictor vector in the pre-treatment period. Intuitively, the method searches for the convex combination of donor countries that reproduces Austria's pre-treatment unemployment dynamics and predictor profile as closely as possible. When Austria and Synthetic Austria match well before 2020, the synthetic control becomes a more credible approximation of the missing counterfactual path after 2020. Good pre-treatment fit is therefore not only a descriptive feature of the model, but a core requirement for interpretation.

A further clarification is important for the interpretation of the predictor set. Variables such as GDP, employment rate, industrial production and inflation may themselves react to the crisis environment and at least indirectly to short-time work policies. For that reason, they are not interpreted here as exogenous right-hand-side variables in a structural

unemployment regression. In this thesis, they function as matching variables rather than structural regressors. Their purpose is to improve pre-treatment comparability between Austria and the donor pool by summarizing the broader macroeconomic and labor-market conditions under which unemployment develops. The empirical objective is therefore not to estimate contemporaneous causal coefficients, but to construct a synthetic control that resembles Austria as closely as possible before treatment.

The temporal ordering of these variables nevertheless matters. Labor-market adjustment is often delayed and unemployment may react only gradually to changes in output, production or demand conditions. This is especially plausible in the context of short-time work, where the immediate response to a shock may be absorbed through hours reductions rather than layoffs. The baseline specification should therefore not be read as implying a simple contemporaneous relationship between all predictors and unemployment. Instead, it represents a parsimonious matching design. To address the timing issue more directly, the robustness analysis includes alternative specifications that incorporate lagged unemployment outcomes from earlier pre-treatment years. In addition, Appendix A reports these lagged variants alongside a reduced specification without additional matching variables beyond the outcome path in order to assess how strongly the results depend on the baseline predictor design.

The choice of predictors also reflects a trade-off between richness and parsimony. On the one hand, a richer predictor set may improve comparability. On the other hand, too many predictors can make the model unnecessarily complex and may reduce stability, especially in a setting with a limited donor pool. For that reason, the baseline specification focuses on a compact set of variables that are economically interpretable and directly relevant for labor-market adjustment. More detailed structural measures, such as sectoral employment shares, are treated as extensions rather than as part of the baseline. This is consistent with the broader design principle of the thesis: the main specification should be transparent, economically plausible and comparable across the different donor-pool variants considered later in the thesis.

Finally, the optimization step should be read together with the later robustness analysis. A good pre-treatment match is necessary, but it is not sufficient on its own. Even when the synthetic control fits Austria well before treatment, the result can still be fragile if donor weights are highly concentrated or if estimated post-treatment gaps change substantially under alternative predictor sets or alternative pre-treatment windows. For that reason, the construction of Synthetic Austria is only the first step. Its credibility is evaluated later through placebo evidence, donor-sensitivity checks and specification changes.

3.5 Identification assumptions in the Kurzarbeit context

As in other SCM applications, identification in this thesis does not rest on a single formal assumption alone, but on a combination of design choices and contextual plausibility. In the present setting, the central identifying idea is that if Synthetic Austria reproduces Austria's unemployment dynamics and predictor profile well before 2020, then post-2020 deviations can be interpreted as evidence of a treatment effect, provided that these

deviations are not mainly driven by other Austria-specific shocks or by instability in the donor-based counterfactual. In the Kurzarbeit context, several assumptions are therefore particularly important.

First, the donor countries should not be materially affected by Austria's Kurzarbeit expansion itself. This is the usual no-spillover requirement. In practical terms, Austria's policy should not strongly alter labor-market outcomes in donor countries through channels such as substantial cross-border labor substitution, institutional coordination or direct policy spillovers. In the present application, this assumption appears plausible enough, although it cannot be guaranteed perfectly. Austria is economically integrated with its neighbors, but its national Kurzarbeit expansion is unlikely to have mechanically changed unemployment trajectories in the donor countries in a way large enough to dominate the SCM results.

Second, anticipation should not play a major role. Donor countries should not adjust their labor-market behavior or statistical reporting in anticipation of Austria's intervention. In a setting with long legislative lead times this could be a serious issue, but in the present case it is less concerning. The onset of the pandemic and the labor-market response in early 2020 were abrupt and the large-scale expansion of Kurzarbeit took place in a crisis environment rather than after a long pre-announced reform period. This does not mean that there was literally no anticipation anywhere, but it makes anticipation less problematic than in many standard policy-evaluation settings.

Third, post-treatment gaps should not be driven mainly by Austria-specific confounding shocks unrelated to Kurzarbeit. This is one of the most important assumptions in this thesis and also one of the hardest to satisfy perfectly. During COVID-19, many policies were implemented at the same time, including lockdown measures, fiscal transfers and other labor-market interventions. It would therefore be too strong to claim that the SCM gap isolates Kurzarbeit in a perfectly clean way. A more realistic interpretation is narrower: the estimated gap captures the effect of Austria's Kurzarbeit expansion within the broader Austrian crisis-policy environment, to the extent that other relevant influences were either broadly similar across countries or indirectly absorbed through pre-treatment matching. This wording is important, because it is more honest and methodologically more defensible than treating the estimate as a fully isolated policy coefficient.

Fourth, the donor pool should remain reasonably stable over the study window. Donor countries should not experience large idiosyncratic institutional changes that fundamentally alter their labor-market adjustment process independently of the pandemic shock. If major donor-specific reforms or measurement breaks occurred during the analysis period, they could distort the synthetic counterfactual and make post-treatment gaps harder to interpret. This assumption is again not absolute, but it becomes more credible when the donor pool is chosen carefully and when highly unstable donor constructions are treated with caution in the robustness analysis.

Fifth, a sufficiently good and sufficiently stable pre-treatment fit is required. In SCM, this is part of the identification logic itself. If Synthetic Austria fails to reproduce Austria's pre-treatment path convincingly, then post-treatment divergence cannot be interpreted with much confidence. This is especially relevant in a labor-market application with quarterly data, possible seasonal patterns and pandemic-related breaks. For that reason, the thesis places strong emphasis on pre-treatment fit and on sensitivity to alternative pre-treatment windows. The goal is not simply to achieve a low RMSPE mechanically, but to ensure that the counterfactual is credible under a stable pre-treatment regime.

Overall, these assumptions imply a deliberately careful interpretation. The SCM estimates in this thesis should not be read as providing a fully model-free and perfectly isolated causal effect of Austrian Kurzarbeit. Rather, they provide disciplined counterfactual evidence on how Austria's unemployment path evolved relative to synthetic comparison units under different donor-pool constructions. This is still highly informative, but it requires that interpretation remains tied to pre-treatment fit, donor composition, placebo behavior and stability across specifications.

3.6 Inference and robustness strategy

SCM is not primarily designed for standard large-sample parametric inference. Instead, inference is usually based on case-specific diagnostics, placebo exercises and structured robustness checks. This is particularly important in the present thesis, because the credibility of the estimated post-2020 gap cannot be judged from one single number alone. A convincing result in this setting requires several things at the same time: a good pre-treatment fit, a meaningful post-treatment divergence, a donor structure that is not excessively fragile and a pattern that does not disappear once plausible alternative specifications are considered. For that reason, the interpretation in this thesis is deliberately based on a broader diagnostic strategy rather than on one formal significance test.

A first element of this strategy is the use of in-space placebo tests. In these exercises, the treatment is reassigned iteratively to donor countries and synthetic controls are constructed for those placebo-treated units using the remaining donor pool. This yields a distribution of placebo gaps and allows Austria's estimated post-treatment deviation to be compared with deviations that also arise for untreated units. The purpose is not to claim classical statistical significance in the usual regression sense, but to ask whether Austria's post-2020 divergence looks unusually large relative to units that were not actually treated. In the context of this thesis, this is especially useful because the pandemic itself generated large labor-market movements across countries, so a simple post-treatment gap is not informative unless it is seen relative to a broader placebo distribution. At the same time, placebo evidence must be interpreted carefully when only a small number of donor units remain after filtering by pre-treatment fit. In such cases, pseudo-p values are informative but should not be overinterpreted.

A second element is the explicit assessment of pre-treatment fit. The root mean squared prediction error (RMSPE) in the pre-treatment period is used to measure how closely

Synthetic Austria reproduces Austria before the intervention. This is a central part of the empirical logic of SCM. If Austria and its synthetic control do not track each other well before 2020, then any post-2020 divergence is difficult to interpret. For that reason, the analysis also reports the ratio of post-treatment RMSPE to pre-treatment RMSPE. A relatively high ratio indicates that the divergence becomes much larger after treatment than before treatment. This is consistent with a treatment-related break, although it is still not sufficient on its own. A high ratio can also reflect temporary volatility, donor-specific post-period dynamics or a model that was already fragile before the intervention. It therefore has to be interpreted together with the size of the average gap, donor concentration and the additional robustness checks.

A third element is donor sensitivity, especially through leave-one-out exercises. These tests examine whether the result depends disproportionately on a single donor country. This is particularly relevant in SCM applications where the donor weights are concentrated. If the estimated post-treatment effect changes sharply once one dominant donor is removed, then the result may be well fitting but still fragile. In the present thesis, this issue is especially important because one of the central model variants deliberately uses a narrow donor pool based on countries with STW = 0 in the Eurostat sense. In such a setting, donor sensitivity is not a side issue but part of the core interpretation. A model with strong pre-fit but extreme donor dependence has to be read more cautiously than a model with somewhat weaker conceptual purity but a broader and more stable donor structure.

A fourth element is specification sensitivity. This includes predictor-one-out exercises as well as alternative pre-treatment windows. The predictor-one-out checks show whether the estimated results are mechanically driven by one specific matching variable. If the sign and broad magnitude of the estimated effect remain similar when individual predictors are excluded, this supports the stability of the result. Alternative pre-treatment windows are equally important. In this thesis, pre-periods such as 2016–2019 and 2015–2019 are considered in addition to the baseline 2017–2019 window. This is necessary because the stable pre-treatment regime is itself a substantive modeling choice. If estimated effects change strongly when the pre-period is extended, this suggests that the result is sensitive to how Austria's pre-COVID trajectory is represented. In the present context, that is not a marginal issue but one of the central empirical diagnostics.

Overall, inference in this thesis is therefore deliberately cumulative. The estimated treatment effect is not judged by one statistic alone, but by a combination of pre-treatment fit, post-treatment divergence, placebo evidence, donor sensitivity and specification robustness. This is also consistent with the broader logic of the thesis as stated in the abstract and in the results discussion: SCM provides informative counterfactual trajectories, but the estimated post-2020 deviations react meaningfully to modeling choices, especially donor composition and pre-treatment length.

More detailed specification checks are reported in Appendix A, including models that incorporate lagged unemployment outcomes from earlier pre-treatment years and a

reduced specification that excludes additional matching variables beyond the lagged outcome path.

3.7 Advantages and limitations of SCM for short-time work evaluation

SCM is well suited for evaluating Kurzarbeit in a cross-country setting for three reasons. First, by matching the pre-treatment outcome path and key predictors, it can absorb time-invariant or slowly moving confounders that are difficult to measure directly (e.g. institutional features or administrative capacity). Second, SCM avoids extrapolation: the convexity constraints ensure that the counterfactual is a weighted average of observed countries, which improves plausibility in an institutional European context. Third, the method is transparent because the implied donor weights can be reported and assessed for substantive comparability.

These strengths come with important limitations. Results depend on data quality and cross-country comparability and STW-related measures may differ in definition or reporting over time. SCM can also be sensitive to donor-pool composition; when few comparable donors exist, weights may become concentrated and estimates fragile. Finally, SCM delivers case-specific evidence since it identifies the Austrian counterfactual under a particular comparison world rather than a general “European average effect” absent additional assumptions.

4 Data

This chapter describes the data used in the empirical analysis and explains how the main variables and comparison groups are constructed. Because the thesis relies on a cross-country evaluation of Austria, data consistency and transparent variable definition are particularly important. The chapter therefore proceeds in several steps. It first introduces the main outcome variable, the quarterly harmonized unemployment rate and explains the time window used in the baseline analysis. It then discusses the donor-country sample and the logic of the donor-pool restrictions, especially the distinction between countries with and without recorded short-time work in the strict Eurostat sense. After that, the chapter presents the matching variables used in the SCM, outlines the data sources and preparation steps and explains in more detail how Eurostat's information on government-supported employment is used to measure short-time work intensity. The chapter concludes by clarifying the classification of countries in the STW = 0 group. Overall, the aim is not only to document the data, but also to make transparent the measurement choices that shape the interpretation of the counterfactuals later in the thesis.

4.1 Target variable: quarterly unemployment rate

The main outcome variable in this thesis is the quarterly harmonized unemployment rate. It is chosen because it is available at quarterly frequency over a sufficiently long horizon and because it is measured in a harmonized way across countries, which is essential for the cross-country SCM design used in this study. In substantive terms, it is also a natural indicator because unemployment is central to both policy debates and public assessments of labor-market stabilization during the COVID-19 crisis.

At the same time, the unemployment rate should be interpreted as a deliberately narrow outcome. In the context of short-time work, labor-market adjustment may occur through reduced working hours, underemployment or the preservation of worker-firm matches without immediately appearing in measured unemployment. Workers on short-time work generally remain classified as employed for statistical purposes, even when effective labor input falls sharply. For that reason, the unemployment rate captures only part of the labor-market role of Kurzarbeit and should be read as a politically salient but incomplete indicator.

For the baseline analysis, the main estimation window covers 2017–2022, with 2017–2019 treated as the pre-intervention period and 2020–2022 as the post-intervention period. This window is chosen to focus on a recent and comparatively stable pre-treatment phase while capturing the main COVID-era years in the post-period.

4.2 Donor countries and sample

The donor pool consists exclusively of European countries. This restriction serves two purposes. First, it improves institutional and economic comparability. Labor-market institutions, welfare systems and crisis policy responses differ widely across the world and extending the donor pool beyond Europe would likely introduce a large amount of variation that is difficult to justify substantively. Second, the European restriction improves data consistency. Even within Europe, information on government-supported

employment and short-time work is not perfectly uniform. Outside Europe, comparable administrative information is often even more limited or unavailable.

Countries are included in the donor pool only if they have sufficiently complete data for the main outcome and matching variables in the relevant pre-treatment period. This is necessary because SCM requires a consistent pre-treatment information set in order to construct a meaningful synthetic control. Countries with major missingness or without a sufficiently comparable macroeconomic profile are therefore excluded. The purpose of this restriction is not to create an artificially narrow sample, but to ensure that the synthetic counterfactual is built from countries that can plausibly contribute to a meaningful comparison.

4.2.1 Donor pool restriction according to short-time work intensity

A key operational step in the data design is the classification of donor countries according to recorded short-time work exposure in the Eurostat administrative data. This classification is used to implement the different comparison worlds analyzed later in Chapter 5.

In the first model variant, Austria is compared only with countries classified as STW = 0 in the strict Eurostat sense. This yields the sharpest possible untreated benchmark within the available data framework. In the second Austria-based variant, the donor pool consists of countries with recorded STW use, which provides a broader peer-group comparison rather than a pure untreated counterfactual. In the validation exercise, selected No-STW countries are treated as placebo-treated units and matched against STW-using donors.

The purpose of this classification is therefore not to declare one donor-pool definition universally superior, but to implement transparent and substantively distinct counterfactual settings within a common empirical framework.

4.3 Data sources and data preparation

The analysis is based almost exclusively on Eurostat. Since the thesis is interested in cross-country comparison and in a transparent distinction between countries with and without recorded STW, it is preferable to use one harmonized source wherever possible rather than combine many national sources with potentially different definitions. Eurostat provides the main labor-market and macroeconomic variables needed for the analysis and also supplies the administrative data used to classify countries according to government-supported employment during COVID-19. In terms of data handling, part of the data is processed directly in R, while some series are prepared in Excel before being imported into R. All relevant transformations, variable constructions and classification decisions are documented so that the final dataset can be traced back to its source and preparation steps. This is particularly important for the STW classification, where the interpretation depends directly on how Eurostat categories are read and translated into the empirical design. The overall objective of the data preparation strategy is therefore not only convenience, but traceability and replicability.

4.4 Data on short-time work and government-supported employment

4.4.1 Data basis and content

The data on short-time work and government-supported employment come from Eurostat's administrative file on the number and proportion of jobs benefiting from governmental measures. The data are based on administrative reporting by national authorities and were assembled by Eurostat in order to provide internationally comparable information on labor-market support during the COVID-19 period. This dataset is especially important for the present thesis because it provides a harmonized basis for distinguishing countries that recorded STW-type support from countries that did not.

The Eurostat file contains several pieces of information, including the total number of jobs applied for, the number of jobs approved, the number of jobs actually used and where available, the share of total jobs in the economy that are supported by government measures. For the present study, the most relevant measure is the proportion of government-supported jobs actually used relative to total employment. This variable is used as the operational measure of STW intensity. It is preferable to mere approvals or applications because it is closer to actual program utilization and therefore more meaningful for classifying the extent of effective support.

4.4.2 Definitions (Eurostat)

Eurostat distinguishes between two main forms of government-supported employment in the COVID-related administrative file. Category (A) covers short-time work and similar arrangements in which employees continue to work reduced hours and receive public compensation for hours not worked. Category (B) covers temporary layoffs in which the employment relationship formally remains in place but employees are temporarily not working and still receive government-supported compensation. The total number of government-supported jobs, category (C), is the sum of these two components.

For the purposes of this thesis, both categories are relevant because they capture labor-market support arrangements that preserve the employment relationship and therefore belong to the broader family of job-retention measures. At the same time, the strict use of Eurostat's definitions is important. The classification employed here is not based on national labels or political descriptions of labor-market policy. It is based on whether Eurostat records positive values in categories (A) and (B). This creates a harmonized administrative benchmark that can be applied consistently across countries, even if national systems differ in legal design or political terminology. Government-supported jobs are therefore understood here as employment relationships in which workers remain attached to employers and receive government-backed financial support either under reduced hours or under temporary non-work arrangements.

4.5 Classification of countries without short-time work (STW = 0)

Based on the harmonized Eurostat data, Romania, Czechia and Norway are classified as countries without short-time work in the period 2020–2022. This classification follows Eurostat's narrow administrative definition of government-supported employment and

should therefore not be confused with a broader substantive judgment that these countries had no labor-market support at all. The point is narrower and more precise: according to the Eurostat file used in this thesis, these countries report zero jobs in categories (A) and (B) over the relevant period. Under the measurement framework applied here, they therefore belong to the $STW = 0$ group.

This distinction is particularly important for Czechia and Norway. In both countries, labor-market support measures did exist during the pandemic, but they were structured differently and do not appear in Eurostat's categories for short-time work or temporary layoffs with continued employment. Instead, they took forms such as direct subsidies, transfers or insurance-based support that fall outside the narrow administrative definition used here. This is exactly why the wording in the thesis should remain careful. The claim is not that these countries had no labor-market intervention. The claim is that they had no recorded STW under the harmonized Eurostat categories that define the empirical comparison in this thesis.

Using this strict classification has two advantages. First, it improves cross-country consistency by avoiding ad hoc national reinterpretations of labor-market policy labels. Second, it helps address the contamination problem discussed earlier. If the donor pool were built without a clear distinction between recorded STW and non-recorded STW, the resulting comparison would be harder to interpret. The strict $STW = 0$ definition therefore strengthens transparency, even if it comes at the cost of a smaller and sometimes more fragile donor pool. This is also one of the main reasons why the thesis later works with multiple model variants rather than insisting on one single donor-pool definition.

5 Results and inference

5.1 Comparison worlds and descriptive setup

The empirical analysis is organized around three comparison worlds. This structure reflects the fact that donor-pool choice is not a purely technical issue in the present setting, but part of the substantive identification problem. If the research question concerns the labor-market effect of Austria's short-time work scheme, then the interpretation of any post-2020 gap depends directly on what type of counterfactual benchmark is used.

V1 compares Austria with countries classified as $STW = 0$ under the strict Eurostat administrative definition. This is the conceptually sharpest "without-STW" benchmark, but it is also based on a narrow donor pool and is therefore more exposed to donor concentration. V2 compares Austria with countries that themselves used STW during the pandemic. This does not provide a pure untreated counterfactual, but it offers a broader peer-group comparison. V3 serves as a validation exercise by treating selected No-STW countries as placebo-treated units and matching them against STW-using donor pools.

To ensure comparability across these models, the same baseline design is maintained throughout: the pre-treatment period covers 2017–2019, the post-treatment period covers 2020–2022 and the same broad predictor architecture is used across specifications. Results are interpreted jointly through pre-treatment fit, post-treatment divergence, donor concentration, placebo evidence and robustness to alternative modeling choices.

Before turning to the model-specific results, two descriptive questions are useful. First, which countries belong to the relevant comparison worlds? Second, how does Austria compare to these groups in the pre-treatment period with respect to unemployment and the main predictor variables? Table Classification of Countries therefore summarizes the donor-pool classification, while Table Pre-Treatment Summary reports pre-treatment summary statistics for Austria and the relevant donor groups.

Table Classification			
Classification of countries across relevant comparison worlds			
Country	STW classification	Average STW 2020–2021	Model-specific role
Austria	STW > 0	0.083	M1: Treated unit M2: Treated unit M3: Not included
Czechia	STW = 0	0.000	M1: Donor country M2: Not included M3: Treated in one validation model
Norway	STW = 0	0.000	M1: Donor country M2: Not included M3: Treated in one validation model
Romania	STW = 0	0.000	M1: Donor country M2: Not included M3: Treated in one validation model
Belgium	STW > 0	0.082	M1: Not included M2: Donor country M3: Donor country
Bulgaria	STW > 0	0.037	M1: Not included M2: Donor country M3: Donor country
Cyprus	STW > 0	0.102	M1: Not included M2: Donor country M3: Donor country
Denmark	STW > 0	0.012	M1: Not included M2: Donor country M3: Donor country
Estonia	STW > 0	0.029	M1: Not included M2: Donor country M3: Donor country
Finland	STW > 0	0.027	M1: Not included M2: Donor country M3: Donor country
France	STW > 0	0.082	M1: Not included M2: Donor country M3: Donor country
Greece	STW > 0	0.031	M1: Not included M2: Donor country M3: Donor country
Hungary	STW > 0	0.019	M1: Not included M2: Donor country M3: Donor country
Ireland	STW > 0	0.115	M1: Not included M2: Donor country M3: Donor country
Italy	STW > 0	0.119	M1: Not included M2: Donor country M3: Donor country
Latvia	STW > 0	0.024	M1: Not included M2: Donor country M3: Donor country
Lithuania	STW > 0	0.042	M1: Not included M2: Donor country M3: Donor country
Malta	STW > 0	0.095	M1: Not included M2: Donor country M3: Donor country
Netherlands	STW > 0	0.112	M1: Not included M2: Donor country M3: Donor country
Poland	STW > 0	0.013	M1: Not included M2: Donor country M3: Donor country
Portugal	STW > 0	0.056	M1: Not included M2: Donor country M3: Donor country
Slovakia	STW > 0	0.089	M1: Not included M2: Donor country M3: Donor country
Slovenia	STW > 0	0.048	M1: Not included M2: Donor country M3: Donor country
Spain	STW > 0	0.025	M1: Not included M2: Donor country M3: Donor country
Sweden	STW > 0	0.028	M1: Not included M2: Donor country M3: Donor country

Notes: The classification refers to the strict Eurostat administrative definition of government-supported jobs. “STW = 0” indicates that no jobs were recorded in Eurostat categories (A) or (B) during 2020–2022. This does not imply the absence of any labour-market support in a broader substantive sense; it only refers to the harmonised Eurostat classification used for cross-country comparability. The average STW intensity reports the simple quarterly mean over 2020–2021.

Table 5.1: Classification of Countries

Table Pre-treatment summary			
Simple averages over 2017–2019 for Austria and the relevant donor groups			
Variable	Austria	No-STW donor pool average (n = 3)	STW donor pool average (n = 21)
Unemployment rate	5.33	3.97	7.49
GDP index	104.81	100.81	101.03
Employment rate	72.92	71.35	68.72
openness	2.05	2.06	2.02
Average hours worked	36.49	38.44	38.17
Industrial production index	111.61	109.12	109.94
Inflation	1.94	2.58	1.69
Notes: The table reports simple averages over the pre-treatment period 2017–2019. It is intended as a descriptive overview of Austria's position relative to the two donor groups before SCM weighting is applied. The purpose is not causal identification, but to illustrate the structural context in which the matching procedure operates.			

Table 5.2: Pre-treatment summary

5.2 Main result I: Austria vs. No-STW countries (V1)

5.2.1 Baseline SCM estimate and pre-treatment fit

The first Austria-based specification compares Austria to the donor pool of countries classified as STW = 0 in the Eurostat sense. Conceptually, this is the sharpest contrast design in the thesis because it comes closest to asking how Austria's unemployment rate would have evolved in a comparison world without observed short-time work. At the same time, it is also a demanding design because the donor pool is very small and therefore raises immediate questions about donor concentration and robustness.

The baseline results indicate a disciplined pre-treatment fit. In V1, the pre-RMSPE is 0.18 and the treated and synthetic unemployment paths track each other closely during 2017–2019. This is an important prerequisite for interpretation because it suggests that the synthetic control reproduces Austria's pre-treatment outcome dynamics reasonably well. In the post-treatment period, however, the post-RMSPE rises to 0.75, producing an RMSPE ratio of 4.12. This indicates a clear increase in post-2020 divergence relative to the pre-treatment fit. At the same time, the average post-treatment effect remains very small: the ATT over 2020–2022 is 0.037. In substantive terms, this implies that V1 does not point to a large and persistent average unemployment effect over the full post-treatment window, even though the post-period deviations become more pronounced in the diagnostic sense. Put differently, the model suggests a visible post-2020 break in fit, but not a large average level shift in unemployment.

Main specification							
Model 1 Austria versus No STW countries							
Model	Treated unit	Donor pool	Pre-treatment RMSPE	Post-treatment RMSPE	RMSPE ratio	Average post-treatment gap	Weight concentration (HHI)
Model 1 Austria versus No STW countries	Austria	Countries without STW	0.183	0.754	4.116	0.037	0.866

Table 5.3: Main SCM specification – V1: AT vs. No-STW donor pool

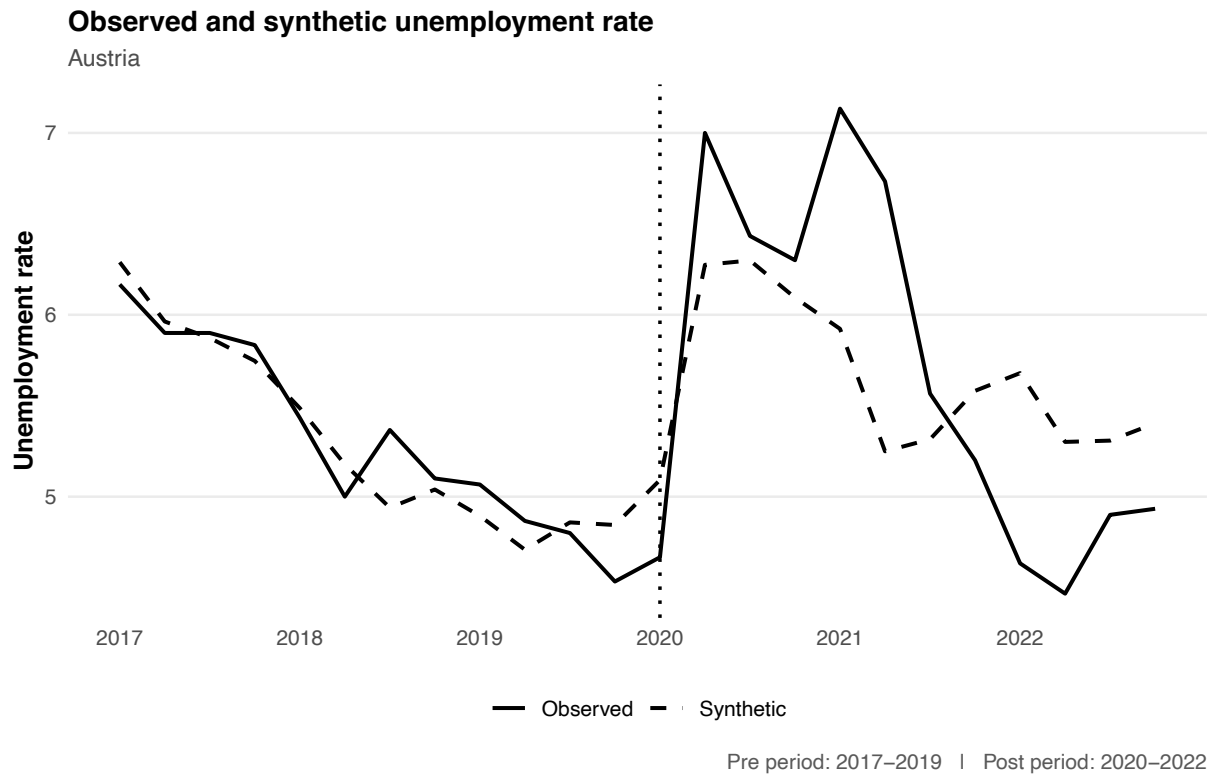


Figure 5.1: Observed and synthetic unemployment rate – V1: AT vs. No-STW donor pool

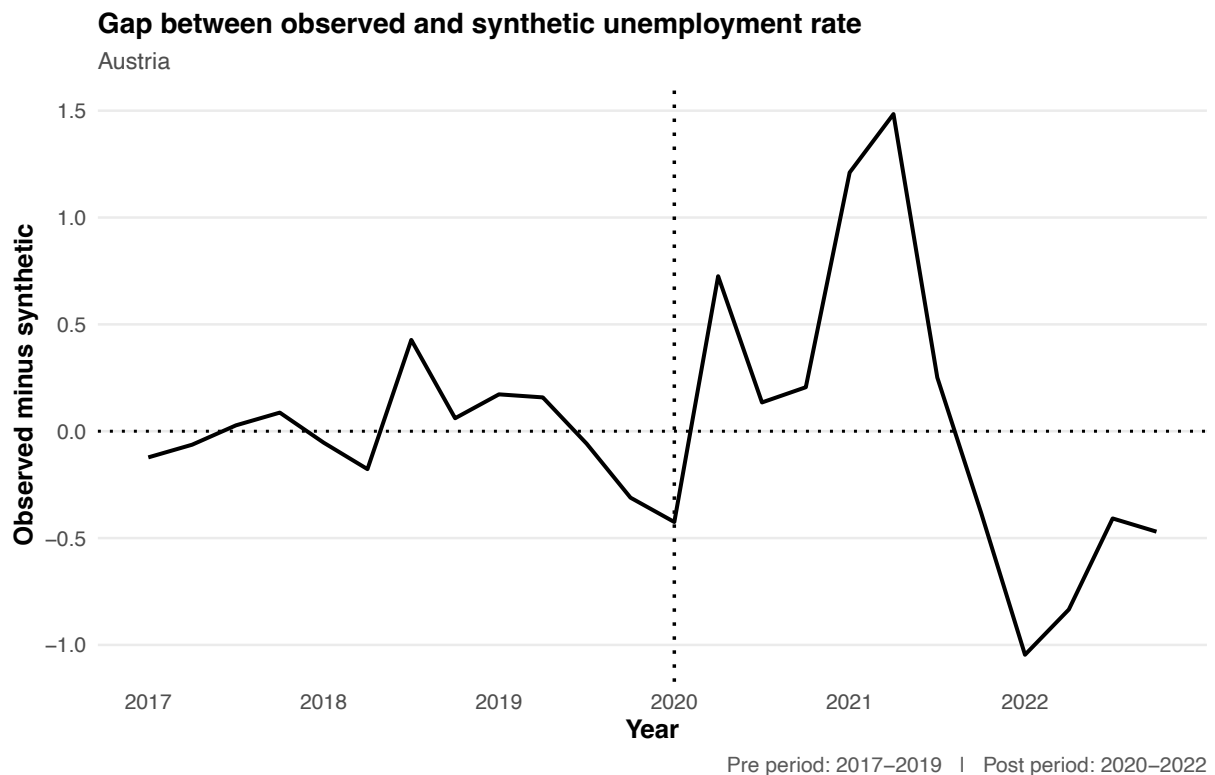


Figure 5.2: Gap observed minus synthetic unemployment rate – V1: AT vs. No-STW donor pool

5.2.2 Donor concentration and weight concentration

The donor structure of V1 is highly concentrated. Romania contributes about 92.81% of the synthetic Austria, Norway about 7.16%, while Czechia receives practically zero weight. The corresponding Herfindahl index is 0.866, indicating a very strong

concentration of donor weights. Methodologically, this matters because the estimated counterfactual is effectively carried by one dominant country. Such a structure is not uncommon in SCM, but it reduces confidence that the result reflects a broad and stable synthetic benchmark rather than the post-period dynamics of one particularly influential donor.

This concentration is central for interpretation. A high RMSPE ratio on its own could easily look persuasive, but in the presence of such dominant donor weights it becomes necessary to ask whether the post-2020 divergence is robust to the exclusion of the main donor and whether similar patterns also arise in alternative comparison worlds. V1 is therefore best understood not as a standalone causal estimate, but as an important contrast design whose meaning depends heavily on the stability diagnostics considered later in the chapter.

Donor weights				
Model 1 Austria versus No STW countries				
Rank	Donor country	Weight	Share (%)	Cumulative share (%)
1	Romania	0.93	92.81	92.81
2	Norway	0.07	7.16	99.97
3	Czechia	0.00	0.03	100.00

Table 5.4: Donor weights – V1: AT vs. No-STW donor pool

5.2.3 Interpretation of V1

Taken on its own, V1 gives a restrained substantive message. It shows that Austria can be matched reasonably well to a synthetic unit composed of countries with STW = 0 and that the post-2020 fit worsens markedly relative to the pre-period. However, the average unemployment effect is close to zero and the entire construction is highly donor-fragile because it depends overwhelmingly on Romania. For that reason, the appropriate interpretation is not that “there was no effect,” but rather that the strict No-STW comparison world does not yield a robust, large and stable average unemployment gap for Austria over 2020–2022. This already points toward the broader conclusion of the thesis: unemployment-based evaluation of short-time work is possible and informative, but its meaning depends strongly on the chosen comparison world and on the stability of the synthetic construction.

Detailed placebo, leave-one-out, predictor-sensitivity and alternative pre-window results for V1 are reported in Appendix A. In the main text, V1 should therefore be read mainly through its baseline fit, donor concentration and substantive comparison with the other model variants.

5.3 Main result II: Austria vs. STW countries (V2)

5.3.1 Baseline SCM estimate and pre-treatment fit

The second Austria-based specification, V2, compares Austria only with countries that themselves recorded positive STW use during the pandemic. This specification therefore does not represent a pure untreated benchmark. Its role in the thesis is different: it serves as a peer-group comparison that asks whether Austria’s post-2020 unemployment dynamics also look unusual relative to other labor markets that used related job-retention

instruments. This makes V2 especially useful as a cross-check to V1. If Austria's divergence also appears in this broader STW-using donor pool, that would strengthen the interpretation that the Austrian pattern is not driven solely by the contrast with No-STW countries.

Empirically, V2 produces a much stronger headline estimate than V1. The pre-treatment fit is good, with a pre-RMSPE of 0.13. In the post-treatment period, post-RMSPE rises sharply to 0.91, corresponding to an RMSPE ratio of 6.86. The average ATT is 0.692, which is substantially larger than in V1. In this comparison world, Austria's unemployment rate in 2020–2022 lies noticeably above the synthetic peer benchmark on average. Relative to the No-STW design, the peer-group specification therefore points toward a much more pronounced post-2020 divergence.

Main specification							
Model 2 Austria versus STW countries							
Model	Treated unit	Donor pool	Pre-treatment RMSPE	Post-treatment RMSPE	RMSPE ratio	Average post-treatment gap	Weight concentration (HHI)
Model 2 Austria versus STW countries	Austria	Countries with STW	0.132	0.908	6.858	0.692	0.204

Table 5.5: Main SCM specification – V2: AT vs. STW donor pool

Observed and synthetic unemployment rate

Austria

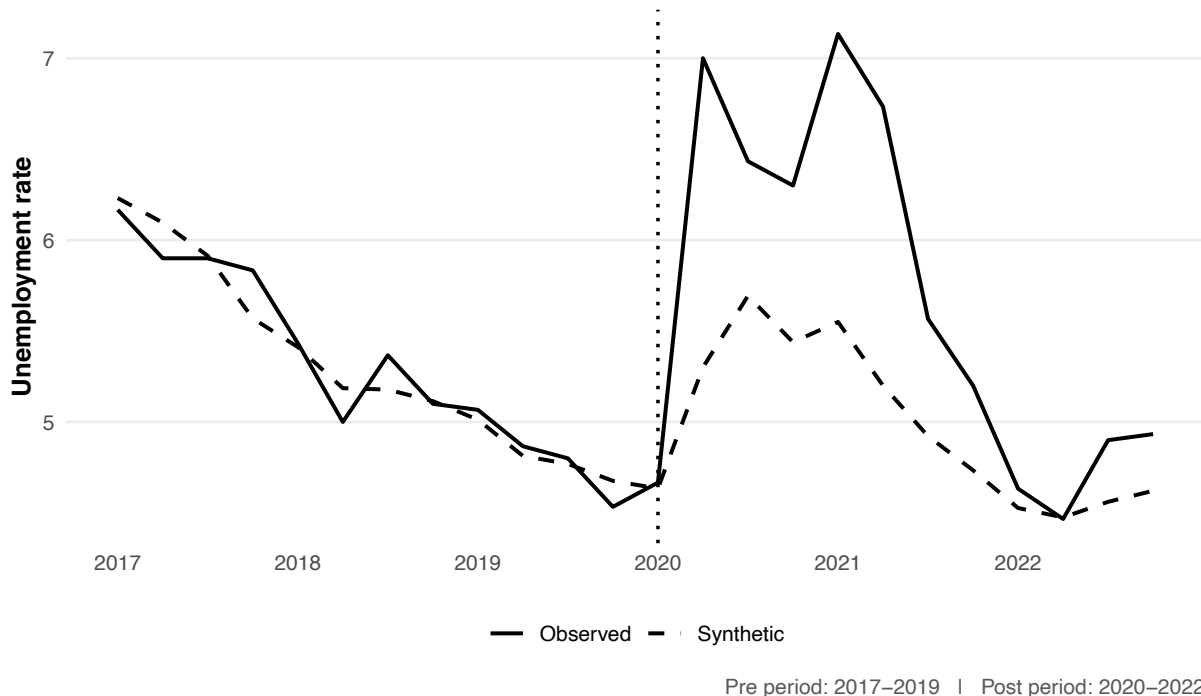


Figure 5.3: Observed and synthetic unemployment rate – V2: AT vs. STW donor pool

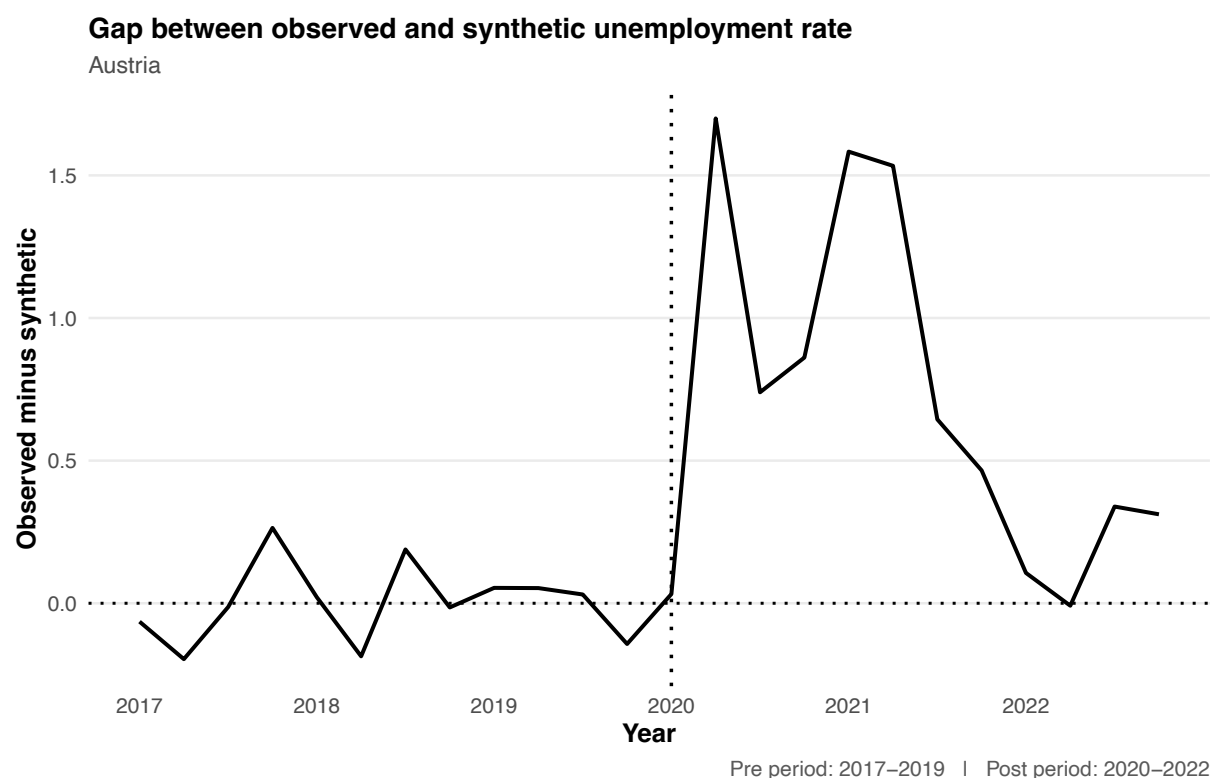


Figure 5.4: Gap observed minus synthetic unemployment rate – V2: AT vs. STW donor pool

5.3.2 Donor structure and weight concentration

A major methodological advantage of V2 lies in its donor structure. Unlike V1, the synthetic Austria in V2 is built from a much broader set of donor countries. The largest weights are assigned to Hungary (33.98%), Poland (17.19%), Denmark (16.95%) and France (16.57%), with smaller contributions from additional countries. The HHI is 0.204, which is far lower than in V1 and indicates a considerably less concentrated synthetic construction. This broader donor mix makes V2 more attractive from a stability perspective because the result is less obviously dominated by a single donor country.

The broader donor base, however, comes at a conceptual cost. Because all donor countries in V2 themselves used STW, the comparison is less suitable for identifying a clean “without-STW” counterfactual. Instead, V2 should be understood as evidence about Austria’s relative position within a group of crisis-response countries. It is therefore methodologically stronger in terms of donor breadth, but conceptually weaker as an untreated benchmark.

Donor weights				
Model 2 Austria versus STW countries				
Rank	Donor country	Weight	Share (%)	Cumulative share (%)
1	Hungary	0.34	33.98	33.98
2	Poland	0.17	17.19	51.17
3	Denmark	0.17	16.95	68.12
4	France	0.17	16.57	84.69
5	Slovenia	0.04	4.03	88.72
6	Bulgaria	0.02	2.31	91.02
7	Sweden	0.01	1.18	92.21
8	Slovakia	0.01	1.11	93.31
9	Belgium	0.01	0.99	94.30
10	Finland	0.01	0.72	95.02
11	Italy	0.01	0.71	95.73
12	Portugal	0.01	0.65	96.38
13	Lithuania	0.01	0.64	97.02
14	Netherlands	0.01	0.59	97.61
15	Other donors	0.02	2.39	100.00

Table 5.6: Donor weights – V2: AT vs. STW donor pool

5.3.3 Interpretation of V2

Substantively, V2 is the strongest estimate in the chapter, but not necessarily the cleanest one. The positive average gap suggests that Austria performed worse in unemployment terms than its synthetic STW-using peer benchmark. At the same time, this result must be interpreted jointly with the later falsification and sensitivity evidence, because the V2 model also shows meaningful instability when the pre-period is partitioned differently and when the pre-treatment window is extended. The correct reading is therefore not that V2 “proves” a large causal unemployment effect of Austrian Kurzarbeit, but that Austria’s post-2020 unemployment dynamics look more distinctive when compared with STW-using peers than when compared with the very narrow No-STW donor pool.

Detailed placebo, leave-one-out, predictor-sensitivity and alternative pre-window results for V2 are reported in Appendix A. In the main text, the key point is that V2 combines a broader donor structure with a materially larger average post-treatment gap, but should still be interpreted cautiously because it is a within-STW comparison rather than a pure untreated benchmark.

5.4 Validation exercise: No-STW countries against STW donors (V3)

The purpose of V3 is to place the Austria-based estimates into a broader comparative context. Instead of treating Austria as the only treated unit, V3 assigns placebo treatment status to selected countries classified as STW = 0 and matches them against donor pools consisting of STW-using countries. If the same modeling environment systematically generated large post-2020 “effects” for these countries as well, then Austria’s results would be less informative. If, by contrast, the V3 results are heterogeneous, this would support the interpretation that the Austria findings are not merely a mechanical artifact of the method.

5.4.1 Romania as placebo-treated unit

Romania provides the strongest pre-treatment fit among all models considered in the chapter. Its pre-RMSPE is 0.091, which is even better than the corresponding values for Austria in V1 and V2. Post-RMSPE rises to 0.45, yielding an RMSPE ratio of 4.99, but the average ATT is only -0.037. Thus, Romania's post-2020 average deviation is very close to zero in substantive terms, despite a sizeable RMSPE ratio. This is important because it shows that large post-/pre-fit ratios do not automatically imply large average unemployment effects.

Romania's donor structure is also notably broader than that of V1. The largest weights are Hungary (44.94%), Lithuania (30.71%), Latvia (20.80%) and Slovakia (3.52%), corresponding to an HHI of 0.341. This is still concentrated, but clearly less extreme than Austria's V1 construction. As a result, Romania offers a useful validation case: the model fits very well, the donor mix is broader and the average post-period deviation remains small. That makes Romania an instructive counterexample to any simple claim that the SCM setup mechanically generates large effects in all such pandemic-era comparisons.

Main specification							
Model 3 Romania versus STW countries							
Model	Treated unit	Donor pool	Pre-treatment RMSPE	Post-treatment RMSPE	RMSPE ratio	Average post-treatment gap	Weight concentration (HHI)
Model 3 Romania versus STW countries	Romania	Countries with STW	0.091	0.452	4.989	-0.037	0.341

Table 5.7: Main SCM specification – V3: RO vs. STW donor pool

Donor weights				
Model 3 Romania versus STW countries				
Rank	Donor country	Weight	Share (%)	Cumulative share (%)
1	Hungary	0.45	44.94	44.94
2	Lithuania	0.31	30.71	75.65
3	Latvia	0.21	20.80	96.46
4	Slovakia	0.04	3.52	99.97
5	Greece	0.00	0.02	99.99
6	Belgium	0.00	0.00	100.00
7	Italy	0.00	0.00	100.00
8	Spain	0.00	0.00	100.00
9	Poland	0.00	0.00	100.00
10	Bulgaria	0.00	0.00	100.00
11	France	0.00	0.00	100.00
12	Cyprus	0.00	0.00	100.00
13	Other donors	0.00	0.00	100.00

Table 5.8: Donor weights – V3: RO vs. STW donor pool

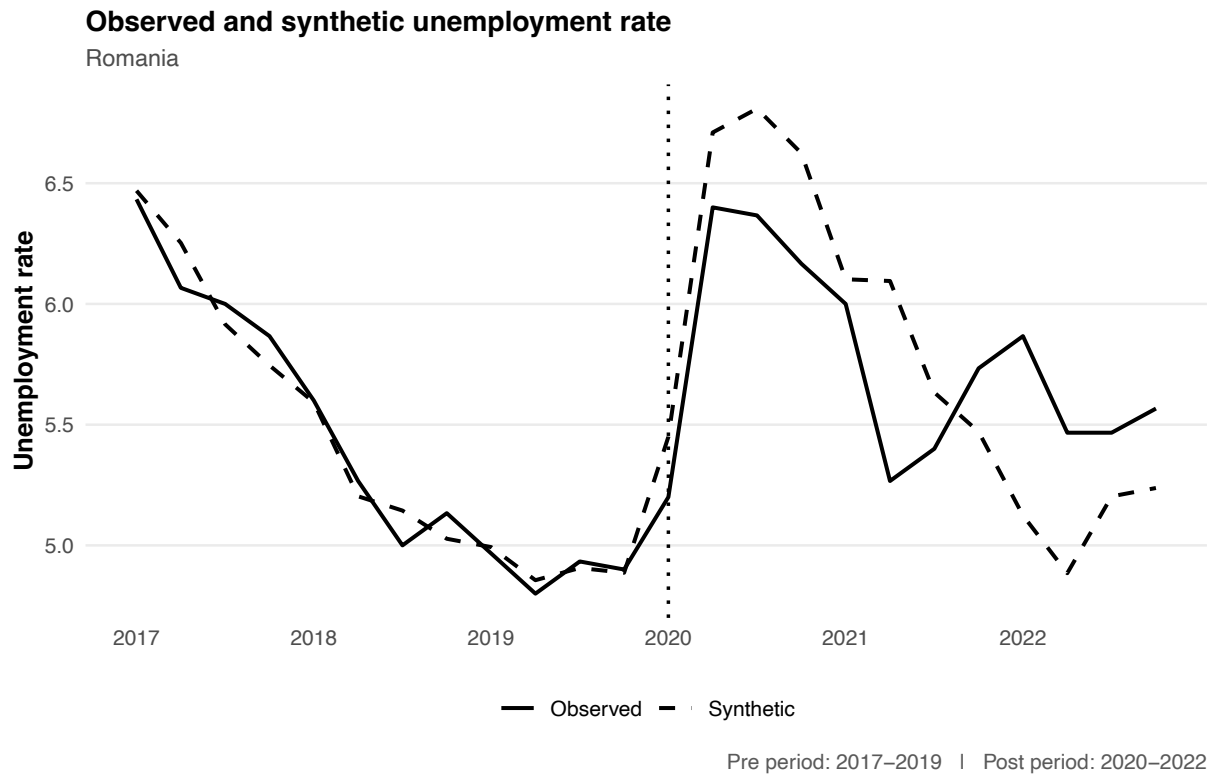


Figure 5.5: Observed and synthetic unemployment rate – V3: RO vs. STW donor pool

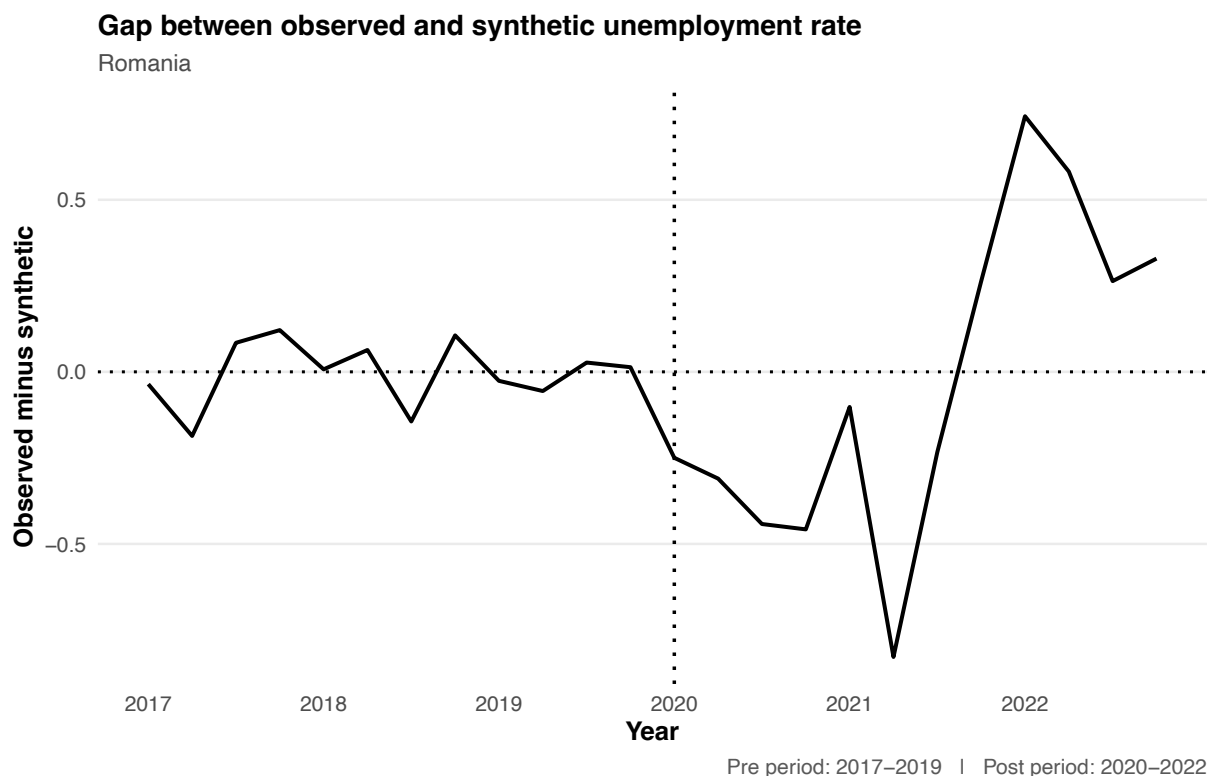


Figure 5.6: Gap observed minus synthetic unemployment rate – V3: RO vs. STW donor pool

5.4.2 Norway as placebo-treated unit

Norway yields a different picture. The pre-treatment fit remains solid, though weaker than in the Romania case, with a pre-RMSPE of 0.17. Post-RMSPE rises to 0.63, producing an RMSPE ratio of 3.84 and the average ATT is -0.433. This is materially larger in

magnitude than the effects in V1 and in the Romania validation case. In substantive terms, Norway's observed unemployment rate lies noticeably below its synthetic path in the post-period. Norway therefore demonstrates that sizeable deviations can arise in this empirical setting, but not in a uniform way and not with the same sign across countries.

Norway's donor weights are again much more concentrated than Romania's: Hungary receives about 72.55%, Sweden 14.26% and Malta 12.94%, with an HHI of 0.563. This makes the Norwegian estimate structurally closer to Austria's V1 case than to Romania's broader construction. The main implication is that the presence of a sizeable ATT in the Norwegian case should not be overread as evidence of a stable structural effect. Rather, it underlines that the pandemic period can generate sizeable country-specific divergences in the SCM setup, especially when donor concentration is high.

Main specification							
Model 3 Norway versus STW countries							
Model	Treated unit	Donor pool	Pre-treatment RMSPE	Post-treatment RMSPE	RMSPE ratio	Average post-treatment gap	Weight concentration (HHI)
Model 3 Norway versus STW countries	Norway	Countries with STW	0.165	0.633	3.836	-0.433	0.563

Table 5.9: Main SCM specification – V3: NO vs. STW donor pool

Donor weights				
Model 3 Norway versus STW countries				
Rank	Donor country	Weight	Share (%)	Cumulative share (%)
1	Hungary	0.73	72.55	72.55
2	Sweden	0.14	14.26	86.81
3	Malta	0.13	12.94	99.75
4	Estonia	0.00	0.17	99.92
5	Netherlands	0.00	0.02	99.94
6	Austria	0.00	0.02	99.96
7	Bulgaria	0.00	0.01	99.97
8	France	0.00	0.01	99.98
9	Poland	0.00	0.00	99.98
10	Ireland	0.00	0.00	99.99
11	Latvia	0.00	0.00	99.99
12	Finland	0.00	0.00	99.99
13	Other donors	0.00	0.01	100.00

Table 5.10: Donor weights – V3: NO vs. STW donor pool

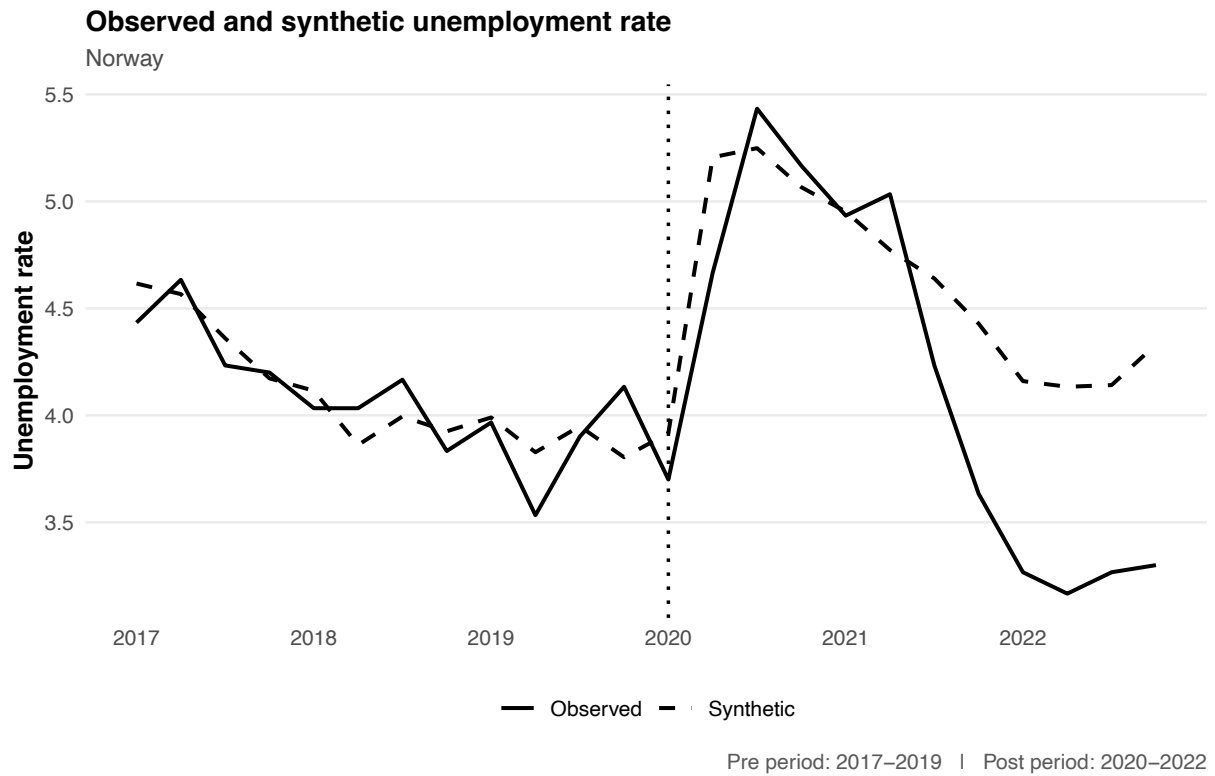


Figure 5.7: Observed and synthetic unemployment rate – V3: NO vs. STW donor pool

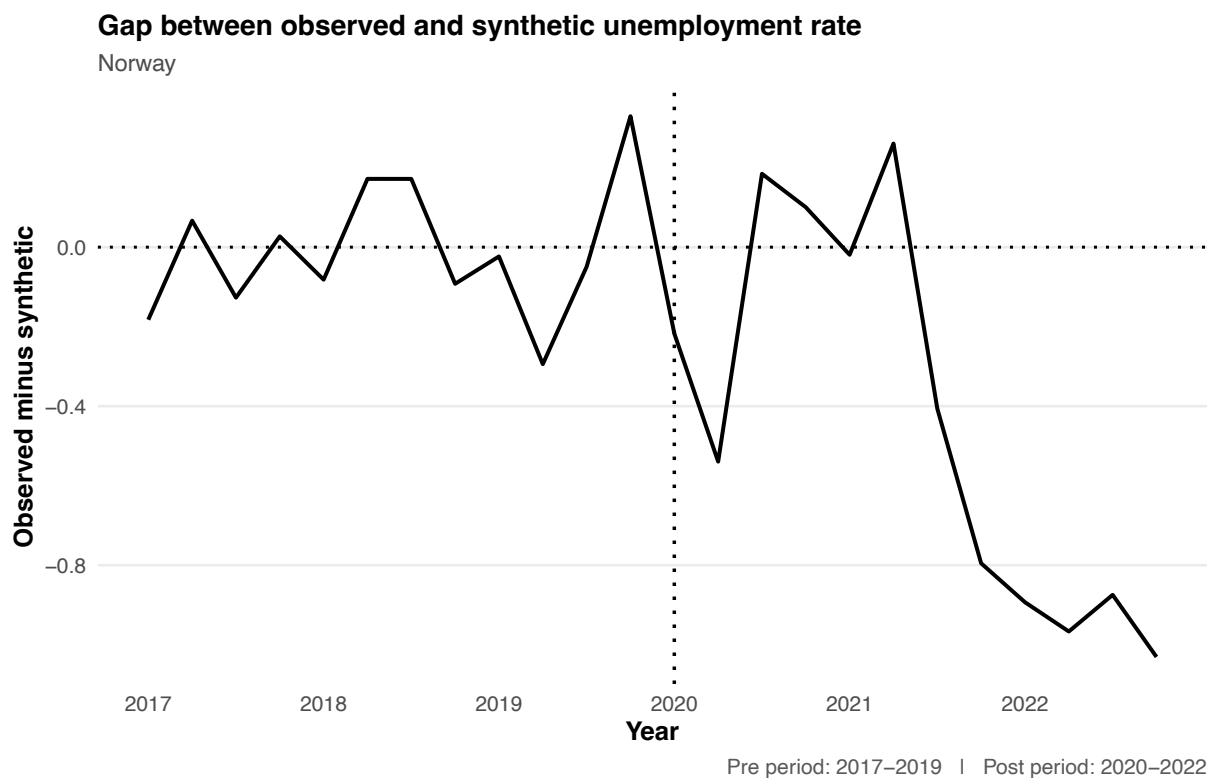


Figure 5.8: Gap observed minus synthetic unemployment rate – V3: NO vs. STW donor pool

5.4.3 Note on Czechia

Czechia was initially considered as a third V3 case, but it is not pursued in the main analysis because the SCM solution collapses onto a single donor country, implying an HHI of 1. In such a setting, the synthetic control ceases to be meaningfully synthetic and

becomes little more than a one-country comparison. Since this makes the estimate extremely fragile and uninformative for inference, Czechia is excluded from the discussion.

5.4.4 What V3 implied for the Austria results

Taken together, the V3 cases reinforce the idea that pandemic-era deviations are heterogeneous rather than mechanical. Romania shows a very strong pre-fit with an average gap close to zero, while Norway shows a clearly negative ATT with much stronger sensitivity to donor choice and pre-window selection. This heterogeneity matters because it weakens any simplistic interpretation of Austria's results as a pure "STW versus no-STW" pattern. Instead, it suggests that the SCM estimates must be interpreted as the product of both crisis heterogeneity and donor-pool design. Austria's results therefore gain credibility not because V3 produces no effects, but because V3 shows that the empirical environment is varied and that model behavior differs across treated units.

5.5 Robustness checks

Overall, the robustness evidence is mixed and should be interpreted as a set of complementary diagnostics rather than as a sequence of formal pass-fail tests. First, the in-space placebos do not provide strong evidence that Austria's estimated effects are uniquely large relative to placebo-treated donor units. This is particularly true in specifications with a small RMSPE-filtered comparison set, where pseudo-p values are mechanically hard to interpret. Second, the in-time placebo exercises show that the models do not uniformly generate artificial treatment-like breaks immediately before 2020, but they also reveal that some specifications can produce noticeable divergence when the pre-period is partitioned differently. Third, the leave-one-out exercises highlight major differences in donor dependence across models: V1 is clearly donor-fragile because it is heavily anchored by Romania, whereas V2 is materially more stable due to its broader donor structure. Fourth, predictor-one-out checks are generally more reassuring than donor-exclusion checks, as they mostly affect the magnitude of estimated gaps rather than their qualitative direction. Finally, the most consistent robustness result across all model families is sensitivity to the length of the pre-treatment window. Extending the pre-period often changes both fit quality and estimated effects in a non-trivial way. Taken together, these results suggest that the baseline estimates are informative, but they should be interpreted with caution: in this application, inference depends not only on pre-treatment fit, but also on donor concentration, placebo behavior and especially on the choice of the stable pre-treatment regime. Detailed robustness tables and figures are reported in Appendix A.

5.6 Synthesis of findings

Taken together, the three model families do not support one single model-invariant estimate of Austrian Kurzarbeit's effect on measured unemployment. Instead, they show that the estimated post-2020 gap depends materially on the comparison world used to construct the counterfactual.

In V1, Austria is compared with countries classified as $STW = 0$ under the strict Eurostat definition. This is the conceptually sharpest benchmark because it comes closest to a “without-STW” comparison. Empirically, however, the model combines a disciplined pre-treatment fit with a very small average post-treatment gap and a highly concentrated donor structure. V1 is therefore informative as a contrast model, but not strong enough to stand on its own as a robust headline estimate.

In V2, Austria is compared with STW-using countries. This model produces the clearest positive post-2020 Austria gap and relies on a broader and less concentrated donor structure. In that sense, it is empirically more stable than V1. At the same time, it is conceptually less sharp, because the donor pool no longer represents a pure untreated benchmark. V2 is therefore best understood as a peer-group comparison that strengthens the descriptive contrast, but does not by itself identify a clean “no-Kurzarbeit” counterfactual.

The validation exercise in V3 adds an important discipline to the interpretation. Romania and Norway do not generate one uniform placebo-treated pattern when matched against STW-using donor pools. This heterogeneity suggests that the empirical environment does not mechanically produce large post-2020 effects for any treated unit. At the same time, it confirms that SCM estimates in this setting remain sensitive to donor composition and model specification.

Appendix A sharpens this interpretation further. For V1, the lagged outcome variants leave the baseline result almost unchanged, which is reassuring with respect to timing structure. However, the reduced specification performs very poorly, implying that the small V1 gap is credible only when the richer baseline matching design is retained. For V2, the appendix evidence is more mixed: the Austria gap remains positive across lagged variants, but its magnitude declines noticeably and the reduced specification weakens the model substantially. The appendix results therefore do not overturn the baseline findings, but they clarify their status: V1 is more robust in direction once the baseline design is preserved, whereas V2 remains substantively suggestive but more specification-sensitive.

The central takeaway of the chapter is therefore not that Austrian Kurzarbeit produced one definitive unemployment effect between 2020 and 2022. Rather, the results show that interpretation depends on a trade-off between conceptual sharpness and empirical stability. The strict No-STW benchmark is cleaner but donor-fragile; the STW-based peer comparison is broader and more stable but less causally sharp and the validation exercise shows that the broader environment is heterogeneous rather than mechanical.

6 Discussion, limitations and policy implications

6.1 Discussion of the main findings

The main findings of this thesis are best understood through one central insight: within the SCM framework applied here, Austrian Kurzarbeit does not produce one single, model-invariant effect on unemployment. Instead, the estimated post-2020 gap varies depending on the counterfactual benchmark that is chosen. This is not merely a technical detail; it is itself part of the substantive result. In this application, the composition of the donor pool directly shapes the meaning of the comparison and, with it, the interpretation of the estimated effect.

This becomes particularly evident when the two Austria-based models are considered jointly. The strict No-STW design is conceptually appealing because it comes closest to a genuine “without Kurzarbeit” counterfactual. At the same time, however, it is based on a relatively narrow donor pool and proves highly sensitive to the inclusion of specific donor countries. The STW-based peer design is broader and empirically more stable, but it is conceptually less clear-cut because the donor countries themselves also relied on job-retention schemes. The thesis therefore points not to a simple distinction between a “good” and a “bad” model, but to a genuine trade-off between conceptual sharpness and empirical stability.

The appendix evidence reinforces this reading. V1 remains highly stable when lagged outcome variants are introduced, but it cannot be reproduced credibly with a reduced outcome-only specification. V2 continues to imply a positive Austria gap, yet its magnitude is more clearly affected by alternative lag structures and by removing the richer matching design. These results matter because they show that the remaining uncertainty does not come from one arbitrary robustness failure. Rather, it comes from a systematic feature of the application: the estimated unemployment effect is sensitive to how the pre-treatment regime is represented and how the counterfactual world is defined.

The broader implication is that the thesis should not be read as delivering one definitive causal number for Austrian Kurzarbeit. Its contribution is more disciplined than that. It provides a transparent counterfactual framework, shows where the Austria results are stable, shows where they are fragile and makes explicit that unemployment-based conclusions during COVID-19 depend materially on modeling choices that are themselves substantively meaningful.

6.2 Limitations

The results should be interpreted in light of several limitations, four of which are especially important for this thesis.

The first limitation concerns the definition of “No-STW” countries. The classification used in this thesis is based on Eurostat’s administrative concept of government- supported jobs and on the STW measure constructed from that source. Countries classified as STW = 0 under this approach may nevertheless have implemented labor-market support measures during the crisis that do not appear as STW in Eurostat’s categories. The interpretation

of V1 and V3 is therefore precise but narrow. These models compare Austria with countries without STW in the Eurostat sense, not with countries without labor-market support more generally. This distinction matters a great deal. It means that the thesis does not claim to compare Austria to a world without intervention, but to a world without recorded STW according to a harmonized administrative definition. That is still useful and transparent, but it is narrower than a broader policy comparison.

The second limitation concerns the outcome variable itself. The unemployment rate is central, internationally comparable and highly relevant in public debate, which makes it a reasonable outcome for the thesis. At the same time, it remains an incomplete measure of labor-market adjustment in the context of short-time work. STW is specifically designed to preserve employment relationships while reducing hours worked. Workers on short-time work remain classified as employed, even if their effective labor input falls substantially. As a result, labor-market slack may appear more strongly in hours worked, underemployment or reduced labor utilization than in the measured unemployment rate. A weak or near-zero unemployment effect therefore does not imply that short-time work was ineffective. It may simply mean that the most important adjustment margins lie elsewhere. This is not a flaw of the empirical design, but it is a limitation of what can be inferred from unemployment alone.

The third limitation concerns heterogeneous COVID shocks and institutional differences across countries. Countries differed substantially in pandemic intensity, lockdown timing, sector composition, tourism exposure, work-from-home feasibility, fiscal support and labor-market institutions. Even with a rich predictor set and good outcome matching, SCM cannot fully absorb all time-varying confounders or all cross-country differences in crisis exposure. This is particularly important in the present context because the Austrian post-2020 gap may partly reflect the broader Austrian crisis environment rather than Kurzarbeit alone. The heterogeneity seen in the validation exercise and the sensitivity patterns in the Austria models are consistent with this interpretation. In other words, the SCM estimates should be understood as disciplined counterfactual evidence within a broader crisis-policy environment, not as perfectly isolated policy coefficients.

The fourth limitation is methodological and concerns the stability of the pre-treatment regime. The robustness analysis shows that some estimated effects change meaningfully when the pre-treatment window is extended or when the donor structure changes. This is a central limitation in a pandemic setting where there is no obviously perfect pre-period and where the treatment occurs in the context of a major structural break. The results therefore depend not only on whether Austria and Synthetic Austria fit each other well before 2020, but also on whether that pre-treatment window is itself a stable and representative basis for the counterfactual. This is one reason why the thesis relies on robustness diagnostics and multiple comparison worlds rather than presenting one preferred model as definitive.

Overall, these limitations do not invalidate the analysis, but they do shape its interpretation. The thesis should not be read as proving one exact unemployment effect

of Austrian Kurzarbeit. Its contribution is narrower: it shows how Austria's post-2020 unemployment path compares to several explicitly constructed counterfactual worlds and it makes visible how much the answer depends on donor-pool definition, donor stability and outcome choice. That is a limitation, but also one of the main strengths of the design.

6.3 Policy implications

The policy implications of this thesis should be understood as conditional rather than universal. Given the sensitivity of the estimates and the limits of unemployment as an outcome, the results do not support a simple statement such as "Kurzarbeit clearly reduced unemployment" or "Kurzarbeit had no effect." A more defensible conclusion is that the labor-market role of short-time work cannot be assessed adequately from the unemployment rate alone and depends strongly on the chosen benchmark.

A first implication is therefore that short-time work should not be evaluated solely on the basis of unemployment. Since STW mainly stabilizes employment relationships and shifts adjustment from layoffs to hours reductions, a broader set of outcomes is needed, including hours worked, underemployment, employment-based indicators and labor input more generally. A modest unemployment effect should therefore not be interpreted automatically as policy failure.

A second implication concerns program design. The results of this thesis, together with the broader literature, suggest that short-time work can be useful in an abrupt downturn, but that targeting and exit design matter. Broad support may be justified in the acute phase of a crisis, but it should be adjusted once the shock becomes more persistent or more heterogeneous across sectors. The policy implication is therefore not to reject STW, but to combine it with clearer targeting and more credible exit rules.

For Austria specifically, the policy lesson is not that Kurzarbeit can be judged as clearly successful or clearly unsuccessful on the basis of unemployment alone. The empirical results show that Austria's unemployment trajectory looks different depending on the benchmark used to construct the counterfactual. This makes simplistic policy conclusions inappropriate. A more defensible implication is that Austrian Kurzarbeit should be evaluated as part of a broader stabilization package and assessed with a wider outcome set that includes labor input, employment preservation and distributional protection, not unemployment alone.

7 Conclusion

This thesis examined whether Austria's extensive expansion of Kurzarbeit during the COVID-19 crisis is reflected in the measured unemployment rate. Using the Synthetic Control Method and a strict Eurostat-based definition of recorded short-time work, the analysis evaluated Austria across different counterfactual comparison worlds rather than relying on one single benchmark.

The results do not support a simple conclusion such as "Kurzarbeit clearly reduced unemployment" or "Kurzarbeit had no meaningful effect." Instead, they show that Austria's estimated post-2020 unemployment gap depends materially on how the counterfactual is constructed. Relative to the strict No-STW benchmark, the average gap is small and donor-fragile. Relative to STW-using peers, Austria appears to perform worse, but this comparison is conceptually less suited to identify a pure untreated counterfactual. The validation exercise further shows that the empirical setting is heterogeneous rather than mechanically generating large effects.

Appendix A strengthens this overall reading. The strict No-STW model remains stable under lagged outcome variants but depends on retaining the richer baseline predictor design. The STW-based Austria model continues to point to a positive gap, yet its magnitude is more sensitive to alternative lag structures and reduced specifications. The most defensible conclusion is therefore not a single headline effect, but a structured interpretation of where the Austria results are robust and where they remain specification-sensitive.

More broadly, the thesis suggests that unemployment is too narrow to serve as the sole metric for evaluating Austrian Kurzarbeit during COVID-19. Kurzarbeit may matter primarily by preserving employment relationships and shifting adjustment to working hours rather than by producing a large and model-invariant reduction in measured unemployment. Future work should therefore combine unemployment with broader labor-market and welfare outcomes in order to assess the program more fully.

The contribution of the thesis is therefore threefold. First, it applies a strict and transparent Eurostat-based STW definition instead of relying on looser national policy labels. Second, it uses SCM to construct an explicit and interpretable counterfactual for Austria rather than depending on a pooled comparison or a parallel-trends assumption. Third, it shows that the choice of donor pool is itself a substantive modeling decision and should be treated as such. In this sense, the thesis contributes less by offering one final causal number and more by clarifying how Austria's unemployment trajectory changes across plausible counterfactual worlds.

At the same time, the findings should be interpreted with appropriate caution. The strict No-STW classification is narrow and administrative in nature, unemployment is only one margin of adjustment and the COVID-19 crisis combined several overlapping shocks and policies. These limitations do not make the analysis uninformative, but they do define its scope. The thesis should therefore be read as a transparent counterfactual study of

Austria's unemployment path during COVID-19, not as a definitive statement on every labor-market effect of Kurzarbeit.

Overall, the thesis shows both the usefulness and the limits of evaluating short-time work through unemployment. The main result is not a single number, but a structured insight: Austria's estimated post-2020 unemployment effect depends materially on how the counterfactual is constructed. Making that dependence visible is itself an important contribution, especially in a policy setting where many countries used related job-retention instruments and where seemingly simple control groups can be misleading.

8 List of References

- Abadie, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, 59(2), 391–425.
- Abadie, A., & Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque Country. *American Economic Review*, 93(1), 113–132.
- Abadie, A., Diamond, A., & 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(490), 493–505.
- Abadie, A., Diamond, A., & Hainmueller, J. (2011). Synth: An R package for synthetic control methods in comparative case studies. *Journal of Statistical Software*, 42(13), 1–17.
- Abadie, A., Diamond, A., & Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, 59(2), 495–510.
- Athey, S., & Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*, 31(2), 3–32.
- Aiyar, S., & Dao, M. C. (2021). The effectiveness of job-retention schemes: COVID-19 evidence from the German states. *IMF Working Paper* 21/242.
- Boeri, T., & Brücker, H. (2011). Short-time work benefits revisited: Some lessons from the Great Recession. *Economic Policy*, 26(68), 697–765.
- Brenke, K., Rinne, U., & Zimmermann, K. F. (2013). Short-time work: The German answer to the Great Recession. *International Labour Review*, 152(2), 287–305.
- Burda, M. C., & Hunt, J. (2011). What explains the German labor market miracle in the Great Recession? *Brookings Papers on Economic Activity*, Spring, 273–335.
- Cahuc, P., & Carcillo, S. (2011). Is short-time work a good method to keep unemployment down? *Nordic Economic Policy Review*, 1(1), 133–165.
- Cahuc, P., Kramarz, F., & Nevoux, S. (2021). The heterogeneous impact of short-time work: From saved jobs to windfall effects. *IZA Discussion Paper No. 14381*.
- Campos, N. F., Coricelli, F., & Moretti, L. (2014). Economic growth and political integration: Estimating the benefits from membership in the European Union using the synthetic counterfactuals method. *IZA Discussion Paper No. 8162*.
- Campos, N. F., Coricelli, F., & Moretti, L. (2019). Institutional integration and economic growth in Europe. *Journal of Monetary Economics*, 103, 88–104.

Eichhorst, W., Marx, P., Rinne, U., et al. (2022). Job retention schemes during COVID-19: A review of policy responses. IZA Policy Paper No. 187.

Fitzenberger, B., & Walwei, U. (2023). Short-time work during the COVID-19 crisis: Lessons learned. IAB-Forschungsbericht 05/2023.

Garoupa, N., & Spruk, R. (2020). Does European Union really matter for long-run growth and development? In search of the missing counterfactual. SSRN Working Paper.

Giupponi, G., & Landais, C. (2023). Subsidizing labour hoarding in recessions: The employment and welfare effects of short-time work. *Review of Economic Studies*, 90(4), 1963–2005.

Huemer, U., Kogler, M., & Mahringer, H. (2021). Kurzarbeit als Kriseninstrument in der COVID-19-Pandemie: Kurzexpertise zum Vergleich der Modelle ausgewählter europäischer Länder. WIFO.

Kagerl, C. (2024). The employment and windfall effects of short-time work: Evidence from Germany. IAB Discussion Paper 14/2024.

OECD. (2021). *OECD Employment Outlook 2021: Navigating the COVID-19 crisis and recovery*. OECD Publishing.

OECD. (2022). *Riding the waves: Adjusting job retention schemes through the COVID-19 crisis*. OECD Publishing.

Tamesberger, D., & Theurl, S. (2021). Design and take up of Austria's coronavirus short-time work model. ICAE Working Paper No. 127, Johannes Kepler University Linz.

WIFO, prospect Unternehmensberatung, & Universität Zürich. (2025). *Evaluierung der COVID-19-Kurzarbeit*. Study commissioned by Arbeitsmarktservice Österreich and the Austrian Ministry of Social Affairs / Labour.

Data Sources (Eurostat)

Eurostat. COVID-19: Jobs benefiting from COVID-19 governmental support measures (file: D1_Job_benefiting_from_Covid19_governmental_support_measures.xlsx).
<https://ec.europa.eu/eurostat/web/covid-19/society-work#Labour%20market>

Eurostat. Unemployment rate – monthly data, seasonally adjusted (file: D2_Unemp_monthly.xlsx).
https://ec.europa.eu/eurostat/databrowser/view/une_rt_m__custom_19371636/default/table

Eurostat. Average usual weekly hours worked in the main job by professional status, full-time/part-time employment and economic activity

(file: D3_Avg_hours_worked.xlsx).

https://ec.europa.eu/eurostat/databrowser/view/lfsq_ewhun2__custom_19373069/default/table

Eurostat. Employment rates by educational attainment level – quarterly data

(file: D4_Employment_attainment_level.xlsx).

https://ec.europa.eu/eurostat/databrowser/view/lfsq_ergaed__custom_19371724/default/table

Eurostat. Employed persons by economic activity – quarterly data

(file: D5_Employment_share.xlsx).

https://ec.europa.eu/eurostat/databrowser/view/lfsq_egan2__custom_19417860/default/table

Eurostat. Gross domestic product (GDP) and main components (output, expenditure and income) – quarterly data

(file: D6_Gdp_data.xlsx).

https://ec.europa.eu/eurostat/databrowser/view/namq_10_gdp__custom_19373549/default/table

Eurostat. HICP – monthly data (annual rate of change)

(file: D7_Inflation_monthly.xlsx).

https://ec.europa.eu/eurostat/databrowser/view/prc_hicp_manr__custom_20015056/default/table

Eurostat. Production in industry – quarterly data (file: D8_Production_Industry.xlsx).

https://ec.europa.eu/eurostat/databrowser/view/sts_inpr_q__custom_20015065/default/table

Eurostat. Gross domestic product (GDP) and main components (output, expenditure and income) – quarterly data

(file: D9_Trade_openness.xlsx).

https://ec.europa.eu/eurostat/databrowser/view/namq_10_gdp__custom_19371914/default/table

Appendix A: Detailed robustness and additional specification checks

A.1 Overview of appendix checks

This appendix reports the detailed robustness and specification checks for the three model variants used in the thesis. V1 refers to Austria versus No-STW countries, V2 refers to Austria versus STW countries and V3 refers to the validation exercise in which selected No-STW countries are treated as placebo-treated units against STW-using donor pools. The appendix is organized by type of robustness check. This structure makes it easier to compare how different models react to the same diagnostic exercise.

A.2 In-space placebos

The in-space placebo results should be interpreted as contextual evidence rather than as mechanical significance tests. For Austria, the placebo comparison in V1 is particularly limited because RMSPE filtering leaves only a very small set of retained placebo units. The resulting pseudo-p of 1 therefore does not indicate proof of no effect, but merely shows that Austria's absolute ATT is not unusually large among the few adequately fitting placebos. The placebo evidence is somewhat more supportive in V2, where 16 units are retained and the pseudo-p falls to 0.25; Austria's estimated effect is thus comparatively large, though still not unique within the placebo distribution. Romania shows a similarly unremarkable placebo profile: with nine units retained after filtering and a pseudo-p again equal to 1, its average ATT of -0.0370 is not extreme relative to the placebo distribution. Norway's placebo evidence is somewhat stronger in substantive terms because its ATT lies more clearly away from zero, yet its pseudo-p remains 0.529, implying that the estimated effect is not unique within the filtered distribution. Overall, the placebo results across models suggest caution rather than strong claims of uniqueness.

Placebo-in-space results					
Model 1 Austria versus No STW countries					
Unit	Unit type	Pre-treatment RMSPE	Post-treatment RMSPE	RMSPE ratio	Average post-treatment gap
Austria	Treated unit	0.183	0.754	4.116	0.037
Norway	Placebo unit	0.288	0.437	1.517	-0.201
Romania	Placebo unit	0.207	0.799	3.868	0.078
RMSPE filter threshold: 0.366. Pseudo p-value: 1.					

Table 5.11: Placebo-in-space results – V1: AT vs. No-STW donor pool

Placebo-in-space results					
Model 2 Austria versus STW countries					
Unit	Unit type	Pre-treatment RMSPE	Post-treatment RMSPE	RMSPE ratio	Average post-treatment gap
Austria	Treated unit	0.132	0.908	6.858	0.692
Spain	Placebo unit	0.240	2.355	9.797	2.224
Poland	Placebo unit	0.263	1.019	3.877	−0.986
Netherlands	Placebo unit	0.067	0.772	11.505	−0.727
France	Placebo unit	0.115	0.770	6.672	−0.656
Latvia	Placebo unit	0.208	0.729	3.502	0.639
Ireland	Placebo unit	0.161	0.840	5.222	0.543
Bulgaria	Placebo unit	0.199	0.727	3.656	−0.539
Lithuania	Placebo unit	0.203	0.772	3.793	0.325
Denmark	Placebo unit	0.120	0.397	3.305	−0.263
Slovenia	Placebo unit	0.220	0.417	1.891	0.241
Finland	Placebo unit	0.195	0.350	1.790	0.236
Belgium	Placebo unit	0.145	0.510	3.528	0.235
Slovakia	Placebo unit	0.119	0.388	3.268	0.134
Portugal	Placebo unit	0.242	0.390	1.609	−0.078
Italy	Placebo unit	0.127	0.428	3.381	0.037
RMSPE filter threshold: 0.265. Pseudo p-value: 0.25.					

Table 5.12: Placebo-in-space results – V2: AT vs. STW donor pool

Placebo-in-space results					
Model 3 Romania versus STW countries					
Unit	Unit type	Pre-treatment RMSPE	Post-treatment RMSPE	RMSPE ratio	Average post-treatment gap
Romania	Treated unit	0.091	0.452	4.989	−0.037
Netherlands	Placebo unit	0.063	0.697	11.008	−0.655
France	Placebo unit	0.115	0.757	6.572	−0.641
Austria	Placebo unit	0.135	0.798	5.934	0.566
Ireland	Placebo unit	0.161	0.839	5.222	0.539
Belgium	Placebo unit	0.150	0.653	4.358	0.524
Denmark	Placebo unit	0.116	0.441	3.797	−0.310
Slovakia	Placebo unit	0.111	0.401	3.596	0.249
Italy	Placebo unit	0.126	0.421	3.339	0.050
RMSPE filter threshold: 0.181. Pseudo p-value: 1.					

Table 5.13: Placebo-in-space results – V3: RO vs. STW donor pool

Placebo-in-space results					
Model 3 Norway versus STW countries					
Unit	Unit type	Pre-treatment RMSPE	Post-treatment RMSPE	RMSPE ratio	Average post-treatment gap
Norway	Treated unit	0.165	0.633	3.836	-0.433
Spain	Placebo unit	0.240	2.364	9.862	2.210
Poland	Placebo unit	0.263	1.019	3.877	-0.986
Latvia	Placebo unit	0.148	0.993	6.705	0.930
Austria	Placebo unit	0.138	0.901	6.532	0.722
Bulgaria	Placebo unit	0.210	0.844	4.020	-0.687
France	Placebo unit	0.115	0.724	6.304	-0.604
Ireland	Placebo unit	0.161	0.848	5.280	0.551
Netherlands	Placebo unit	0.050	0.532	10.577	-0.485
Lithuania	Placebo unit	0.193	0.802	4.163	0.426
Slovakia	Placebo unit	0.115	0.494	4.276	0.371
Denmark	Placebo unit	0.117	0.430	3.674	-0.282
Finland	Placebo unit	0.196	0.369	1.879	0.256
Slovenia	Placebo unit	0.220	0.418	1.898	0.246
Belgium	Placebo unit	0.146	0.508	3.479	0.222
Portugal	Placebo unit	0.242	0.415	1.710	-0.117
Italy	Placebo unit	0.126	0.404	3.201	0.016
RMSPE filter threshold: 0.33. Pseudo p-value: 0.529.					

Table 5.14: Placebo-in-space results – V2: NO vs. STW donor pool

A.3 In-time placebos

The in-time placebo exercises are particularly informative in the present thesis because they assess whether the models generate treatment-like discontinuities even before the actual intervention date. For Austria, the results for V1 are somewhat mixed but not strongly alarming. The 2018Q1 placebo generates some volatility, with a placebo RMSPE ratio of 2.73 and an average placebo gap of 0.0408, whereas the 2019Q1 placebo is clearly weaker, with a ratio of 1.12 and an average ATT close to zero (-0.0093). Taken together, these findings suggest that V1 does not mechanically produce a pronounced break immediately prior to 2020, although some divergence can arise when the pre-treatment period is split relatively early. V2, by contrast, provides the more cautionary Austria specification in this dimension. Here, both placebo dates generate non-trivial divergence during the pre-treatment period, indicating that part of the post-2020 gap may reflect underlying pre-trend differences or limitations in donor comparability rather than a purely post-2020 discontinuity. Romania's in-time placebo results are comparatively reassuring: the 2018Q1 split shows some volatility, but not a persistent treatment-like break, while the 2019Q1 placebo is weaker still. Norway appears more problematic, as both placebo dates continue to show noticeable divergence before 2020. Overall, the in-time placebo evidence supports a cautious interpretation of the estimated effects, particularly for specifications in which pre-treatment divergence is not fully negligible.

In-time placebo summary					
Model 1 Austria versus No STW countries					
Fake treatment date	Pre-treatment RMSPE	Placebo post-period RMSPE	RMSPE ratio	Average placebo gap	Weight concentration (HHI)
2018Q1	0.080	0.218	2.730	0.041	0.846
2019Q1	0.176	0.197	1.119	-0.009	0.865

Table 5.15: In-time placebo summary – V1: AT vs. No-STW donor pool

In-time placebo summary					
Model 2 Austria versus STW countries					
Fake treatment date	Pre-treatment RMSPE	Placebo post-period RMSPE	RMSPE ratio	Average placebo gap	Weight concentration (HHI)
2018Q1	0.117	0.351	3.001	-0.241	0.328
2019Q1	0.125	0.366	2.932	-0.304	0.296

Table 5.16: In-time placebo summary – V2: AT vs. STW donor pool

In-time placebo summary					
Model 3 Romania versus STW countries					
Fake treatment date	Pre-treatment RMSPE	Placebo post-period RMSPE	RMSPE ratio	Average placebo gap	Weight concentration (HHI)
2018Q1	0.052	0.182	3.505	0.047	0.322
2019Q1	0.106	0.081	0.764	-0.005	0.246

Table 5.17: In-time placebo summary – V3: RO vs. STW donor pool

In-time placebo summary					
Model 3 Norway versus STW countries					
Fake treatment date	Pre-treatment RMSPE	Placebo post-period RMSPE	RMSPE ratio	Average placebo gap	Weight concentration (HHI)
2018Q1	0.129	0.262	2.033	0.170	0.419
2019Q1	0.126	0.233	1.852	0.023	0.528

Table 5.18: In-time placebo summary – V3: NO vs. STW donor pool

Model 1 Austria versus No STW countries – In-time placebo checks
Two fake treatment dates inside the pre-treatment period

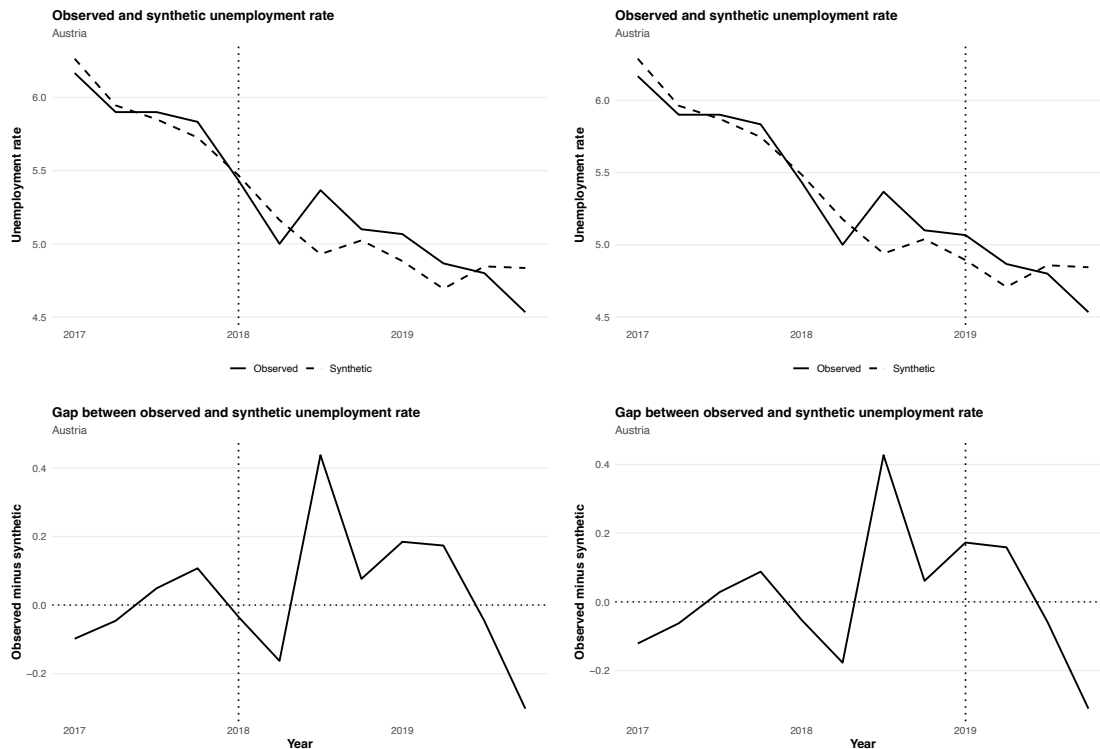


Figure 5.9:
Left: In-time placebo (fake treat. 2018Q1) – V1: AT vs. No-STW donor pool
Right: In-time placebo (fake treat. 2019Q1) – V1: AT vs. No-STW donor pool

Model 2 Austria versus STW countries – In-time placebo checks
Two fake treatment dates inside the pre-treatment period

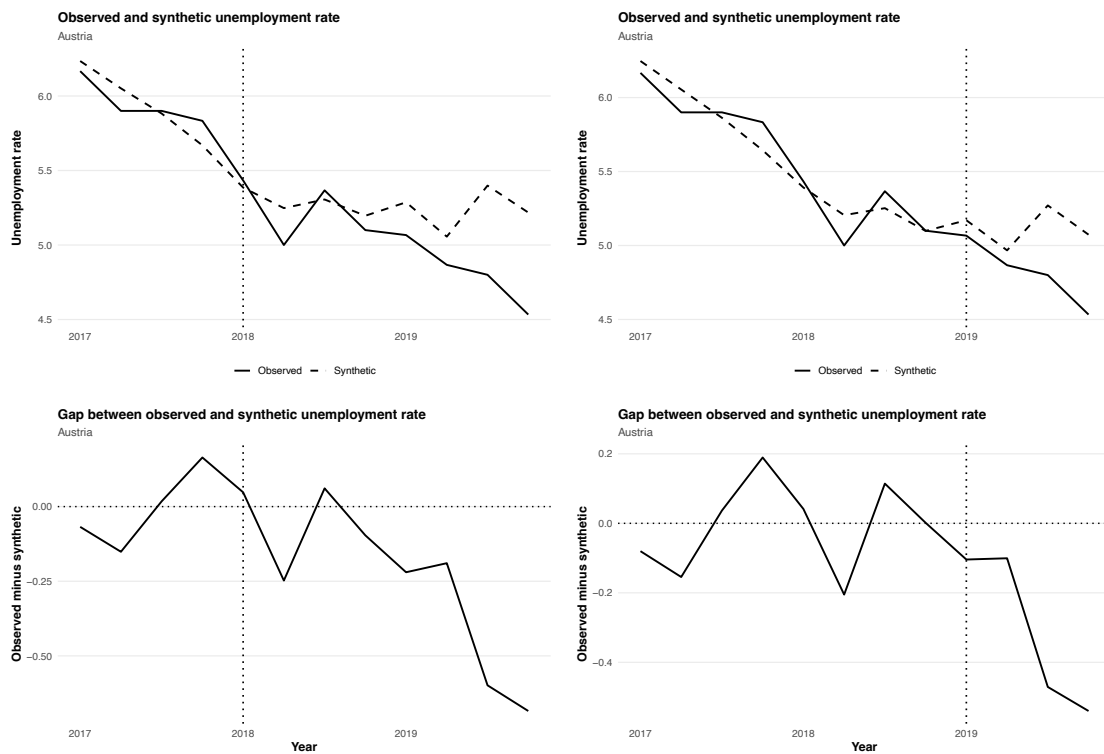


Figure 5.10:
Left: In-time placebo (fake treat. 2018Q1) – V2: AT vs. STW donor pool
Right: In-time placebo (fake treat. 2019Q1) – V2: AT vs. STW donor pool

Model 3 Norway versus STW countries – In-time placebo checks
Two fake treatment dates inside the pre-treatment period

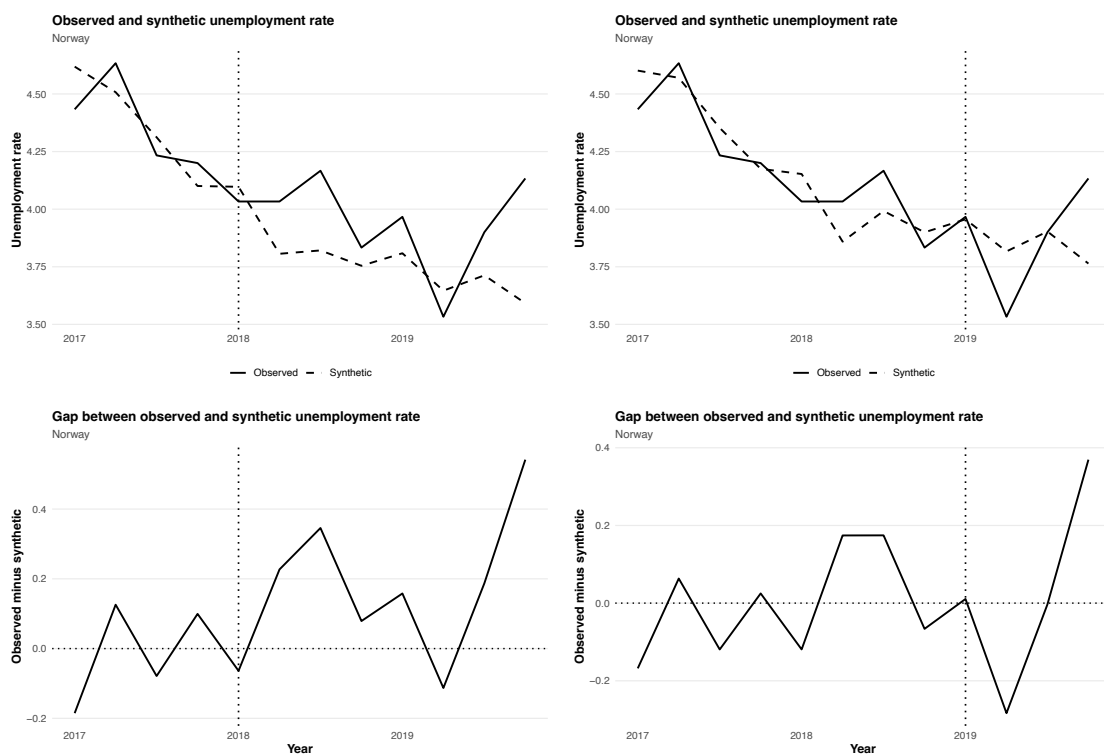


Figure 5.11:
Left: In-time placebo (fake treat. 2018Q1) – V3: RO vs. STW donor pool
Right: In-time placebo (fake treat. 2019Q1) – V3: RO vs. STW donor pool

Model 3 Romania versus STW countries – In-time placebo checks
Two fake treatment dates inside the pre-treatment period

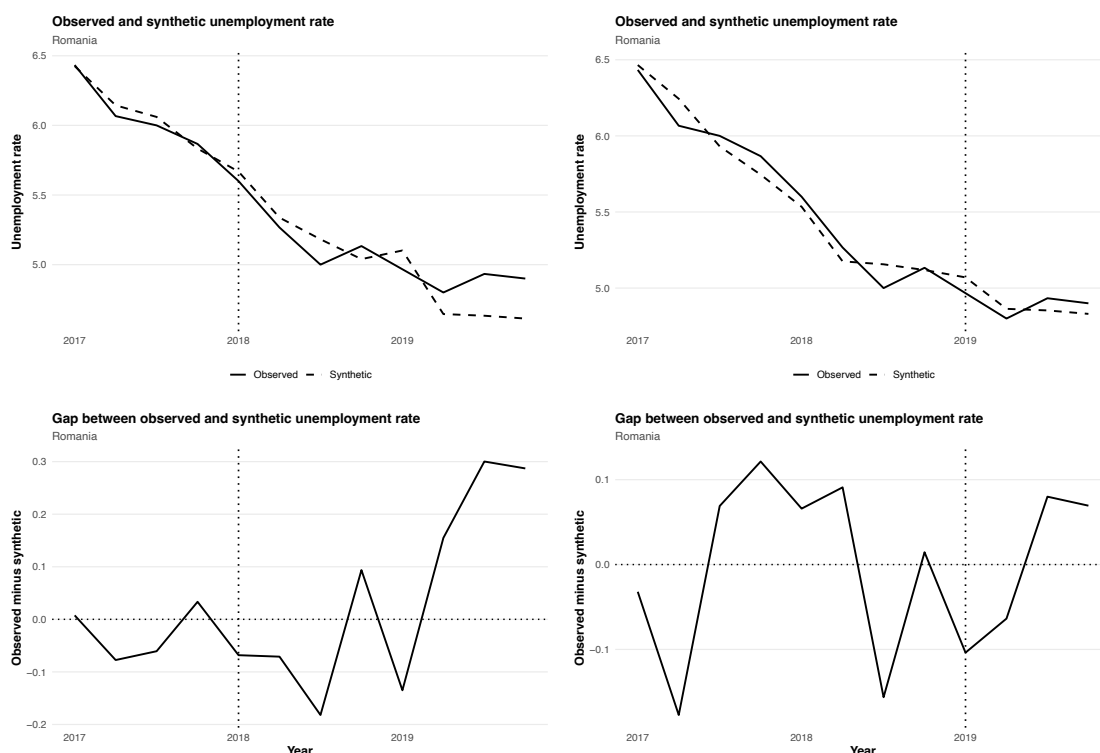


Figure 5.12:
Left: In-time placebo (fake treat. 2018Q1) – V3: NO vs. STW donor pool
Right: In-time placebo (fake treat. 2019Q1) – V3: NO vs. STW donor pool

A.4 Leave-one-out robustness

The leave-one-out diagnostics reveal the clearest structural differences across the model specifications. For Austria, V1 proves highly donor-fragile. Excluding Romania causes pre-treatment fit to deteriorate sharply, with pre-RMSPE increasing from 0.1832 to 1.2895, while the estimated ATT rises from 0.0372 to 1.5139. By contrast, excluding Norway or Czechia alters the results only marginally. This indicates that the V1 estimate depends overwhelmingly on Romania and is therefore not robust to the removal of its dominant donor. V2 performs considerably better in this dimension. Although Hungary remains an important contributor, the estimated ATT stays positive and substantial even when individual top donors are excluded, making V2 markedly more stable than V1. This constitutes one of the strongest arguments for treating V2 as a serious benchmark, even if it is weaker in terms of causal purity. Romania itself appears more stable than V1, but not fully robust: excluding Hungary shifts the ATT from a small negative value to 0.5837, while dropping Lithuania or Latvia also changes the estimate materially. Norway likewise remains highly sensitive, as excluding Hungary turns the ATT positive at 0.3566, whereas excluding Sweden moves the estimate close to zero. Overall, the leave-one-out evidence suggests that donor dependence is a central concern in several specifications, but substantially less so in V2 than in V1.

Leave-one-out robustness					
Model 1 Austria versus No STW countries					
Case	Number of donors	Pre-treatment RMSPE	Post-treatment RMSPE	RMSPE ratio	Average post-treatment gap
Baseline	3	0.183	0.754	4.116	0.037
Romania	2	1.289	1.568	1.216	1.514
Norway	2	0.187	0.790	4.218	0.013
Czechia	2	0.183	0.755	4.124	0.033

Table 5.19: Leave-one-out robustness – V1: AT vs. No-STW donor pool

Leave-one-out robustness					
Model 2 Austria versus STW countries					
Case	Number of donors	Pre-treatment RMSPE	Post-treatment RMSPE	RMSPE ratio	Average post-treatment gap
Baseline	21	0.132	0.908	6.858	0.692
Hungary	20	0.122	0.575	4.727	0.332
Poland	20	0.135	0.795	5.908	0.559
Denmark	20	0.118	0.904	7.681	0.665

Table 5.20: Leave-one-out robustness – V2: AT vs. STW donor pool

Leave-one-out robustness					
Model 3 Romania versus STW countries					
Case	Number of donors	Pre-treatment RMSPE	Post-treatment RMSPE	RMSPE ratio	Average post-treatment gap
Baseline	22	0.091	0.452	4.989	–0.037
Hungary	21	0.070	0.671	9.554	0.584
Lithuania	21	0.098	0.365	3.723	0.024
Latvia	21	0.094	0.482	5.147	0.314

Table 5.21: Leave-one-out robustness – V3: RO vs. STW donor pool

Leave-one-out robustness					
Model 3 Norway versus STW countries					
Case	Number of donors	Pre-treatment RMSPE	Post-treatment RMSPE	RMSPE ratio	Average post-treatment gap
Baseline	22	0.165	0.633	3.836	–0.433
Hungary	21	0.219	0.561	2.558	0.357
Sweden	21	0.185	0.457	2.463	–0.076
Malta	21	0.176	0.708	4.024	–0.507

Table 5.22: Leave-one-out robustness – V3: NO vs. STW donor pool

A.5 Predictor sensitivity

The predictor-one-out checks are generally more reassuring than the leave-one-out results, but they still reveal meaningful differences across specifications. For Austria, V1 appears relatively stable in this dimension. Omitting individual predictors leaves pre-treatment fit almost unchanged, and the estimated ATT remains within a narrow range of roughly 0.03 to 0.04. This is encouraging, as it suggests that the V1 result is not mechanically driven by any single matching variable. V2 is likewise fairly robust in qualitative terms. Although the omission of individual predictors can reduce the magnitude of the estimated ATT – for instance when GDP is excluded – the effect remains predominantly positive. This indicates that predictor sensitivity in V2 mainly affects the size of the estimated effect rather than its overall direction. Romania also appears mostly stable under predictor-one-out variants, although excluding the production index or inflation shifts the estimated effect meaningfully into positive territory, suggesting that

these variables play a comparatively important role in that specification. Norway likewise remains negative across the predictor-one-out checks, but the magnitude of the estimated effect becomes smaller when GDP, the employment rate, or openness are excluded. Overall, the predictor-one-out evidence suggests that sensitivity to individual matching variables is less severe than donor dependence and mainly influences effect magnitude rather than the broader qualitative interpretation of the models.

Sensitivity analysis					
Model 1 Austria versus No STW countries					
Check	Detail	Pre-treatment RMSPE	Post-treatment RMSPE	RMSPE ratio	Average post-treatment gap
Baseline	as selected	0.183	0.754	4.116	0.037
Predictor omitted	openness	0.183	0.755	4.123	0.033
Predictor omitted	Inflation	0.183	0.755	4.123	0.034
Predictor omitted	Industrial production index	0.183	0.755	4.122	0.034
Predictor omitted	Employment rate	0.183	0.755	4.123	0.033
Predictor omitted	GDP index	0.183	0.755	4.122	0.034
Predictor omitted	Average hours worked	0.183	0.755	4.119	0.037
Alternative pre period	2016–2019	0.290	0.722	2.487	0.224
Alternative pre period	2015–2019	0.524	0.818	1.560	0.589

Table 5.23: Sensitivity analysis – V1: AT vs. No-STW donor pool

Sensitivity analysis					
Model 2 Austria versus STW countries					
Check	Detail	Pre-treatment RMSPE	Post-treatment RMSPE	RMSPE ratio	Average post-treatment gap
Baseline	as selected	0.132	0.908	6.858	0.692
Predictor omitted	Employment rate	0.130	0.758	5.837	0.524
Predictor omitted	Average hours worked	0.134	0.678	5.041	0.420
Predictor omitted	Inflation	0.135	0.803	5.959	0.573
Predictor omitted	openness	0.135	0.798	5.906	0.566
Predictor omitted	GDP index	0.140	0.517	3.681	0.201
Predictor omitted	Industrial production index	0.142	0.810	5.724	0.594
Alternative pre period	2016–2019	0.145	0.555	3.821	0.258
Alternative pre period	2015–2019	0.201	0.522	2.597	0.129

Table 5.24: Sensitivity analysis – V2: AT vs. STW donor pool

Sensitivity analysis					
Model 3 Romania versus STW countries					
Check	Detail	Pre-treatment RMSPE	Post-treatment RMSPE	RMSPE ratio	Average post-treatment gap
Baseline	as selected	0.091	0.452	4.989	-0.037
Predictor omitted	Average hours worked	0.091	0.460	5.077	-0.047
Predictor omitted	Employment rate	0.091	0.460	5.077	-0.047
Predictor omitted	openness	0.091	0.459	5.076	-0.047
Predictor omitted	GDP index	0.091	0.455	5.010	-0.044
Predictor omitted	Industrial production index	0.095	0.407	4.294	0.203
Predictor omitted	Inflation	0.095	0.384	4.045	0.148
Alternative pre period	2016–2019	0.092	0.499	5.426	0.269
Alternative pre period	2015–2019	0.103	0.516	5.022	0.240

Table 5.25: Sensitivity analysis – V3: RO vs. STW donor pool

Sensitivity analysis					
Model 3 Norway versus STW countries					
Check	Detail	Pre-treatment RMSPE	Post-treatment RMSPE	RMSPE ratio	Average post-treatment gap
Baseline	as selected	0.165	0.633	3.836	-0.433
Predictor omitted	Industrial production index	0.165	0.637	3.852	-0.437
Predictor omitted	Inflation	0.166	0.617	3.723	-0.413
Predictor omitted	Average hours worked	0.167	0.594	3.565	-0.383
Predictor omitted	GDP index	0.170	0.503	2.967	-0.255
Predictor omitted	Employment rate	0.170	0.529	3.105	-0.286
Predictor omitted	openness	0.173	0.473	2.737	-0.204
Alternative pre period	2016–2019	0.212	0.432	2.037	-0.042
Alternative pre period	2015–2019	0.456	0.473	1.036	0.017

Table 5.26: Sensitivity analysis – V3: NO vs. STW donor pool

A.6 Alternative pre-treatment windows

The most important common robustness finding in this chapter concerns sensitivity to the choice of the pre-treatment time window. For Austria, V1 reacts markedly to extensions of the pre-treatment period. When the window is expanded to 2016–2019, pre-treatment fit deteriorates, with pre-RMSPE rising to 0.2903, while the estimated ATT increases to 0.2243. Extending the window further to 2015–2019 worsens pre-treatment fit again and raises the ATT to 0.5894. V2 shows the same broad pattern in a somewhat different form, as extending the pre-treatment period substantially reduces the estimated ATT. Romania exhibits a similar sensitivity, though expressed through a change in sign: longer pre-treatment windows produce substantially larger positive ATTs of 0.2690 and 0.2402. Norway’s pre-window diagnostics are even more striking, as the baseline negative ATT shrinks toward zero and eventually changes sign when longer pre-treatment periods are used. This repeated pattern across all model families is too systematic to be treated as a minor robustness detail. Instead, it should be emphasised as a central empirical finding of the thesis: in this application, identification is highly sensitive to the choice of the pre-treatment regime considered sufficiently stable for synthetic control estimation.

Model 1 Austria versus No STW countries – Alternative pre-period checks
Two alternative pre-treatment windows

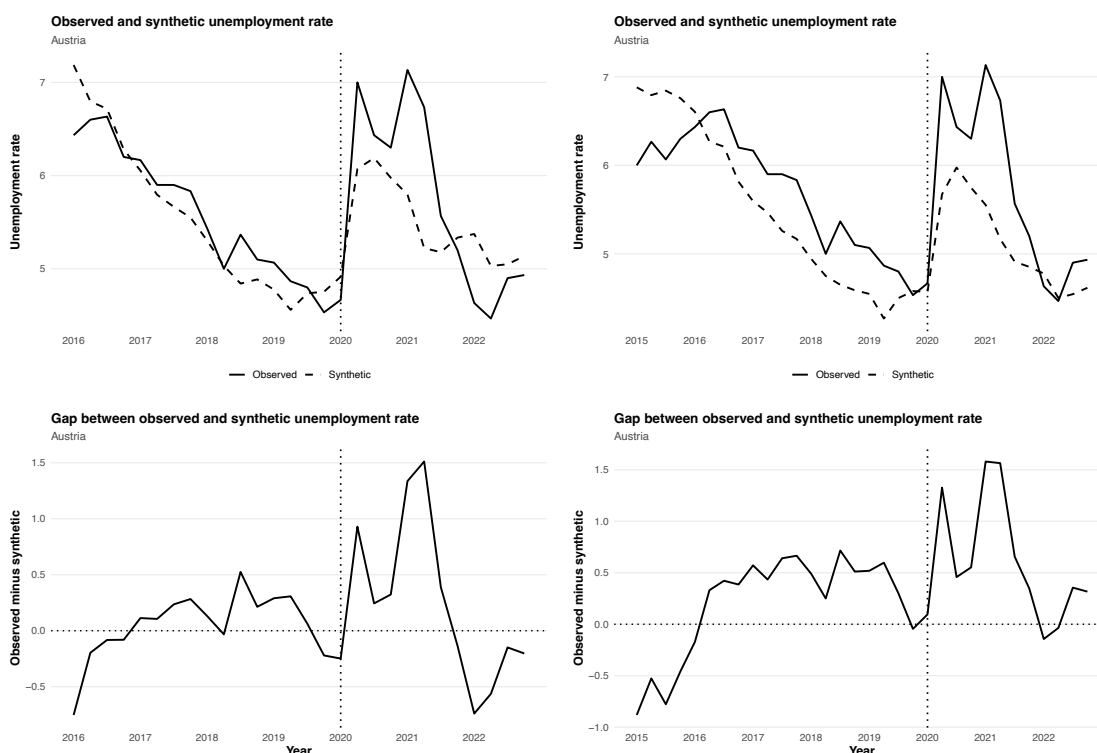


Figure 5.13:
Left: Alt. pre-window (2015–2019) – V1: AT vs. No-STW donor pool
Right: Alt. pre-window (2016–2019) – V1: AT vs. No-STW donor pool

Model 2 Austria versus STW countries – Alternative pre-period checks
Two alternative pre-treatment windows

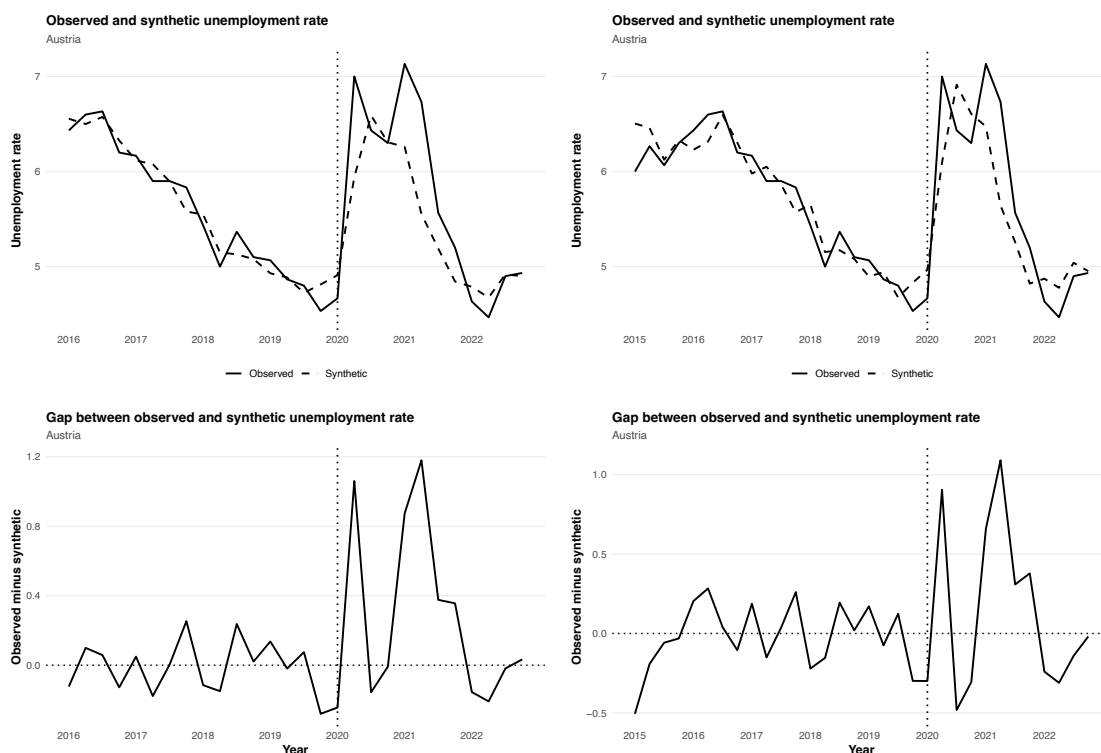


Figure 5.14:
Left: Alt. pre-window (2015–2019) – V2: AT vs. STW donor pool
Right: Alt. pre-window (2016–2019) – V2: AT vs. STW donor pool

Model 3 Romania versus STW countries – Alternative pre-period checks
Two alternative pre-treatment windows

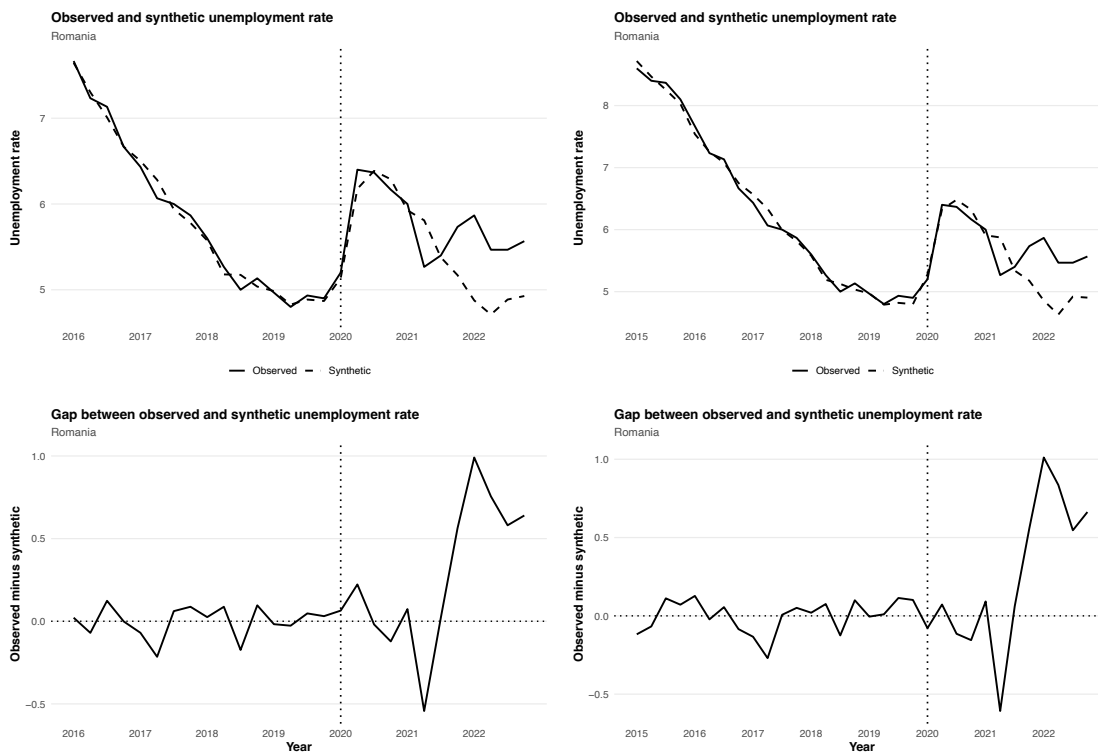


Figure 5.15:

Left: Alt. pre-window (2015–2019) – V3: RO vs. STW donor pool
Right: Alt. pre-window (2016–2019) – V3: RO vs. STW donor pool

Model 3 Norway versus STW countries – Alternative pre-period checks
Two alternative pre-treatment windows

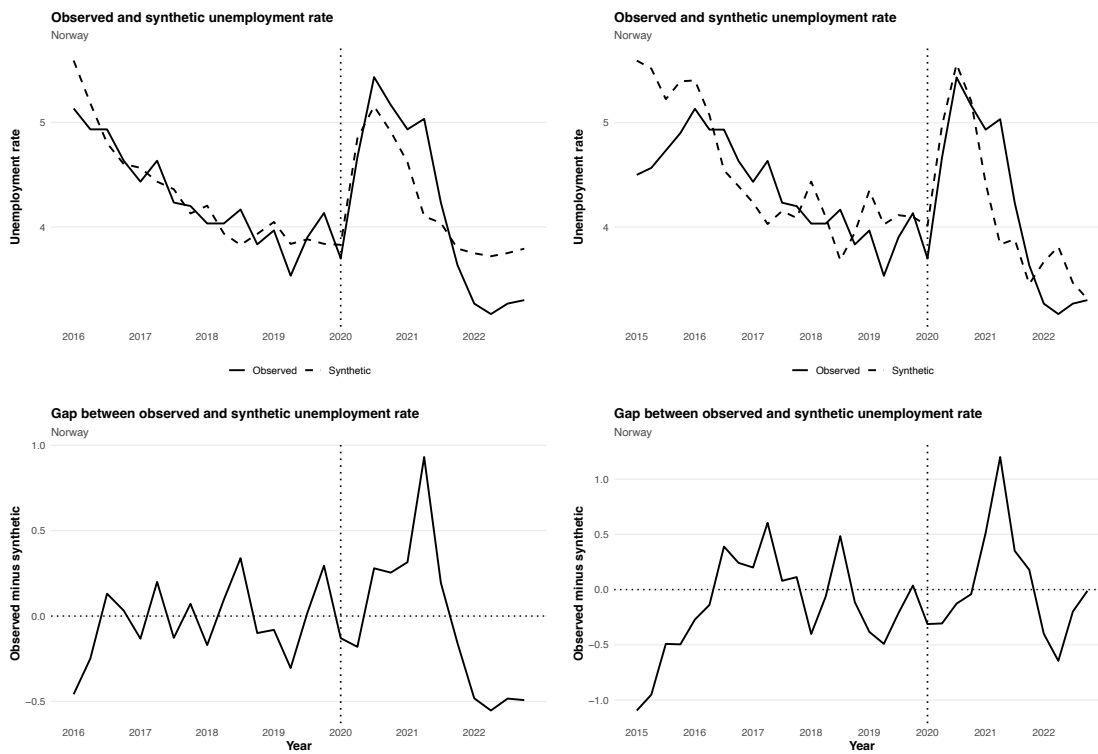


Figure 5.16:

Left: Alt. pre-window (2015–2019) – V3: NO vs. STW donor pool
Right: Alt. pre-window (2016–2019) – V3: NO vs. STW donor pool

A.7 Lagged outcome variants

This subsection addresses the timing concern discussed in Chapter 3 more directly by varying the temporal structure of the outcome information used in the matching process. Rather than relying only on the baseline setup, additional specifications are estimated that incorporate lagged unemployment information from earlier pre-treatment years. The purpose of these variants is not to turn the SCM into a dynamic time-series model, but to examine whether the main post-treatment results depend materially on how the pre-treatment outcome path is represented. This is particularly relevant in the present context because labor-market adjustment may be delayed and because unemployment may react gradually rather than contemporaneously to crisis conditions.

For V1 (Austria versus No-STW countries), the lagged outcome variants are highly reassuring. All three variants produce virtually identical results relative to the baseline. Pre-treatment RMSPE remains unchanged at 0.183 in all cases, the RMSPE ratio stays between 4.112 and 4.119, and the average post-treatment gap remains between 0.036 and 0.039. The deviations from baseline are therefore negligible: the change in the average post-treatment gap is at most 0.002, and the change in the RMSPE ratio remains within ± 0.004 . This means that the central V1 result is essentially invariant to whether the matching design additionally incorporates lagged outcome information from 2019 alone, from 2018 and 2019, or from 2017 to 2019. In substantive terms, the conclusion remains unchanged: the estimated average post-treatment gap in V1 is very small, and this finding is not driven by a particular temporal representation of the pre-treatment outcome path.

For V2 (Austria versus STW countries), the picture is more mixed and therefore more informative. The lagged outcome variants still point to a positive post-treatment gap, but they reduce the estimated magnitude relative to the baseline. The baseline model has a pre-treatment RMSPE of 0.132, an RMSPE ratio of 6.858 and an average post-treatment gap of 0.692. Under the lagged outcome variants, pre-treatment RMSPE increases moderately to between 0.138 and 0.165, the RMSPE ratio declines to between 3.659 and 4.905 and the average post-treatment gap falls to values between 0.189 and 0.579. The strongest attenuation occurs in the specification that includes lagged outcomes for 2018 and 2019, where the ATT declines by 0.504 relative to baseline. This means that the qualitative V2 result survives, since the estimated gap remains positive in all variants, but its magnitude is clearly sensitive to how the pre-treatment outcome dynamics are structured.

Taken together, the lagged outcome variants strengthen the interpretation in two ways. First, they show that V1 is highly stable with respect to the timing structure of the pre-treatment outcome path. Second, they show that V2 is more timing-sensitive: the direction of the result remains the same, but the size of the estimated post-treatment gap depends meaningfully on the lag structure. This does not invalidate the baseline V2 model, but it does imply that the stronger Austria-versus-STW result should be interpreted more cautiously than the baseline table alone might suggest for V1.

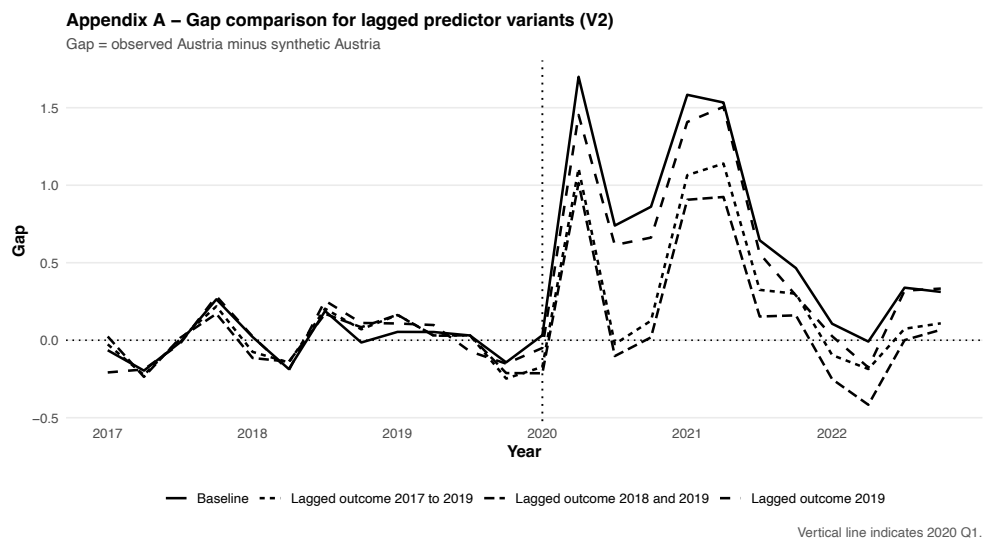
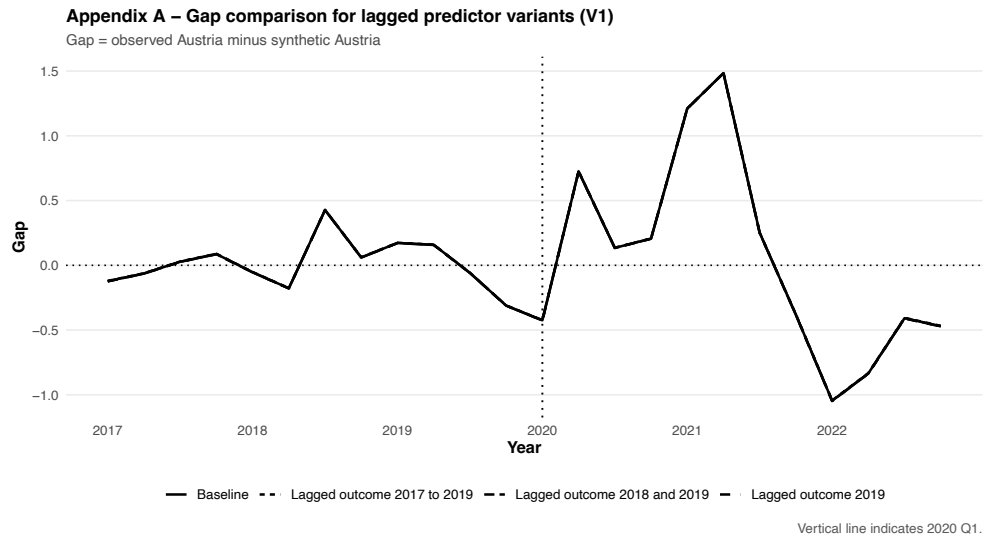


Figure 5.17:
Left: Gap comparison for lagged outcome variants (V1)
Right: Gap comparison for lagged outcome variants (V2)

Appendix A — Lagged predictor variants					
V1: Austria versus No-STW countries					
Variant	Pre-treatment RMSPE	Post-treatment RMSPE	RMSPE ratio	Average post-treatment gap	HHI
Lagged outcome 2019	0.183	0.755	4.119	0.036	0.868
Lagged outcome 2018 and 2019	0.183	0.754	4.117	0.036	0.867
Lagged outcome 2017 to 2019	0.183	0.753	4.112	0.039	0.865

Table 5.27: Lagged outcome variants V1

Appendix A — Lagged predictor variants					
V2: Austria versus STW countries					
Variant	Pre-treatment RMSPE	Post-treatment RMSPE	RMSPE ratio	Average post-treatment gap	HHI
Lagged outcome 2019	0.165	0.807	4.905	0.579	0.240
Lagged outcome 2018 and 2019	0.138	0.504	3.659	0.189	0.160
Lagged outcome 2017 to 2019	0.148	0.575	3.896	0.314	0.165

Table 5.28: Lagged outcome variants V2

A.8 Reduced specification without additional matching variables

This subsection reports a reduced specification in which the SCM is estimated without additional matching variables beyond the outcome path. The purpose of this exercise is to assess how strongly the baseline results depend on the richer macroeconomic matching design. In this reduced specification, the synthetic control is driven primarily by the pre-treatment behavior of unemployment itself rather than by the broader set of matching variables used in the baseline model. This provides a useful benchmark against which the baseline specification can be evaluated.

For V1, the reduced specification performs substantially worse than the baseline. Pre-treatment RMSPE increases from 0.183 to 2.231, which means that the synthetic control no longer reproduces Austria credibly before treatment. At the same time, the RMSPE ratio falls sharply from 4.116 to 1.095, and the average post-treatment gap rises from 0.037 to 2.356. The change in the estimated post-treatment gap is therefore extremely large at 2.319. This pattern strongly suggests that the reduced specification breaks down not because it reveals a better counterfactual, but because it fails to match Austria adequately in the pre-treatment period. In other words, for V1 the richer baseline predictor set is not optional refinement, but necessary for constructing a meaningful synthetic control.

For V2, the reduced specification also weakens the model substantially. Pre-treatment RMSPE rises from 0.132 to 0.868, the RMSPE ratio collapses from 6.858 to 0.581, and the average post-treatment gap declines from 0.692 to 0.112. The reduction in ATT relative to baseline is therefore 0.580. This result should not be interpreted as stronger evidence against the baseline V2 estimate. Rather, it reflects the fact that the reduced model no longer fits Austria sufficiently well before treatment. As in V1, removing the additional matching variables leads to a synthetic construction that is much less credible as a counterfactual.

Taken together, the reduced-specification results support the broader methodological logic of the thesis. The richer baseline predictor design is not merely adding complexity without purpose. Instead, it is what allows the synthetic control to reproduce Austria reasonably well in the pre-treatment period. Especially in V1, the collapse in fit is so large that the reduced specification should be treated mainly as a diagnostic benchmark rather than as a competing model. In V2, the same logic applies: the smaller post-treatment gap under the reduced specification is not evidence of a better model, but mainly the consequence of a much weaker pre-treatment match.

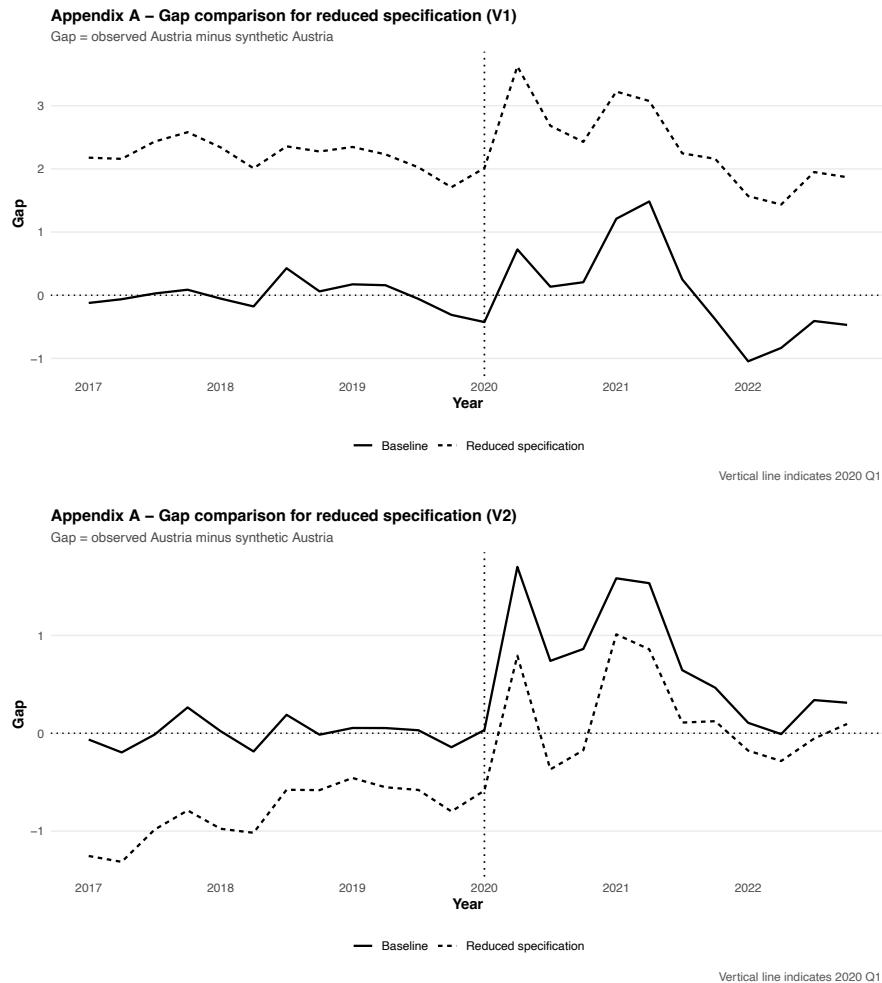


Figure 5.18:
Left: Gap comparison for reduced specification V1
Right: Gap comparison for reduced specification V2

A.9 Summary comparison to the baseline model

The appendix checks reported in A.7 and A.8 provide a fuller basis for interpreting the two main Austria models. Their purpose is not to replace the baseline specifications, but to show more clearly where the baseline findings are stable and where they are sensitive to modeling decisions. In particular, the lagged outcome variants address the timing concern raised in Chapter 3, while the reduced specification addresses the question of how strongly the results depend on the inclusion of additional matching variables.

For V1, the overall message is very clear. The lagged outcome variants leave the main result almost completely unchanged: pre-treatment fit remains identical, the RMSPE ratio changes only trivially, and the average post-treatment gap stays close to zero. By contrast, the reduced specification performs extremely poorly, with a large deterioration in pre-treatment fit and a dramatic rise in the estimated post-treatment gap. This means that the substantive V1 conclusion is stable with respect to timing structure, but not with respect to removing the broader matching design. In short, V1 is robust in direction and magnitude once the baseline predictor structure is maintained, but it cannot be reproduced credibly with an outcome-only design.

For V2, the comparison to baseline is more mixed. The lagged outcome variants still produce a positive Austria gap, but one that is clearly smaller than in the baseline, especially when more lagged years are included. This means that the qualitative direction of the V2 result survives, but its magnitude is timing-sensitive. The reduced specification weakens the model much more strongly still, with sharply worse pre- treatment fit and a much smaller post-treatment gap. The appropriate interpretation is therefore not that the baseline V2 estimate disappears under scrutiny, but that it is more specification-sensitive than V1. In substantive terms, V2 continues to suggest that Austria’s unemployment path lies above its STW-based synthetic peer benchmark, but the exact strength of that result depends materially on how the pre- treatment outcome dynamics are represented and whether the richer matching design is retained.

Overall, the appendix checks reinforce the broader interpretation of the thesis. The strict No-STW model (V1) remains highly stable once alternative lag structures of the outcome are introduced, but it still depends on a sufficiently rich matching design.

The STW-based peer model (V2) continues to point to a positive Austria gap, but this result is more sensitive to modeling choices and should therefore be interpreted with greater caution. These appendix results do not overturn the baseline findings.

Rather, they help clarify which parts of the main argument are robust and where the remaining uncertainty comes from.

Appendix A — Comparison to baseline							
V1: Austria versus No-STW countries							
Section	Variant	Pre-treatment RMSPE	Δ vs baseline	RMSPE ratio	Δ vs baseline	Average post-treatment gap	Δ vs baseline
Baseline	Baseline	0.183	0.000	4.116	0.000	0.037	0.000
B.1 Lagged predictor variants	Lagged outcome 2019	0.183	0.000	4.119	0.003	0.036	−0.001
B.1 Lagged predictor variants	Lagged outcome 2018 and 2019	0.183	0.000	4.117	0.001	0.036	−0.001
B.1 Lagged predictor variants	Lagged outcome 2017 to 2019	0.183	0.000	4.112	−0.004	0.039	0.002
B.2 Reduced specification	Reduced specification	2.231	2.048	1.095	−3.021	2.356	2.319

Table 5.29: Comparison to baseline V1

Appendix A — Comparison to baseline							
V2: Austria versus STW countries							
Section	Variant	Pre-treatment RMSPE	Δ vs baseline	RMSPE ratio	Δ vs baseline	Average post-treatment gap	Δ vs baseline
Baseline	Baseline	0.132	0.000	6.858	0.000	0.692	0.000
B.1 Lagged predictor variants	Lagged outcome 2019	0.165	0.032	4.905	−1.953	0.579	−0.114
B.1 Lagged predictor variants	Lagged outcome 2018 and 2019	0.138	0.005	3.659	−3.199	0.189	−0.504
B.1 Lagged predictor variants	Lagged outcome 2017 to 2019	0.148	0.015	3.896	−2.962	0.314	−0.378
B.2 Reduced specification	Reduced specification	0.868	0.735	0.581	−6.278	0.112	−0.580

Table 5.30: Comparison to baseline V2

Appendix B: Extensive pre-treatment summary

Appendix Table A2

Pre-treatment means for Austria and individual Model 2 donor countries

Variable	Austria	Belgium	Bulgaria	Cyprus	Denmark	Estonia	Spain	Finland	France	Greece	Hungary	Ireland	Italy	Lithuania	Latvia	Malta	Netherlands	Poland	Portugal	Sweden	Slovenia	Slovakia
Unemployment rate	5.33	6.14	6.17	8.89	5.32	5.22	15.54	7.63	8.96	19.82	3.59	5.85	10.60	6.53	7.48	4.02	5.06	4.03	7.68	6.72	5.37	6.76
GDP index	104.81	102.73	99.80	97.58	100.06	98.98	109.99	101.25	106.20	107.61	99.56	87.85	109.28	95.48	101.28	98.58	101.68	97.24	105.99	99.81	100.88	99.83
Employment rate	72.92	64.29	68.35	68.53	74.12	74.30	62.27	71.67	65.20	54.97	69.31	68.62	58.56	71.92	71.40	70.69	77.07	67.33	69.28	77.12	70.76	67.39
openness	2.05	2.04	2.06	1.85	2.03	1.97	2.16	2.04	2.15	2.08	1.98	1.84	2.10	2.00	1.95	1.77	1.97	1.99	2.13	2.05	2.09	2.10
Average hours worked	36.49	38.60	40.26	39.10	34.35	38.10	37.57	36.61	37.21	41.03	39.63	37.05	37.06	38.66	38.64	39.85	32.05	39.97	39.41	37.08	39.27	39.97
Industrial production index	111.61	110.12	107.12	126.07	108.66	113.42	105.52	110.70	103.42	107.92	110.49	100.33	105.65	115.21	115.58	102.14	102.79	116.17	105.62	109.24	119.97	112.52
Inflation	1.94	1.94	2.09	0.67	0.84	3.10	1.51	1.04	1.52	0.81	2.91	0.62	1.08	2.83	2.74	1.49	1.86	1.64	1.00	1.88	1.73	2.24

Notes: The table reports simple averages over the pre-treatment period 2017–2019. Each country column shows the country-specific pre-treatment mean before SCM weighting is applied.