



The Ronald O. Perelman Center for Political
Science and Economics (PCPSE)
133 South 36th Street
Philadelphia, PA 19104-6297

pier@econ.upenn.edu
<http://economics.sas.upenn.edu/pier>

PIER Working Paper 22-026

A Stage-Based Identification of Policy Effects

CHRISTIAN ALEMÁN
Université Libre de Bruxelles
ECARES

ALEXANDER LUDWIG
Goethe University Frankfurt
ICIR, UAB and CEPR

CHRISTOPHER BUSCH
LMU Munich
CESifo

RAÜL SANTAEULÀLIA-LLOPIS
University of Pennsylvania
UAB, BSE and CEPR

October 28, 2022

A Stage-Based Identification of Policy Effects*

Christian Alemán Université Libre de Bruxelles ECARES	Christopher Busch LMU Munich CESifo	Alexander Ludwig Goethe University Frankfurt ICIR, UAB and CEPR
	Raül Santaèulàlia-Llopis University of Pennsylvania UAB, BSE and CEPR	

October 28, 2022

Abstract

We develop a method that identifies the effects of policy implemented nationwide—i.e. across all regions at the same time. Starting point is the insight that outcome paths can be tracked over *stages* using a reference path. The *stage* of a regional outcome path is defined as its location on the support of a reference path. It is formally the result of a normalization that maps the time-path of regional outcomes onto a reference path using pre-policy data only. Intuitively, our normalization seeks to reshape the structural parameters that determine the outcome path of non-reference regions into those of a reference region—a phenomenon that we show with an example for which we can derive exact identification. Since regions can differ by *stage* at any point in time, our normalization uncovers heterogeneity in the *stage* at the time of policy implementation—even in instances where the implementation occurs at the same time across regions. We use this *stage* variation at the time of policy implementation to identify the policy effects: a *stage-leading* region delivers the counterfactual path inside an identification window in which non-leading regions are subject to policy whereas the leading region is not. Our identification assumption is that the normalization conducted using pre-policy data holds post policy, i.e. the normalization coefficients reshaping the regional pre-policy outcome paths into those of a reference region are unaffected by policy. We validate our method with Monte-Carlo experiments on model-generated data that detect bounds for a successful identification. We use our method to evaluate the effectiveness of public health stay-home policies (i.e. the national lockdown against Covid-19 in Spain), the effects of oral contraceptives (i.e. the 1960 FDA nationwide approval of oral contraceptives in the U.S.) on women's fertility and college education and the effects of growth policy (e.g. German Reunification). We further show how our method can be applied to non-nationwide policy—i.e. untreated regions and staggered rollouts—and discuss the implications of spillovers across regions.

Keywords: Policy Effects, Identification, Stages

JEL Classification: C01, H00, E01, E22, E25

*We thank Francesco Agostinelli, Brant Callaway, Xu Cheng, Hanming Fang, Frank Schorfheide and participants at seminars and conferences for helpful comments and suggestions. Raül Santaèulàlia-Llopis acknowledges financial support from the AGAUR 2020PANDE00036 "Pandemias" Grant 2020-2022, from the Spanish Ministry of Economy and Competitiveness, through the Proyectos I+D+i 2019 Retos Investigacion PID2019-110684RB-I00 Grant, Europa Excelencia EUR2021-122011 and the Severo Ochoa Programme for Centres of Excellence in R&D (CEX2019-000915-S).

Contents

1	Introduction	1
2	A Stage-Based Method to Identify Policy Effects	8
2.1	Normalization Procedure	9
2.2	Measuring the Policy Effects	13
2.3	An Analytical Example with Exact Identification	16
3	Method Performance	19
3.1	Can Our Method Recover the True Policy Effects?	19
3.1.1	Public Health Policy Against a Pandemic	20
3.1.2	The Pill and Women's Choice	24
3.1.3	Growth policy and structural transformation	28
3.2	A Monte Carlo Analysis: Bounds to Method Performance	32
3.3	Confounding Factors	35
3.3.1	Time-Varying Latent Heterogeneity	35
3.3.2	Confounding Policies	36
3.4	Inference	38
3.4.1	Placebo Diagnosis	39
3.4.2	Stochastic Component	40
4	Applications	43
4.1	The Effects of the Spanish <i>Confinamiento</i> Against Covid-19	43
4.2	The Effects of the 1960 FDA Approval of Oral Contraceptives in the U.S.	46
4.3	The Effects of the German Reunification	50
5	Heterogenous Policy Effects by Stage	52
6	Further Discussion	54
6.1	Non-Nationwide Policy	54

6.1.1	Untreated Regions	55
6.1.2	Staggered Rollout	57
6.2	Spillovers	60
7	Conclusion	61
A	Further Examples: Policy After the Peak	66
A.1	Policy After the Peak: One Region	66
A.2	Policy After the Peak: Two Regions	69
B	Tracking Population in the Econ-Epi Model: Further Details	72
C	More on Confounding Factors	73
C.1	More on Time-Varying Latent Heterogeneity	73
C.2	More on Confounding Policies	73
D	Further Details on Inference with an Stochastic Component	75
E	Further Results and Robustness on our Applications	77
E.1	Identified Policy Effects Without the Trend-Extraction Step	77
E.2	Placebo Diagnosis for the Applications	77
F	Extrapolating to Find the the Policy Effects for the Control Region	80
G	More on Policy Effects with Spillovers	83

1 Introduction

Motivation The empirical assessment of a policy requires the notion of a credible counterfactual. To construct this counterfactual, standard empirical strategies critically rely on cross-regional heterogeneity in the time of policy implementation as source of identification—e.g. the existence of one untreated region or a staggered rollout.¹ Further, the credibility of the counterfactual requires the pre-policy paths of the outcome of interest to be similar across regions with differences not exceeding a constant gap over time—i.e. the so-called parallel trends assumption.² However, many relevant policy contexts violate these conditions. First, a large set of policies are implemented nationwide—i.e. carried out to all regions at the same time. Since the nationwide nature of a policy eliminates the source of identification for standard empirical strategies, these strategies are unworkable in such policy contexts. Second, the pre-policy outcome paths can be non-linear and differ across regions—e.g. in their starting date, speed or magnitude, in a manner that precludes the parallel trends assumption to hold. We illustrate these challenges in panel (a) of Figure 1 with the hypothetically different outcome paths of two regions that are subject to a nationwide policy implemented across all regions at time t_p . Our goal in this paper is to provide an identification strategy for such policy contexts.

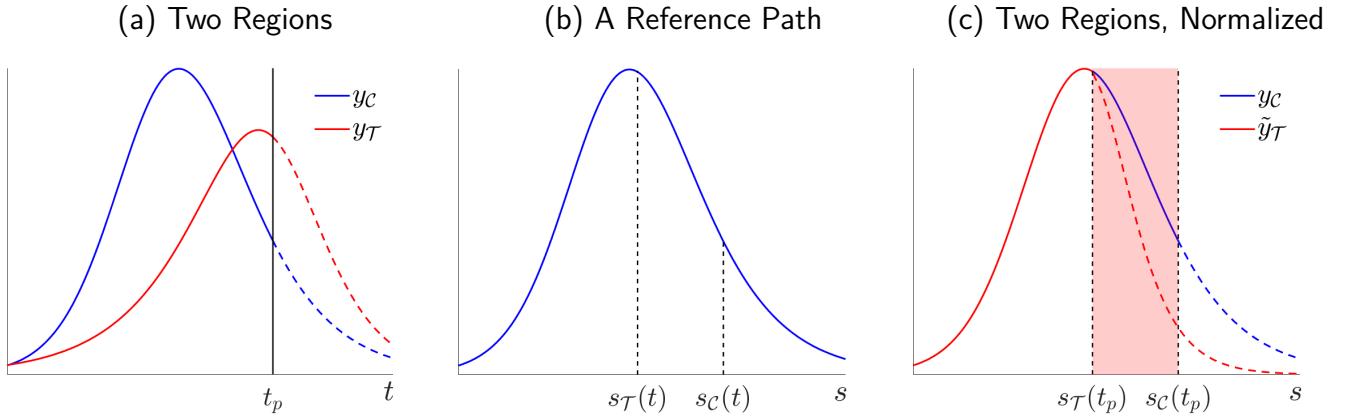
Idea We develop the idea that outcome paths can be tracked in terms of stages—as opposed to time. Precisely, consider a reference outcome path that evolves over stages as in panel (b) of Figure 1. Then, we define a stage of a regional outcome in time t as its location on the support of the reference outcome path.³ In this context, we note that there can be cross-regional heterogeneity in stages at any given point in time. For example, in our illustration, the outcome of region C is in a more advanced stage than the outcome of region T at a time t . For this reason, there is potential cross-regional heterogeneity in stages at the time of policy implementation, t_p . We argue that the cross-regional heterogeneity in stages at the time of policy implementation serves to identify policy effects even when policy is implemented in all regions at the same time. In order to exploit this potential cross-regional heterogeneity in stages at the time of policy implementation, we propose a stage-based identification (SBI) strategy—a method—based on mapping regional pre-policy outcome paths (in the time dimension) onto a reference outcome path (in the stage dimension).

¹See, among others, Angrist and Krueger (1999), Blundell and MacCurdy (1999) and Imbens and Rubin (2015). See also the more recent discussions in Athey and Imbens (2017) and Card (2022).

²This assumption is relaxed in Abadie (2005) when there exist untreated regions and, more recently, in Callaway and Sant'Anna (2021) for staggered policies. See also recent related work by Rambachan and Roth (2021).

³Hence, if a regional outcome path is chosen as reference path, the stage for the outcome of that reference region is time itself. As we discuss in Section 2.2, the choice of the reference path is innocuous for the identification of policy effects.

Figure 1: A Stage-Based Identification for Nationwide Policy: An Illustration



Notes: Panel (a) shows the hypothetical time paths of an outcome variable for a region C and another region T where a nationwide policy is implemented in both regions at the same time t_p . Panel (b) shows a reference path for the outcome of interest. In this illustration we choose the outcome path of region C as reference. Our identification strategy consists of a normalization that uses pre-policy data to map the non-reference outcome path, y_T , onto the reference outcome path, y_C . Panel (c) shows the result of our identification which is a normalized path of the non-reference region, \hat{y}_T , that is identical—up to minimization error—to the outcome path of the reference region before policy implementation in the stage dimension.

The Method Our method consists of a normalization mapping the pre-policy outcome paths of non-reference regions onto the outcome path of a reference region, for which the stage is defined as the calendar time. The normalization involves a scale factor adjustment together with a time-to-stage transformation. For this transformation we use a parametric function of time with monomial basis and an associated set of undetermined normalization coefficients.⁴ We determine these coefficients together with the scaling factor as the solution to a problem that minimizes the Euclidean distance between the pre-policy outcome path of the reference region and the normalized outcome path of the non-reference region before policy is implemented. As a result of our normalization, the regional outcome paths are identical before policy implementation—up to minimization error; see panel (c) of Figure 1 in which we use the outcome path of region C as reference and normalize the outcome path of region T to the reference.⁵

We provide an interpretation of our normalization using an analytical example with exact identification of the normalization coefficients. In particular, we show that the role of the normal-

⁴Note that under monomials, the normalization coefficients are easily interpreted. For example, the coefficient associated with the monomial of degree zero is a time-shifter for the entire outcome path of the non-reference region; the coefficient associated with the monomial of degree one adjust the speed of the outcome paths contracting (or expanding) time; and the coefficient associated with the monomial of degree two adjusts for asymmetries of the speed adjustment. We provide a detailed discussion in Section 2.

⁵The choice of the reference path anchors the (absolute) policy effects to the magnitude of that reference region (e.g. total number of deaths prevented by a stay-home policy against Covid-19). At the same time, the reference path is irrelevant for measuring policy effects in relative terms (e.g. proportion of deaths prevented by a stay-home policy against Covid-19); see our detailed discussion Section 2.2.

ization coefficients is to reshape the structural parameters that determine the outcome path of the non-reference region into those of the reference region before policy is implemented. In this manner, our identification strategy seeks to explicitly control for the cross-regional differences in structural parameters—*independently* of their (un)observability. More generally, in cases without exact identification where the structural parameters exceed the number of normalization coefficients, the role of the normalizing coefficients is to minimize the difference in the underlying structural parameters that determine the differential behavior across regions.

The identification of policy effects emerges because our normalization uncovers heterogeneity in the time of policy implementation. For example, in our illustration, policy is implemented at an earlier stage in region \mathcal{T} than in region \mathcal{C} . The resulting stage heterogeneity in the time of policy implementation opens an identification window in the stage dimension where a stage-leading region—in our example, region \mathcal{C} —is not subject to policy whereas a non-leading region—in our example, region \mathcal{T} —is subject to policy; see pink shaded area in panel (c) of Figure 1. Our identification of policy effects relies on the assumption that the normalization coefficients determined using only pre-policy data are unaffected by policy. That is, under our identification assumption we apply our pre-policy normalization coefficients onto post-policy (in terms of stages) data with the notion that, absent policy, the normalized outcome path of the non-reference region would be identical to the outcome path of the reference region. In this manner, our method captures the policy effects as the area between the actual path with policy minus the counterfactual path without policy inside the identification window.

Method Performance We apply our identification strategy to model-generated data in order to assess whether our identified policy effects recover the true policy effects that emerge from the model. We focus on three nationwide policies that resemble our empirical applications: a stay-home policy against a pandemic using a model where economic activity shapes and is shaped by the pandemic in equilibrium; the approval of oral contraceptives in a model with women fertility and education choices; and the removal of an institutional barrier to economic growth in a model of structural transformation. Here, note that the data available for a policy evaluator consists of the actual outcome paths—i.e. the regional paths before policy implementation pieced together with the associated outcome paths that are subject to policy after the policy is implemented. In particular, the true counterfactual outcome path that would occur if policy were absent throughout is not available outside of the model. Applying our identification strategy to the same set of model-generated data that would be available to a policy evaluator, we find that our method can successfully identify the true effects of nationwide policy generated by the model.

At the same time, we detect that the success of our method is bounded. With a Monte

Carlo study, we numerically characterize the bounds within which our method is able to recover the true (model-generated) effects of policy. Precisely, we randomly draw structural parameters to create a large number of artificial reference outcome paths. The artificial reference paths differ on how similar they are to the non-reference path which depends on the draw of structural parameters. Applying our method to the artificial paths we find that, in order for our method to work, the pre-policy paths of the reference regions cannot be too different from the outcome path of the non-reference region with precise bounds established via the Monte Carlo analysis. Under the interpretation that our normalization reshapes the structural parameters that determine the outcome path of the non-reference region, our method requires the structural parameters not to be too dissimilar across regions before policy implementation.

We further assess how our method fares in the presence of confounding factors. First, we explore the presence of time-varying latent heterogeneity that can differ across regions. Second, we asses the introduction of an additional and different policy before the policy under evaluation is actually implemented which generates behavioral change not attributable to the evaluated policy. Using model-generated data, we find that our identification strategy is robust to these new scenarios with behavioral change before policy implementation as long as these modifications keep the non-reference region sufficiently close to the reference region. The reason is that our method—normalization—seeks to control for the cross-regional differences in the behavior of the outcome path before policy is implemented—including potential behavioral changes generated from confounding factors. Again, for the method to work, the pre-policy regional outcome paths cannot be too different—that is, confounding factors cannot drive the regional outcome paths across regions too far from each other.

Three Applications We apply our identification strategy to study the effects of nationwide policy in three different applications. First, we study the effectiveness of the stay-home policy implemented nationwide in response to the first wave of the Covid-19 pandemic in the Spring of 2020 in Spain. We use the flow of deaths as our outcome of interest at the regional level—autonomous community, in Spain. Our method finds a leading region (Madrid) that is subject to the policy at a later stage than the rest of Spain at the time of policy implementation. Hence, the path of the flow of deaths in Madrid serves as counterfactual for the rest of Spain in the stage dimension. We find that the stay home policy significantly reduces the amount of deaths by 24.7% in the rest of Spain inside an identification window of seven days. In other words, had the stay-home policy not been implemented, there would have been 1,734 more deaths. Second, we assess the effects of the Food and Drugs Administration (FDA) approval of oral contraceptives (the pill) in the United States in 1960. We focus on two outcomes of interest: the crude fertility rates (number of births per 10,000) and the proportion of college women by cohort. In terms of

crude fertility rates, the leading region at the time that the pill was approved is West Virginia. We find that the pill reduces the crude fertility rate by 8.36% in the rest of the United States. In terms of the proportion of college women, we find Washington DC serves as the counterfactual leading region. In this context, we find that the pill increased the proportion of college women by 24.9% during the decade that followed the FDA approval. Third, we study the effect of the German reunification in 1990 on income per capita of West Germany. Our method finds that Hessen is the leading region in the stage dimension in West Germany. That is, Hessen is subject to reunification at a later stage than the rest of West Germany. Then, using the GDP per capita path of Hessen as counterfactual, we find that the German Reunification significantly reduces the income per capita of the rest of West Germany by 3.29% in a window of approximately 7 years.

Heterogeneous effects across stages. Note that our identification strategy applies to any pair of regions within a country. Focusing on our Covid-19 application, we find substantial heterogeneity in stages across Spanish regions at the time of policy implementation. For the non-leading regions, the introduction of the stay-home policy in terms of stages varies from about two weeks (e.g. Murcia) to just two days (e.g. Basque Country) earlier than the stage of the leading region at the time policy implementation. Further, we find that the stage at which policy is implemented systematically influences the effects of stay-home increasing the earlier the stage of policy implementation is (e.g. the number of prevented deaths is 65% in Murcia and 12% in the Basque Country). We further create an artificial set of regions using the the power set of the non-reference regions and reach similar insights.

Non-nationwide policy. We show that our method can also be applied to non-nationwide policy in the form of one untreated region or staggered rollout. In the case where there is one untreated region, our method delivers a right-open identification window bounded from below with the stage at which the policy is implemented in the treated region and unbounded from above. In the case of a staggered rollout where the policy is implemented across all regions but at different points in time, our method delivers an identification window that—for a pair of regions—is bounded from below with the stage at which the policy is implemented first across regions and bounded from above with the stage in which the policy is implemented last across regions. In the case of a staggered rollout of the policy, it is worth noting that our identification strategy can endogenously flip what region is considered control and what region is considered treatment with respect to the standard procedure in other empirical strategies. This flipping occurs when the region that implements the policy first in time happens to be at a later stage than that of regions that implement the policy later in time.

Related literature. Our method is directly related to the standard empirical strategies designed for settings that resemble natural experiments. These strategies rely on a difference-in-differences methodology in order to generate the counterfactual path (or potential outcome as in [Imbens and Rubin, 2015](#)) that serves as control for a treated region—i.e. the region subject to policy. We emphasize two main differences in our method. First, a critical common factor in previous strategies is that the source of identification relies on the heterogeneity in the time of policy implementation across regions either with the existence of one untreated region (e.g., [Card, 1990](#); [Card and Krueger, 2000](#)) or a staggered policy adoption (e.g., [Athey and Imbens, 2021](#); [Borusyak et al., 2021](#)).⁶ This is not the case in our method. Precisely, our main point of departure with respect to previous work is that our method is able to deliver identification of policy effects for contexts in which the cross-regional heterogeneity in the time of policy implementation is absent. In this paper, we provide a new identification that uncovers cross-regional heterogeneity in a stage dimension—rather than time—of the outcome of interest at the time of policy implementation. Using this stage heterogeneity at the time of policy implementation, our strategy is able to identify the effects of nationwide policy.

Second, a relevant concern for the standard empirical strategies is that there might be cross-regional differences in the pre-policy determinants of the outcome of interest that also determine the outcome paths after policy. Hence, the policy effects are only credibly identified after netting out of these determinants, an idea that is typically conveyed through the parallel trends assumption.⁷ In this context, there is a growing discussion on how to identify effects when parallel trends do not exactly hold. [Abadie \(2005\)](#) conditions the parallel trends to a set of observables using propensity scores ([Heckman et al., 1998](#)). This idea is extended to staggered rollout policy in [Callaway and Sant'Anna \(2021\)](#). Recently, [Rambachan and Roth \(2021\)](#) discuss how much the trends before policy implementation can differ from the trends after policy in order to still being able to identify causal effects. Our approach to this question is rather different. We design our method to minimize the distance in the pre-policy paths across regions with a normalization of the outcome variable that maps pre-policy non-reference paths onto a reference path. Therefore, to conduct causal inference, our identification does not require a parallel trend assumption. Basically, our method reshapes pre-policy paths directly aiming to control for the cross-regional differences in the pre-policy determinants of outcomes—e.g. region-specific structural parameters, whether these determinants are observable or not. At the same time, we show that, for our method to succeed, the pre-policy outcome paths before normalization can not be too different—with bounds that can be numerically constructed for the normalizing coefficients. In our method, the identification assumption is that the normalization mapping outcome paths onto a reference

⁶See also the recent discussion in [Goodman-Bacon \(2021\)](#).

⁷See [Bertrand et al. \(2004\)](#) among others.

path conducted using only pre-policy data also holds post policy.

Our work also speaks to the synthetic control group approach ([Abadie and Gardeazabal, 2003](#); [Abadie et al., 2010](#)).⁸ Two main differences—analogous to our previous discussion—stand out. First, as in other empirical strategies, with a synthetic control the policy evaluator requires the existence of one untreated region (or group) for identification. Our source of identification is new and differs from the synthetic control approach in that it builds on regional heterogeneity of stages at the time of policy implementation—i.e. an untreated region is not required for identification in our method. For this reason, unlike in the synthetic control group approach—or other methods for that matter, we can apply our identification strategy to nationwide policy occurring across all regions at the same time. Second, our method does not require the use of observable potential determinants of cross-regional differences in the outcome path in order to generate the counterfactual. Instead, the counterfactual in our method is constructed using solely the paths of the outcome of interest. At last, since our method can also be applied in non-nationwide policy, we also use the United States and an aggregate consisting of the same set of OECD countries (excluding Germany) used in [Abadie et al. \(2014\)](#) as controls for West Germany. We find that the policy effects that emerge from using our stage-based identification are not significantly different from those obtained using the synthetic control group approach.

Finally, our definition of stages relates to previous work in [Iorio and Santaeulàlia-Llopis \(2010, 2016\)](#) where country-specific HIV epidemic paths are normalized to a reference path in order to define stages of the epidemic in a comparable manner across time and space. We depart from that work in that we use our normalization to a reference path as base for identifying the effects of policies that aim to alter the path of the outcome of interest. For this reason, our normalization coefficients are obtained using strictly pre-policy outcome paths. More broadly, our work relates to analysis of the demographic transition (e.g. [Greenwood et al., 2005](#)) and structural transformation (e.g. [Galor and Weil, 2000](#); [Hansen and Prescott, 2002](#); [Gollin et al., 2002](#); [Herrendorf et al., 2014](#); [Cervellati and Sunde, 2015](#)) where, typically, the level of income per capita summarizes the stage of development in cross-country comparisons. In contrast, rather than replacing time for an observable (e.g. income per capita), we conduct a time-to-stage normalization of the outcome of interest which can be income per capita itself; see our evaluation of the German reunification on the income per capita of West Germany, for example.

The rest of the paper is structured as follows. We discuss our identification strategy in [Section 2](#). We assess the performance of our method using model generated data in alternative policy contexts in [Section 3](#). We conduct empirical applications in [Section 4](#). We discuss heterogenous

⁸[Doudchenko and Imbens \(2017\)](#) use a joint framework for difference-in-difference and synthetic control groups.

policy effects by stage in Section 5. We provide further discussion on non-nationwide policy and spillover effects in Section 6. Section 7 concludes.

2 A Stage-Based Method to Identify Policy Effects

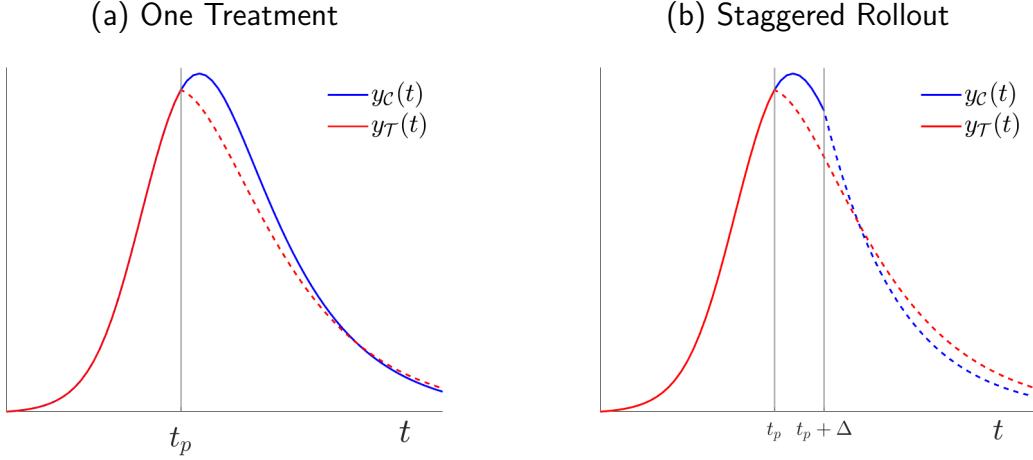
To contextualize our contribution, we first briefly discuss how standard empirical strategies identify policy effects.⁹ Consider a scenario in which, absent any policy intervention, the time path of an outcome $y_r(t)$ is identical across two regions $r \in \{\mathcal{C}, \mathcal{T}\}$.¹⁰ Now assume that a policy is implemented only in region \mathcal{T} at some date t_p which affects the outcome path in that region after policy implementation. Illustratively, we plot a hypothetical outcome path of a treated region $y_{\mathcal{T}}(t)$ before policy implementation (solid red) and after policy implementation (dashed red) in panel (a) of Figure 2. We also show a hypothetical outcome path for a region where the policy is not implemented, $y_{\mathcal{C}}(t)$ (solid blue). This scenario is ideal because the pre-policy outcome paths are identical across regions which warrants the use of region \mathcal{C} as control for region \mathcal{T} . That is, the outcome path $y_{\mathcal{C}}(t)$ provides a useful counterfactual (without policy) to assess the effects of policy on $y_{\mathcal{T}}(t)$ after t_p . The effects of policy are captured by the difference between $y_{\mathcal{C}}(t)$ and $y_{\mathcal{T}}(t)$ in the interval (t_p, ∞) . To this scenario, we can further add the implementation of the same policy to region \mathcal{C} at some later date $t_p + \Delta$ with $\Delta > 0$; see panel (b) of Figure 2. Under this staggered rollout of the policy, the effects of policy on region \mathcal{T} are identified using region \mathcal{C} as counterfactual within the interval $(t_p, t_p + \Delta]$. In that interval, region \mathcal{T} is subject to the policy whereas region \mathcal{C} is not—which—together with the fact that the outcome paths are identical across regions before t_p —warrants identification.

In this context, we find worth noting that the standard identification strategies of policy effects just described—either in the case with only one treated region or with the staggered rollout, fundamentally rely on two principles. First, the behavior of the outcome path before policy implementation must be credibly similar (the so-called ‘parallel trends) across regions—which is accomplished in our exemplified scenarios by assumption. Second, there must be variation in the time of policy implementation across regions which serves as source of identification. Unfortunately, many policy contexts violate these conditions which poses a challenge for the standard identification of policy effects. First, the path of the outcome variable before policy is implemented often differs across regions. In particular, outcome paths can differ by starting date, evolve at different speed and show different magnitude. Second, a large set of policies are implemented nationwide—i.e. carried out to *all regions at the same time*. The nationwide nature

⁹See comprehensive discussions in, for example, [Imbens and Rubin \(2015\)](#) and [Card \(2022\)](#).

¹⁰We use *regions* in the description of the method due to the applications presented below. *Region* can be used interchangeably with *group* or *unit* throughout.

Figure 2: Ideal Policy Scenarios with Two Regions: Standard Identification Strategies



Notes: Denote with $y_C(t)$ and $y_T(t)$ the outcome paths of, respectively, region \mathcal{C} and \mathcal{T} . Solid lines depict outcome paths before policy implementation and dashed lines after policy. The identified policy effects are captured by $\frac{\int_{\rightarrow t_p}^h (y_C(t) - y_T(t)) dt}{\int_{\rightarrow t_p}^h y_T(t) dt}$ with $h = \infty$ in the one-treatment case and $h = t_p + \Delta$ in the staggered rollout.

of a policy shatters the source of identification used in standard strategies. We illustrate these two challenges in panel (a) of Figure 3 where a nationwide policy is implemented in a context where the outcome path of region \mathcal{C} starts earlier, evolves at a faster speed and reaches a larger magnitude than in region \mathcal{T} .

To address these challenges, we propose an identification strategy based on the idea that outcome paths can be tracked in terms of stages—as opposed to time. Since regions can be at different stages at the same time t , there is potential variation in terms of stages at the time of policy implementation, t_p . Formally, our strategy consists of a normalization that maps regional outcome paths onto a reference outcome path using only pre-policy data. This normalization serves two purposes at once. First, the mapping generates identical pre-policy paths across regions. Second, the mapping uncovers variation in terms of stages at the time of policy implementation.

2.1 Normalization Procedure

We anchor the notion of stages of an outcome of interest $y_r(t)$ to a reference path—which can be the outcome path of a specific region. Then, we define the stage of a non-reference region as its location on the support of the reference path. Precisely, in our method, the stage is the result of a normalization that maps the outcome of a non-reference region $y_r(t)$ onto the reference

path. We now describe this normalization procedure and provide a formal definition of the stages afterwards.

In a first step of our normalization, we define a stage-time transformation as follows:

$$s = s_r(t; \psi_{-0}) = \begin{cases} t & \text{if } r = \mathcal{T} \\ \sum_{k=1}^n \psi_k B_k(t) & \text{if } r = \mathcal{C} \end{cases} \quad (1)$$

where $s_r : t \rightarrow \mathcal{S} = \mathbb{R}$ for $r = \{\mathcal{C}, \mathcal{T}\}$. The set $\psi_{-0} = \{\psi_1, \psi_2, \dots, \psi_n\}$ is a vector collecting a set of n unknown coefficients to be determined and $B_k(t)$ is a known basis function. Without loss of generality, we treat the outcome path of region \mathcal{T} as the reference path.¹¹ The stage for the outcome of the reference region—here, $y_{\mathcal{T}}(t)$ —is time itself, i.e. $s = s_{\mathcal{T}}(t; \psi_{-0}) = t$. Instead, the stage-time transformation implies that the stage for the outcome path of the non-reference region—here, $y_{\mathcal{C}}(t)$ —is $s = s_{\mathcal{C}}(t; \psi_{-0}) = \sum_{k=1}^n \psi_k B_k(t)$.

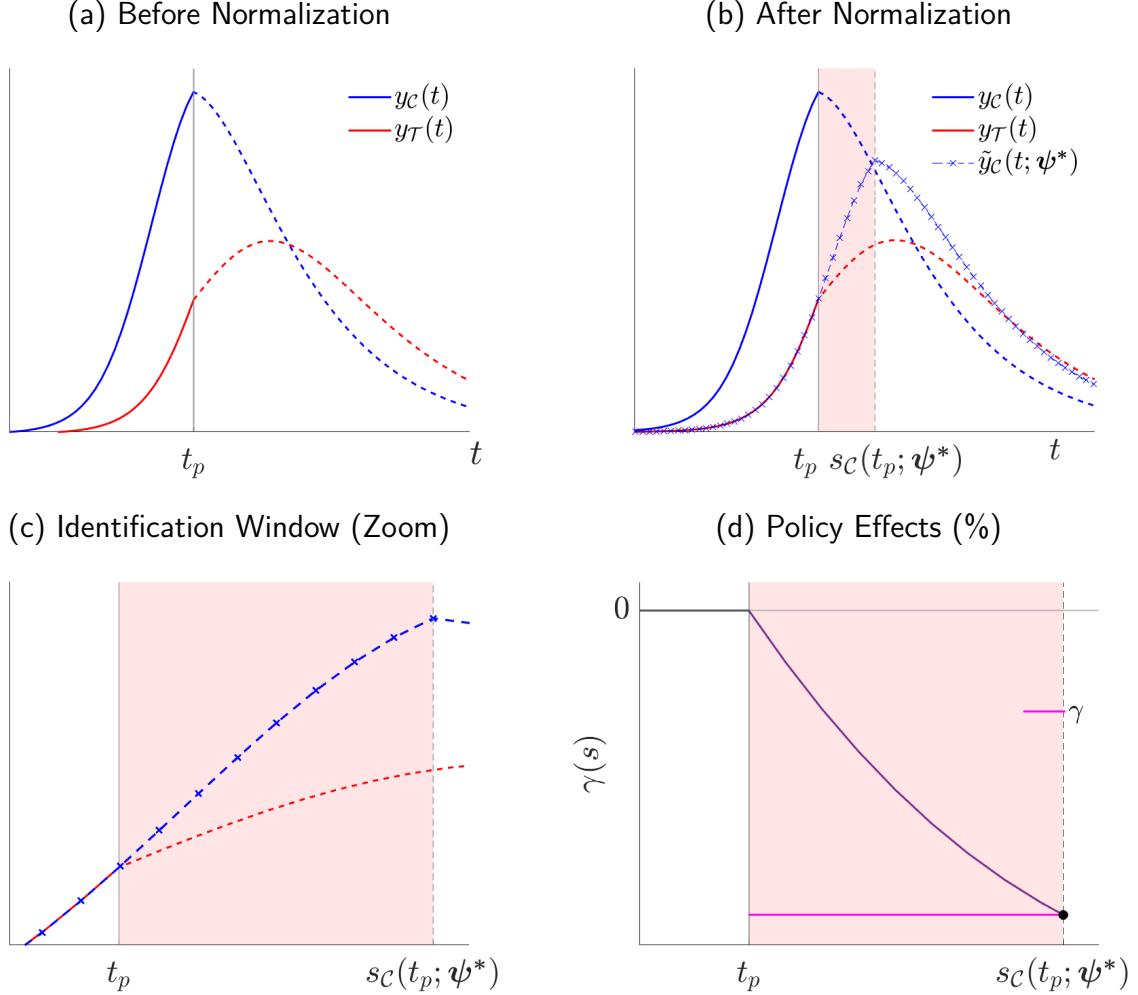
We choose the monomial basis for our benchmark normalization, i.e. $B_k(t) = t^{k-1}$. A nice feature of the monomial basis is that it delivers a straightforward interpretation of the coefficients behind the stage-time transformation. The parameter ψ_1 shifts the entire outcome path forward (with $\psi_1 > 0$) or backwards (with $\psi_1 < 0$) in time, adjusting for different start dates. The parameter ψ_2 adjusts the speed in a constant way across periods. If $\psi_2 < 1$ the stage-time transformation contracts (in time) the outcome path of region \mathcal{C} , whereas with $\psi_2 > 1$ the stage-time transformation expands (in time) the outcome path of region \mathcal{C} . That is, if $\psi_2 < 1$, then region \mathcal{T} is permanently faster than region \mathcal{C} , i.e., in one time-period region \mathcal{C} advances by more than one stage, and vice versa for $\psi_2 > 1$. Further, allowing for the stage-time transformation to be quadratic (i.e. $\psi_3 > 0$) captures the notion that the relative speed across the regions can change over time: for example, the outcome path of region \mathcal{T} 's might initially be slower than region \mathcal{C} 's, then catch up, and eventually move faster. While in principle the mapping can be of a higher degree, we consider benchmarks up to a quadratic mapping.

In a second step, we solve for the unknown coefficients ψ_{-0} together with a scaling parameter ψ_0 that shapes the magnitude of the normalized outcome path. Precisely, our criteria is to determine the set of coefficients $\psi = \{\psi_0, \psi_{-0}\}$ as those that minimize the (log) difference between the outcome path of the reference region, $y_{\mathcal{T}}(s) = y_{\mathcal{T}}(t)$, and the normalized path of the non-reference region, $\tilde{y}_{\mathcal{C}}(s) = \tilde{y}_{\mathcal{C}}(t; \psi) = \psi_0 y_{\mathcal{C}}(s(t; \psi_{-0}))$, that is:

$$\min_{\{\psi\}} \|\ln \tilde{y}_{\mathcal{C}}(s) - \ln y_{\mathcal{T}}(s)\|_{\mathbb{C}(s)}, \quad (2)$$

¹¹There choice of the reference region is innocuous, see our discussion in Section 2.2.

Figure 3: A Stage-Based Identification of Policy Effects: A Nationwide Policy



Notes: In panel (d), we report the policy effects γ together with the interim cumulative effects of policy, $\gamma(s)$, as defined in Section 2.2.

where $\|\cdot\|$ is the euclidean distance defined on the interval of stages,

$$\mathbb{C}(s) = \{\mathbb{C}_r(s)\} = \begin{cases} [s_r(t_0; \psi_{-0}), s_T(t_p; \psi_{-0})] & \text{if } s_T(t_p; \psi_{-0}) \leq s_C(t_p; \psi_{-0}) \\ [s_r(t_0; \psi_{-0}), s_C(t_p; \psi_{-0})] & \text{if } s_T(t_p; \psi_{-0}) > s_C(t_p; \psi_{-0}) \end{cases} \quad (3)$$

for $r = \{\mathcal{C}, \mathcal{T}\}$ and note that $s_T(t_p; \psi_{-0}) = t_p$. That is, the interval $\mathbb{C}(s)$ ensures that the minimization (2) only uses the outcome paths up to the first policy implementation—in terms of stages—across regions. Note that since $\mathbb{C}(s)$ contains ψ_{-0} , which region ends up considered as treatment or control is endogenous to our normalization. If $t_p \leq s_C(t_p; \psi)$, then policy arrives to region \mathcal{T} at an earlier stage than to region \mathcal{C} which sets \mathcal{T} as the treated region and \mathcal{C} as control.

The opposite occurs if $t_p > s_C(t_p; \psi)$. Now, we can define the stages of an outcome, $y_r(t)$, for any region r and time t .

Definition 1. *The stage of an outcome $y_r(t)$ of region r at time t is $s_r(t; \psi_{-0}^*)$ where ψ_{-0}^* is the solution to the minimization of (2) subject to (1) and (3).*

Thus, the stages formally emerge as the result of our normalization procedure that maps the outcome path of a non-reference region onto the outcome path of the reference region before policy is implemented. To gain some intuition, we exemplify our method using a nationwide policy that affects the outcome path of two regions, $y_C(t)$ and $y_T(t)$. Before policy implementation at time t_p , the outcome path of region C (solid red) differs from region T (solid blue) in that it starts earlier, grows faster and is larger. We also show the outcome paths after policy implementation for, respectively, region C (dashed blue) and T (dashed red); see panel (a) of Figure 3. Our normalization procedure—i.e. the minimization of (2) subject to (1) and (3)—delivers a normalized outcome path for the non-reference region, $\tilde{y}_C(t; \psi^*)$ (cross-dashed blue), that maps onto the outcome path of the reference region before policy implementation; see panel (b) of Figure 3.^{12,13}

In this manner, our strategy accomplishes two goals. First, the normalization makes the pre-policy paths of both regions identical—up to a minimization error—before policy is implemented in terms of stages, i.e. for $s \in \mathbb{C}(s)$. Second, the normalization uncovers stage variation in the time of policy implementation across regions. Precisely, in our example, we find that $t_p < s_C(t_p; \psi^*)$, that is, policy is implemented first—in terms of stages—in region T which sets that region as treated and region C as control.

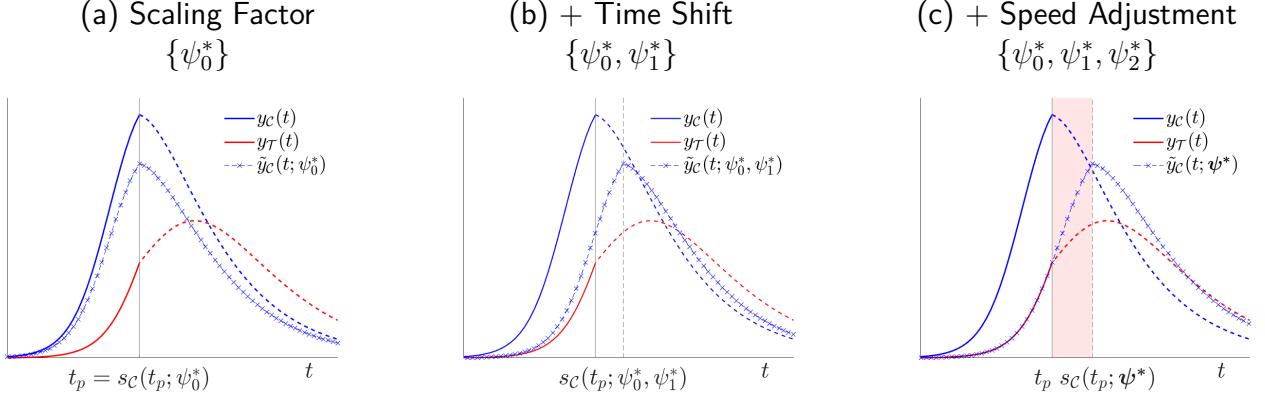
We further unpack the effects of each of the coefficients $\{\psi^*\}$ on the path of the non-reference region in Figure 4. Since these coefficients are jointly determined in our minimization, we provide a non-orthogonal decomposition where we sequentially add the effects of each parameter. In panel (a), the scaling factor, $\psi_0^* < 1$, proportionally shifts down the entire outcome path of the non-reference region C throughout its support. In panel (b), the additional time shifter, $\psi_1^* > 0$, moves the outcome path to the right delaying the outcome's take off. In panel (c), adding the speed adjustment, $\psi_2^* > 1$, decreases the pace of the normalized outcome.¹⁴ We further explore

¹²Note that outcome variables are almost invariably observed on discrete dates. For these cases, since our mapping can generate stages $s_T(t; \psi)$ that are non-integer values—i.e. not on discrete dates—we interpolate between $y_C(fl(s_T(t; \psi)))$ and $y_C(cl(s_T(t; \psi)))$, where $fl(\cdot)$ and $cl(\cdot)$ denote the integer floor or integer ceiling, respectively.

¹³In the Appendix A, we provide additional examples that include nationwide policies that are implemented when a region has already surpassed the peak whereas other regions have not reached the peak yet.

¹⁴Note that in the realm of standard empirical strategies one can partly address the time shift, ψ_1 . This requires a choice by the researcher to fix the region-specific start dates of the outcome path of interest. For example, for the analysis of a Covid-19 containment policy this has been suggested in an event study design by Liu et al.

Figure 4: Unpacking the Normalization



Notes: Normalization step by step

this decomposition with additional undetermined coefficients in the Appendix.¹⁵

2.2 Measuring the Policy Effects

To measure the effects of policy, we apply our normalization of the outcome path of non-reference regions on the post-policy data using the parameters ψ^* obtained with pre-policy data. Precisely, applying ψ^* to the post-policy data of the non-reference region we note that the normalized path of the non-reference region, $\tilde{y}_C(s)$, and the path of the reference region, $y_T(s)$, overlap during an interval of stages—an identification window—in which one of the regions is subject to policy whereas the other region is not. This gives rise to our identification assumption.

Identification Assumption 1. *The normalization of the outcome paths of non-reference regions using $\{\psi^*\}$ —i.e. the solution the minimization of (2) subject to (1) and (3)—holds for $s \in \mathbb{D}(s)$ where,*

$$\mathbb{D}(s) = \begin{cases} [s_T(t_p; \psi_{-0}^*), s_C(t_p; \psi_{-0}^*)] & \text{if } s_T(t_p; \psi_{-0}^*) \leq s_C(t_p; \psi_{-0}^*) \\ [s_C(t_p; \psi_{-0}^*), s_T(t_p; \psi_{-0}^*)] & \text{if } s_T(t_p; \psi_{-0}^*) > s_C(t_p; \psi_{-0}^*) \end{cases}. \quad (4)$$

In other words, our identification assumes that, absent policy, the normalized non-reference outcome path, $\tilde{y}_C(s)$, is identical to the reference path, $y_T(s)$, for all stages inside the identification window defined by (4). Our identification of policy effects rests in this assumption which we

(2021) and Glogowsky et al. (2021). In contrast, our method endogenously finds the appropriate time shifter (ψ_1) together with a speed adjustment (ψ_2).

¹⁵For example, adding a quadratic term to the time-to-stage transformation (i.e. $\psi_3 > 0$) helps capturing asymmetries that are potentially relevant if the policy occurs after the peak; see Appendix.

assess in the context of an analytical example with exact identification in Section 2.3 and also through a placebo test with model-generated data without exact identification in Section 3.4.1.

Following our illustrative example, we show the identification window as the shaded (pink) area in panel (b) of Figure 3. There, since the reference region is \mathcal{T} , then $s_{\mathcal{T}}(t_p; \psi_{-0}^*) = t_p$. Further, since we find that $t_p \leq s_{\mathcal{C}}(t_p; \psi_{-0}^*)$, then $\mathbb{D}(s) = [t_p, s_{\mathcal{C}}(t_p; \psi_{-0}^*)]$. That is, the identification window goes from the stage where the policy was implemented in region \mathcal{T} , i.e. t_p , to the stage where the policy is implemented in region \mathcal{C} , i.e. $s_{\mathcal{C}}(t_p; \psi_{-0}^*)$.

Policy effects In the identification window, the outcome path of the reference region $y_{\mathcal{T}}(t)$ is affected by policy whereas the normalized outcome path of the non-reference region $\tilde{y}_{\mathcal{C}}(s)$ is not. Therefore, $\tilde{y}_{\mathcal{C}}(s)$ serves as a counterfactual for $y_{\mathcal{T}}(s)$ for all stages $s \in \mathbb{D}(s)$.¹⁶ We measure the effects of policy as,

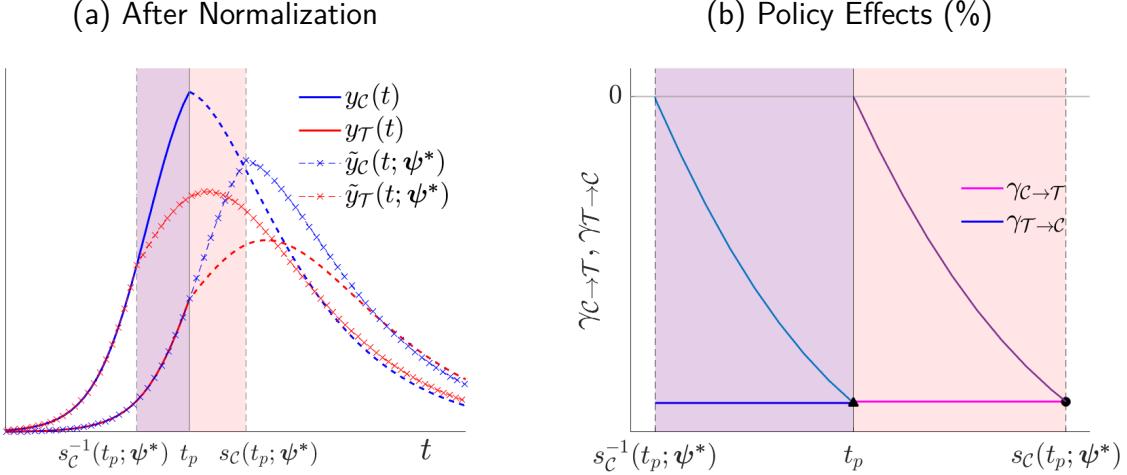
$$\gamma = \frac{\int_{\mathbb{D}(s)} (y_{\mathcal{T}}(s) - \tilde{y}_{\mathcal{C}}(s)) ds}{\int_{\mathbb{D}(s)} \tilde{y}_{\mathcal{C}}(s) ds}, \quad (5)$$

that is, a relative measure with respect to the scenario without policy. Precisely, the denominator of equation (5) captures the entire area below the counterfactual outcome path without policy for region \mathcal{T} —i.e. $\tilde{y}_{\mathcal{C}}(s)$ —in $\mathbb{D}(s)$, whereas the numerator of equation (5) is the absolute difference between the actual outcome path subject to policy—i.e. $y_{\mathcal{T}}(s)$ —and the counterfactual path without policy which captures the area between the path with policy $y_{\mathcal{T}}(s)$ (dashed red) and the counterfactual path without policy $\tilde{y}_{\mathcal{C}}(s)$ (cross-dashed blue) inside the identification window; see panel (b) of Figure 3, which we zoom in panel (c) of the same Figure.

For example, consider the assessment of the effects of a stay-home policy on the flow of deaths in a pandemic in Figure 3. Then, the denominator of equation (5) measures the counterfactual number of deaths that would have occurred in region \mathcal{T} had the policy not been in place, and the numerator of equation (5) measures the number of deaths prevented by the policy, i.e. the difference between the actual number of deaths that occur with policy in region \mathcal{T} and the counterfactual number of deaths that would have occurred without policy in that region. Therefore, if the counterfactual number of deaths without policy is one hundred—i.e. $\int_{\mathbb{D}(s)} \tilde{y}_{\mathcal{C}}(s) ds = 100$ —and the policy prevents twenty deaths—i.e. $\int_{\mathbb{D}(s)} (y_{\mathcal{T}}(s) - \tilde{y}_{\mathcal{C}}(s)) ds = -20$, then the policy effects are $\gamma = -\frac{20}{100} = -0.2$. That is, the effect of policy is a reduction of twenty per cent in the total number of deaths that would have otherwise occurred in the absence of policy. We show the policy effects γ together with the interim cumulative effects, $\gamma(s)$, in panel (d) of Figure 3. Precisely, $\forall s \in \mathbb{D}(s)$, we define $\gamma(s) = \frac{\int_{\underline{s}}^s (y_{\mathcal{T}}(\mathfrak{s}) - \tilde{y}_{\mathcal{C}}(\mathfrak{s})) d\mathfrak{s}}{\int_{\underline{s}}^s \tilde{y}_{\mathcal{C}}(\mathfrak{s}) d\mathfrak{s}}$ where $\mathfrak{s} \in [\underline{s}, s]$ with $\underline{s} = \min_{s \in \mathbb{D}(s)} s$. Further,

¹⁶The opposite occurs if $t_p > s_{\mathcal{C}}(t_p; \psi_{-0}^*)$.

Figure 5: Irrelevance of the Reference Region



Notes: In panel (b), we report the policy effects γ together with the interim cumulative effects of policy, $\gamma(s)$, as defined in Section 2.2.

note that if $s = \max_{s \in \mathbb{D}(s)} s$, then $\gamma(s) = \gamma$.

The (ir)relevance of the reference region So far, we used a specific region, \mathcal{T} , as reference. Would the use of the alternative region \mathcal{C} as reference change the policy effects γ ? The answer is no.

Theorem 2. *The policy effects are invariant to the choice of the reference region. More generally, for any set of coefficients $\phi = \{\phi_k\}_{k=0}^n$,*

$$\gamma = \frac{\int_{\mathbb{D}(s)} (y_T(s) - \tilde{y}_C(s)) ds}{\int_{\mathbb{D}(s)} \tilde{y}_C(s) ds} = \frac{\int_{\mathbb{D}(g(s; \psi_{-0}))} (\phi_0 y_T(g(s; \psi_{-0})) - \phi_0 \tilde{y}_C(g(s; \psi_{-0}))) dg(s; \psi_{-0})}{\int_{\mathbb{D}(g(s; \psi_{-0}))} \phi_0 \tilde{y}_C(s) dg(s; \psi_{-0})}, \quad (6)$$

with $g(s; \phi_{-0}) = \sum_{k=1}^n \phi_k B_k(s)$ and basis functions $B_k(s) = s^{k-1}$.

Proof. First, note that the scaling factor ϕ_0 cancels out from the numerator and denominator in (6). Second, note that for any $s \in \mathcal{S}$, the stage-time transformation through $g(s; \phi_{-0}) = \sum_{k=1}^n \phi_k B_k(s)$ and basis functions $B_k(s) = s^{k-1}$ does not modify the image of $y_r(s)$, that is, $y_r(s) = y_r(g(s; \phi_{-0}))$. Therefore, since $\mathbb{D}(s) \subset \mathcal{S}$, the equivalence result in (6) holds.

This theorem states that the mapping from \mathcal{C} to \mathcal{T} delivers the same policy effects γ as the mapping from \mathcal{T} to \mathcal{C} . To see this we need to undo the first normalization resulting from the

minimization of (2) subject to (1) and (3)—i.e. ψ^* . That is, given ψ^* , we must solve for ϕ_{-0}^* in $s_C^{-1}(t; \psi^*) = g(s; \phi_{-0}^*)$ with $\phi_0^* = (\psi_0^*)^{-1}$. In this manner, the resulting coefficients $\{\phi^*\}$ map the normalized path for region C to its original path and also normalize the path of region T mapping it onto the original path of region C before t_p ; see panel (a) in Figure 5. We keep the same mapping post-policy using our identification assumption. Then, the corresponding identification window is $\mathbb{D}(g(s, \phi_{-0}^*))$ (purple shaded area). Note that size of the identification window changes—since $\phi_2^* < 0$ the window contracts, but the relative size of the numerator and denominator in (6) is the same as per our theorem. That is, even though $|t_p - s_C^{-1}(t_p; \psi^*)| \neq |t_p - s_C(t_p; \psi^*)|$, the policy effects are identical independently of the reference region, i.e. $\gamma = \gamma_{C \rightarrow T} = \gamma_{T \rightarrow C}$; see panel (b) in Figure 5.¹⁷ Finally, note the absolute effects captured by the numerator in (6) can differ across mappings by the scaling factor—if that factor is different from one. For example, if the outcome of interest is the flow of deaths and the policy under evaluation is a stay-home policy, the numerator in (6) in the mapping from C to T captures the amounts of deaths prevented by the policy in region T at the time (and stage) of policy implementation and the numerator in (6) in the mapping from T to C captures the amounts of deaths prevented in region C had the policy been implemented at an earlier stage $s_C^{-1}(t_p; \psi^*)$ in region C .

2.3 An Analytical Example with Exact Identification

Here, we provide an interpretation of our normalization procedure. Precisely, using an analytical example, we show that our normalization—which gives rise to our identification strategy—reshapes the structural parameters determining the outcome path of the non-reference region into the structural parameters that determine the pre-policy outcome path of the reference region. In this manner, our normalization implies that the normalized outcome path of the non-reference region inherits the pre-policy determinants of the outcome path of the reference region.

Consider a scenario in which the process generating the outcome paths and the effects of policy can be explicitly parameterized. In particular, let the path of a region r be determined by

$$y_r(t) = (1 - \gamma \mathbf{1}_{t > t_p}) g(t; \Theta_r), \quad (7)$$

where γ captures the effects of a nationwide policy implemented at t_p for regions $r = \{C, T\}$ ¹⁸

¹⁷This irrelevance implies that we could use an aggregate—or cross-regional average path—as reference.

¹⁸The assumption of constant policy effects is innocuous. More generally, we can let the the effects of policy to be region- and time-specific without consequences for our insights.

and, absent policy, the outcome path of region r is generated by $g(t; \Theta_r) = \frac{\partial G(t; \Theta_r)}{\partial t}$ with

$$G(t; \Theta_r) = \frac{\theta_{0,r}}{1 + \exp(-\theta_{2,r}(t - \theta_{1,r}))}, \quad (8)$$

where $\Theta_r = \{\theta_{0,r}, \theta_{1,r}, \theta_{2,r}\}$ is the set of region-specific structural parameters that determine the behavior of the outcome path in region r . In other words, the pre-policy outcome paths are determined by $g(t; \Theta_r)$, whereas the outcome paths that emerge after policy is implemented are determined by $(1 - \gamma_t)g(t; \Theta_r)$. Further, note that $G(t; \Theta_r)$ is the logistic function which captures the cumulative effects of $g(t; \Theta_r)$. We find this functional shape useful for the intuitive interpretation of its parameters in the context of our stage-time normalization. First, $\theta_{0,r}$ determines the magnitude of the outcome path for region r . Second, $\theta_{1,r}$ pins down the take-off date of the outcome path in region r . Third, $\theta_{2,r}$ determines the pace—or speed—of the outcome path in region r . In panel (a) of Figure 6, we plot $y_r(t)$ using region-specific structural parameters and effects of a nationwide policy implemented in t_p assuming that the outcome path in region \mathcal{C} (blue) starts earlier with $\theta_{1,C} < \theta_{1,T}$, grows faster with $\theta_{2,C} > \theta_{2,T}$ and shows a larger magnitude $\theta_{0,C} > \theta_{0,T}$ than the outcome path in region \mathcal{T} (red).

We now apply our normalization to identify the effects of nationwide policy in this example. Picking \mathcal{T} as reference region, we minimize (2) subject to (1) and (3) using the assumed parametric shapes in (7) and (8). The resulting first order conditions imply:

$$\psi_0^* \theta_{0,C} = \theta_{0,T} \quad (9)$$

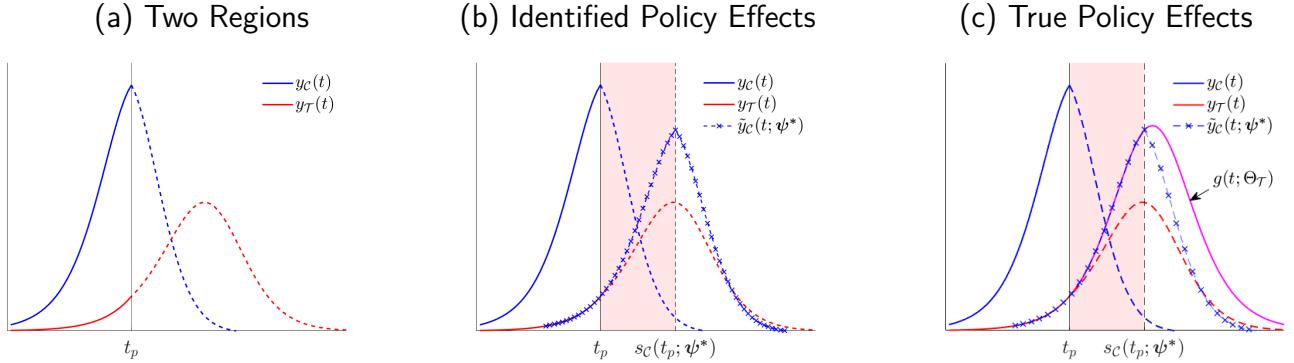
$$\psi_1^* - \theta_{1,C} = -\theta_{1,T} \quad (10)$$

$$\psi_2^* \theta_{2,C} = \theta_{2,T}, \quad (11)$$

for $s \in \mathbb{C}(s)$.¹⁹ This is a system of three equations that characterizes the solution of the three unknown coefficients $\psi^* = \{\psi_0^*, \psi_1^*, \psi_2^*\}$. Several results emerge. First, the magnitude of the normalized outcome path of the non-reference region \mathcal{C} is identical to the magnitude of the reference path, i.e. $\tilde{\theta}_{0,C} = \psi_0^* \theta_{0,C} = \theta_{0,T}$. Second, the normalized starting date for the normalized outcome path of the nonreference region is identical to that of the reference region, i.e. $\tilde{\theta}_{1,C} = \theta_{1,C} - \psi_1^* = \theta_{1,T}$. Third, the speed of the normalized outcome path of the non-reference region \mathcal{C} is identical to the magnitude of the reference path, i.e. $\tilde{\theta}_{2,C} = \psi_2^* \theta_{2,C} = \theta_{2,T}$. That is, our normalization coefficients $\{\psi^*\}$ reshape the structural parameters in the non-reference region

¹⁹Note that since it is clear that region \mathcal{C} is at a more advance stage than region \mathcal{T} at the time of policy implementation, $\mathbb{C}_T(s) = [\theta_{1,T}, t_p]$ for region \mathcal{T} and $\mathbb{C}_C(s) = [\psi_1 + \psi_2 \theta_{1,C}, \psi_1 + \psi_2 t_p]$ for region \mathcal{C} .

Figure 6: Our Identification Strategy Reshapes Structural Parameters



Notes: The outcome paths are determined by (7) and the generalized logistic function (8) with parameters $\Theta_C = \{4.00, 0.15, 35.00\}$ and $\Theta_T = \{3.50, 0.14, 55.00\}$ respectively. We assume that $\gamma_t = \alpha_0(e^{-\alpha_1(t-t_p)} - 1)$ with $\alpha_0 = -2$ and $\alpha_1 = 0.01$ which implies the policy effect is $\gamma = -0.238$, i.e. a 23.8% reduction in the number of deaths that would have occurred otherwise without policy.

\mathcal{C} so as to be exactly identical to the structural parameters in the reference region \mathcal{T} , that is,

$$\tilde{\Theta}_C = \{\tilde{\theta}_{0,C}, \tilde{\theta}_{1,C}, \tilde{\theta}_{2,C}\} = \{\theta_{0,T}, \theta_{1,T}, \theta_{2,T}\} = \Theta_T,$$

for all $s \in \mathbb{C}(s)$.

Two implications emerge from the normalization procedure. First, the normalized outcome path of the non-reference region, $\tilde{y}_C(s) = \psi_0^* y_C(\psi_1^* + \psi_2^* s)$, is exactly identical to the outcome path of the reference region, $y_T(s)$,

$$\tilde{y}_C(s) = g(t; \tilde{\Theta}_C) = g(t; \Theta_T) = y_T(s), \quad (12)$$

for all stages before policy implementation, i.e. $s \in \mathbb{C}(s)$. We show this graphically in panel (b) of Figure 6 where the normalized outcome path of the non-reference region $\tilde{y}_C(s)$ (cross-dashed blue) exactly overlaps with the path of the reference region $y_T(s)$ (solid red) before policy. Second, the normalization uncovers stage heterogeneity in the time of policy implementation. Precisely, at the time of policy implementation the non-reference region is at a more advanced stage than the reference region, i.e. $t_p = s_C(t < t_p; \psi^*)$.

In this context, it is worth noting that applying our normalization $\{\psi^*\}$ —obtained using only pre-policy data when solving (9)–(11)—to the post-policy data, the normalized outcome path of the non-reference region \mathcal{C} is also identical to the actual outcome path of reference region \mathcal{T} after policy. That is, in our example with exact identification, equation (12) holds post policy, i.e. for all $s \notin \mathbb{C}(s)$, and, in particular, for an identification window that emerges between the stage in

which the reference region \mathcal{T} enters the policy and the stage in which the non-reference region \mathcal{C} enters policy, i.e. $\mathbb{D}(s) = [t_p, s_{\mathcal{C}}(t_p; \psi^*)]$. In that identification window, in which the outcome of the reference region $y_{\mathcal{T}}(s)$ is subject to policy and the normalized outcome of the non-reference region $\tilde{y}_{\mathcal{C}}(t)$ is not, the normalized path $\tilde{y}_{\mathcal{C}}(t)$ is exactly identical to the outcome path of the region \mathcal{T} absent policy, i.e. $g(t; \Theta_{\mathcal{T}})$; see panel (c) of Figure 6. Therefore, the identified policy effects that are defined using the difference between $\tilde{y}_{\mathcal{C}}(s)$ and $y_{\mathcal{T}}(s)$ are identical to the true policy effects that are captured by the difference between $g(t; \Theta_{\mathcal{T}})$ and $y_{\mathcal{T}}(s)$. In other words, the normalized outcome path of region \mathcal{C} , i.e. $\tilde{y}_{\mathcal{C}}(s)$, provides a useful no-policy counterfactual for the outcome path of reference region \mathcal{T} . In this manner, the fact that $\tilde{y}_{\mathcal{C}}(s) = g(t; \Theta_{\mathcal{T}})$ in $\mathbb{D}(s)$ justifies our identification assumption A.1 in Section 2.2

More generally, we are interested in policy contexts where the structural parameters that determine behavior are potentially unknown—and a larger set—than the set of normalizing coefficients ψ . In Section 3, we assess our method performance in such contexts with non-exact identification.

3 Method Performance

Here, we assess whether our stage-based identification method is able to recover the true policy effects that emerge from theoretical models. First, we implement our identification strategy on model-generated data and compare our stage-based identified effects with the true effects in Section 3.1. Second, we conduct a Monte Carlo analysis—with a wider spectrum of outcome paths—in order to provide bounds to the performance of our method in Section 3.2. Third, we assess how our method fares in the presence of confounding factors—e.g. time-varying heterogeneity and confounding policy in Section 3.3. Fourth, we conduct a placebo diagnosis in Section 3.4.1. Finally, we discuss the presence of measurement error and conduct inference in Section 3.4.2.

3.1 Can Our Method Recover the True Policy Effects?

To address this question, we use three alternative policy contexts that include: a public health policy against a pandemic using a model where economic activity in the form of hours worked shapes and is shaped by a pandemic; the effects of the approval of the pill in a model of women career and fertility choices; and an economic growth policy using a model of structural transformation.

3.1.1 Public Health Policy Against a Pandemic

Consider an economy with many individuals that is unexpectedly hit by an epidemic at time $t_o = 1$ with an initial number of infections $I_1 > 0$. We normalize the pre-pandemic population, N_0 , to one. A benevolent social planner solves:

$$\max_{\{c_t \geq 0, h_t \in [0,1]\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \delta^t \Pi_{\tau=0}^t \phi_{\mathcal{P}}(h_{\tau-1}) u(c_t, h_t; \omega) \quad (13)$$

where the felicity function is strictly concave in both arguments consumption c_t and leisure $1 - h_t \in [0, 1]$. The parameter ω measures the individual value of life. We assume the period utility takes the form $u(c, h; \omega) = \log(c) - \kappa \frac{h^{1+\frac{1}{\nu}}}{1+\frac{1}{\nu}} + \omega$. Our planner is subject to an aggregate resource constraint $N_t c_t = w_t h_t N_t$ where w_t is the implicit price (marginal product) of labor using technology $Y_t = z h_t N_t$, i.e. $w_t = z$. The timing of the model is such that individuals work and consume and, when working, there is infection risk. After working and consuming, the infected face death or recovery. In this manner, individuals are either susceptible S_t , infected I_t , recovered R_t or dead D_t . The total population alive is $N_t = S_t + I_t + R_t$.²⁰ Using $X_{G,t} = G_{t+1} - G_t$ for $G = \{S, I, R, D\}$, the planner's belief is that law of motion of the population structure is:²¹

$$X_{S,t} = -\lambda_{\mathcal{P}}(h_t) \beta \frac{I_t}{N_t} S_t \quad (14)$$

$$X_{I,t} = (1 - \gamma) \lambda_{\mathcal{P}}(h_t) \beta \frac{I_t}{N_t} S_t - \gamma I_t \quad (15)$$

$$X_{R,t} = (1 - \zeta) \gamma \tilde{I}_t \quad (16)$$

$$X_{D,t} = \zeta \gamma \tilde{I}_t \quad (17)$$

where $\tilde{I}_t = \lambda_{\mathcal{P}}(h_t) \beta \frac{I_t}{N_t} S_t + I_t$.^{22,23} The parameter β captures features like density, occupation-industry composition, age-health structure of the population or pollution (among others) which can differ across locations. Then, conditional on randomly meeting an infected individual at rate $\frac{I_t}{N_t}$, the planner's belief is that individuals get infected with probability $\lambda_{\mathcal{P}}(h_t) = \xi_{\mathcal{P}} h_t^{\alpha}$ which depends on the choice of h_t in a strictly concave fashion with $\alpha \in (0, 1)$. The parameter $\xi_{\mathcal{P}}$ captures the planner's beliefs on how much economic activity affects the probability of infection. However, the actual infections arise from the true probability function $\lambda(h_t) = \xi h_t^{\alpha}$ where ξ can

²⁰There is no fertility in this economy. That is, the evolution of the population is solely determined by survival.

²¹Without beliefs, the population law of motion is essentially that in, for example, Atkeson (2020).

²²We allow for new infections to transit to death in the same period t . In this manner, h_t has an immediate effect on the survival rate between t and $t+1$. This assumption is innocuous and we use it to ease the exposition of the trade-off between economic activity and public health. We can easily accommodate lagged effects.

²³In the Appendix, we show how to derive (14)-(17) from a population matrix projection model where infections occur before death and recovery.

differ from the planner's beliefs. If $\xi_{\mathcal{P}} < \xi$, then the planner underestimates the actual effects of h_t on infections and deaths. In contrast, if $\xi_{\mathcal{P}} > \xi$, then the planner overestimates these effects.

Then, given N_t , S_t and I_t and the population law of motion (14)-(17), the planner's belief is that the survival rate between t and $t + 1$ is $\phi_{\mathcal{P}}(h_t) = 1 - \frac{X_{D,t}}{N_t}$, which can differ from the actual survival rate, $\phi(h_t)$. Note that since the planner chooses h_t based on her beliefs on the infection process, she is not able to forecast the true population dynamics.²⁴ In this context, we introduce an unexpected population shock that fully corrects for the forecast error *ex-post*. Precisely, whereas at period t the planner's belief is that population between t and $t + 1$ changes following the survival probability function $\phi_{\mathcal{P}}(h_t)$ —and chooses h_t accordingly, at period $t + 1$ we allow the planner to learn the actual survival rate without being aware that $\phi(h_t)$ determines it. That is, we define the unexpected shock at $t + 1$ —or equivalently the one-period ahead forecast error at t —as the difference $\phi(h_t) - \phi_{\mathcal{P}}(h_t)$, which occurs at every period t . This setting implies that looking backwards the planner observes the actual evolution of the population $N_t = \prod_{\tau=0}^t \phi(h_{\tau-1})$, whereas looking forward the planner always commits a forecast error.²⁵ Analogously to N_t , I_t and S_t are also updated by the actual infections with an unexpected shock equal to $\lambda(h_t) - \lambda_{\mathcal{P}}(h_t)$.

The amount of economic activity h_t is determined by the following Euler condition

$$\underbrace{\frac{\partial u(c_t, h_t; \omega)}{\partial c_t}_w}_{\text{Marginal Benefit of Working: Consumption Gain}} - \underbrace{\frac{\partial u(c_t, h_t; \omega)}{\partial h_t}}_{\text{Marginal Cost of Working: Loss of Leisure}} = \underbrace{\delta \frac{\partial \phi_{\mathcal{P}}(h_t)}{\partial h_t} u(c_{t+1}, h_{t+1}; \omega)}_{\text{Marginal Cost of Working: Loss of Lives}} \quad \forall t, \quad (18)$$

which states that the marginal benefit of working (more consumption) needs to equate its marginal costs consisting of an intratemporal component (disutility from working) and an intertemporal component (loss of lives). Note the Euler equation is a first order difference equation in h_t . We need a terminal condition to solve for the path of h_t . We separately solve for the pre-pandemic equilibrium at $t = 0$ (actually, for any $t \leq 0$) before the unexpected arrival of the pandemic at $t = 1$. In this pre-pandemic era there are no infections and, hence, $\phi(h_0) = 1$. That is, the equilibrium labor supply sets the right hand side of the Euler equation (18) to zero in which case h_0 simply solves an intratemporal trade-off. The same equilibrium emerges after the pandemic at some large $t = T$ which delivers a terminal condition $h_T = h_0$. Given h_T (or h_0), we can easily solve for the optimal labor path $\{h_t\}_{t=1}^{T-1}$ during the epidemic using standard techniques. Last, since there is an unexpected population shock every period t , the planner reoptimizes the

²⁴Precisely, given N_t , S_t and I_t , the actual survival rate between t and $t + 1$ is the result of plugging the h_t that the planner's chooses—under her beliefs on the infection process $\lambda_{\mathcal{P}}(h_t)$ —into the population law of motion (14)-(17) after replacing the planner's belief by the true probability of infection $\lambda(h_t)$.

²⁵See [Aleman et al. \(2022\)](#) for a framework with learning where mistakes in the forecast error of infections decrease over time in the context of the HIV epidemic.

entire sequence $\{h_t\}_{t=1}^{T-1}$ every period after the realization of the shock; see our Appendix for our complete solution algorithm.

In what follows, we assess whether our stage-based identification of policy effects (as described in Section 2) is able to recover the true effects of policy.

True (model-generated) policy effects We focus on two modeled regions that differ in the underlying parameter values that govern model behavior. Let us denote the set of model parameters as $\Theta = \{\delta, \omega, z, \beta, \gamma, \zeta, \kappa, \nu, \xi_{\mathcal{P}}, \alpha, I_1\}$. We show the equilibrium response of hours without policy intervention for region \mathcal{C} (solid blue) and region \mathcal{T} (solid red) in panel (a) of Figure 7. In particular, we assume that region \mathcal{C} underestimates the effect that economic activity has on infections (i.e. $\xi_{\mathcal{P}}$) by less than region \mathcal{T} which reduces economic activity (i.e. h) earlier and also by a larger amount in region \mathcal{C} than region \mathcal{T} . This earlier and stronger endogenous response in terms of hours of region \mathcal{C} affects out outcome of interest, i.e. the epidemic path of deaths, by reducing the peak of deaths and flattening the curve in region \mathcal{C} relative to region \mathcal{T} ; see panel (b) in Figure 7. We also assume that region \mathcal{C} has higher odds of encountering infected individuals at work (i.e. higher β) which advances and increases the peak of deaths for region \mathcal{C} relative to region \mathcal{T} . Further, we assume that region \mathcal{C} has a lower disutility of work κ which implies a larger pre- and post-pandemic level of hours worked for region \mathcal{C} than region \mathcal{T} .²⁶

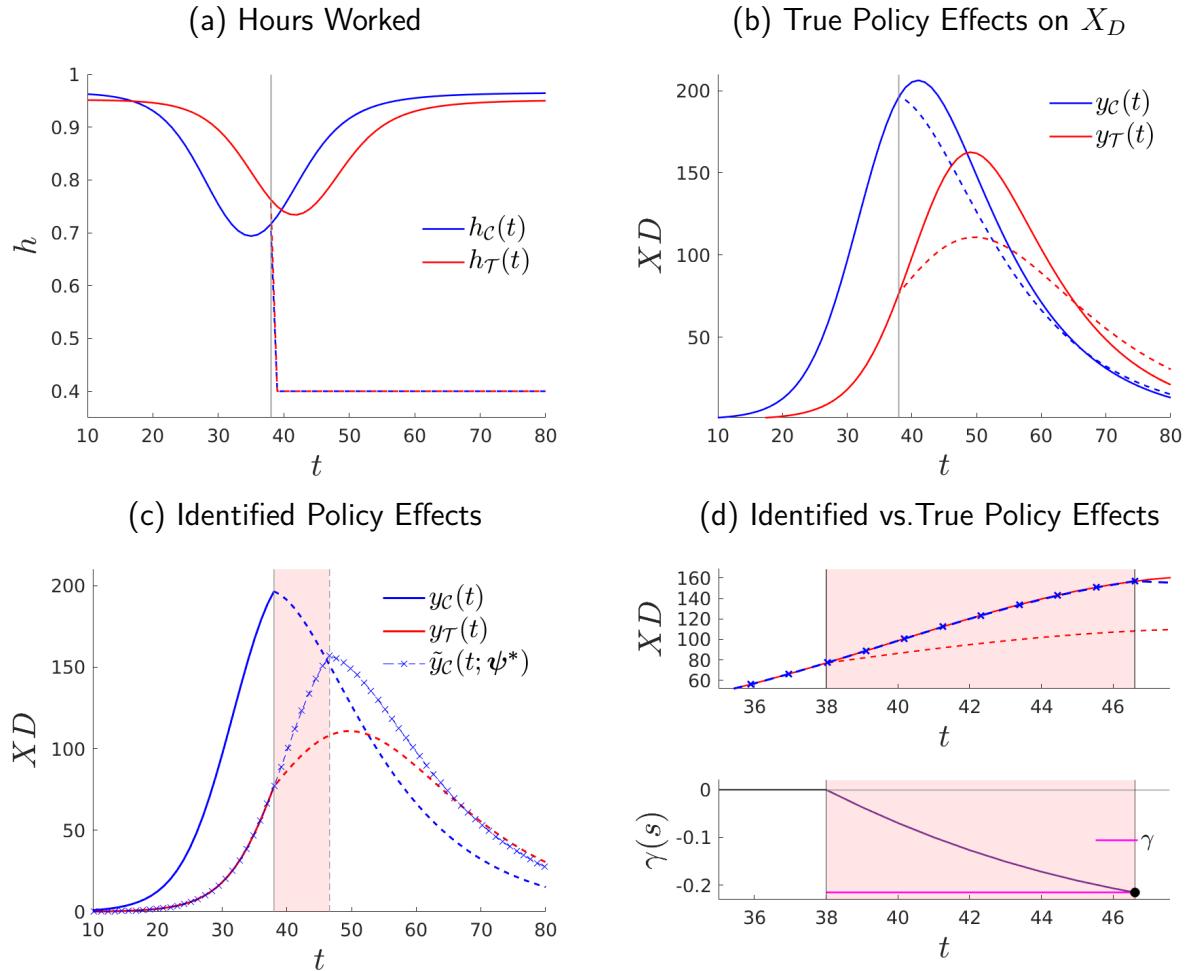
In this context, we introduce a nationwide public health policy²⁷ that imposes an upper bound on hours worked, $h < \bar{h} = 0.5$, from (day) $t_p = 38$ to $t_f = 250$. Since, without policy, households in both regions would work more hours than those imposed by the policy, the policy is binding in both \mathcal{C} and \mathcal{T} —see, respectively, the blue and red dashed lines that emerge after $t_p = 38$ in panel (a) of Figure 7. In this manner, the imposed policy resembles a stay-home policy or national lockdown. The lower economic activity imposed by the policy has consequences for the flow of deaths. With policy, the flow of deaths peaks earlier and by a lower magnitude in both \mathcal{C} and \mathcal{T} —see, respectively, the blue and red dashed lines that emerge after $t_p = 38$ in panel (b) of Figure 7. The difference between the flow of deaths with policy (dashed lines) and the flow of deaths without policy (solid lines) after t_p captures the true effects of policy as generated from the model which is easily computed.

Unfortunately, the counterfactual paths of the flow of deaths without policy after policy implementation (i.e. the solid lines after t_p) are not available outside of the model. That is, from the perspective of an evaluator that wants to assess the policy effects, the data available

²⁶The notes of Figure 7 provides a full account of the assumed differences in model parameters across regions. We further conduct *ceteris-paribus* comparative statics for each model parameter in the Appendix.

²⁷From now on we refer to any world with no policy with the subscript “np”.

Figure 7: Stage-Based Identification of Model-Generated Policy Effects: A Nationwide Public Health Policy Against a Pandemic



Notes: We assume that some parameters defer across regions: $\Theta_C = \{\beta = 0.509, \zeta = 0.0010, \kappa = 1.05, \xi = 0.20, I_0 = 1\}$ and $\Theta_T = \{\beta = 0.501, \zeta = 0.0008, \kappa = 1.07, \xi = 0.19, I_0 = 6\}$. The rest of the model parameters are identical across regions, $\{\delta = 0.95, \omega = 560400, z = 64, \beta = 0.501, \alpha = 0.65\}$. Further, the parameters associated to the policy are $\bar{h} = 0.4, t_p = 38, t_f = 250$.

for the policy evaluation consists of the path without policy (solid lines) for all periods before t_p pieced together with the path with policy (dashed lines) for all periods after t_p . We now apply our stage-based identification strategy on these same data that would be available to the policy evaluator.

Stage-based identified policy effects The policy effects identified in this manner are in panel (c) of Figure 7. In particular, we map the flow of deaths in region \mathcal{C} (solid blue line) onto those of region \mathcal{T} (solid red line) using *only* pre-policy data following the normalization described in Section 2. The result of our identification is $\tilde{y}_{\mathcal{C}}(t; \psi^*)$ (blue line with cross markers) which provides our candidate counterfactual without policy for \mathcal{T} in the identification window $\mathbb{D}(s) = [t_p, s_{\mathcal{C}}(t_p; \psi^*)]$ (shaded pink area). In this context, do our identified policy effects recover the true policy effects generated by the model? To answer this question we compare our candidate counterfactual $\tilde{y}_{\mathcal{C}}(t; \psi^*)$ (blue line with cross markers in panel (c) of Figure 7) with the true counterfactual (red line in panel (b) of Figure 7) in the identification window. We zoom in this comparison in panel (d) of Figure 7. Our main result is that the identified policy effects are not significantly different from the true effects. Precisely, the identified total number of lives saved is $\int_{\mathbb{D}(s)} (\tilde{y}_{\mathcal{C}}(s) - y_{\mathcal{T}}(s)) ds = 248.54$ in a window of $s_{\mathcal{C}}(t_p; \psi^*) - t_p = 8.60$ days, whereas the true policy effects are 250.72 lives saved. Therefore, the policy prevented $\gamma = -21.50\%$ of the total deaths that would have occurred had the policy not been implemented, whereas the true effect is $\gamma_{\text{true}} = -21.64\%$. This implies that the identified policy effects catch the true policy effects with an error of $\varepsilon(\gamma) = \left| \left(\frac{\gamma}{\gamma_{\text{true}}} - 1 \right) \times 100 \right| = 0.68\%$.

3.1.2 The Pill and Women's Choice

We now pose a framework in which the introduction of oral contraceptives (the pill) has effects on women's human capital, sexual and fertility choices. The pill provides access to a technology that reduces unwanted pregnancies at the time (age) where human capital decisions are taken. We assume that each cohort- t of women derives utility from consumption, children and sexual intercourse x by choosing the amount of human capital, h , sex, x , and pill usage, o , that solves:

$$\max_{\{h, o, x\}} c + \kappa n + \zeta x - \iota o \quad (19)$$

where $c \geq 0$ is consumption, $n \geq 0$ are children that generate joy by a relative factor κ , sexual intercourse provides utility with relative factor ζ and taking the pill carries a utility cost ι such as

a social norm.²⁸ Women are subject to a resource constraint,

$$c + qh = (1 - \tau(n))w(1 + z_t e(h)) \quad (20)$$

where q is the relative price of human capital (e.g. tuition fees or job training), $\tau(n) \in [0, 1]$ captures an additional cost of accumulating human capital associated with the presence of children with $\tau_n(n) > 0$ and $\tau_{nn}(n) < 0$. In terms of earnings, w is a constant base wage and $z_t e(h)$ is an endogenous human capital wage premium with two components. First, there is skill-biased technical change (SBTC), $z_t = z_0 \prod_{\tau=1}^{\tau=t} (1 + \gamma_\tau)$, where $z_0 > 0$ and $\gamma_t > 0$ captures cohort- t growth in SBTC. Second, there is a mapping from human capital h to a rate $e(h) \in [0, 1]$ with $e_h(h) > 0$ and $e_{hh}(h) < 0$. We interpret $e(h)$ as the fraction of educated (e.g. college completed) women in the economy that benefit from the SBTC.²⁹ Finally, children production is given by,

$$n = \phi(x)[1 - \mathbf{1}_{t_p} g(o)], \quad (21)$$

where access to the pill is captured by the policy dummy $\mathbf{1}_{t_p}$ that is equal to zero if a cohort t does not have access to the pill, and equal to one otherwise. Hence, if women do not have access to the pill, then the amount of children is solely determined by the probability of pregnancy $\phi(x) \in [0, 1]$ where we assume that all pregnancies end in a child, that is $n = \phi(x)$. If women have access to the pill, then the probability of pregnancy is adjusted downward by the pill effectiveness in preventing pregnancy, $g(o) \in [0, 1]$. We assume that larger use of the pill—e.g. higher adherence to follow protocol, increases the effectiveness of the pill. That is, $g_o(o) > 0$ with $g_{oo}(o) < 0$.³⁰

Note that we can plug the resource constraint (20) and child production (21) into women's objective function (19). The first order condition of h implies,

$$\text{FOC}(h) : \underbrace{q}_{\text{Marginal Cost of Human Capital}} = \underbrace{(1 - \tau(n))wz_t e_h(h)}_{\text{Marginal Benefit of Human Capital}}, \quad (22)$$

where the marginal cost of human capital is the price of education qualified by the cost of children, and the marginal benefit of human capital are the premium returns from increases in human capital. Note that the marginal benefit of human capital is convex in h whereas the

²⁸The analysis in Goldin and Katz (2002) encompasses career choice and marriage delays. In their case, they model the increase in a counterpart of our h as the utility lost from abstinence and/or forgone home production.

²⁹The mapping of the outcome variable from h to $e(h)$ is innocuous for our analysis and we use it merely for exposition convenience. In particular, if the outcome $e(h)$ is a rate we can interpret it as the fraction of educated women (e.g. college degree completion) in the population.

³⁰We note that lawful access to the pill does not necessarily fully determine use which is also likely to be affected by social norms Goldin and Katz (2002). Further, access to the pill can—at the same time—shape social norms (Fernández-Villaverde et al., 2014).

marginal cost of human capital is concave in n , hence, an increase in the number of children results in a decrease in human capital. Therefore, a technology that reduces n can enhance human capital.

The first order condition for sexual intercourse x is:

$$FOC(x) : \underbrace{\tau_n(n)\phi_x(x)(1 - \mathbf{1}_{t_p}g(o))w(1 + z_t e(h))}_{\text{Marginal Cost of Intercourse}} = \underbrace{\zeta + \kappa_t\phi_x(x)(1 - \mathbf{1}_{t_p}g(o))}_{\text{Marginal Benefit of Intercourse}}, \quad (23)$$

Where the marginal benefit takes into account the additional utility that agents get from sex, and children. The marginal cost reflects the cost of additional children in terms of human capital. The first order condition for pill use o is:

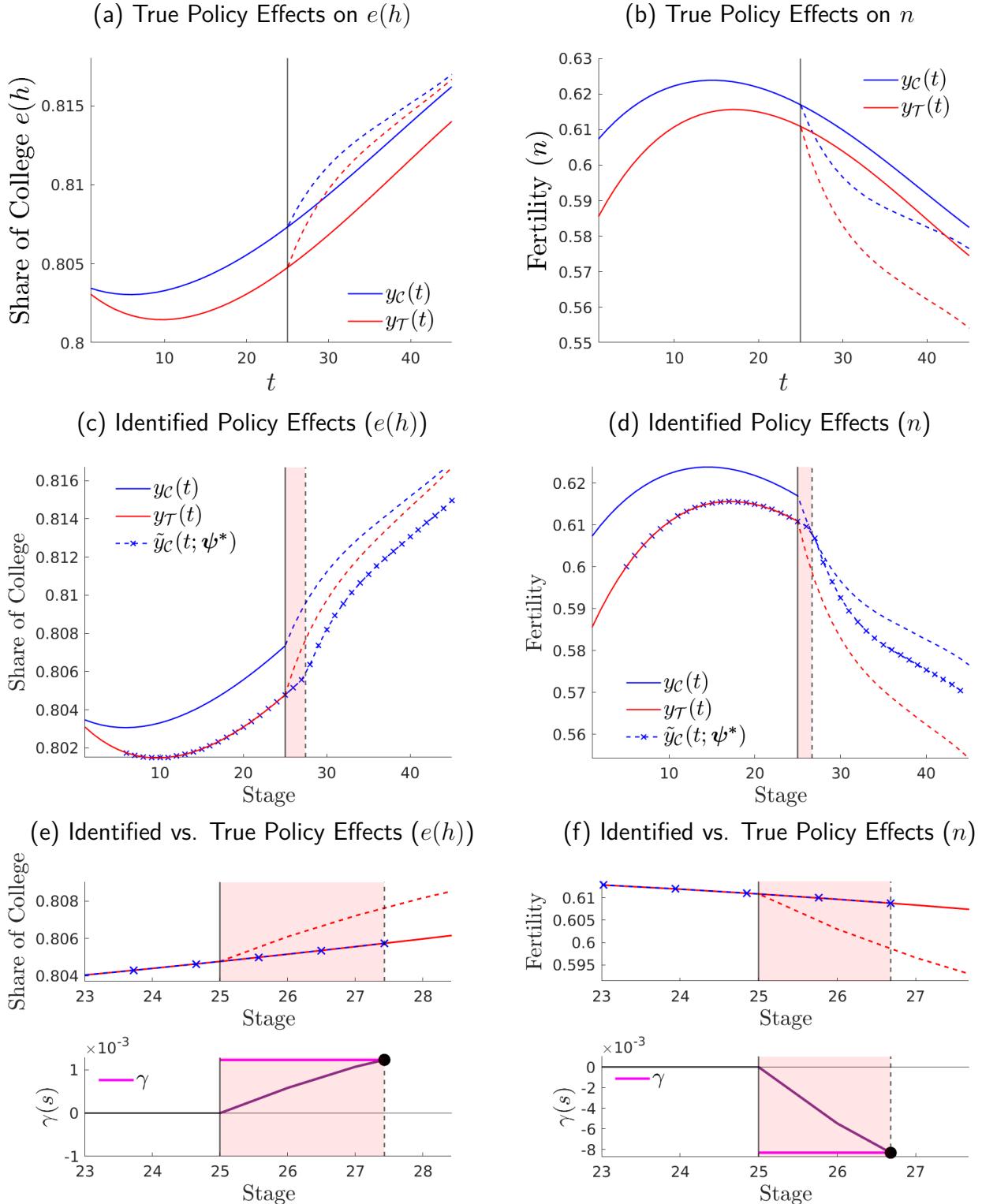
$$FOC(o) : \underbrace{\tau_n(n)\phi(x)\mathbf{1}_{t_p}g(o)w(1 + z_t e(h))}_{\text{Marginal Benefit of Pill}} = \underbrace{\kappa_t\phi(x)\mathbf{1}_{t_p}g_o(o) + \iota}_{\text{Marginal Cost of Pill}} \quad (24)$$

where the marginal cost of the pill is a reduction utility derived from children and the marginal benefit of the pill is a reduction in the price of human capital.

True (model-generated) policy effects In Figure 8, we show the equilibrium path for the women schooling choices in panel (a) and fertility choices in panel (b). We show the model-generated paths in a scenario without the pill (solid lines) and in a scenario in which the government grants women legal access to the pill technology (dashed lines). We do this separately for region \mathcal{C} (blue) and region \mathcal{T} (red). We also report the behavior of pill use and sexual intercourse in the Appendix. Regions differ in the model parameters $\Theta = \{\kappa, \xi, q, w, z, \gamma, \theta_x, \theta_h, \theta_o\}$. In particular, we allow for the returns to human capital to be larger and grow faster in region \mathcal{C} than in region \mathcal{T} which explains the larger levels of human capital in region \mathcal{C} than in region \mathcal{T} . This also explains the lower amount of sexual intercourse and children in region \mathcal{C} than in region \mathcal{T} . Further, we exogenously shape the SBTC parameter γ such that the endogenous human capital path is S-shaped for both regions. We also choose an exogenous path for the relative utility derived from children, κ , in order for endogenous fertility to display a boom and bust.

After the pill technology is legalized, i.e. with $\mathbf{1}_{t_p} = 1$ for all for all cohorts of women $t_p \geq 25$, there is an endogenous increase in the use of the new technology (see Appendix). The pill sustains a higher amount of sex with a lower amount of children and hence higher human capital. That is, the pill reduces births in both regions (see dashed lines panel (a) in Figure 8). By reducing optimal fertility, the pill reduces the cost of acquiring human capital which increasing the share of women that enter college (see dashed lines in panel (b) in Figure 8).

Figure 8: Stage-Based Identification of Model-Generated Policy Effects: Introduction of the Pill



Notes: We plot outcomes for region \mathcal{C} (red) and \mathcal{T} (blue) without policy (solid lines) and with policy (dashed lines). The policy is the introduction of the pill for all periods $t_p \geq 25$. We show the equilibrium path of pill use ($g(o)$) in panel (a) and sexual intercourse (x) in panel (b). The parameter values that we choose for region \mathcal{C} are $\Theta_{\mathcal{C}} = \{\xi = 8, q = 3.2, w = 64, z_0 = 1, \gamma = 0.1\%, 2\bar{g}_x = 0.5, \theta_h = 0.4, \theta_o = 0.43, \iota_{t,\mathcal{C}}, \kappa_{t,\mathcal{C}}\}$. The parameter values for region \mathcal{T} are $\Theta_{\mathcal{T}} = \{\kappa_{t,\mathcal{T}}, \xi = 8, q = 3.2, w = 63, z_0 = 1, \gamma = 0.1\%, \theta_x = 0.5, \theta_h = 0.4, \theta_o = 0.43, \iota_{t,\mathcal{T}}, \kappa_{t,\mathcal{T}}\}$.

Stage-based identified policy effects In a context in which the policy evaluator is not provided with the true counterfactual path without policy (solid lines for the periods after t_p), can our stage-based identification recover the model-generated effects of the pill? We implement our method using region \mathcal{T} as reference. Hence we apply our normalization mapping the outcome path of region \mathcal{T} (solid blue line) onto that of region \mathcal{C} (solid red line) using only pre-policy data as described in Section 2. Our identification results in a candidate counterfactual $\tilde{y}_{\mathcal{C}}(t; \psi^*)$ (blue line with cross markers) for a window that goes from t_p to $s_{\mathcal{C}}(t_p; \psi^*)$ (shaded pink area); see panel (c) and (d) of Figure 8 for, respectively, human capital and children. We provide a zoomed comparison of our identified counterfactual $\tilde{y}_{\mathcal{C}}(t; \psi^*)$ and the true effects of policy in panel (e) and (f) of Figure 8 for, respectively, human capital and children. The main result is that the identified policy effects are not significantly different from the true effects. The stage-based identification finds that the pill increases the proportion of women going to college $e(h)$ by $\gamma = 0.12\%$ for the window between t and $s_{\mathcal{C}}(t_p; \psi^*)$, a figure that is identical to the true policy effects with $\gamma_{\text{true}} = 0.12\%$. Our identification also implies that the pill reduces the overall amount of children by $\gamma = 0.83\%$ in the identification window which is also identical to the true policy effects with $\gamma_{\text{true}} = 0.83\%$. This implies that the identified policy effects catch the true policy effects with an error $\varepsilon(\gamma)$ of 0.18% for human capital and of 0.23% for fertility.

3.1.3 Growth policy and structural transformation

Consider an economy with structural transformation from an agricultural to a manufacturing sector, respectively, $i = \{a, m\}$. An infinitely-lived representative agent chooses sectoral allocations of consumption $\{c_{at}, c_{mt}\}_{t=0}^{\infty}$, labor $\{n_{at}, n_{mt}\}_{t=0}^{\infty}$ and next period capital $\{k_{t+1}\}_{t=0}^{\infty}$:

$$\max_{\{c_{at}, c_{mt}, n_{at}, n_{mt}, k_{t+1}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t (u(c_{at} - \bar{c}_a) + \kappa v(c_{mt}))$$

subject to a budget constraint,

$$p_{at}c_{at} + c_{mt} + k_{t+1} = \sum_{i \in \{a, m\}} w_{it}n_{it} + r_t k_t + (1 - \delta)k_t + \pi_t(\ell)$$

where $\beta \in (0, 1)$ is the discount factor, κ is the utility weight of manufacturing consumption relative to agriculture and \bar{c}_a a subsistence level. The non-homothetic preferences are a force behind the structural transformation of the economy. The household is endowed with one unit of time, i.e. $n_{at} + n_{mt} = 1 \forall t$ that is allocated to either agriculture or manufacturing and receive, respectively, wage rates $\{w_{at}, w_{mt}\}$. The capital's return is r_t and capital depreciates at rate δ . The manufacturing good is the numeraire and p_a is the price of agriculture relative

to manufacturing, exogenously given. Land ℓ is fixed and inelastically supplied by the household that receives pure rents from renting it to agricultural firms $\pi(\ell)$. We assume that the felicity functions satisfy strict concavity.

There is one representative firm per sector and we assume competitive markets. The agricultural firm produces output y_a with labor n_a and land. Agricultural firms solve:

$$\max_{n_{at}} \pi_t(l) = (1 - \tau)p_{at}y_{at} - w_{at}n_{at} \quad \text{s.t.} \quad y_{at} = z_{at}n_{at}^\phi \ell^{1-\phi},$$

where ϕ is the labor share in agriculture. Since land is fixed, the agricultural technology exhibits decreasing returns to scale which is an additional force shifting resources out of agriculture as the economy grows.³¹ We assume that there are inefficient institutions in the agricultural sector that are captured by the parameter τ that taxes agricultural revenue. Manufacturing firms produce output y_{mt} with labor n_{mt} and capital k_t solving:

$$\max_{n_{mt}, k_{t+1}} y_{mt} - w_{mt}n_{mt} - r_t k_t \quad \text{s.t.} \quad y_{mt} = z_{mt}n_{mt}^\alpha k_t^{1-\alpha},$$

where α is the labor share in manufacturing. We assume that total factor productivity (TFP) differs by sector following $z_{it} = z_{i,0}(1 + \gamma_i)^t$ for $i = \{a, m\}$ with $\gamma_a < \gamma_m$. That is, productivity in the manufacturing sector grows at a faster rate than that of the agricultural sector.

There are three first order conditions for the household problem.³² First, an intratemporal condition governing the substitution across consumption goods:

$$FOC(c_{at}) : u_{c_{at}}(c_{at}) \underbrace{\frac{1}{p_{at}}(-1)}_{\frac{\partial c_{at}}{\partial c_{mt}}} + \kappa v_{c_{mt}}(c_{mt}) = 0 \quad (25)$$

Second, an intertemporal Euler condition for k_{t+1} governing the trade off between one additional

³¹The ability of non-homothetic preferences to generate structural change is studied in, for example, [Gollin et al. \(2002\)](#). The role of technological choice in generating structural change—from a decreasing returns to scale technology to a constant returns to scale, is studied in [Hansen and Prescott \(2002\)](#); in the context of a one-good economy. Our identification of policy effects is innocuous as to which force drives structural change.

³²Note that we can isolate c_m from the budget constraint and plug it into the objective function plus use $n_{mt} = 1 - n_{at}$. This implies that we can maximize the objective function in terms of the sequences of three unknowns $\{c_{at}, n_{at}, k_{t+1}\}_{t=0}^\infty$.

unit of consumption today versus tomorrow's consumption,

$$FOC(k_{t+1}) : \underbrace{u_{ca}(c_{at})}_{\frac{\partial c_{at}}{\partial k_{t+1}}} \underbrace{\frac{1}{p_{at}}(-1) + \beta u_{ca}(c_{at+1})}_{\frac{\partial c_{at+1}}{\partial k_{t+1}}} \underbrace{\frac{1}{p_{at+1}}(1 + r_{t+1} - \delta)}_{\frac{\partial c_{at+1}}{\partial k_{t+1}}} = 0, \quad (26)$$

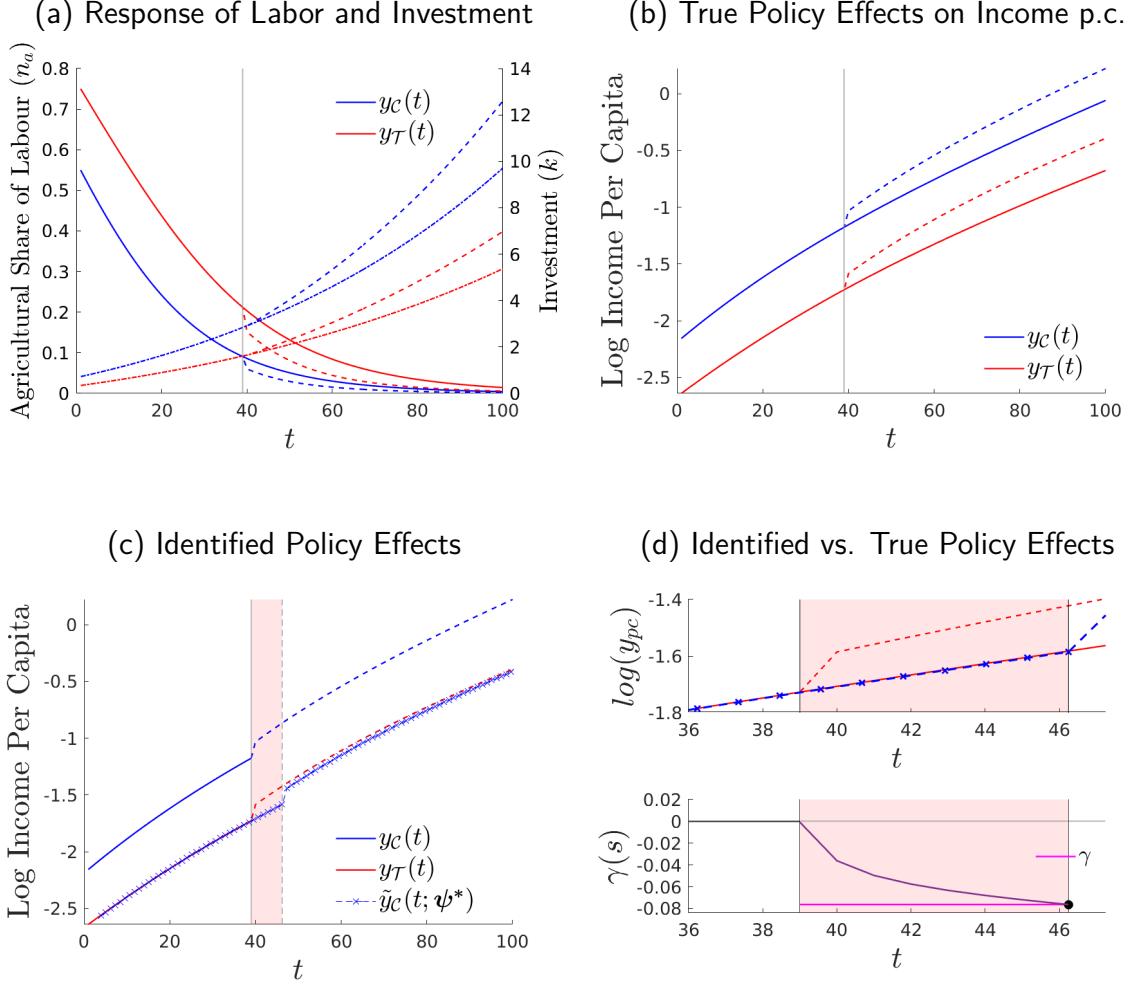
and note that we can rewrite this intertemporal condition in terms of c_m using (25). Third, an intratemporal condition for n_a equates wages across sectors,

$$FOC(n_{at}) : \underbrace{u_{ca}(c_{at})(w_{at} - w_{mt})}_{\frac{\partial c_{at}}{\partial n_{at}}} = 0 \quad (27)$$

These allocations need to satisfy the marginal product conditions arising from the firms' problems in competitive markets, that is, $w_{at} = \phi \frac{p_{at} y_{at}}{n_{at}}$, $w_{mt} = \alpha \frac{y_{mt}}{n_{mt}}$ and $r_t = (1 - \alpha) \frac{y_{mt}}{k_t}$. We solve the economy by guessing the sequences of factor prices $\{w, r\}_{t=0}^{\infty}$ with $w_t = w_{at} = w_{mt}$. Given these prices, we find the allocations c_{at}, k_{t+1} and n_{at} that solve the following first order conditions (25)-(27) with $p_{at} = \frac{w_t}{\phi z_{at}} \left(\frac{n_{at}}{\ell} \right)^{1-\phi}$.³³ There is market clearing in labor and capital, and aggregate consistency. Note that the intertemporal Euler condition (26) is a second order different equation in $\{k_t, k_{t+1}, k_{t+2}\}$ at every period t . We use as initial and terminal conditions the corresponding stationarized economies at $t = 0$ and at a large T with negligible agricultural share of labor.

³³Note that without the distortion τ , if $\phi = \alpha$ and if we had the same factor inputs in the production of both goods, then the equality of the ratio of factor inputs prices across sectors would imply that the ratio of factor inputs must be identical across sectors. In turn, this would imply a standard result for the pricing of agricultural goods, $p_{at} = \frac{z_{mt}}{z_{at}}$, which renders the price of agricultural good as exogenous. The fact that we allow for ϕ to differ from α and that we have different factor inputs differ across sectors both prevents the standard result. Indeed, in our case, the price of agricultural goods depends endogenously on n_{at} .

Figure 9: A Stage-Based Identification of Model-Generated Policy Effects: A Growth Policy



Notes: The decreasing solid lines dotted dashed lines (left axis) represent the agricultural labor share of output. The increasing dash-dotted lines (right axis) are capital investment. For region \mathcal{T} , we choose, $n_{a,0} = 0.45, z_{a,0} = 0.15, z_{m,0} = 0.17, \gamma_a = 0.007, \gamma_m = 0.0073$. For region \mathcal{C} , we choose, $n_{a,0} = 0.65, z_{a,0} = 0.145, z_{m,0} = 0.145, \gamma_a = 0.007, \gamma_m = 0.0072$. Common parameters between both regions are $\beta = 0.98, \alpha = 0.6, \phi = 0.8, \kappa = 2, \delta = 0.02$. Further, we assume that the felicity functions are logs, that is, $u(c_a - \bar{c}_a) = \ln(c_a - \bar{c}_a)$ and $v(c_m) = \ln c_m$.

True (model-generated) policy effects We consider two regions that potentially differ in model parameters $\Theta = \{\beta, \bar{c}_a, \kappa, \delta, z_{a,0}, \gamma_a, z_{m,0}, \gamma_m, \phi, \alpha, \tau\}$. In particular, we allow for the total factor productivity in the manufacturing sector to be larger in region \mathcal{C} than in region \mathcal{T} . At the same time, to ease the exposition, we keep the total factor productivity in the agricultural sector identical across regions. The larger productivity the manufacturing of region \mathcal{C} generates a larger amount of investment, lower agricultural share of labor and, ultimately, higher income

per capita in region \mathcal{C} than in region \mathcal{T} at any point in time. In panel (a) of Figure 9, we show the equilibrium paths for the agricultural share of labor (declining solid lines, left axis) and capital (increasing solid lines, right axis) for region \mathcal{C} (blue) and region \mathcal{T} . We show the equilibrium path for income per capita in panel (b). The model is able to generate an agricultural share that declines over time whereas, at the same time, capital and income per capita increase asymptotically reaching a balanced growth path with a trifling agricultural share; a phenomenon well documented by the macro-development literature.³⁴ In this context, we introduce an unexpected nationwide growth policy that removes the institutional constraint τ in the agricultural sector in both regions. Precisely, we set $\tau = 0$ after t_p in both regions. The endogenous responses of capital and the agricultural share of labor are in (dashed lines in) panel (a) of Figure 9. Removing the constraint in the agricultural sector accelerates investment (and capital) and the decline in agricultural sector. The reallocation to the non-agricultural sector increases income per capita in the economy, see (dashed lines) in panel (b) of Figure 9. The effects of the policy differ across regions being relatively larger for region \mathcal{T} that is behind in terms of aggregate development with respect to region \mathcal{C} (higher agricultural share, lower capital and lower income per capita).

Stage-based identified policy effects Figure 9 shows the results of applying our method to the model-generated data. We focus on the effects on income per capita. The policy evaluator is not provided with the true counterfactual path without policy (solid lines for the periods after t_p). That is, the information available for the policy evaluator is the path of income per capita without policy (i.e. solid lines) for the periods before t_p and the path of income per capita with policy (i.e. dashed lines) for the periods after t_p . Under these same data constraints, we implement our stage-based identification mapping the outcome path in region \mathcal{C} (solid blue line) onto the outcome path in region \mathcal{T} (solid red line) using only pre-policy data as described in Section 2. We plot the resulting counterfactual candidate $\tilde{y}_{\mathcal{C}}(t; \psi^*)$ (blue line with cross markers) for a window that goes from t_p to $s_{\mathcal{C}}(t_p; \psi^*)$ (shaded pink area) in panel (c) of Figure 9. We compare the identified counterfactual $\tilde{y}_{\mathcal{C}}(t; \psi^*)$ and the true effects of policy in panel (d) of Figure 9. Our stage-based identification finds that the growth policy increases income per capita by $\gamma = 7.63\%$ for the window between t_p and $s_{\mathcal{C}}(t_p; \psi^*)$ whereas the true policy effects are $\gamma_{\text{true}} = 7.51\%$. That is, the identified policy effects catch the true policy effects with an error of $\varepsilon(\gamma) = 1.57\%$.

3.2 A Monte Carlo Analysis: Bounds to Method Performance

The performance analysis in Section 3.1 shows that our identification strategy can recover the true policy effects. But, can our strategy always do the job? In order to address this question we

³⁴See a comprehensive description in [Herrendorf et al. \(2014\)](#).

numerically characterize the bounds within which our method is able to recover the true effects of policy with a Monte Carlo experiment.

We focus this analysis on the benchmark economic model with an endogenous pandemic described in Section 3.1.1. Specifically, we randomize a set of the structural parameters of the benchmark non-reference region \mathcal{C} in order to simulate a large number $m \in \mathcal{M} = \{1, \dots, m, \dots, M\}$ of reference outcome paths. In this manner, how much the outcome paths of the simulated reference regions differ from that of the non-reference region depends on the randomized subset of structural parameters. Precisely, we randomize a subset of the parameter space in the modeled region \mathcal{C} that includes β , ζ , κ and t_o assuming that these parameters are uniformly and independently distributed. We draw a total of $M = 381,000$ quadruplets $(\beta, \zeta, \kappa, t_o)$.³⁵ Each draw generates a new reference outcome path, which implies a total number of M simulated reference outcome paths. In panel (a) of Figure 10, we show the epidemic path of our benchmark regions \mathcal{C} and \mathcal{T} as described in Section 3.1.1. In the same panel, we also show one of the simulated reference regions, $y_{\mathcal{T}(m)}(t)$, that starts later, grows slower and reaches a lower magnitude than the benchmark reference region, $y_{\mathcal{T}}(t)$, and, therefore, is further away from the non-reference region, $y_{\mathcal{C}}(t)$.³⁶

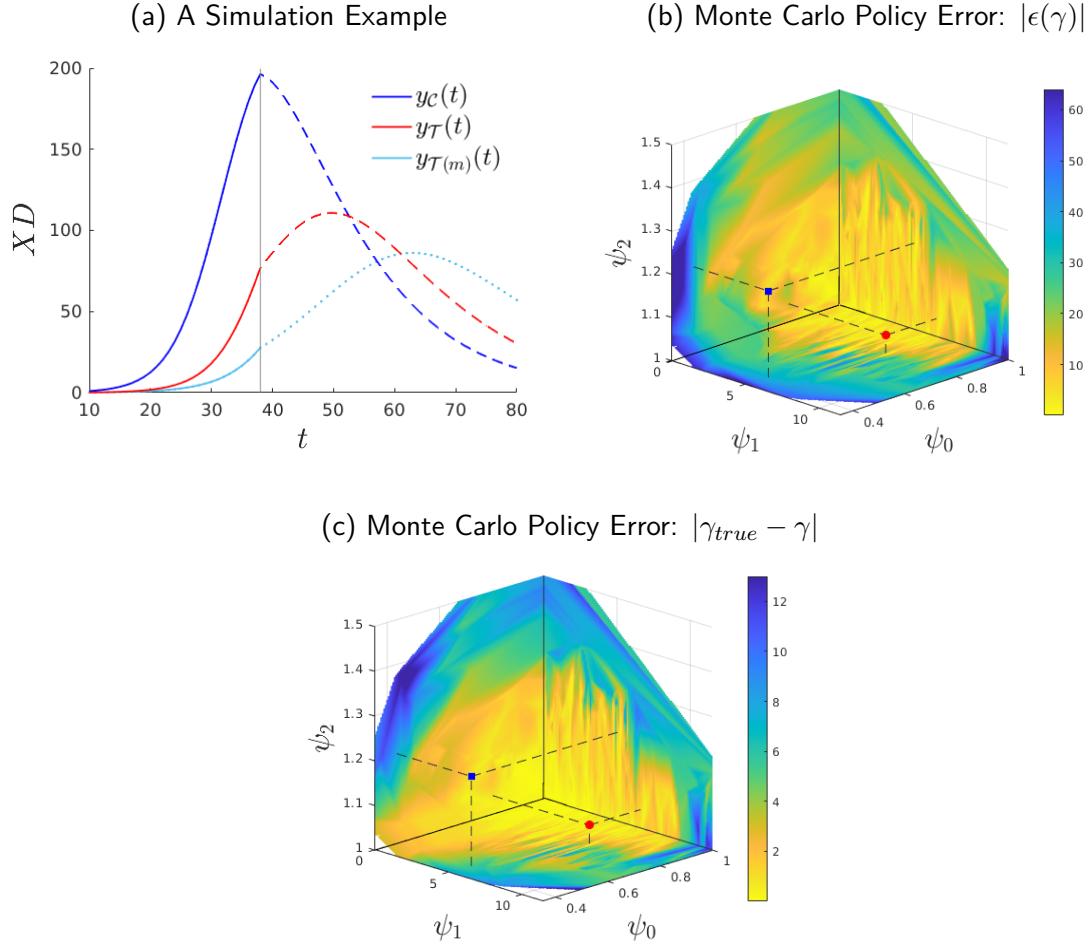
In this context, in order to assess the ability of our method in identifying the true policy effect we study the policy error across all simulations. Precisely, for each simulation, we apply our stage-based identification strategy mapping the non-reference region, $y_{\mathcal{C}}(t)$, onto the simulated reference path, $y_{\mathcal{T}(m)}(t)$. This implies that we find a set of normalization coefficients $\psi^*(m) = \{\psi_0^*(m), \psi_1^*(m), \psi_2^*(m)\}$ per simulation $m \in \mathcal{M}$. Then, for each simulation, we measure the policy error as the (absolute) value of the policy effect identified by our method with respect to the (model-generated) policy effect; i.e. $\varepsilon(\gamma)(m) = \left| \left(\frac{\gamma(m)}{\gamma_{\text{true}}(m)} - 1 \right) \times 100 \right|$. In panel (b) of Figure 10, We plot the policy errors of each of our simulations $\psi^*(m)$ that belong to the vector space $\Phi^q = \Phi_0^q \times \Phi_1^q \times \Phi_2^q = \{\psi_0(m) < 1.0\} \times \{\psi_2(m) > 0.0\} \times \{\psi_2(m) > 1.0\} \subset \Phi = \mathbb{R}^3$. Precisely, we restrict our plot to the vector space $\{\psi_0(m) \in (0.350, 1.000)\} \times \{\psi_1(m) \in (0.000, 10.000)\} \times \{\psi_2(m) \in (1.000, 1.500)\} \subset \Phi^q$ because it suffices to capture the policy error of the outcome path for the benchmark reference region $y_{\mathcal{T}}(t)$ described in Section 3.1.1.³⁷

³⁵We are currently increasing this number to, at least, triple the current size.

³⁶Note that there can be cases where the simulated reference region flips the role of control and treatment across regions. In the Appendix, we show the simulated epidemic path of a reference region $y_{\mathcal{T}}(t)$ that happens to start earlier than that of the non-reference region $y_{\mathcal{C}}(t)$ up to a point where the simulated reference region becomes the control region for the non-reference region.

³⁷Our insights do not change with alternative choices of the vector space; see Appendix.

Figure 10: Bounds to Method Performance: A Monte Carlo Analysis



Notes: We assume that $\{\beta, \zeta, \kappa, t_0\}$ are uniformly and independently distributed. The simulations are drawn from the intervals $[\beta^{lb}, \beta^{ub}] \times [\zeta^{lb}, \zeta^{ub}] \times [\kappa^{lb}, \kappa^{ub}] \times [t_0^{lb}, t_0^{ub}] = [0.5, 0.9] \times [0.001, 0.008] \times [1.05, 1.89] \times [-10, 10]$ where the superindices *lb* and *ub* denote, respectively, the lower and upper bounds of each parameter space. We pick the bounds of the uniform distribution in a manner that our simulations generate sufficiently different outcome paths of the reference region in order to assess the performance of our method. We constructed a total of $M = 381,000$ simulations though not all the simulations fall in the vector space ψ^* in panel (b). Precisely, the hyperplane (ψ_0, ψ_1) has 3,698 simulations, the hyperplane (ψ_0, ψ_2) has 17,504 simulations and the hyperplane (ψ_1, ψ_2) has 3,698 simulations. Panels (b) and (c) show values from an evenly spaced 200×200 grid on each hyperplane. We approximate the values of the grid through linear interpolation of the simulated data.

Our main result is that the success of our method in identifying the true policy effects is bounded. To see this, first, note that the centroid in the vector space Φ , i.e. $\psi_c^* = (1.000, 0.000, 1.000)$, implies that the outcome path of the simulated reference region, $y_{\mathcal{T}(m)}(t)$, and the outcome path of the non-reference region, $y_{\mathcal{C}}(t)$, are identical.³⁸ Second, note that if the outcome path of a simulated reference region, $y_{\mathcal{T}(m)}(t)$, and the outcome path of the non-reference region, $y_{\mathcal{C}}(t)$, are similar—in that our identification strategy delivers a set of normalization coefficients that is in a neighborhood of the centroid $\mathcal{N}(\psi_c^*) \subset \Phi^q$ —then the policy error is small; see panel (b) of Figure 10. To see this, note that policy errors with values of $\varepsilon(\gamma) \leq 5\%$ emerge in a bounded neighborhood (approximately) $\mathcal{N}(\psi_c^*) = \{0.764, 1.000\} \times \{0, 6.776\} \times \{1.000, 1.210\}$ which we depict (yellow area) around the centroid. Here, note that our benchmark reference outcome path $y_{\mathcal{T}}(t)$ falls in that neighborhood with a set of normalization coefficients $\psi^* = \{0.803, 6.592, 1.041\}$ and a policy error $\varepsilon(\gamma) = 0.68\%$ (red marker). Third, moving away from the centroid increases the policy error. For example, the simulated reference outcome path $y_{\mathcal{T}(m)}(t)$ in panel (a) of Figure 10 implies a set of normalization coefficients $\psi^*(m) = \{0.436, 5.083, 1.200\}$ that fall outside of the neighborhood $\mathcal{N}(\psi_c^*)$ and deliver a larger policy error of 36.04%. We further reconduct our exercise using an alternative measure of the policy error defined as $|\gamma - \gamma_{\text{true}}|$ in panel (c) of Figure 10 reaching similar insights.

Summing up, our analysis shows that in order for our method to identify the true policy effect, the pre-policy outcome paths between the reference and non-reference region cannot be too different in the vector space of the normalization coefficients. The specific set of the coefficients $\psi^*(m) \in \mathcal{N}(\psi_c^*)$ that bounds the identification will depend on the model.

3.3 Confounding Factors

Here, we assess the implications of confounding factors for our method. First, we focus on the presence of time-varying latent heterogeneity in Section 3.3.1. Second, we assess the presence of additional confounding policy interventions that occur before the policy under evaluation is implemented in Section 3.3.2.

3.3.1 Time-Varying Latent Heterogeneity

Using the economic model with an endogenous pandemic described in Section 3.1.1, we formalize the time-varying latent heterogeneity across two regions as an underlying time-varying structural parameter present in one region but not the other. Precisely, we consider a scenario in which one region, \mathcal{T} , learns about the process of infection faster than the other region, \mathcal{C} , before policy

³⁸Indeed, exactly at the centroid the policy effects are not identified because $y_{\mathcal{C}}(t) = y_{\mathcal{T}(m)}(t)$ and there is no heterogeneity in stages at the time of policy implementation.

implementation. That is, we allow for the beliefs on the infection process, $\xi_{\mathcal{P}}$, to gradually move closer to the actual ξ in region \mathcal{T} but not in region \mathcal{C} ; see Figure 28 in the Appendix.³⁹ These cross-regional differences in beliefs induce pre-policy behavioral change that differs across regions. Since region \mathcal{T} learns faster about the process of infection, there is a larger behavioral change in this region before policy intervention in the form of an earlier and larger reduction in the hours worked compared with other regions; see panel (a) in Figure 11. Note that in our exemplification the pre-policy behavioral change is rather large—after behavioral change the drop in hours worked before policy implementation in region \mathcal{T} becomes actually larger than that of region \mathcal{C} . In this scenario, we show the implied true policy effects on the flow of deaths in panel (b), the identified policy effects in panel (c) and a comparison between true and identified effects in panel (d) of Figure 11.⁴⁰ The main finding is that our method is robust to this type of heterogeneity. The estimated percentage of lives saved is $\gamma = 12.24\%$ which is close to the true effects, $\gamma_{\text{true}} = 12.66\%$.

Nevertheless, although the introduced time-varying heterogeneity is large, in the sense that it is able to revert the largest endogenous economic response across regions, it can definitely be larger. In that context, analogously to our discussion in Section 3.2, the robustness of our method to time-varying heterogeneity is bounded by how far this heterogeneity drives the cross-regional outcome path of one region away from the other region. That is, our method is robust to time-varying heterogeneity if this heterogeneity does not make the path of the outcome of interest sufficiently different across regions.⁴¹ At the same time, note that the time-varying heterogeneity can also play the opposite role if it actually drives the outcome path of one region closer to the other region.

3.3.2 Confounding Policies

Similarly to the case with time-varying latent heterogeneity, our method is robust to confounding policies as long as these policies do not push the regional outcome paths sufficiently away from each other, i.e. beyond the feasibility bounds described in Section 3.2.

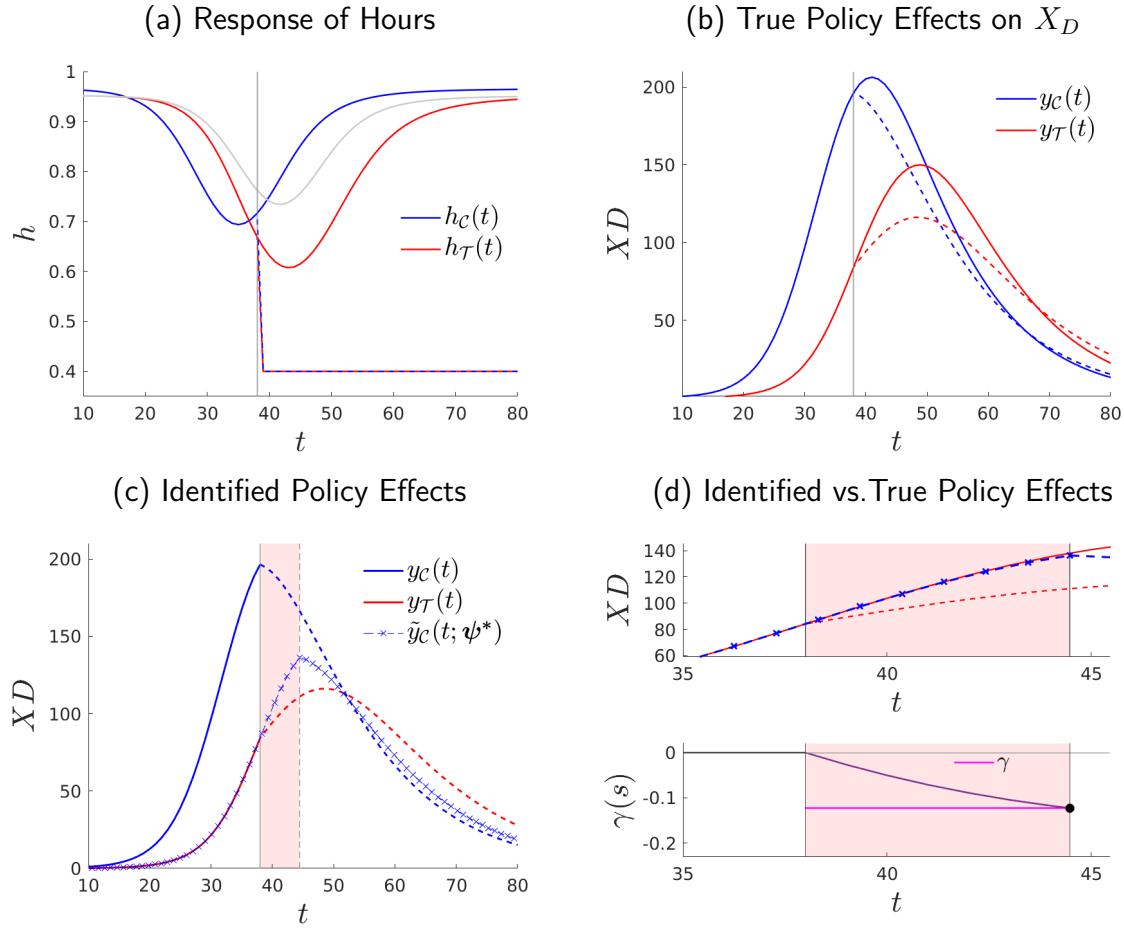
We illustrate this result with a scenario in which an additional confounding policy is introduced in region \mathcal{T} right before the actual nationwide stay-home policy under evaluation is implemented

³⁹The time-varying latent heterogeneity does not need to be gradual. Sudden changes in the form of jumps in $\xi_{\mathcal{P}}$ do not change our insights; see Appendix.

⁴⁰Note that from the perspective of our method, the introduction of time-varying latent heterogeneity does not change the aim of our normalization which is to reduce the difference in structural parameters across regions which, in the current scenario, includes a time-varying component.

⁴¹In the Appendix we show examples where the time-varying heterogeneity generates regional differences in the outcome paths that our method is not able to overcome.

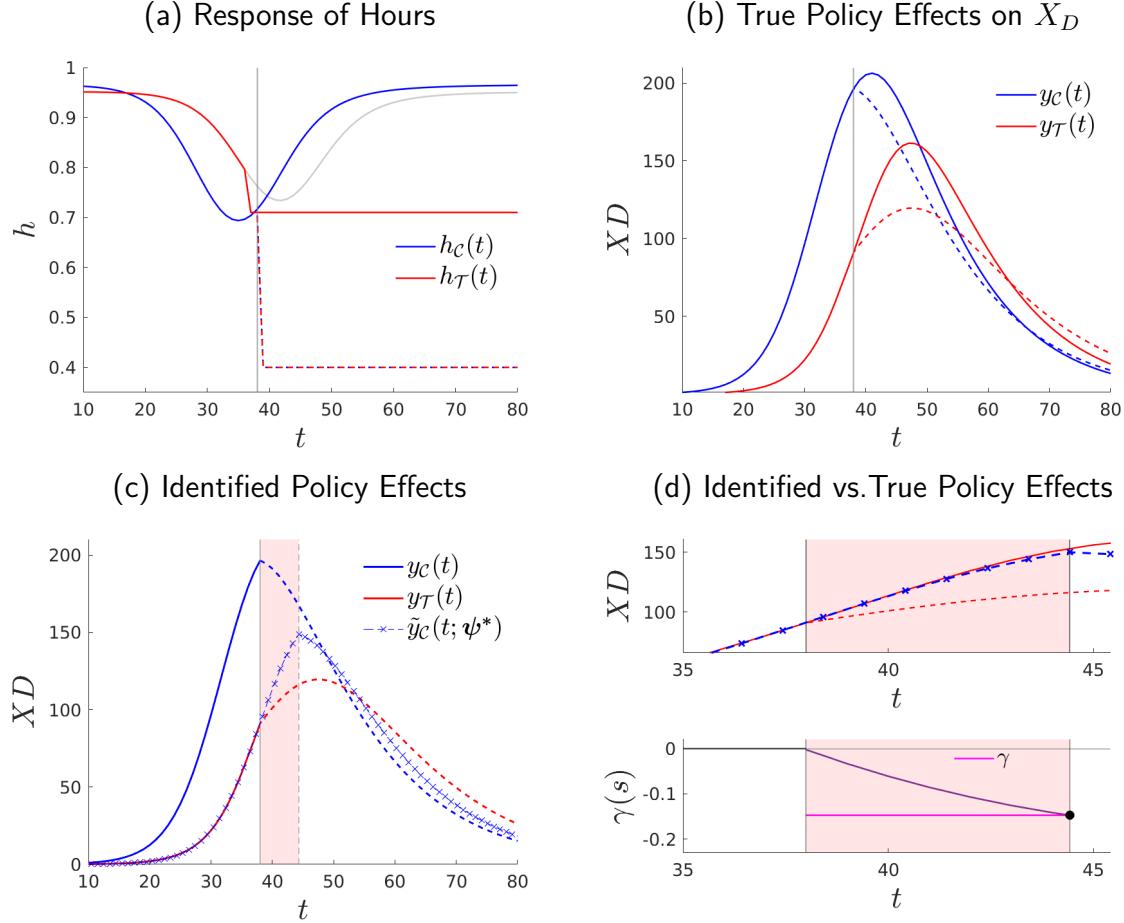
Figure 11: Stage-Based Identification of Policy Effects: Time-Varying Latent Heterogeneity



Notes: Where $\bar{h} = 0.4$, $t_p = 38$, $t_f = 250$, $\gamma = -12.23\%$, $\epsilon(\gamma) = -3.3\%$.

in period t_p . Precisely, we assume that this additional policy occurring before t_p imposes a less strong constraint on hours worked $\underline{h} = 0.71$, see panel (a) of Figure 12. That is, in region \mathcal{T} hours drop to 0.71 one period before the nationwide stay-home policy is put in place, which farther constrains hours to 0.40 from t_p onward. Under this scenario, we use our stage-based method to assess the nationwide stay-home policy introduced at t_p purposefully ignoring the presence of the additional policy introduced before t_p in region \mathcal{T} . Our method identifies the effects of policy to be a 14.71% of lives saved in the identification window which is close to the true effects in that window, 15.59%. This implies a policy error of -5.64%. Again, the policy error increases with the effect of the additional confounding policy on hours and the outcome of interest which implies that the robustness of our method the confounding policy depends on the strength of that confounding policy. We re-conduct the exercise but imposing the additional confounding policy one period t_p in region \mathcal{C} , $\underline{h} = 0.65$, without further insights; see our Appendix. In fact, adding

Figure 12: Stage-Based Identification of Policy Effects: With Confounding Policy in \mathcal{T}

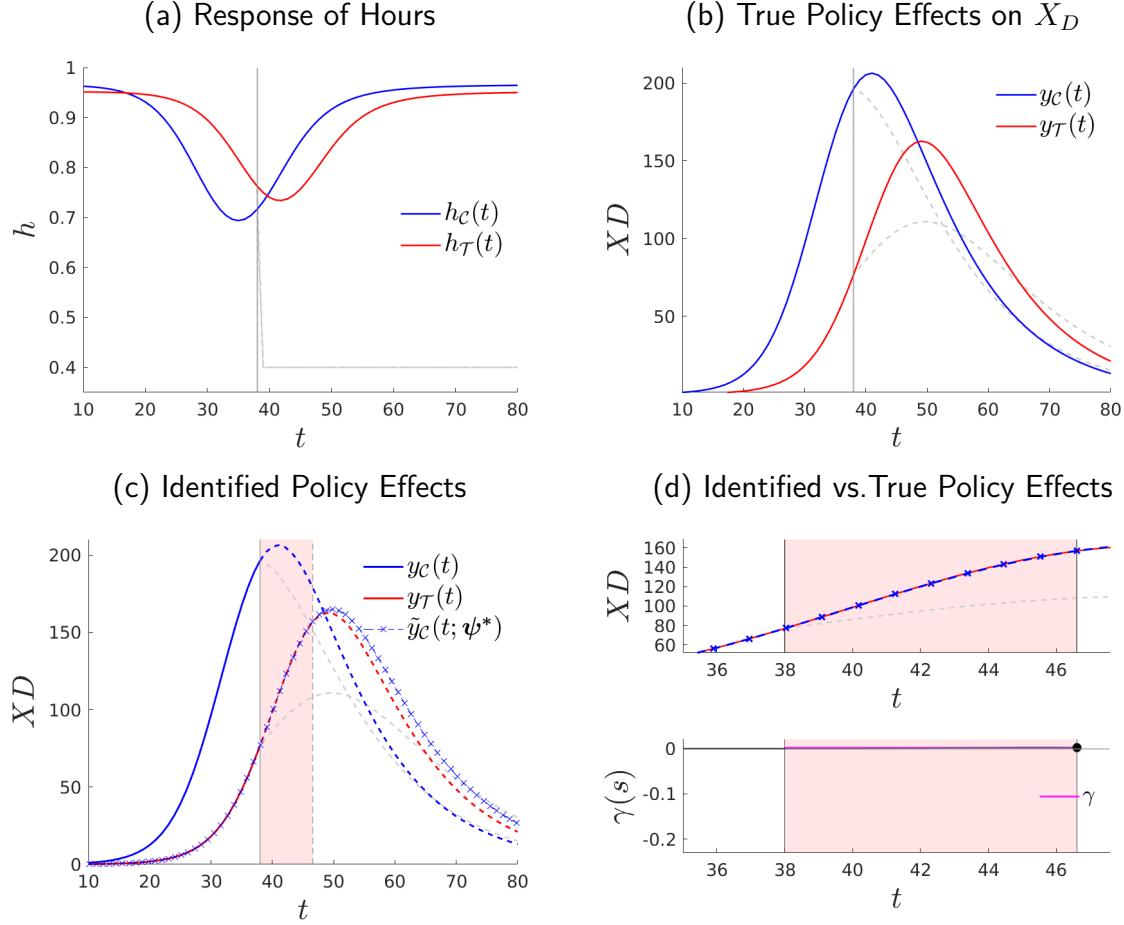


this policy before t_p to \mathcal{C} moves the outcome path of region \mathcal{C} closer to that of region \mathcal{T} which lowers the policy error to 2.61%, with an identified policy effect of 18.41% lives saved compared to a true policy effect of 17.94% lives saved.

3.4 Inference

We show how to conduct inference on our stage-based identification method in two ways. First, we conduct a placebo diagnosis in order to assess how our method performs in the evaluation of nonexistent policy effects. Second, we assess our method when the outcome path of interest is subject to a stochastic component which we use to conduct inference.

Figure 13: Stage-Based Identification of Policy Effects: A Placebo Test



Notes: Where $\bar{h} = 0.4$, $t_p = 38$, $t_f = 250$, $\gamma = 0.19\%$.

3.4.1 Placebo Diagnosis

Here, we assess whether our method identifies policy effects when the policy effects are non-existent. In such scenario, a successful diagnosis is one in which our method identifies the effects of policy to be nil, as they truly are. To conduct this assessment, we apply our method to model-generated data from models that are not subject to policy. Again, we use as benchmark the economic model with endogenous pandemic as calibrated in Section 3.1.1 with the relevant difference that we do not impose a policy at time t_p . That is, the true effects of policy are zero. Under such scenario, the paths for the flow of deaths for region C (solid blue) and region T (solid red) are as depicted in panel (a) of Figure 13. For reference, we also plot in light colors (dashed lines) the outcome paths of what would be the effect of policy had there been one implemented at t_p as we did in Section 3.1.1.

We now purposefully assume that there is policy at some period t_p , when there is actually none. The results of applying our identification strategy to this scenario are in panel (b) of Figure 13. Three remarks are in order. First, note that the pre-policy paths for a fictitious policy at time t_p are identical. Second, applying the normalizing coefficients to post-policy paths shows that the identified counterfactual matches the actual outcome path—i.e. an outcome path without policy. That is, our method correctly identifies that in this scenario without policy the policy effects are null—or quantitatively negligible, $\gamma = 0.19\%$.⁴² This leads us to conclude that our method is able to identify the true zero effects of policy. We find similar insights after re-conducting this exercise for different values of t_p ; see Appendix.

3.4.2 Stochastic Component

In empirical applications, the outcome path of interest might be subject to fluctuations due to the presence of a stochastic component. The stochastic component can capture measurement error.⁴³ Alternatively, the stochastic fluctuations might be genuine but yet of higher frequency than the outcome path of interest—e.g. autocorrelated business cycles when the object of interest is the growth path. We use this stochastic component to conduct inference.

To address these issues, we artificially embed model-generated data with a stochastic component. We start with the presence of classical measurement error and the outcome paths are,

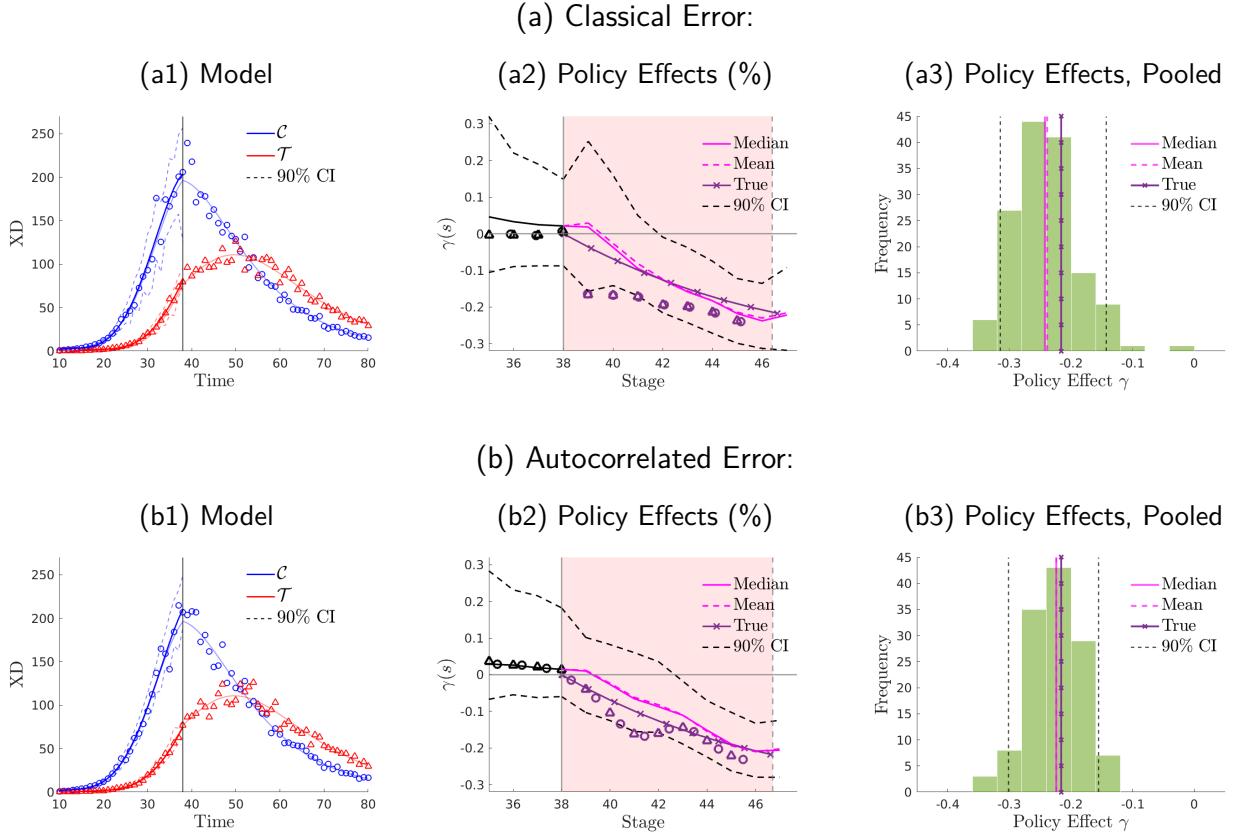
$$\hat{y}_r(t) = y_r(t) + u_r(t) \quad \text{with} \quad u_r(t) \sim N(0, \sigma_{u,r}^2), \quad (28)$$

for each region $r = \{\mathcal{C}, \mathcal{T}\}$, where $\hat{y}_r(t)$ is the outcome path observable to the policy evaluator, $y_r(t)$ is the unobservable true outcome path and the innovations $u_r(t)$ capture measurement error that follows a normal distribution with zero mean and variance σ_r^2 . In the case where the stochastic component is autocorrelated, we replace the innovations in our specification (28) with $u_r(t) = \rho u_r(t - 1) + v_r(t)$ where $v_r(t) \sim N(0, \sigma_{v,r}^2)$. We show the outcome path $\hat{y}_r(t)$ by region embedded with classical measurement error from one simulation of (28) in panel (a.1) of Figure 14. In what follows, we treat that one simulation $\hat{y}_r(t)$ (circle markers) as the actual data. Note that, although through the lens of the model we know the true outcome path $y_r(t)$ (solid lines), the policy evaluator only observes $\hat{y}_r(t)$.

⁴²We also conduct a placebo diagnosis using our model on oral contraceptives as described in Section 4.2. Likewise, we find that the identified policy effects are close to zero. For the proportion of women going to college $\gamma = -0.0002\%$ and for fertility $\gamma = -0.0018\%$.

⁴³See a general discussion on measurement error in regression analysis in, for example, [Hausman \(2001\)](#). Further, [Hyslop and Imbens \(2000\)](#) discuss biases when the mismeasured data are optimal predictions of true values.

Figure 14: Stage-Based Identification of Model-Generated Policy Effects: Inference



Notes: We use the benchmark calibration in Section 3.1.1. The top panels (a) introduce classical error in our model with $\{\sigma_C^2, \sigma_T^2\} = \{0.008, 0.008\}$. The bottom panels (b) introduce autocorrelated error with $\{\rho_C, \rho_T\} = \{0.13, 0.13\}$ and $\{\sigma_C^2, \sigma_T^2\} = \{0.008, 0.008\}$.

In this context, in order to identify policy effects, we add a smoothing—or trend-extraction—step that precedes our entire methodology described in Section 2. We apply this additional step to the observed pre-policy outcome paths by region, i.e. $\hat{y}_r(t)$ with $t < t_p$. The idea is to smooth $\hat{y}_r(t)$ in order to purge the pre-policy outcome paths of the stochastic fluctuations—of higher frequency than the object of interest—defined as deviations from a smooth estimand. In what follows, we define $\hat{\hat{y}}_r(t)$ as the estimand of $y_r(t)$. To obtain $\hat{\hat{y}}_r(t)$ we use Chebyshev polynomials which we apply to the observed data $\hat{y}_r(t)$. In this manner, we recover estimates for the error terms $u_r(t)$.

We use the empirical distribution of the recovered error terms $u_r(t)$ to conduct inference. Precisely, we reallocate the recovered errors $u_r(t)$ on the initial estimand $\hat{\hat{y}}_r(t)$ across periods using a bootstrap procedure without replacement. We redo this procedure for $B = 1,000$ times to construct an equal number of $\hat{y}_{r,b}(t)$ simulations with $b \in B$. In panel (a1) of Figure 14, we show the median, mean and 90% confidence intervals across the simulated paths $\hat{y}_{r,b}(t)$. Then

for each bootstrap sample, we recover a different estimand $\hat{\bar{y}}_{r,b}(t)$ in the trend-extraction step. Therefore, there are as many stage-based identified policy effects $\gamma_b(s)$.

We show the identified policy effects $\gamma_b(s)$ in panel (b1) of Figure 14. Here, note that the heterogeneity in the stage-identified policy effects arising from our simulations is partly due the differences in the size of the identification window, i.e. heterogeneity in $s_{C,b}(t; \psi^*)$ —assuming C leads T in stages at t_p . For this reason, in panel (b1) we focus on the simulations that are in the neighborhood of the mean window size, i.e. in $N(p) = \{s_{C,b} \in S : |s_{C,b} - p| < rp\}$ with $p = \sum_b \frac{s_{C,b}}{B}$ and $r = 0.05$. Two results emerge. First, after normalization our method generates outcome paths that are not significantly different before policy implementation. Second, the identified policy effects—using data with measurement error—are not significantly different from the true (model-generated) policy effects without measurement error (purple line with crossed markers). Precisely, the identified mean policy effect (dashed magenta line) is 22.58%—within a 90 percent confidence interval of [11.71, 31.58]—which is not significantly different from the true (model-generated) policy effect without measurement error, i.e. 21.50%. The median policy effect (solid magenta line) is of similar size, 23.1%. However, the significance of the identified policy effect can be affected by the size of the measurement error. In the Appendix, we show how the accuracy of the identified policy effect depends on the size of the measurement error. We find that the identified mean policy effect is similar when we unrestricted our analysis to the mean window size, 23.89%, but shows somewhat wider confidence intervals; see panel (a3) of Figure 14. We also conduct robustness of our methodology using a wider set of smoothers and find similar insights—our alternative smoothers include B-splines, cubic splines, moving averages and the Hodrick-Prescott filter; see Appendix.⁴⁴

We re-conduct our analysis assuming that the stochastic component is autocorrelated, see panel (b1) of Figure 14. There, we show the outcome paths that emerge from bootstrapping the empirical distribution. In this case, to keep the empirical autocorrelation structure of the error terms—including potentially temporal differences in the cross-sectional variance—we use a block bootstrap procedure that increases the sampling weight of preceding error terms in a pre-specified window Carlstein (1986).⁴⁵ With autocorrelated measurement error, we also find that our identification strategy is able to recover policy effects that are not significantly different from

⁴⁴An altogether alternative way to conduct inference with the recovered estimates for the error terms $u_r(t)$ is to estimate the sample variance of the errors, i.e. $\hat{\sigma}_r$. Then, under a normality assumption on the error term in (28), we simulate $Q = 1,000$ paths of errors and, hence, the same number of pre-policy outcome paths onto which we apply the smoothing step in order to recover a simulation-specific estimand $\hat{\bar{y}}_{r,q}(t)$. Since the estimand $\hat{\bar{y}}_{r,q}(t)$ differs by simulation $q \in Q$, each simulation delivers an stage-based identified policy effect, γ_q . The results under this different inference are in the Appendix. Overall, we find similar insights with an identified mean policy effect of 21.3% [14.53, 28.21] that is not significantly different from the true (model-generated) policy effect.

⁴⁵We select a block window of size 5

the true (model-generated) policy effects, see panel (b2) and (b3) of Figure 14.

Finally, in order to assess the role of the smoother on the identification of the true policy effects, we perform our identification strategy directly on the observed data $\hat{y}_r(t)$ —i.e. the pre-policy outcome paths (circle markers) in panel (a1) of Figure 14. That is, we conduct the normalization by mapping directly the outcome path $\hat{y}_C(t)$ onto $\hat{y}_{\mathcal{T}}(t)$ without the smoothing step. There is a unique identified policy effect (purple markers) that we show in panel (a2) of Figure 14. The policy effect identified using the observed data $\hat{y}_r(t)$ also replicates the true policy effect. This suggests that the smoother serves for the sole purpose of conducting inference, but it does affect the identified policy effect.

4 Applications

We use our strategy to identify the policy effects in a set of empirical applications associated with nationwide policies. First, we assess the effects of stay-home policies on the flow of deaths during the Covid-19 pandemic in Section 4.1.⁴⁶ Second, we assess the effects of the approval of oral contraceptives on fertility and women’s college education Section 4.2. Third, we study the effects of the German reunification on income per capita in Section 4.3.

4.1 The Effects of the Spanish *Confinamiento* Against Covid-19

Like many countries, Spain pursued non-pharmaceutical public health policies in response to the Covid-19 pandemic. On March 14, 2020, the Spanish government announced a nationwide stay-at-home policy—enacted the following day—which locked down all non-essential workers in all regions of Spain. Indicative of its strictness, the public debate referred to the policy as confinement. The strictest measures were lifted on May 2 when the first wave of the epidemic flattened out. Here, we apply our stage-based identification strategy to assess the effects of this policy intervention on the course of the pandemic. As outcome of interest, we focus on the daily flow of deaths attributed to Covid-19.⁴⁷ We use two Spanish regions to assess the nationwide policy: Madrid and an artificially created region Rest of Spain (RoSPA) which is composed of all Spanish regions without Madrid. We label Madrid as region C and RoSPA as region \mathcal{T} .

We show the daily flow of Covid-19 deaths (per million inhabitants) for Madrid (blue circles)

⁴⁶The Covid-19 has generated lots of empirical work assessing public health policies against the pandemic; see, for example, Fang et al. (2020) for a careful study of the early mobility restrictions in China and Liu et al. (2021) for the provision of density forecasts with Bayesian techniques for a panel of countries and regions.

⁴⁷Although the data on daily deaths is potentially imperfectly measured, we regard daily deaths as less prone to measurement error than daily infections (cases), especially during the onset of the pandemic, since testing was only gradually introduced. Our choice of daily deaths as outcome variable for our analysis reflects this belief.

and the RoSPA (red triangles) in panel (a) of Figure 15; Instituto de Salud Carlos III (ISCIII). In order to mitigate potential measurement error on the reported deaths, we smooth the pre-policy data using as benchmark Chebyshev polynomials separately by region as described in Section 3.4.2.⁴⁸ Note that we add a lag parameter to the policy date, reflecting that a policy that aims at reducing infections will show an effect on the flow of deaths with a delay. We set a lag parameter of $\tau = 12$ which implies that the policy is (effectively) implemented on March 27; and conduct robustness in the Appendix. Hence, we smooth the data starting the day on which we observe the first death until the policy is effectively implemented on March 27. The resulting smoothed daily flow of deaths for Madrid (solid blue) and RoSPA (solid red) are also in panel (a) of Figure 15.

There are clear differences in the path of the flow of deaths between Madrid and RoSPA. First, the figure of one death (per million inhabitants) is reached in March 08 for Madrid and March 14 for the RoSPA. Second, by March 14 the daily flow of deaths in Madrid is 9.3 deaths (per million inhabitants) whereas this figure is 1.2 for the RoSPA. Furthermore, at the (effective) time of policy implementation, the flow of deaths is reaching a peak in Madrid at 50 deaths (per million inhabitants), whereas the peak in RoSPA is smaller at 16 deaths (per million inhabitants) and occurs about a week after that in Madrid. That is, the flow of deaths starts at an earlier date, it raises more rapidly and reaches a larger peak in Madrid than in the RoSPA.

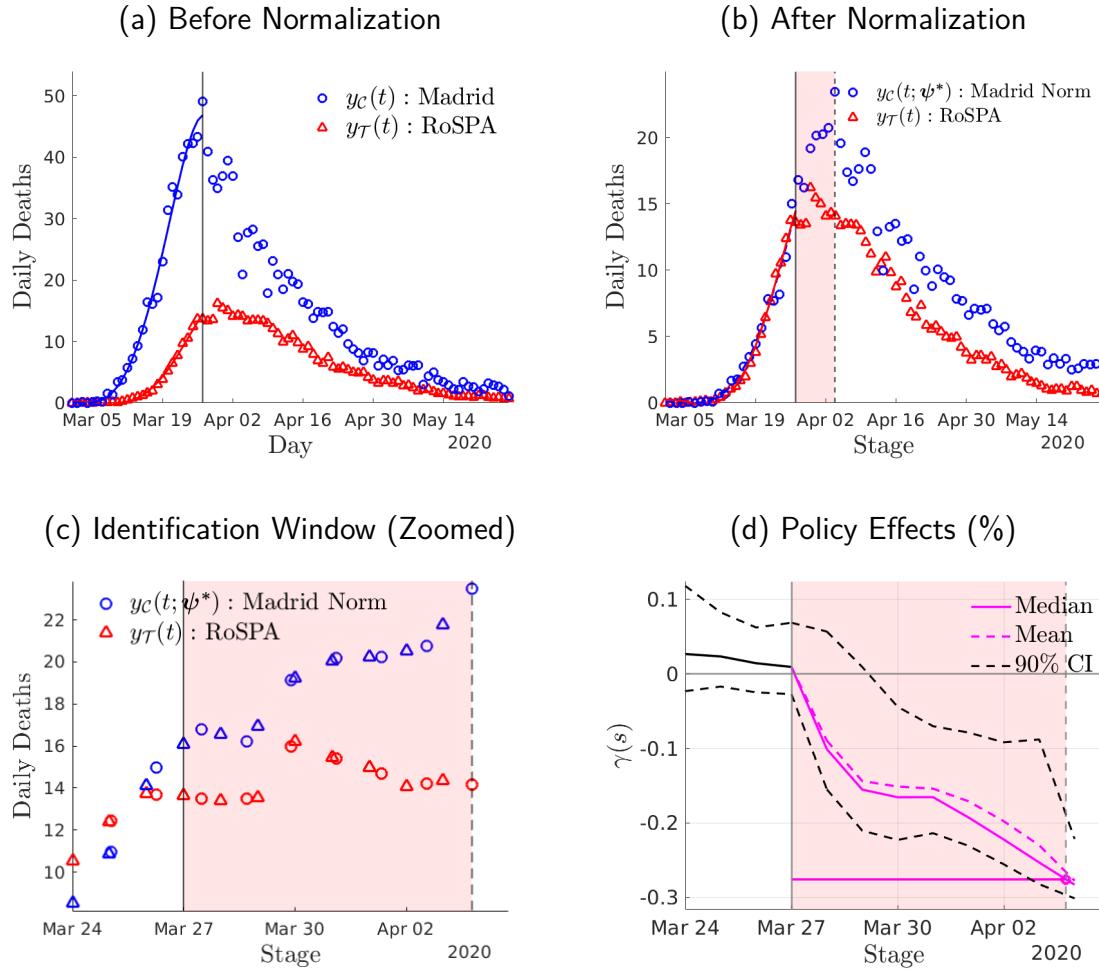
Normalization. We now apply our stage-based identification strategy following the normalization described in Section 2. Picking the RoSPA as reference region, we map the flow of deaths in the region of Madrid ($y_C(t)$, solid blue circles) onto the flow of deaths in the RoSPA ($y_T(t)$, solid red circles) using *only* pre-policy data. The result of our identification is a normalized path for the region of Madrid $y_C(t; \psi^*)$ tracked on the stage domain; see panel (b) of Figure 15. Precisely, we find $\psi_1 = -0.14$ [-0.24,-0.04], $\psi_2 = 1.21$ [1.16,1.24] and $\psi_0 = 0.47$ [0.39,0.53] which, respectively, delays the start, slows down the growth and lowers the peak of daily deaths in Madrid.⁴⁹

Our normalization shows that the policy is implemented in Madrid at a later stage than RoSPA, i.e. $s_C(t_p; \psi^*) > t_p$. That is, Madrid leads the epidemic. At the same time, the normalization unveils an identification window $\mathbb{D}(s) = [t_p, s_C(t_p; \psi^*)]$ (shaded pink area) in which the region of Madrid is not yet under the effect of policy whereas RoSPA is. Therefore, the normalized path for Madrid provides a counterfactual without policy for RoSPA in $\mathbb{D}(s)$. We find that the window runs

⁴⁸We use a Chebyshev polynomial of degree 6 with 5 Chebyshev nodes. We also report the identified policy effects for an alternative set of smoothers.

⁴⁹Precisely, our normalization expedited the start of the epidemic for Madrid in $\psi_0 = 0.14$ days, it slows down the growth of the epidemic—taking $\psi_2 = 1.21$ days for Madrid to cover one day for RoSPA in the stage-domain—and it lowers the peak by $(1-\psi_0)*100=53\%$. Thus, only adjusting for different start dates (as in [Glogowsky et al., 2021](#)) would imply wrong estimates.

Figure 15: The Effects of the Spanish *Confinamiento* Against Covid-19



Notes: Panel (a) shows the daily Covid-19 deaths for Madrid, region \mathcal{C} , and for an artificial region \mathcal{T} that aggregates the rest of Spain. We use a Chebyshev smoother (solid lines) of degree 6 with 5 knots. Panel (b) shows the results of our normalization using region \mathcal{T} as reference and mapping the pre-policy outcome paths of region \mathcal{C} onto region \mathcal{T} . Panel (c) zooms the identification window. Panel (d) shows the policy effect $\gamma(s)$ as defined in equation (5). We show the mean, median and 90% confidence interval bands from bootstrapped simulations constructed as described in Section 3.4.2. We further show the policy effects identified without the implementation of the smoothing step as described in Section 3.4.2. We estimate a significant auto-correlation coefficient for the residuals, thus perform block-bootstrap, $\rho_{\mathcal{C}} = 0.37$ $\rho_{\mathcal{T}} = 0.65$, respectively. We choose a block window of 5 days

from the time where the policy is effectively implemented in the reference region RoSPA at t_p (i.e. March 27) to the time where Madrid enters the policy in terms of stages, $s_{\mathcal{C}}(t_p; \psi^*) = t_p + 7.7$ days (i.e. during April 03). Further, note that as a result of our normalization the daily deaths across regions before the policy is implemented are not significantly different from each other in the stage domain, which warrants the counterfactual; see panel (d) of Figure 15.

Policy Effect. We zoom the outcome paths in the identification window in panel (c) of Figure 15. We show the policy effects (as defined in (5)) in panel (d) where we restrict the attention to the bootstrap simulations within the neighborhood of the mean window size—12% distance. The (mean) identified total number of lives saved (per million inhabitants) is $\int_{\mathbb{D}(s)} (\tilde{y}_C(s) - y_T(s)) ds = 36.92$ within approximately one week after policy implementation. Then, the total amount of lives saved by the policy in RoSPA is 1734 during that first week. In this manner, the policy prevented $\gamma = -24.71\%$ of the total deaths that would have occurred in RoSPA had the policy not been implemented. These effects are significant with 90% confidence intervals [-29.71,-19.30]. The median effect is similar, -26.45%. Further, unrestriciting the window size, we find similar significant policy effects with mean -22.11% and median -22.95% Last, conducting our assessment without the smoothing step, we find that the policy prevented 25.61% of the deaths in RoSPA during approximately the first week (see purple markers in panel (d) of Figure 15).

4.2 The Effects of the 1960 FDA Approval of Oral Contraceptives in the U.S.

In 1960, the first hormonal birth control pill (oral contraceptive) was approved in the U.S. by the Food and Drug Administration (FDA). The use of pill was approved for use by women above the age of majority. Access to the pill has potential benefits to women's education attainment and fertility decision. In a seminal paper, Goldin and Katz (2002) use state-level variation in the age of majority in order to assess how women in that threshold change schooling and career choices.⁵⁰ Since our stage-based identification strategy does not require state-level variation, we can directly assess the effects of the federal approval of oral contraceptives on the entire population of women. We focus on two outcome variables. First, we study the pill effects on women's fertility choices focusing on (crude) birth rates. Second, we study the pill effects on women college choices—precisely, the share of women that at age 25 have completed college⁵¹

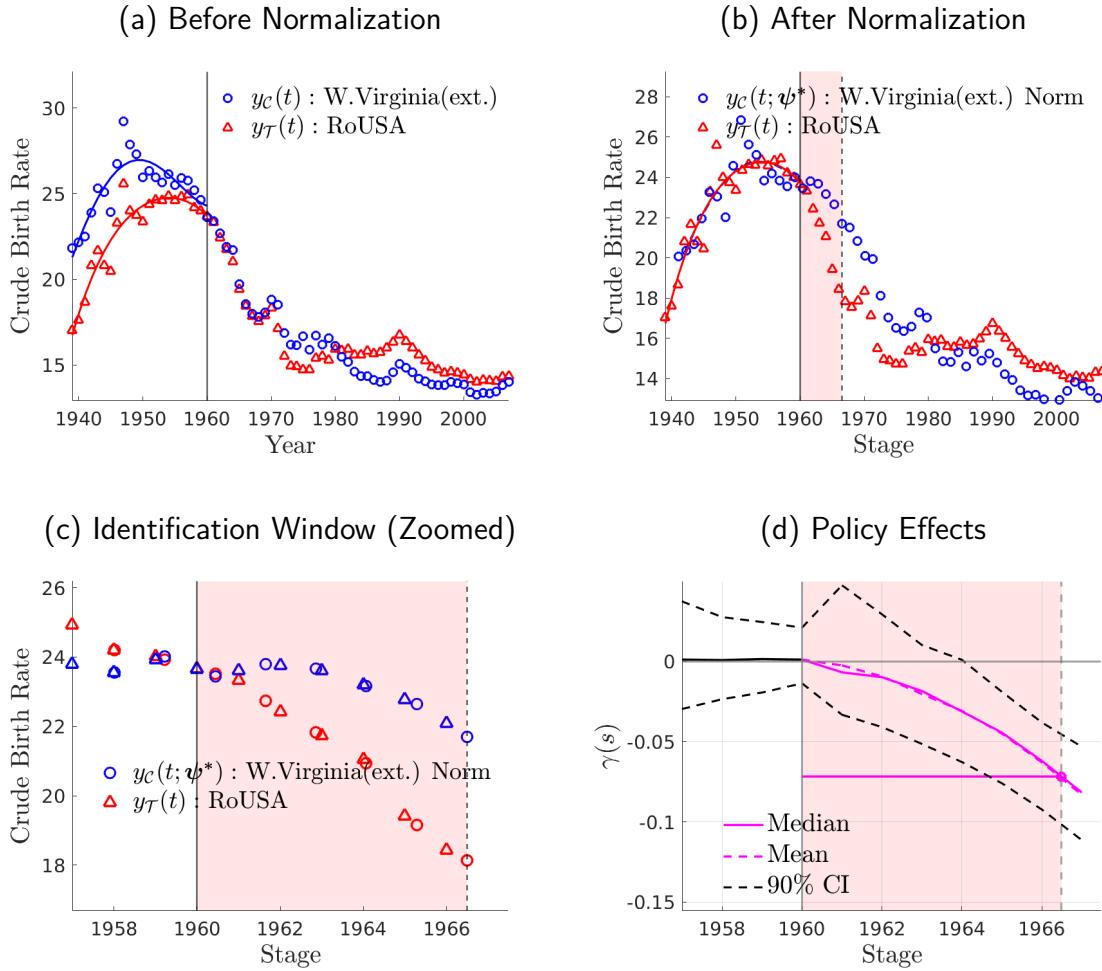
To measure the share of women of a certain age with college attainment we use decennial CENSUS data from IPUMS starting in 1940 up to 1980. In the absence of information on the year of graduation, we construct the historical series by using cohort information by CENSUS year. For example when using CENSUS data for 1960, the share of college women of age 25 in 1959 will be the share of a woman age 26 who reported (already) having attained college by 1960.⁵² After computing the historical series per CENSUS year we compute the average across

⁵⁰Further, Bailey (2006) uses state-level variation in the age of majority to assess the effects of the pill on the timing of first births and women's labor force participation. Greenwood and Guner (2010) use an equilibrium matching model to assess the effects of oral contraceptives on premarital sex and how it is perceived in society.

⁵¹We conduct robustness on the choice of the age in the Appendix.

⁵²Later completion and death could hinder the precision of our measure, however after comparing the historical

Figure 16: The Effects of the 1960 FDA Approval of Oral Contraceptives: Crude Birth Rate



Notes: Panel (a) shows the crude birth rate for a region \mathcal{C} which consists of a set of states leading the fertility bust (West Virginia, Idaho, Nevada and Arkansas) and a region \mathcal{T} that aggregates the rest of the United States. We use a Chebyshev smoother (solid lines) of degree 5 with 3 knots. Panel (b) shows the results of our normalization using region \mathcal{T} as reference and mapping the pre-policy outcome paths of region \mathcal{C} onto region \mathcal{T} . Panel (c) zooms the identification window. Panel (d) shows the policy effect $\gamma(s)$ as defined in equation (5). We show the mean, median and 90% confidence interval bands from bootstrapped simulations constructed as described in Section 3.4.2. We further show the policy effects identified without the implementation of the smoothing step as described in Section 3.4.2. We estimate a non-significant auto-correlation coefficient for the residuals, $\rho_{\mathcal{C}} = 0.31$ $\rho_{\mathcal{T}} = 0.18$, respectively.

CENSUS series.

In terms of the crude birth rate, we find that West Virginia leads the rest of the U.S. In order to increase the sample size, we further add the next three leading regions (Idaho, Nevada and Arkansas) in order construct an artificial region—population weighted average. We also

series from various census years, the measure doesn't seem to be suffering from these problems

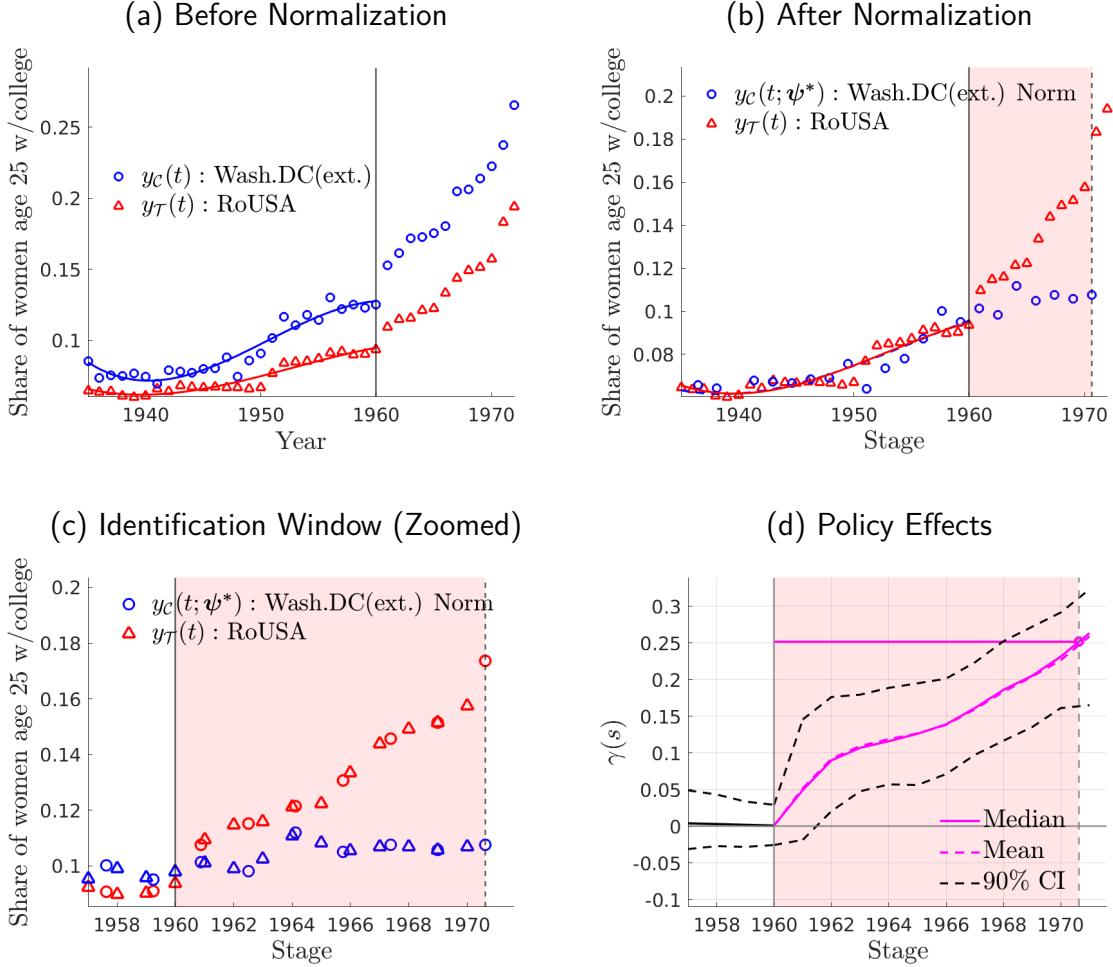
construct an artificial region that consists of the rest of the U.S. The crude birth rate shows a clear inverted-U shape pattern typically labeled as baby boom and baby bust (Greenwood et al., 2005); see panel (a) in Figure 16. We find differential patterns across states. In particular, the crude birth rate in the leading regions peaks in the second half of the 1940s and in 1960 is already busting and close the levels in 1940. Instead, the crude birth rate in the RoUSA peaks in the second half of the 1950s at somewhat lower level and, on average, it has barely started to decline by the early 1960.

In terms of women's college completion, our method identifies that the leading state is Washington D.C. Since this state has a relatively small population size, we construct an artificial region—population weighted average—that additionally includes the next three leading states in terms of women college completion (Massachusetts, Colorado and Connecticut). We also construct an artificial region that consists of the remaining states in the U.S. (RoUSA). We find that the proportion of women of age 25 with completed college attainment has more tripled over a span of twenty years raising from 8% in 1950 to 26% in 1970 in the leading states; see panel (a) of Figure 17. In the RoUSA, the proportion of women of age 25 with completed college attainment shows a much larger relative increase from 2% in 1950 to 15% in 1970.

Normalization. Picking the rest of the U.S. as reference region, we apply our normalization by mapping the pre-policy crude birth rates in the leading region of West Virginia (extended) ($y_C(t)$, solid blue circles) onto the preo-policy crude births rates in RoUSA ($y_T(t)$, solid red circles). Our identification strategy results in a normalized path for West Virginia (extended) $y_C(t; \psi^*)$ tracked on the stage domain; see panel (b) of Figure 16. Our normalization implies the estimates $\psi_1 = 1.85 [-0.53, 9.96]$, $\psi_2 = 1.21 [0.41, 1.58]$ and $\psi_0 = 0.91 [0.89, 0.94]$ which, respectively, delays the start, slows down the growth and lowers the peak of the baby boom.⁵³ A result of the normalization is that West Virginia (ext.) leads the crude births rate path. Therefore, the normalized path for West Virginia (ext.) provides a counterfactual without policy for RoUSA. Precisely, there is an identification window in terms of stages, $\mathbb{D}(s) = [t_p, s_C(t_p; \psi^*)]$ (shaded pink area), in which West Virginia (ext.) is not subject to the 1960 FDA approval of oral contraceptives, whereas RoUSA is. The window runs from 1960 to the time where West Virginia enters the policy in terms of stages, which is approximately in 1970. Further, note that as a result of our normalization the share of women of age 25 that completed college is not significantly different between West Virginia (ext.) and RoUSA before policy implementation in the stage domain; see panel (d) of Figure 16.

⁵³The fact that the policy happens after peak of the crude birth rate can provide a role for an asymmetry parameter in the stage-time transformation, i.e. adding the monomial basis $\psi_3 t^2$. However, at the time of policy implementation the decline in the crude birth rate for the non-leading region has barely started and when we introduce an asymmetry parameter ψ_3 we find that is not significantly different from zero.

Figure 17: The Effects of the 1960 FDA Approval of Oral Contraceptives: Women College



Notes: Panel (a) shows the proportion of women of age 25 that completed college for a region \mathcal{C} which consists of a set of states leading women's college completion (Washington D.C., Massachusetts, Colorado and Connecticut) and a region \mathcal{T} that aggregates the rest of the United States. We use a Chebyshev smoother (solid lines) of degree 4 with 3 knots. Panel (b) shows the results of our normalization using region \mathcal{T} as reference and mapping the pre-policy outcome paths of region \mathcal{C} onto region \mathcal{T} . Panel (c) zooms the identification window. Panel (d) shows the policy effect $\gamma(s)$ as defined in equation (5). We show the mean, median and 90% confidence interval bands from bootstrapped simulations constructed as described in Section 3.4.2. We further show the policy effects identified without the implementation of the smoothing step as described in Section 3.4.2. We estimate a non-significant auto-correlation coefficient for the residuals, $\rho_{\mathcal{C}} = 0.21$, $\rho_{\mathcal{T}} = 0.64$, respectively.

We proceed analogously for the proportion of women of age 25 that completed college; see Figure 17. In this case, we apply our normalization by mapping the pre-policy share of college women in the leading region of Washington DC (extended) ($y_{\mathcal{C}}(t)$, solid blue circles) onto the associated preo-policy path in RoUSA ($y_{\mathcal{T}}(t)$, solid red circles). Our identification strategy generates a normalized path for Washington DC (extended) $y_{\mathcal{C}}(t; \psi^*)$; see panel (b) of Figure 16. Here, the normalizing parameters are $\psi_1 = -5.65$ [-11.17, 1.84], $\psi_2 = 1.62$ [1.07, 1.73] and $\psi_0 = 0.85$

[0.76,0.92].

Policy Effects. We zoom the identification window for the crude birth rates, which spans for approximately seven years, in panel (c) and the associated policy effects in panel (d) of Figure 16. We find that the policy significantly reduced by $\gamma = -8.36\%$ the number of births (per 10,000 inhabitants) that would have otherwise occurred without the pill. The median policy effects are very similar with a -6.94% reduction. In the previous effects, the window size is restricted to be in the neighborhood of the mean window size—12% distance. However, unrestricting the window size, we also find significant effects of similar size $\gamma = -7.53\%$. Further, without the smoothing step, the FDA approval of oral contraceptives implies a reduction of -14.81% in the number of births, which is not significantly different from our mean bootstrapped effects.

The policy effects on the share of women of age 25 with completed college education are in panel (d) of Figure 17. The FDA approval of oral contraceptives significantly increased the share of women of age 25 with a completed college degree by $\gamma = 24.69\%$ during the decade that followed the policy implementation. The median policy effects are almost identical $\gamma = 24.00\%$. Unrestricting the window size, we also find significant effects of $\gamma = 19.19\%$. Finally, if we do not conduct the smoothing step, then we find that the policy effects are also of similar size with the pill increasing in the share of women of age 25 that complete college education by 18.25% .

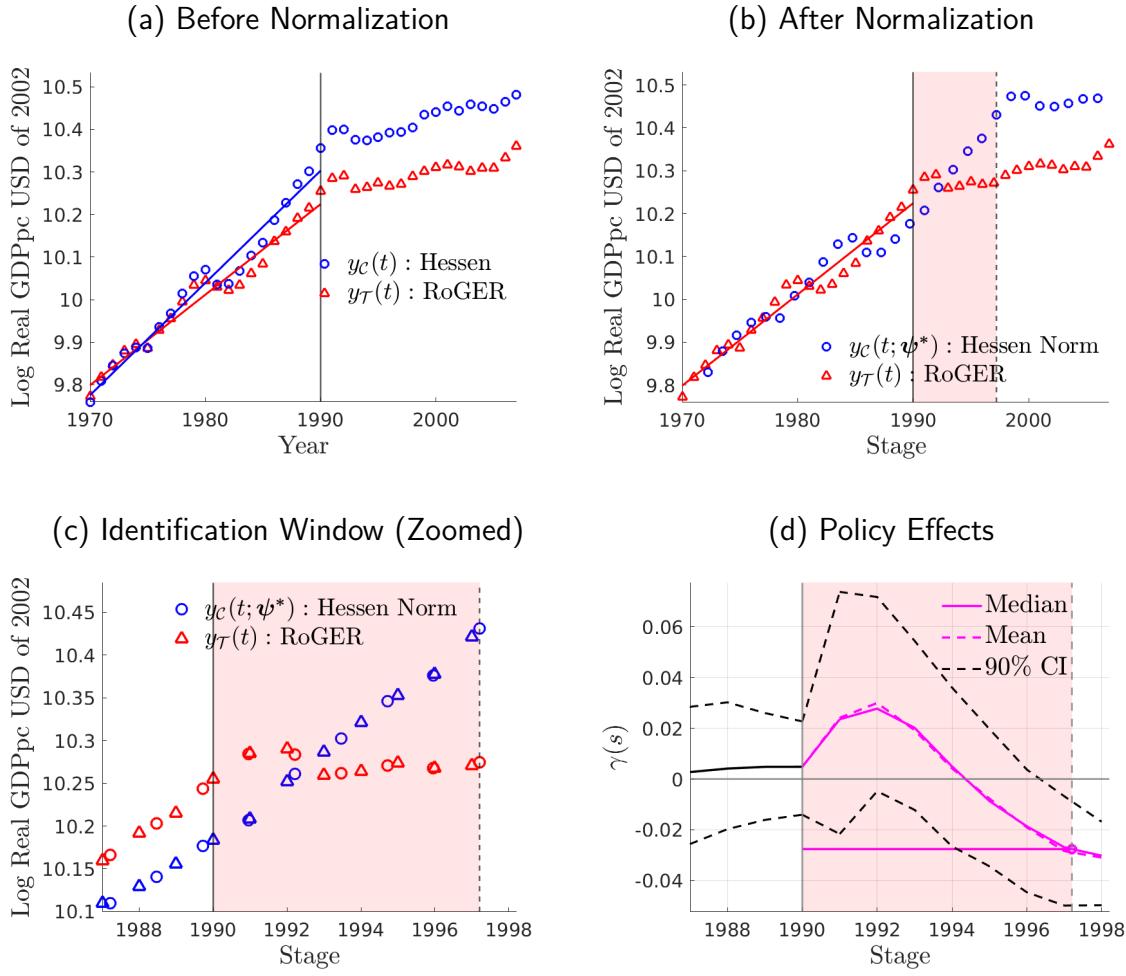
4.3 The Effects of the German Reunification

In 1990, after the fall of the Berlin wall in 1989, the German Democratic Republic was abolished and integrated fully into the Federal Republic of Germany. Given large differences between the West German states and the East German states, the political and economic integration came at some cost—the size of which is subject to debate. [Abadie et al. \(2014\)](#) study the consequences of the German reunification for West Germany and forming a counterfactual path for GDP per capita using a Synthetic Control Group (SCG) approach. Here, we apply our stage-based identification method to the same context, and construct a counterfactual for the evolution of GDP per capita in West Germany had it not been for the reunification. In contrast with [Abadie et al. \(2014\)](#), our counterfactual is constructed using the GDP per capita paths of West German regions only. Precisely, our identification of the effects of the German Reunification relies on the heterogeneity across West German regions in the stage of the growth path at the time of the reunification.⁵⁴

We focus on the effects of the German Reunification on GDP per capita. Here, we use the region of Hessen, which our normalization pins down as the leading region in West Germany,

⁵⁴In Section 6.1.1, we provide a detailed discussion on the comparison between our proposed stage-based identification (SBI) and the SCG approach.

Figure 18: The Effects of the German Reunification on GDP per capita



Notes: Panel (a) shows the GDP per capita of region \mathcal{C} , Hessen, that leads West Germany and a region \mathcal{T} that aggregates the rest of West Germany. We use a Chebyshev smoother (solid lines) of degree 3 with 2 knots. Panel (b) shows the results of our normalization using region \mathcal{T} as reference and mapping the pre-policy outcome paths of region \mathcal{C} onto region \mathcal{T} . Panel (c) zooms the identification window. Panel (d) shows the policy effect $\gamma(s)$ as defined in equation (5). We show the mean, median and 90% confidence interval bands from bootstrapped simulations constructed as described in Section 3.4.2. We further show the policy effects identified without the implementation of the smoothing step as described in Section 3.4.2. We estimate a significant auto-correlation coefficient for the residuals, thus perform block-bootstrap, $\rho_{\mathcal{C}} = 0.78$ $\rho_{\mathcal{T}} = 0.74$, respectively. We choose a block window of 3 years

and an artificially created region Rest of West Germany (RoGER) which is composed of all West German regions excluding Hessen. We label Hessen as region \mathcal{C} and RoGER as region \mathcal{T} .

Normalization. Picking RoGER as reference region, we apply our normalization by mapping the pre-policy GDP per capita in the leading region of Hessen ($y_{\mathcal{C}}(t)$, solid blue circles) onto the preo-policy GDP per capita in RoGER ($y_{\mathcal{T}}(t)$, solid red circles). Our identification strategy

results in a normalized path for Hessen $y_C(t; \psi^*)$ tracked on the stage domain; see panel (b) of Figure 16. The normalizing parameters are $\psi_1 = 1.98$ [1.51, 6.69], $\psi_2 = 1.24$ [1.13, 1.48] and $\psi_0 = 1.00$ [1.00, 1.01]. A result of the normalization is that Hessen leads RoGER. Hence, the normalized path for Hessen provides a counterfactual without policy for RoGER. We find an identification window in terms of stages, $\mathbb{D}(s) = [t_p, s_C(t_p; \psi^*)]$ (shaded pink area), running from approximately seven years in which Hessen is not subject to the German Reunification but RoGER is. Further, note that as a result of our normalization the GDP per capita path is not significantly different between Hessen and RoGER before the German Reunification in the stage domain; see panel (d) of Figure 18.

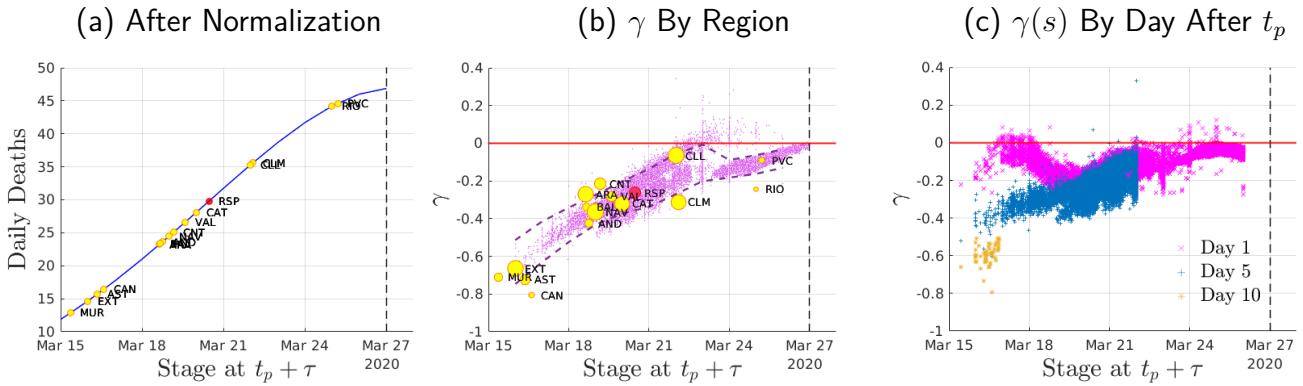
Policy Effects. We zoom the identification window for the GDP per capita in panel (c) and the associated policy effects in panel (d) of Figure 16. We find that the German Reunification significantly reduced the GDP per capita of RoGER by $\gamma = 3.39\%$ compared to the GDP that it would have otherwise attained without the Reunification. The median policy effects are very similar with a 3.30% reduction in GDP per capita. In the previous effects, the window size is restricted to be in the neighborhood of the mean window size—12% distance. However, unrestricting the window size, we also find significant effects of similar size $\gamma = 3.29\%$. Further, without the smoothing step, the German Reunification generates a reduction of 4.82% in the GDP per capita of RoGER, which is not significantly different from our mean bootstrapped effects.

5 Heterogenous Policy Effects by Stage

So far, we have only entertained the idea of policy evaluation with two regions (or sets of regions) in both our models (Section 3) and applications (Section 4). We now expand our analysis to assess the policy effects for multiple regions and stages, including an extension that builds the power set of treated regions.

We use the available outcome paths of multiple regions—which might differ by stage at the time of policy implementation—in order to measure the heterogeneous effects of policy by stage. We focus the analysis on the stay-home nationwide policy implemented in the first wave of Covid-19 in Spain; see Section 4.1. First, we apply our stage-based identification by region. In particular, since Madrid leads all other regions we pick Madrid as reference and separately map each region to Madrid. In this manner all regional mappings are conducted with respect to the same reference path, Madrid. The result of this mapping is in panel (a) of Figure 19. Note that since we picked Madrid as reference region, the calendar time is the stage for Madrid. Then, for example, for the region of Murcia (MUR) policy implementation occurs at a stage corresponding to approximately twelve days earlier than Madrid—that is, at the time of policy implementation

Figure 19: Heterogeneous Effects



Notes: We have a total of 17 region (comunidad autonoma) names: Andalucia (AND), Aragon (ARA), Asturias (AST), Baleares (BAL), Canarias (CAN), Cantabria (CNT), Castilla-La Mancha (CLM), Castilla y Leon (CLL), Catalunya (CAT), Ceuta(CEU), Valencia (VAL), Extremadura (EXT), Galicia (GAL), Madrid (MAD), Melilla (MEL), Murcia (MUR), Navarra (NAV), Pais Vasco (PVC), La Rioja (RIO). We exclude GAL from the analysis due to the fact that we find positive (yet, non-significant) effects of the policy on the flow of deaths. The size of the yellow is the stock of deaths per thousand inhabitants accumulated during the identification window. In panel (b), we report the policy effects γ (see Section 2.2) by region where the (yellow) marker size is the flow of deaths at the time of policy implementation. In addition to the policy effects by region, we also report the policy effects for each hybrid region constructed for each element in the the power set $2^{16} - 1$ of regions (tiny markers) in panel (b). In panel (b), the 90% CI's exclude the top 5% and bottom 5% of policy effects by stage in rolling windows of 2 stages/days. In panel (c) we show the interim effect $\gamma(s)$ (see Section 2.2) by stage for day 1, day 5 and day 10 after policy implementation.

Murcia is at a stage of the epidemic that corresponds to where Madrid was 12 days earlier. In stages, the closest to Madrid in terms of the epidemic is the Basque Country (PVC) that lags Madrid for approximately two days.

Second, we find heterogeneous policy effects by region (yellow markers) that we plot in panel (b) of Figure 19. Clearly, the effects of policy are largest for the regions when the stage at the time of policy implementation differs the most with respect to the stage in which Madrid enters the policy. For example, in the region of Murcia, the policy effect reaches $\gamma = -0.65$ and, hence, the policy prevented 65% of the deaths that would have otherwise occurred in Murcia. In contrast, in the Basque Country, which is closest to Madrid in terms of stages at the time of policy implementation, the policy effects are $\gamma = -0.12$ and, hence, the policy prevented 12% of the deaths that would have otherwise occurred in the Basque Country. To further complete our exploration of the heterogeneous effects of policy by stage, we construct hybrid regions from the power set of the treated regions, i.e. a total of $2^{16} - 1 = 131,072$ hybrid regions, that we separately map using our stage-based identification strategy to Madrid.⁵⁵ We report the policy

⁵⁵Precisely, a hybrid path between region A and region B is constructed as the weighted sum of the flow of deaths per capita in each region.

effects (tiny purple markers) associated with each of these hybrid regions (with 90% confidence intervals) in panel (b) of Figure 19. Overall, we find similar insights as the policy effects are largest in instances where the the stage at the time of policy implementation is farthest away from the stage at which Madrid implemented the policy.

Third, we explore what drives the differences in policy effects by stage. An obvious candidate to determine these differences is the size of the identification window—i.e. the closest a region is to Madrid in terms of stages at the time of policy implementation, the smaller is the identification window. At the same time, differences in policy effects can emerge within the same horizon into the policy within the identification window.

To assess this question, we isolate the effects of policy by the number of stages within the identification window. Here, note that since we picked Madrid as reference, the stage for Madrid is the actual calendar time (i.e. days). In those terms, we find substantial heterogeneity across identification windows by stage. For example, one day into the policy (i.e inside the identification window) at a stage of approximately 10 days before Madrid enters policy (e.g. March 18) the policy effect is below 10%, whereas one day into the policy at a stage of approximately 7 days before Madrid enters policy (e.g. March 12) the policy effect is above 10%, and one day into the policy at a stage of approximately 3 days before Madrid enters policy (e.g. March 24) is again below 10%; see the magenta markers in panel (c) of Figure 19. We also show differences across stages in the policy effects for the cases of five and ten days into the policy. That is, not only the size of the identification window matters (i.e. how close in stages a given region is to Madrid at the time of policy implemenation) but there are also differences in policy effects driven by the differential within-window policy effects across identification windows.

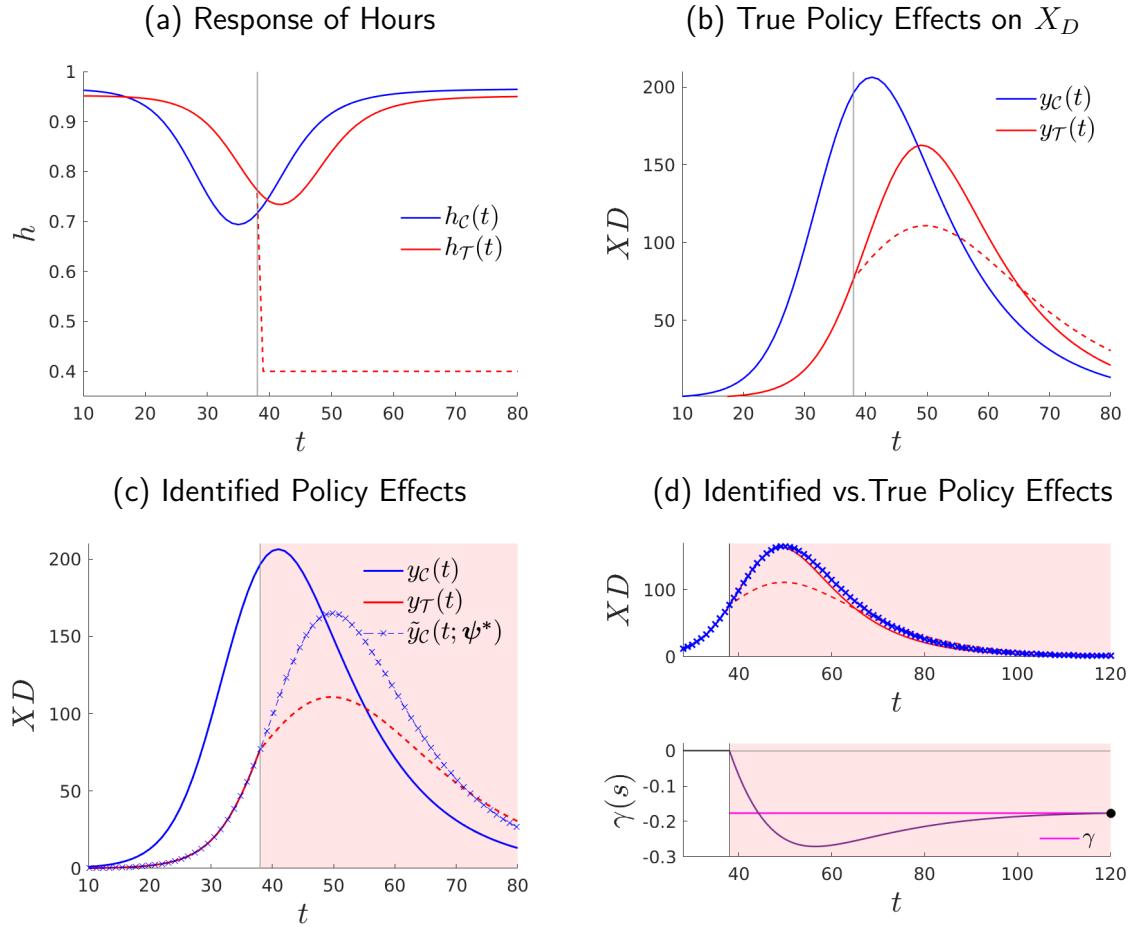
6 Further Discussion

We provide two additional discussions in this Section. First, we show that our method can be applied to non-nationwide policy including cases where there are untreated regions or a staggered rollout of the policy in section 6.1. Interestingly, we show that in the context of a staggered rollout policy, our identification strategy can endogenously uncover a flip between standardly defined control region and treatment region. Second, we show how the presence of spillovers across regions affects our identification of policy effects in Section 6.2.

6.1 Non-Nationwide Policy

Here, we show that our identification strategy works in scenarios where there are regions that never receive the policy intervention or scenarios where there is a staggered rollout of the policy.

Figure 20: Stage-Based Identification of Policy Effects: Untreated Regions



6.1.1 Untreated Regions

Consider a scenario with two regions where one region, e.g. \mathcal{T} , receives the policy intervention at period t_p and the other region, e.g. \mathcal{C} , is never treated. To illustrate this scenario we use our benchmark model with endogenous pandemics as described in Section 3.1.1. Precisely, we introduce the stay-home policy that puts an upper bound on hours worked in region \mathcal{T} , but not in region \mathcal{C} ; see panel (a) of Figure 20. The implications for the outcome of interest, the flow of deaths, is displayed in panel (b) of Figure 20. Note that only the region that is subject to policy, i.e. region \mathcal{T} , shows policy effects. That is, only in region \mathcal{T} the outcome path with policy (dashed red line) differs from the model counterfactual path without policy (solid red line) after the policy is implemented at t_p .

In order to identify the policy effects, we need to modify the set on which the normalization must be conducted. In particular, picking region \mathcal{T} as reference, our normalization parameters are the solution to the minimization of (2) subject to (1) and

$$\mathbb{C}(s) = \begin{cases} [t_0, t_p] & \text{if } r = \mathcal{T} \\ [s_{\mathcal{C}}(t_0; \psi_{-0}), s_{\mathcal{C}}(t_f; \psi_{-0})] & \text{if } r = \mathcal{C} \end{cases} \quad (29)$$

for $r = \{\mathcal{C}, \mathcal{T}\}$ where t_0 denotes the first period of observed data and t_f the last.

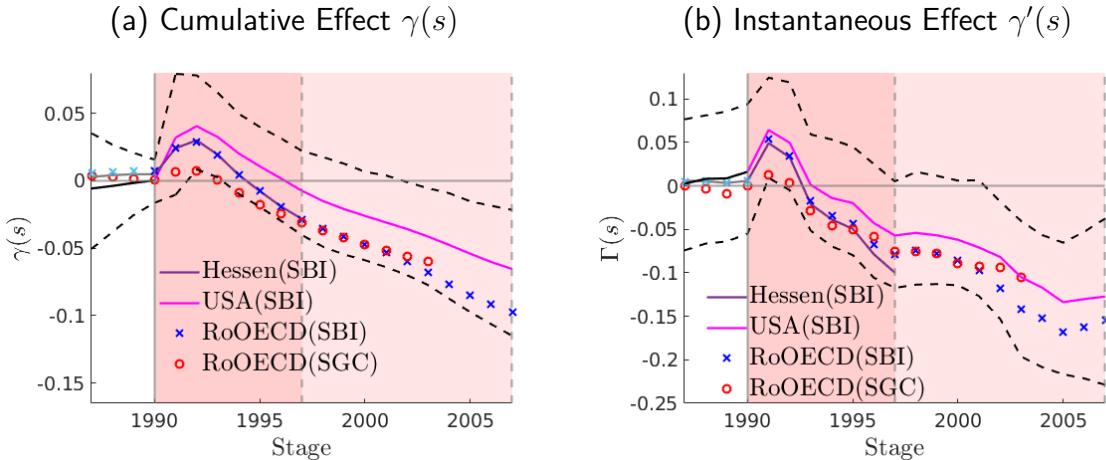
The results of our normalization in panel (c) of Figure 20 highlighting the identification window (pink shaded area). Note that since only region \mathcal{T} is treated, the identification window is,

$$\mathbb{D}(s) = \begin{cases} [t_p, s_{\mathcal{C}}(t_f; \psi_{-0}^*)] & \text{if } s_{\mathcal{C}}(t_f; \psi_{-0}^*) > t_p \\ \emptyset & \text{if } s_{\mathcal{C}}(t_f; \psi_{-0}^*) < t_p \end{cases} \quad (30)$$

Hence, there is a policy effect if and only if the stage of normalized series evaluated at the last period of observed data $y_{\mathcal{C}}(t_f; \psi_{-0}^*)$ —i.e. $s_{\mathcal{C}}(t_f; \psi_{-0}^*)$ —falls beyond the period of policy implementation t_p . Otherwise, the identification window is the empty set because the treated region leads throughout the entire sample. In our model-generated example, the normalization shows that region \mathcal{C} covers stages beyond that of region \mathcal{T} at the time of policy implementation which implies that $\mathbb{D}(s) = [t_p, s_{\mathcal{C}}(t_f; \psi_{-0}^*)]$ and we can assess policy effects. We show the identified policy effects in panel (d) of Figure 20, which we find that overlap with the true (model-generated) policy effects.

We further exemplify how to use stage-based identification in cases in which not all regions are treated by re-conducting our assessment of the German Reunification. Here, we take West Germany as the treated region and use as potential controls the United States and an aggregate consisting of the same set of OECD countries (that excludes Germany) studied in [Abadie et al. \(2014\)](#). Hence, this exercise also serves as means for comparison between our proposed stage-based identification and the synthetic group control approach ([Abadie and Gardeazabal, 2003](#)). In order to apply the stage-based identification we pick West Germany as reference region. Then we conduct our normalization by mapping the GDP per capita path of the U.S. and the OECD aggregate onto the GDP per capita path of West Germany. We show the policy effects that emerge from our stage-based identification in panel (a) of Figure 21. To ease the comparison with [Abadie et al. \(2014\)](#), we also show in panel (b) of Figure 21 the instantaneous policy effects by stage defined as, abusing of notation, $\gamma'(s) = \frac{y_{\mathcal{T}}(s) - \tilde{y}_{\mathcal{C}}(s)}{y_{\mathcal{C}}(s)}$. That is, $\gamma'(s)$ measures the change in GDP per capita of region \mathcal{T} (West Germany) relative to the counterfactual region \mathcal{C} (e.g. the U.S. or the rest of the OECD) at any given stage s due to policy. We also reproduce the results

Figure 21: The Effects of the German Reunification: Stage-Based Identification (SBI) and Synthetic Control Group (SCG)



Notes: The outcome variable is real GDP per capita in USD of 2002. The plotted 90% confidence intervals correspond to the United States.

from using Hessen as leading region in the context of nationwide policy within Western Germany reported in Section 4.3. For inference, we show the 90% confidence intervals associated to the USA constructed as described in Section 3.4.

Our main finding is that the policy effects that emerge from using our stage-based identification either for the U.S. (or the aggregate of the rest of the OECD that excludes Germany) are not significantly different from those obtained using the synthetic control group in [Abadie et al. \(2014\)](#). In particular, we find that the instantaneous policy effects imply a loss of income per capita for West Germany due to the Reunification of 12.73% when compared to the United States and of 15.44 when compared to the rest of the OECD in 2007. These figures are, respectively, 10.51% and 14.13% in 2003 which are not significantly different from the effects of reunification 10.04% obtained in [Abadie et al. \(2014\)](#). Further, we also find that within the shorter window that emerges when the counterfactual from the stage-based identification strategy is Hessen, the results under the alternative counterfactuals are not significantly different from the results obtained with Hessen.

6.1.2 Staggered Rollout

Now, consider a scenario with two regions where one region, e.g. \mathcal{T} , receives the policy intervention at period $t_p^{\mathcal{T}} = t_p$ and the other region, e.g. \mathcal{C} , receives the policy at a later date $t_p^{\mathcal{C}} = t_p + \Delta$

with $\Delta > 0$. This scenario exemplifies the staggered rollout of policy.⁵⁶ In panel (a) of Figure 22, we introduce the same stay-home policy that we discussed in our benchmark example in Section 3.1.1, i.e. an upper bound on hours worked. We show the true (model-generated) policy effects of the rollout policy on the daily deaths by region in panel (b) of Figure 22.

To apply our stage-based identification in this context, we pick as reference region \mathcal{T} and minimize (2) subject to (1) and

$$\mathbb{C}(s) = \{\mathbb{C}_r(s)\} = \begin{cases} [s_r(t_0; \psi_{-0}), s_{\mathcal{T}}(t_p^{\mathcal{T}}; \psi_{-0})] & \text{if } s_{\mathcal{T}}(t_p^{\mathcal{T}}; \psi_{-0}) > s_{\mathcal{C}}(t_p^{\mathcal{C}}; \psi_{-0}) \\ [s_r(t_0; \psi_{-0}), s_{\mathcal{C}}(t_p^{\mathcal{C}}; \psi_{-0})] & \text{if } s_{\mathcal{T}}(t_p^{\mathcal{T}}; \psi_{-0}) > s_{\mathcal{C}}(t_p^{\mathcal{C}}; \psi_{-0}) \end{cases} \quad (31)$$

for $r = \{\mathcal{C}, \mathcal{T}\}$. That is, we generalize the set on which the normalization must be conducted in order to accommodate for differences in the time of policy implementation across regions.

The result of our normalization is in panel (c) of Figure 22. Note that in the case of a staggered rollout, the identification window (pink shaded area) is more general than in the case of the nationwide policy,

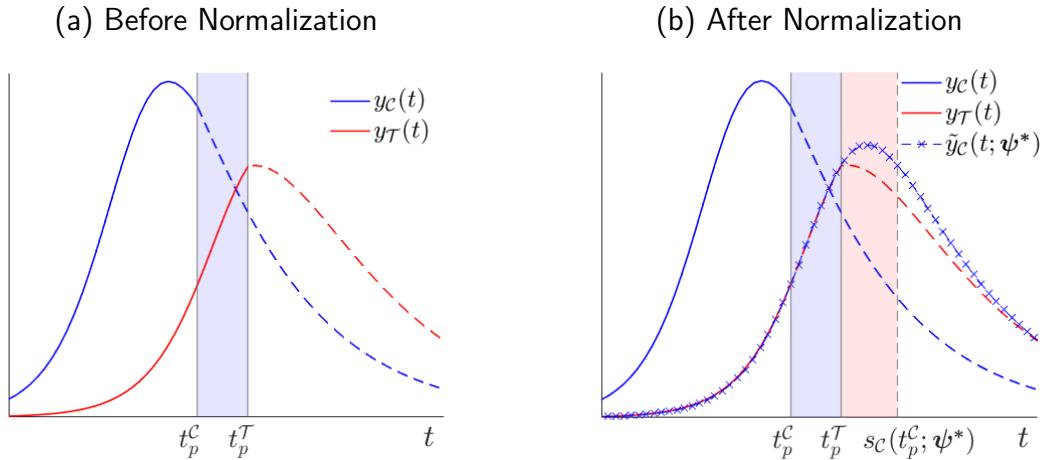
$$\mathbb{D}(s) = \begin{cases} [s_{\mathcal{T}}(t_p^{\mathcal{T}}; \psi_{-0}^*), s_{\mathcal{C}}(t_p^{\mathcal{C}}; \psi_{-0}^*)] & \text{if } s_{\mathcal{T}}(t_p^{\mathcal{T}}; \psi_{-0}^*) \leq s_{\mathcal{C}}(t_p^{\mathcal{C}}; \psi_{-0}^*) \\ [s_{\mathcal{C}}(t_p^{\mathcal{C}}; \psi_{-0}^*), s_{\mathcal{T}}(t_p^{\mathcal{T}}; \psi_{-0}^*)] & \text{if } s_{\mathcal{T}}(t_p^{\mathcal{T}}; \psi_{-0}^*) > s_{\mathcal{C}}(t_p^{\mathcal{C}}; \psi_{-0}^*) \end{cases}. \quad (32)$$

Our normalization delivers $t_p^{\mathcal{T}} \leq s_{\mathcal{C}}(t_p^{\mathcal{C}}; \psi_{-0}^*)$ and, therefore, the identification window is $\mathbb{D}(s) = [t_p^{\mathcal{T}}, s_{\mathcal{C}}(t_p^{\mathcal{C}}; \psi_{-0}^*)]$. That is, the stage at which the policy is implemented in region \mathcal{C} is more advanced than the stage at which the policy is implemented in region \mathcal{T} —even though the policy was implemented earlier in time in region \mathcal{C} than in region \mathcal{T} . We show the identified policy effects in panel (d) of Figure 22 which are aligned with the true (model-generated) policy effects.

We emphasize two main differences of our stage-based identification with respect to the standard empirical methods for the assessment of rollout policies as implemented in our illustration in Figure 22. First, in standard strategies the identification of the effects of rollout policies relies on the heterogeneity in the time of policy implementation across regions. Precisely, in our model-generated example the standard identification strategy corresponds to the window $\mathbb{G}(t) = [t_p^{\mathcal{C}}, t_p^{\mathcal{T}}]$ (shaded purple area) where $t_p^{\mathcal{C}} < t_p^{\mathcal{T}}$ and, hence, region \mathcal{T} serves as control (without policy) for region \mathcal{C} in all periods $t \in \mathbb{G}(t)$. Instead, our identification resides on the stage domain using heterogeneity in the stage at the time of policy implementation—including cases in which policy implementation occurs at different dates for different regions. Specifically, in our illustration our strategy shows that the control and treatment regions that emerge from the stage-based

⁵⁶See the recent analysis in Goodman-Bacon (2021) for a careful assessment of DiD strategies that can be used under different timings and scenarios of staggered rollout policies.

Figure 22: Stage-Based Identification of Policy Effects: Staggered Rollout, With a Control and Treatment Flip



Notes:

identification—which catch the true (model-generated) effects—actually reverts the control and treatment that would be used in standard empirical strategies. Second, a related difference is that the outcome paths of regions \mathcal{C} and \mathcal{T} can be far from showing parallel trends before $\mathbb{G}(t)$ which is a requirement to warrant the policy assessment in the context of standard empirical strategies. Indeed, within the realm of empirical strategies that use heterogeneity in the time of policy implementation, there is an exciting debate regarding with more flexible forms of parallel trend assumptions that can overcome some of these obstacles (e.g. [Callaway and Sant'Anna, 2021](#); [Rambachan and Roth, 2021](#)). In our illustration, note that the outcome path of region \mathcal{T} is monotonically increasing before $\mathbb{G}(t)$, whereas the outcome path of region \mathcal{C} shows a non-monotonicity by first rising, reaching a peak and then already declining before entering $\mathbb{G}(t)$. This behavior of the outcome paths before policy implementation—in calendar time—makes standard strategies largely unworkable. However, as we show, our stage-based identification—which aims to minimize the regional differences of the pre-policy outcome paths in the stage domain—does not rely upon the parallel trends requirement in order to provide a credible policy assessment—credible in that our method delivers normalized regional pre-policy outcome paths in the stage domain that are not significantly different from each other and in that our method performs well in terms of statistical inference (see Section 3.4).

6.2 Spillovers

So far, we have focused on policy effects in environments where regional outcomes do not affect each other. Here, we assess the performance of our method in the presence of spillovers using our benchmark econ-epi model in Section 3.1.1 where a region \mathcal{C} leads a region \mathcal{T} . In that context, we introduce the idea that the deaths of the leading region affect the beliefs of the probability of infection in the non-leading region. This implies that the fact that the epidemic occurs in the leading region endogenously potentially alters the behavior of the non-leading region. We formalize this idea by setting the beliefs of the non-leading region in our model as,

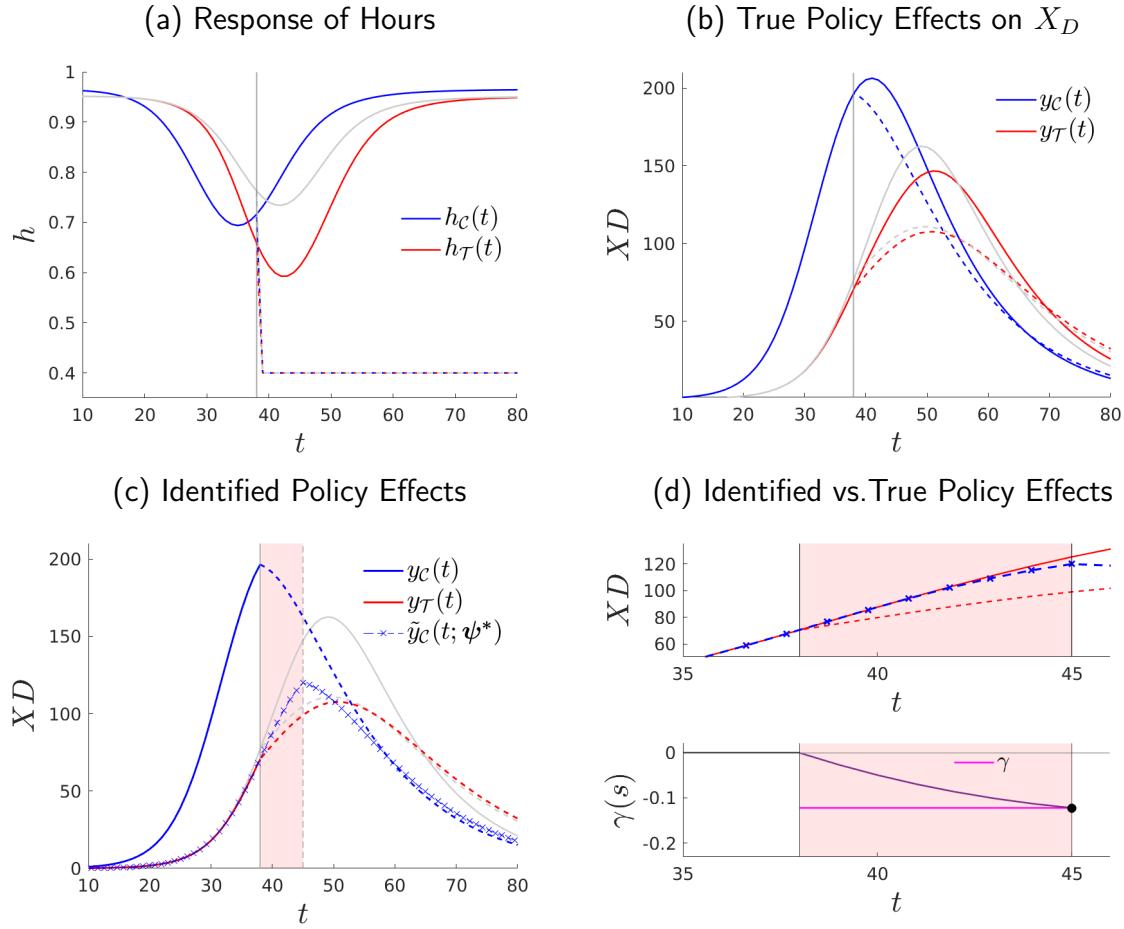
$$\xi_{\mathcal{P}, \mathcal{T}} = \tilde{\xi}_{\mathcal{P}, \mathcal{T}} \left(1 + \left(\exp \left(- \frac{X D_{\mathcal{C}, t-1}}{N_{\mathcal{C}, t-1}} \right)^{-\sigma} \right) \right) \quad (33)$$

with $\sigma > 0$. That is, the beliefs on the probability of infection in region \mathcal{T} are now affected by deaths from regions \mathcal{C} . In particular, the beliefs on infection increase by an amount equal to the CDF of the Frechet distribution of region \mathcal{C} 's deaths per capita, $\frac{X D_{\mathcal{C}, t-1}}{N_{\mathcal{C}, t-1}}$.

We find that in the presence of the spillover, the hours in region \mathcal{T} (red solid line) react earlier and drop by a larger amount than in the benchmark economy without the spillover (gray solid line), see panel (a) in Figure 23. Precisely, we use as region without spillovers the benchmark \mathcal{T} region in Section 3.1.1. The larger decrease in the response of hours for region \mathcal{T} in the presence of spillovers has consequences for deaths in region \mathcal{T} , see panel (b) in Figure 23. In particular, with spillovers, the growth of the flow of deaths slows down reaching a later and lower amount of deaths (red dashed line) than without spillovers (gray dashed line). Note that this is also the case without policy where the spillovers alleviate the pandemic in region \mathcal{T} (solid red line) in comparison with region \mathcal{T} without spillovers (gray solid line). For this reason, the true policy effects potentially differ by the strength of the spillover.

Can our method recover the true policy effects in a context with spillovers? In panel (c) of Figure 23, we show the identified policy effects resulting from our method using region \mathcal{T} as reference region. We find that the identified policy effects recover the true policy effects closely; see the zoomed panel (d) in Figure 23. That is, our method recovers the true policy effects in contexts with spillovers as long as the spillover effects do not make the regional outcome paths too different—in the sense of our Monte Carlo experiment in Section 3. Finally, consider another non-leading region that although it is not directly subject to policy it is indirectly subject to policy implemented in the leading region to the extend that there exist spillovers. In this case, our method recovers the effects that the policy implemented in the leading region has on the outcome path of non-leading region solely through the spillover effects. Further, note that this

Figure 23: Stage-Based Identification of Policy Effects: Spillovers



Notes: Where $\bar{h} = 0.4$, $t_p = 38$, $t_f = 250$, $\tilde{\xi}_{P,r} = 0.2$, $\sigma = 30$, $\gamma = -12.21\%$, $\epsilon(\gamma) = 9.32\%$.

environment—with one leading region subject to policy and two non-leading regions subject to spillovers that only differ in that one is subject to policy and the other one is not—is sufficient to separately identify the effects of policy implemented in the same non-leading region versus the effects of spillovers which are also affected by policy implemented in the leading region; see our Appendix. However, in reality, it is likely that spillovers generate feedback effects across all regions (e.g. Chodorow-Reich, 2020), which entails an analysis that is beyond the scope of this paper.

7 Conclusion

We develop a novel empirical methodology to evaluate policy in contexts where all regions are intervened at the same time, i.e. nationwide policy. Our method consists of a normalization that

maps the time-path of regional outcomes onto a reference path—using only pre-policy data. This entails a time-to-stage transformation where the *stage* of a regional outcome is defined as its location on the support of the reference path. Since the normalized regions can differ by *stage* at any point in time, our normalization uncovers heterogeneity in the *stage* at the time of policy implementation—even in instances where the implementation occurs at the same time across regions. We use this *stage* variation at the time of policy implementation to identify the policy effects: a *stage-leading* region delivers the counterfactual path inside a window in which non-leading regions are subject to policy when the leading region is not. We show how our method works also in the evaluation of non-nationwide policy with untreated regions and with staggered rollouts. We show several empirical applications including the effects of public health stay-home policies (i.e. the national lockdown against Covid-19 in Spain), the effects of oral contraceptives (i.e. the FDA nationwide approval of oral contraceptives in 1960 in the U.S.) on women’s fertility and college education choices and the effects of growth policy (e.g. German Reunification).

References

- Abadie, A. (2005). Semiparametric Difference-in-Differences Estimators. *Review of Economic Studies*, 72(1):1–19.
- Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California’s Tobacco Control Program. *Journal of the American Statistical Association*, 105(490):493–505.
- Abadie, A., Diamond, A., and Hainmueller, J. (2014). Comparative Politics and the Synthetic Control Method. *American Journal of Political Science*, 59(2):495–510.
- Abadie, A. and Gardeazabal, J. (2003). The Economic Costs of Conflict: A Case Study of the Basque Country. *American Economic Review*, 93(1):113–132.
- Aleman, C., Iorio, D., and Santaularia, R. (2022). A quantitative theory of the hiv epidemic: Education, risky sex and asymmetric learning. Working paper, Barcelona School of Economics.
- Angrist, J. D. and Krueger, A. B. (1999). Empirical strategies in labor economics. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 3 of *Handbook of Labor Economics*, chapter 23, pages 1277–1366. Elsevier.
- Athey, S. and Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*, 31(2):3–32.

- Athey, S. and Imbens, G. W. (2021). Design-based analysis in difference-in-differences settings with staggered adoption. *Journal of Econometrics*.
- Atkeson, A. (2020). What will be the economic impact of covid-19 in the us? rough estimates of disease scenarios. Working Paper 26867, National Bureau of Economic Research.
- Bailey, M. J. (2006). More Power to the Pill: The Impact of Contraceptive Freedom on Women's Life Cycle Labor Supply*. *The Quarterly Journal of Economics*, 121(1):289–320.
- Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How Much Should We Trust Differences-In-Differences Estimates? *The Quarterly Journal of Economics*, 119(1):249–275.
- Blundell, R. and MacCurdy, T. (1999). Labor supply: A review of alternative approaches. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 3 of *Handbook of Labor Economics*, chapter 27, pages 1559–1695. Elsevier.
- Borusyak, K., Jaravel, X., and Spiess, J. (2021). Revisiting Event Study Designs: Robust and Efficient Estimation. Working papers, University College London.
- Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-Differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230.
- Card, D. (1990). The impact of the mariel boatlift on the miami labor market. *ILR Review*, 43(2):245–257.
- Card, D. (2022). Design-based research in empirical microeconomics. *American Economic Review*, 112(6):1773–81.
- Card, D. and Krueger, A. B. (2000). Minimum wages and employment: A case study of the fast-food industry in new jersey and pennsylvania: Reply. *The American Economic Review*, 90(5):1397–1420.
- Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *The Annals of Statistics*, 14(3):1171–1179.
- Cervellati, M. and Sunde, U. (2015). The economic and demographic transition, mortality, and comparative development. *American Economic Journal: Macroeconomics*, 7(3):189–225.
- Chodorow-Reich, G. (2020). Regional data in macroeconomics: Some advice for practitioners. *Journal of Economic Dynamics and Control*, 115.
- Doudchenko, N. and Imbens, G. W. (2017). Balancing, regression, difference-in-differences and synthetic control methods: A synthesis.

- Fang, H., Wang, L., and Yang, Y. (2020). Human mobility restrictions and the spread of the novel coronavirus (2019-ncov) in china. *Journal of Public Economics*, 191:104272.
- Fernández-Villaverde, J., Greenwood, J., and Guner, N. (2014). From Shame To Game In One Hundred Years: An Economic Model Of The Rise In Premarital Sex And Its De-Stigmatization. *Journal of the European Economic Association*, 12(1):25–61.
- Galor, O. and Weil, D. N. (2000). Population, technology, and growth: From malthusian stagnation to the demographic transition and beyond. *American Economic Review*, 90(4):806–828.
- Glogowsky, U., Hansen, E., and Schächtele, S. (2021). How effective are social distancing policies? evidence on the fight against covid-19. *PLOS ONE*, 16(9):1–12.
- Goldin, C. and Katz, L. F. (2002). The Power of the Pill: Oral Contraceptives and Women's Career and Marriage Decisions. *Journal of Political Economy*, 110(4):730–770.
- Gollin, D., Parente, S., and Rogerson, R. (2002). The role of agriculture in development. *American Economic Review*, 92(2):160–164.
- Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277.
- Greenwood, J. and Guner, N. (2010). Social change: The sexual revolution. *International Economic Review*, 51(4):893–923.
- Greenwood, J., Seshadri, A., and Vandenbroucke, G. (2005). The Baby Boom and Baby Bust. *American Economic Review*, 95(1):183–207.
- Hansen, G. D. and Prescott, E. C. (2002). Malthus to solow. *American Economic Review*, 92(4):1205–1217.
- Hausman, J. (2001). Mismeasured variables in econometric analysis: Problems from the right and problems from the left. *The Journal of Economic Perspectives*, 15(4):57–67.
- Heckman, J. J., Ichimura, H., and Todd, P. (1998). Matching As An Econometric Evaluation Estimator. *Review of Economic Studies*, 65(2):261–294.
- Herrendorf, B., Rogerson, R., and Valentinyi, A. (2014). Growth and structural transformation. In *Handbook of Economic Growth*, volume 2, chapter 06, pages 855–941. Elsevier, 1 edition.
- Hyslop, D. R. and Imbens, G. W. (2000). Bias from Classical and Other Forms of Measurement Error. NBER Technical Working Papers 0257, National Bureau of Economic Research, Inc.

Imbens, G. W. and Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press.

Iorio, D. and Santaëulàlia-Llopis, R. (2010). Education, HIV Status, and Risky Sexual Behavior: How Much Does the Stage of the HIV Epidemic Matter? Working papers, washington university in st. louis.

Iorio, D. and Santaëulàlia-Llopis, R. (2016). Education, HIV Status, and Risky Sexual Behavior: How Much Does the Stage of the HIV Epidemic Matter? Working papers, Barcelona Graduate School of Economics.

Liu, L., Moon, H. R., and Schorfheide, F. (2021). Panel forecasts of country-level Covid-19 infections. *Journal of Econometrics*, 220(1):2–22.

Rambachan, A. and Roth, J. (2021). A More Credible Approach to Parallel Trends. Technical report.

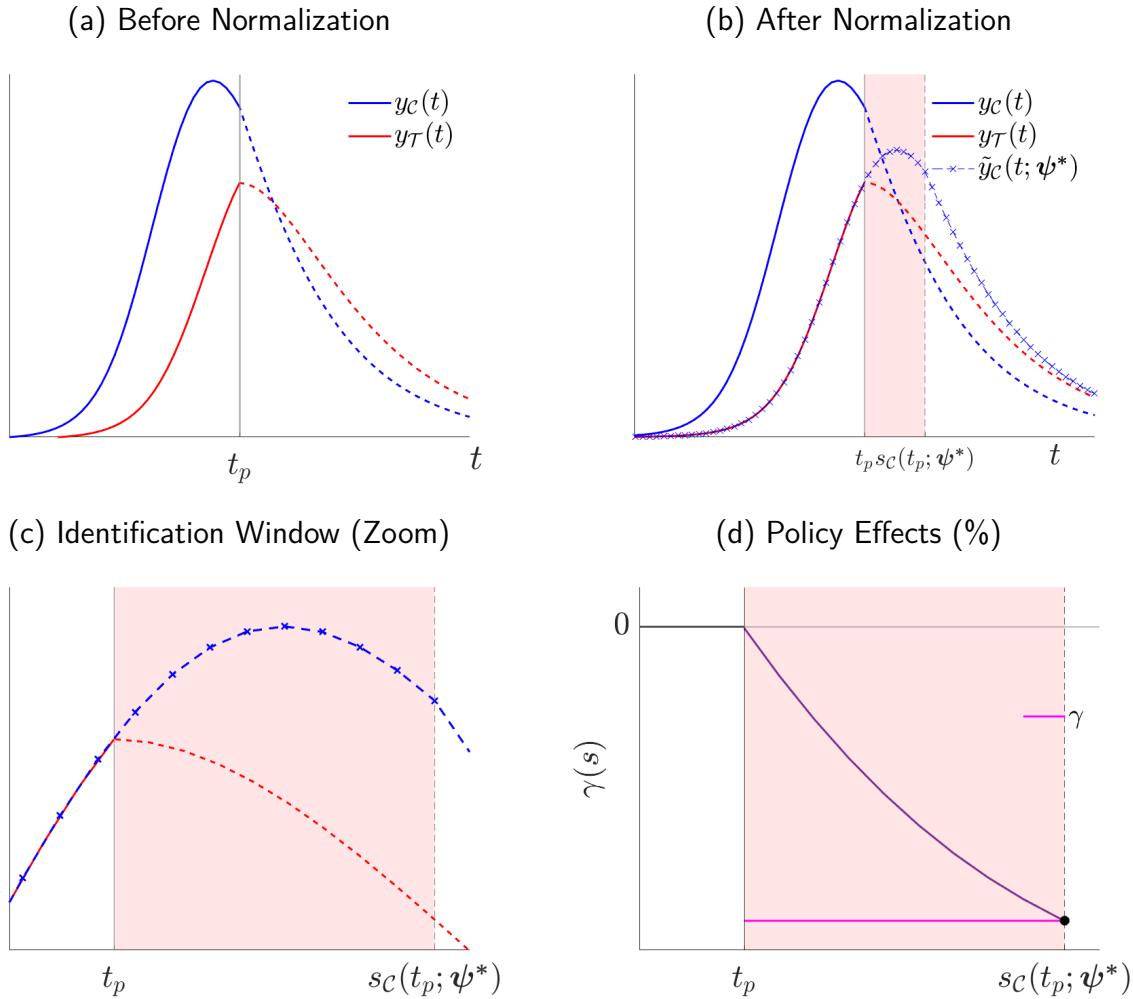
A Further Examples: Policy After the Peak

Here, we study cases in which the policy is implemented after the peak in addition to our before-the-peak benchmark in Section 2.

A.1 Policy After the Peak: One Region

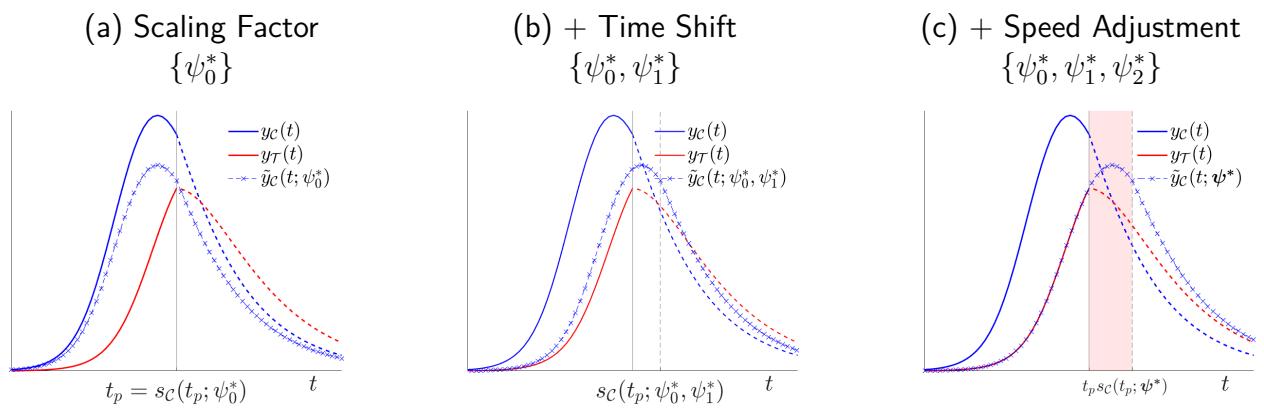
Here we discuss a scenario in which the nationwide policy is implemented before the peak in region \mathcal{T} and after the peak in region \mathcal{C} ; see panel (a) of Figure 24. Note that in this case the outcome paths of region \mathcal{T} and region \mathcal{C} show clearly differentiated (not parallel) trends with region \mathcal{C} monotonically increasing and region \mathcal{T} displaying a non-monotonic path. This makes standard empirical strategies such unworkable. In contrast, applying our method mapping region \mathcal{C} onto region \mathcal{T} using pre-policy data only generates the normalized outcome path $\tilde{y}_{\mathcal{C}}(t; \psi^*)$; see panel (b) in Figure 24. Hence, our methodology opens a window in stages between t_p and $s_{\mathcal{C}}(t; \psi^*)$ in which the effects of policy are identified. We zoom the identification window in panel (c) of Figure 24 and the associated policy effects in panel (d) of Figure 24. In Figure 25, we further unpack the contribution of each normalization coefficient in generating the normalized path.

Figure 24: A Stage-Based Identification of Policy Effects: A Nationwide Policy



Notes: See the notes in Figure 3.

Figure 25: Unpacking the Normalization



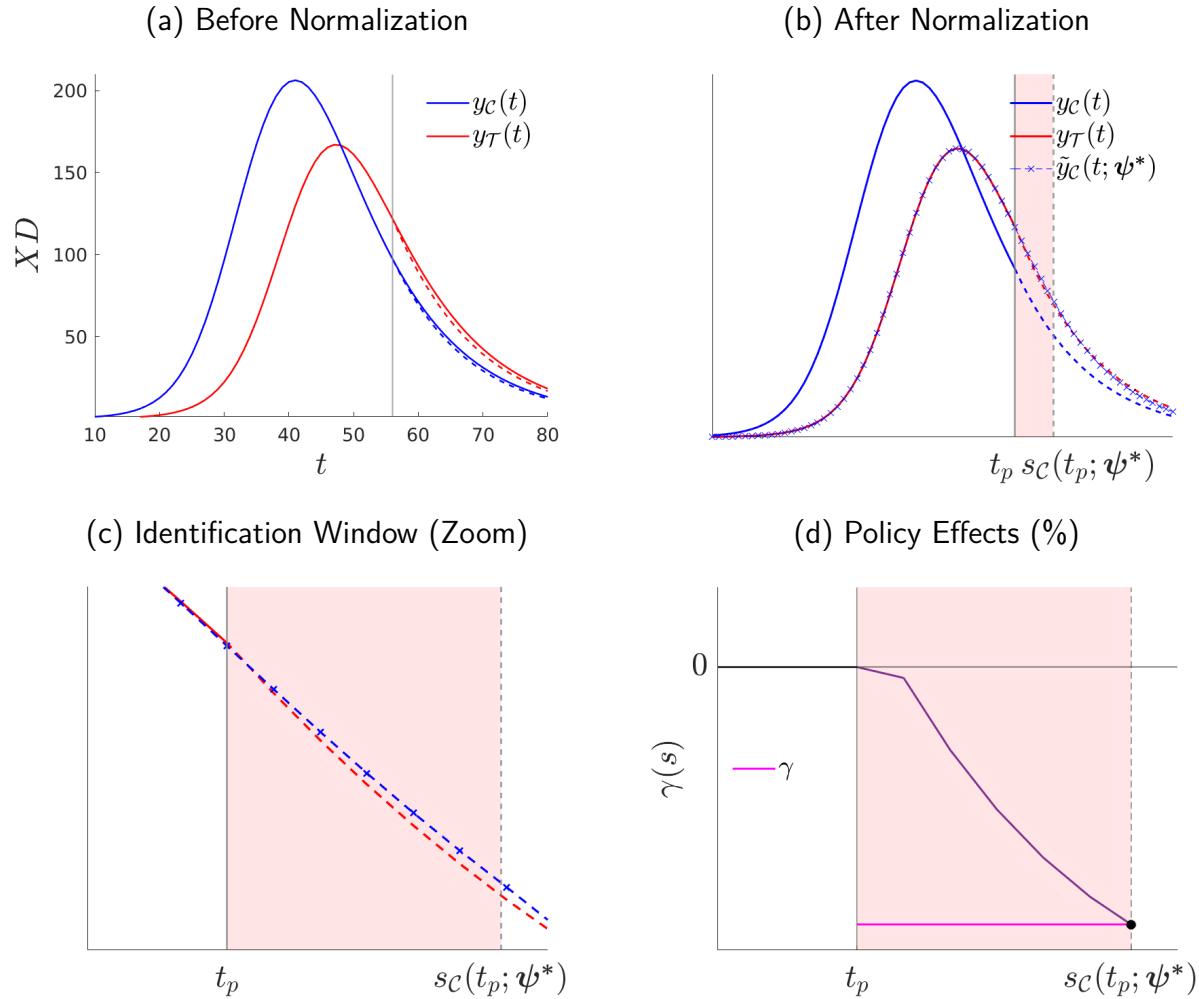
Notes: See the notes in Figure 4.

A.2 Policy After the Peak: Two Regions

Here we discuss a scenario in which the nationwide policy is implemented before the peak in region \mathcal{T} and also in region \mathcal{C} ; see panel (a) of Figure 26. Applying our method mapping region \mathcal{C} onto region \mathcal{T} using pre-policy data only generates the normalized outcome path $\tilde{y}_{\mathcal{C}}(t; \psi^*)$; see panel (b) in Figure 26. Hence, our methodology opens a window in stages between t_p and $s_{\mathcal{C}}(t; \psi^*)$ in which the effects of policy are identified. We zoom the identification window in panel (c) of Figure 26 and the associated policy effects in panel (d) of Figure 26.

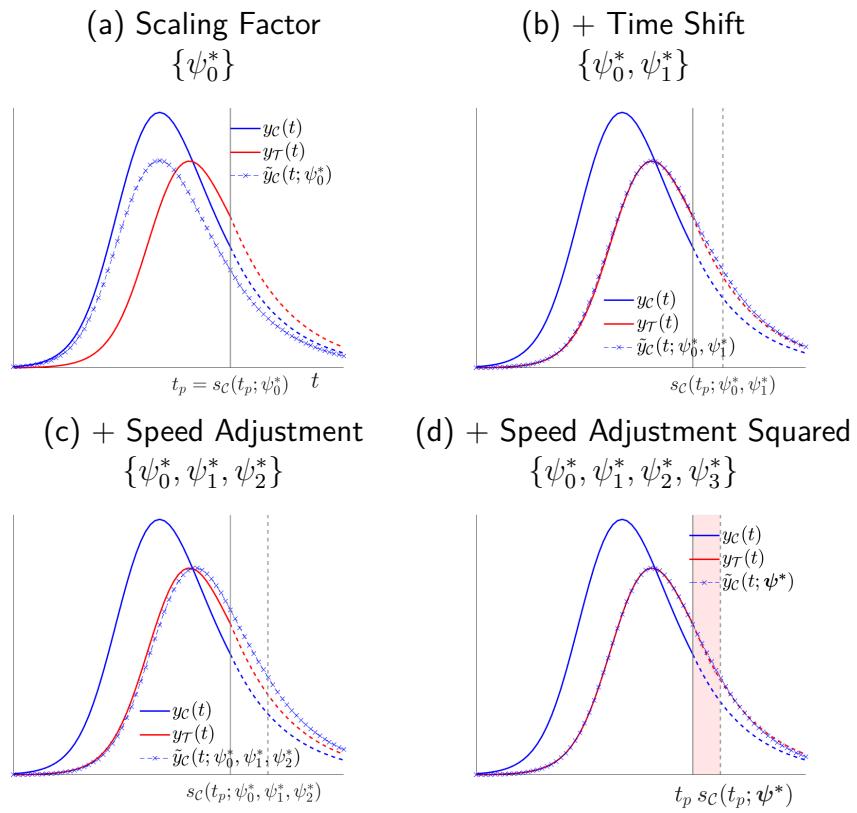
In Figure 27, we further unpack the contribution of each normalization coefficient in generating the normalized path. Note that in this scenario, since the outcome paths of the two regions is already affected by the potential asymmetry in which the outcome paths increase and decrease, the additional parameter ψ_3 that asymmetrically shapes time into stages plays a role.

Figure 26: A Stage-Based Identification of Policy Effects: A Nationwide Policy



Notes: See the notes in Figure 3.

Figure 27: Unpacking the Normalization



Notes: See the notes in Figure 4.

B Tracking Population in the Econ-Epi Model: Further Details

In order to keep track of the population in the econ-epi model in Section 3.1.1, we use a population projection matrix. That is, the evolution of the population structure in our econ-epi model follows this projection matrix model,

$$\begin{bmatrix} S_{t+1} \\ I_{t+1} \\ R_{t+1} \\ D_{t+1} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 - \gamma & 0 & 0 \\ 0 & (1 - \zeta)\gamma & 1 & 0 \\ 0 & \zeta\gamma & 0 & 1 \end{bmatrix}}_{\text{Death \& Recovery Process}} \underbrace{\begin{bmatrix} 1 - \lambda_P(h_t)\beta \frac{I_t}{N_t} & 0 & 0 & 0 \\ \lambda_P(h_t)\beta \frac{I_t}{N_t} & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\text{Infection Process}} \begin{bmatrix} S_t \\ I_t \\ R_t \\ D_t \end{bmatrix} \quad (34)$$

where note that infections occur before death or recovery. Hence, within the same period new infections will be subject to death and recovery. First, after infection, the population structure is:

$$\begin{bmatrix} \tilde{S}_t \\ \tilde{I}_t \\ \tilde{R}_t \\ \tilde{D}_t \end{bmatrix} = \underbrace{\begin{bmatrix} 1 - \lambda_P(h_t)\beta \frac{I_t}{N_t} & 0 & 0 & 0 \\ \lambda_P(h_t)\beta \frac{I_t}{N_t} & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\text{Infection Process}} \begin{bmatrix} S_t \\ I_t \\ R_t \\ D_t \end{bmatrix} = \begin{bmatrix} -\lambda_P(h_t)\beta \frac{I_t}{N_t} S_t + S_t \\ \lambda_P(h_t)\beta \frac{I_t}{N_t} S_t + I_t \\ R_t \\ D_t \end{bmatrix}$$

Second, after death and recovery:

$$\begin{bmatrix} S_{t+1} \\ I_{t+1} \\ R_{t+1} \\ D_{t+1} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 - \gamma & 0 & 0 \\ 0 & (1 - \zeta)\gamma & 1 & 0 \\ 0 & \zeta\gamma & 0 & 1 \end{bmatrix}}_{\text{Death \& Recovery Process}} \begin{bmatrix} \tilde{S}_t \\ \tilde{I}_t \\ \tilde{R}_t \\ \tilde{D}_t \end{bmatrix} = \begin{bmatrix} \tilde{S}_t \\ (1 - \gamma)\tilde{I}_t \\ (1 - \zeta)\gamma\tilde{I}_t + R_t \\ \zeta\gamma\tilde{I}_t + D_t \end{bmatrix} = \begin{bmatrix} -\lambda_P(h_t)\beta \frac{I_t}{N_t} S_t + S_t \\ (1 - \gamma) \left(\lambda_P(h_t)\beta \frac{I_t}{N_t} S_t + I_t \right) \\ (1 - \zeta)\gamma\tilde{I}_t + R_t \\ \zeta\gamma\tilde{I}_t + D_t \end{bmatrix}$$

Then, using $X_{G,t} = G_{t+1} - G_t$ for $G = \{S, I, R, D\}$ we can write (34) as:

$$\begin{bmatrix} X_{S,t} \\ X_{I,t} \\ X_{R,t} \\ X_{D,t} \end{bmatrix} = \begin{bmatrix} -\lambda_P(h_t)\beta \frac{I_t}{N_t} S_t \\ (1 - \gamma)\lambda_P(h_t)\beta \frac{I_t}{N_t} S_t - \gamma I_t \\ (1 - \zeta)\gamma\tilde{I}_t \\ \zeta\gamma\tilde{I}_t \end{bmatrix} = \begin{bmatrix} -\lambda_P(h_t)\beta \frac{I_t}{N_t} S_t \\ (1 - \gamma)\lambda_P(h_t)\beta \frac{I_t}{N_t} S_t - \gamma I_t \\ (1 - \zeta)\frac{\gamma}{1 - \gamma} I_{t+1} \\ \zeta\frac{\gamma}{1 - \gamma} I_{t+1} \end{bmatrix} \quad (35)$$

where the stock of infected (before death and recovery) is $\tilde{I}_t = \lambda_P(h_t)\beta \frac{I_t}{N_t} S_t + I_t$, and note that $\tilde{I}_t = \frac{I_{t+1}}{1 - \gamma}$.⁵⁷

⁵⁷Or, capturing the new infections as $NI(h_t) = \lambda_P(h_t)\beta \frac{I_t}{N_t} S_t$. we can also write:

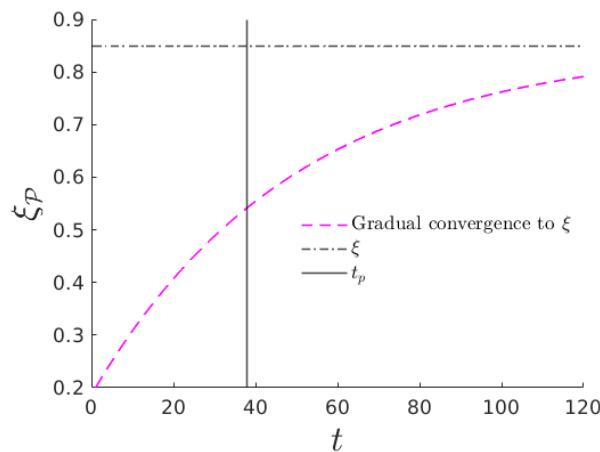
$$\begin{bmatrix} X_{S,t} \\ X_{I,t} \\ X_{R,t} \\ X_{D,t} \end{bmatrix} = \begin{bmatrix} -\lambda_P(h_t)\beta \frac{I_t}{N_t} S_t \\ (1 - \gamma)\lambda_P(h_t)\beta \frac{I_t}{N_t} S_t - \gamma I_t \\ (1 - \zeta)\gamma \left(\lambda_P(h_t)\beta \frac{I_t}{N_t} S_t + I_t \right) \\ \zeta\gamma \left(\lambda_P(h_t)\beta \frac{I_t}{N_t} S_t + I_t \right) \end{bmatrix} = \begin{bmatrix} -NI(h_t) \\ (1 - \gamma)NI(h_t) - \gamma I_t \\ (1 - \zeta)\gamma(NI(h_t) + I_t) \\ \zeta\gamma(NI(h_t) + I_t) \end{bmatrix}$$

C More on Confounding Factors

C.1 More on Time-Varying Latent Heterogeneity

In Figure 28, we show the assumed path for the beliefs on $\xi_{\mathcal{P}}$ in region \mathcal{T} that generates the time-varying latent heterogeneity studied in Section 3.3.1. That is, we assumed that beliefs on the probability of infection conditional on hours work $\xi_{\mathcal{P}}$ (magenta dashed line) converge from below to the true probabilities ξ (dashed gray line). That is, a structural parameter $\xi_{\mathcal{P}}$ that is unobserved to the policy evaluator evolves over time before and after policy implementation.

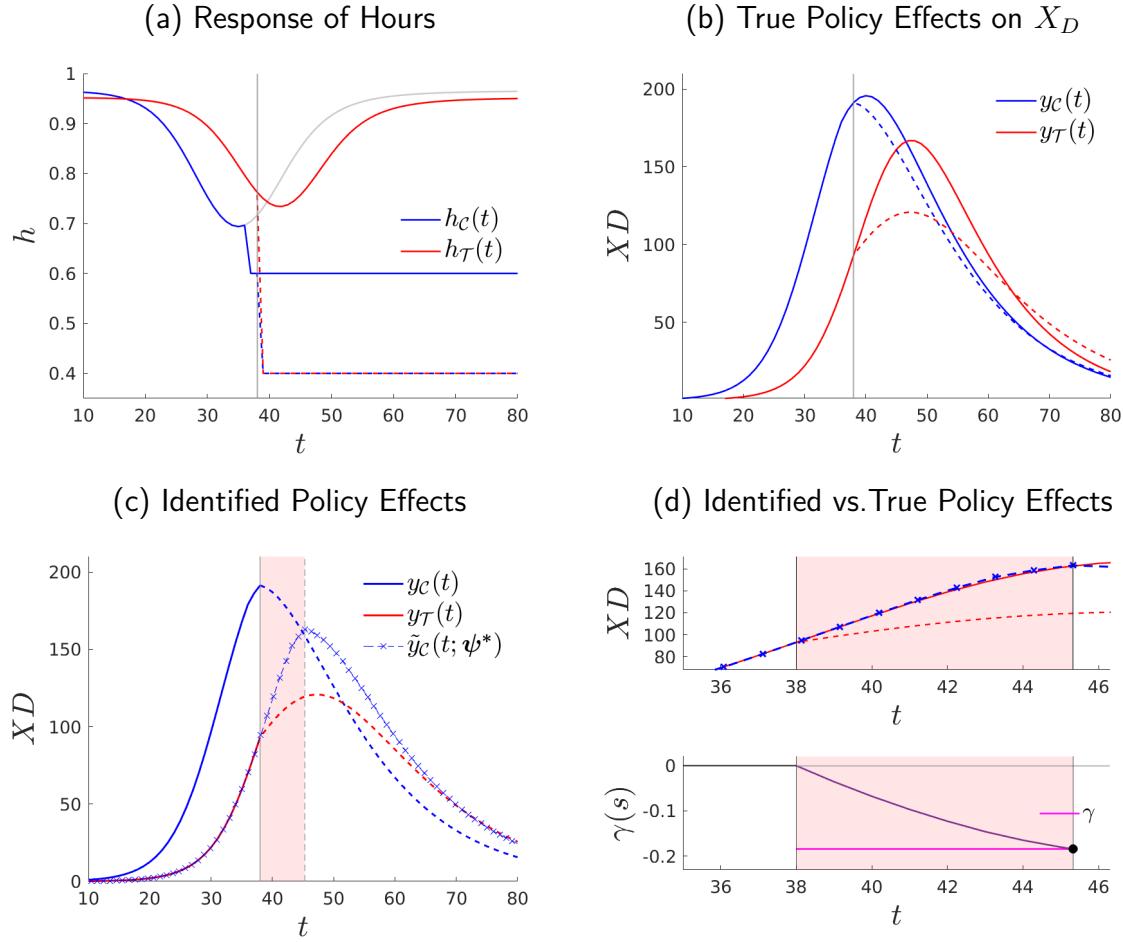
Figure 28: Time-Varying Latent Heterogeneity: Assumed Path of Beliefs $\xi_{\mathcal{P}}$ for region \mathcal{T}



C.2 More on Confounding Policies

Here, we show how our method performs where the confounding policy that happens before the stay-home policy is implemented occurs in region \mathcal{C} and not in region \mathcal{T} , i.e. the opposite case studied in our Section 3.3.2. In this case, we find that the policy effect is recovered with an error of 2.34%. Again, we can make this error larger if the confounding policy drives the outcome path of region \mathcal{C} further away from the outcome path of region \mathcal{T} .

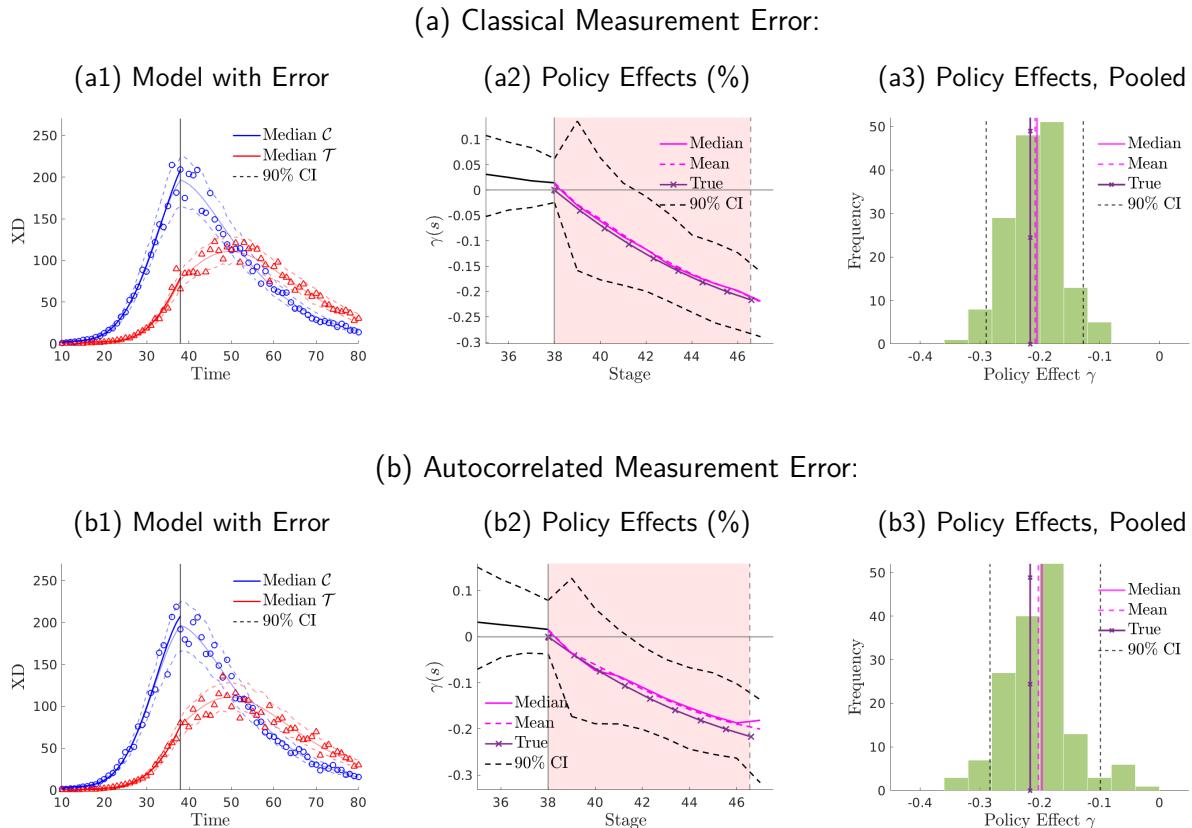
Figure 29: Stage-Based Identification of Policy Effects: With Confounding Policy in \mathcal{C}



D Further Details on Inference with an Stochastic Component

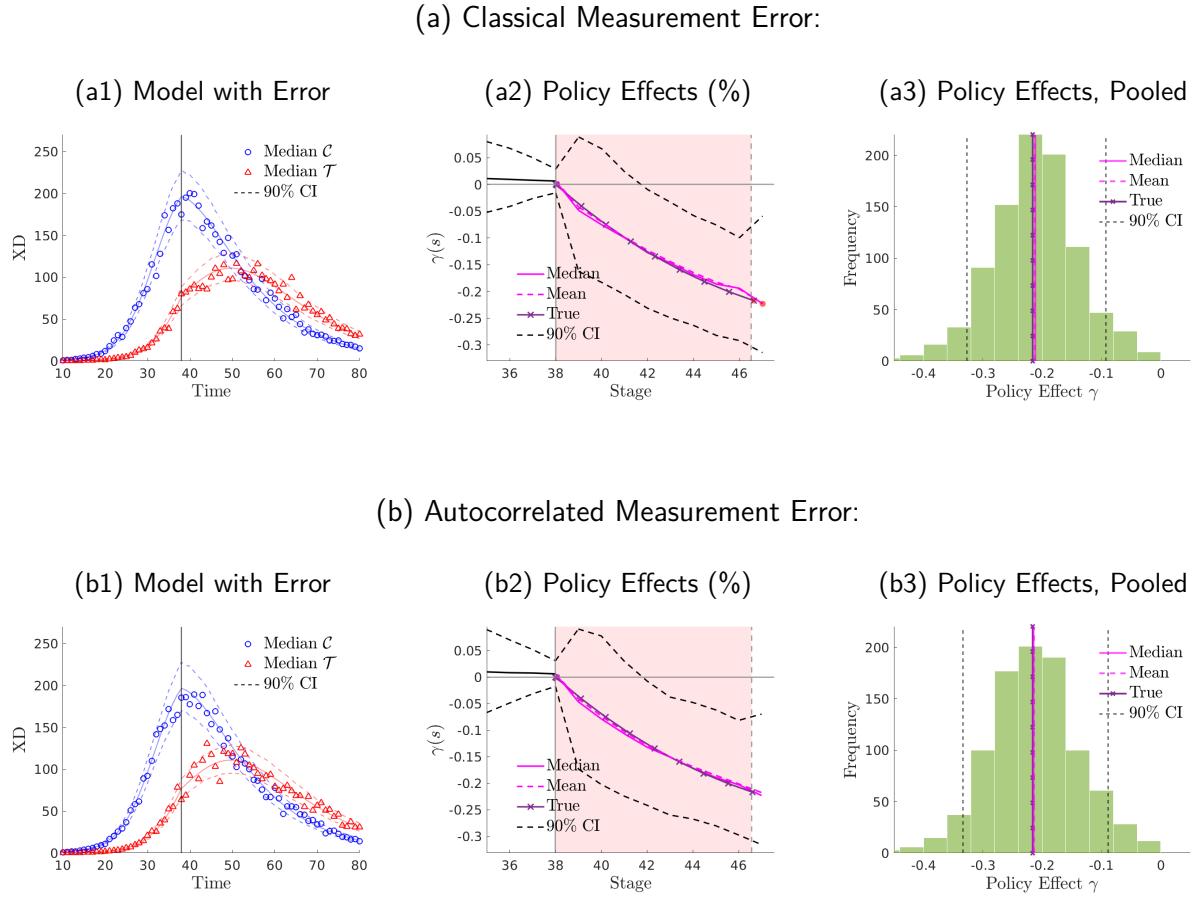
An altogether alternative way to conduct inference with the recovered estimates for the error terms $u_r(t)$ is to estimate the sample variance of the errors, i.e. $\widehat{\sigma}_r$. Then, under a normality assumption on the error term in (28), we simulate $Q = 1,000$ paths of errors and, hence, the same number of pre-policy outcome paths onto which we apply the smoothing step in order to recover a simulation-specific estimand $\widehat{y}_{r,q}(t)$. Since the estimand $\widehat{y}_{r,q}(t)$ differs by simulation $q \in Q$, each simulation delivers an stage-based identified policy effect, γ_q . We show the results of this different inference in Figure 30. Overall, we find similar insights with an identified mean policy effect of 21.12% [14.41,28.23] and 19.61% [12.32,29.35] with classical ME and with auto-correlated ME, respectively. The recovered policy effect is not significantly different from the true (model-generated) policy effect. In Figure 31, we show the policy effects from directly using the $Q = 1,000$ simulations of data $\widehat{y}_{r,q}(t)$, that is, without applying the smoothing step. The identified mean policy effect obtained without the smoothing step is 20.92% [7.88,30.32] and 21.55% [7.51,30.81] with classical ME and with auto-correlated ME respectively.

Figure 30: Stage-Based Identification of Model-Generated Policy Effects: Inference



Notes: We use the benchmark calibration in Section 3.1.1. the top panels (a), we introduce classical measurement error in our model with $\{\sigma_C^2, \sigma_T^2\} = \{0.01, 0.01\}$. In the bottom panels (b), we introduce non-classical measurement error with $\{\rho_C, \rho_T\} = \{0.13, 0.13\}$ and $\{\sigma_C^2, \sigma_T^2\} = \{0.01, 0.01\}$.

Figure 31: Stage-Based Identification of Model-Generated Policy Effects: Inference, No Smoother



Notes: We use the benchmark calibration in Section 3.1.1. the top panels (a), we introduce classical measurement error in our model with $\{\sigma_C^2, \sigma_T^2\} = \{0.008, 0.008\}$. In the bottom panels (b), we introduce non-classical measurement error with $\{\rho_C, \rho_T\} = \{0.13, 0.13\}$ and $\{\sigma_C^2, \sigma_T^2\} = \{0.008, 0.008\}$.

E Further Results and Robustness on our Applications

First, we assess the robustness of our results when we do incorporate the smoothing step to our benchmark SBI algorithm in Section E.1. Second, we conduct further inference using a placebo diagnosis in each of our applications in Section E.2.

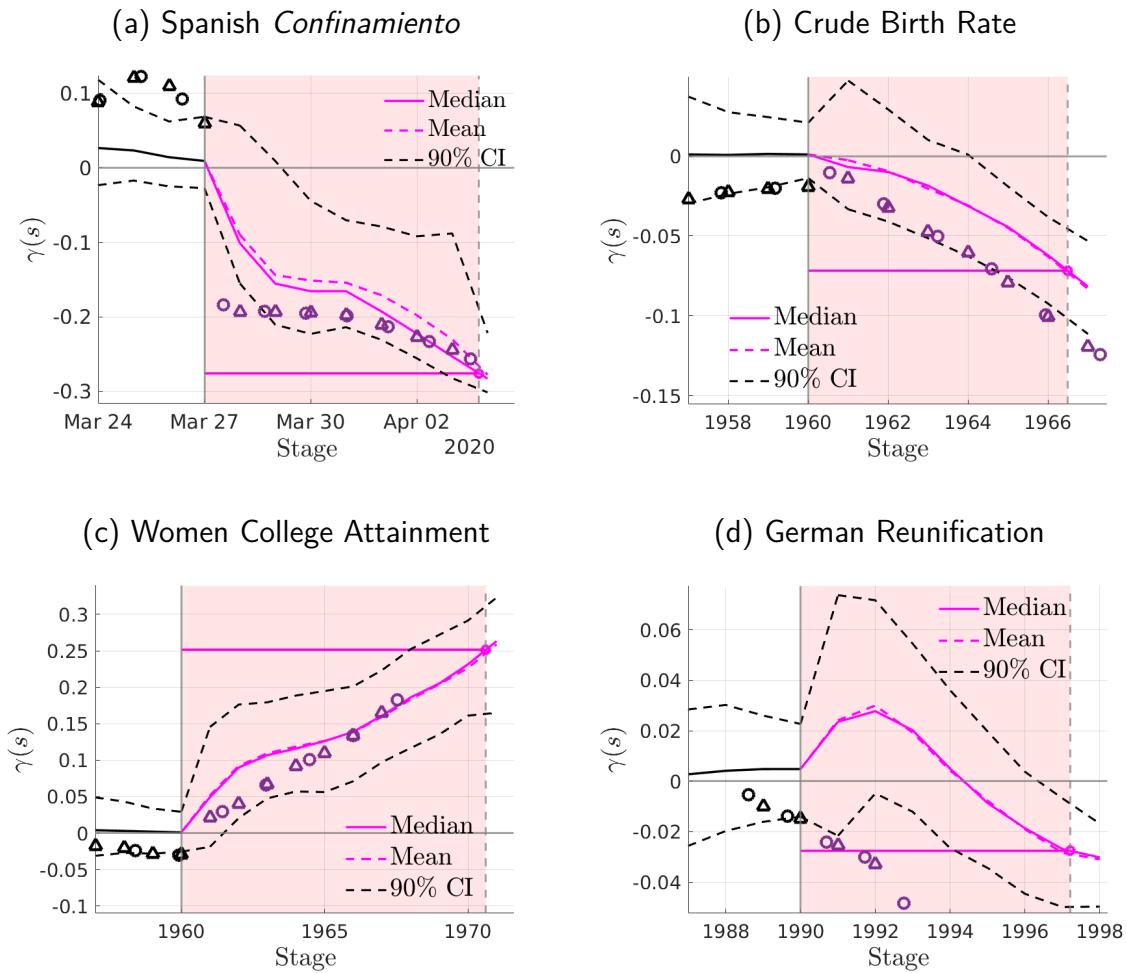
E.1 Identified Policy Effects Without the Trend-Extraction Step

In Figure 32, we compare the policy effects of all our empirical applications in which we used a trend-extractin step in order to conduct inference in Section 4 with a counterpart where we skip the trend-extraction step and minimize the original data points. Our exercise shows that the trend extraction does not affect the identified effects of the Spanish Confinamiento on daily deaths (see panel (a) of Figure 32) and the effects of the 1960 FDA approval of oral contraceptives on the crude birth rate (see panel (b) of Figure 32) and women college attainment (see panel (c) of Figure 32). However, we find that the identified policy effects without the trend-extraction step differ from those with the trend-extraction in the context of the German Reunification; see panel (d) in Figure 32. The reason for this differential result is that in the case of the German Reunification the outcome variable (GDP per capita) shows fluctuations at a larger frequency (business cycles) than the frequency in which we are ultimately interested in for the evaluation of the German Reunification. Since all regions share the same aggregate fluctuations, if we do not purge GDP per capita from the business cycle fluctuations our algorithm that aims to minimizes the distance between the GDP per capita across regions (to some reference region) will be drawn to map the business cycle fluctuations of all regions to that of the reference region. For this reason, and since our interest is not the business cycle but the longer run, we extract the trend from the GDP removing the higher frequency fluctuations of GDP. Therefore, in addition to be useful to conduct inference using the deviations from trend as an stochastic component, the trend-extraction step can also serve the purpose of removing fluctuations that are of higher frequency than the onex in which the policy evaluator is interested in.

E.2 Placebo Diagnosis for the Applications

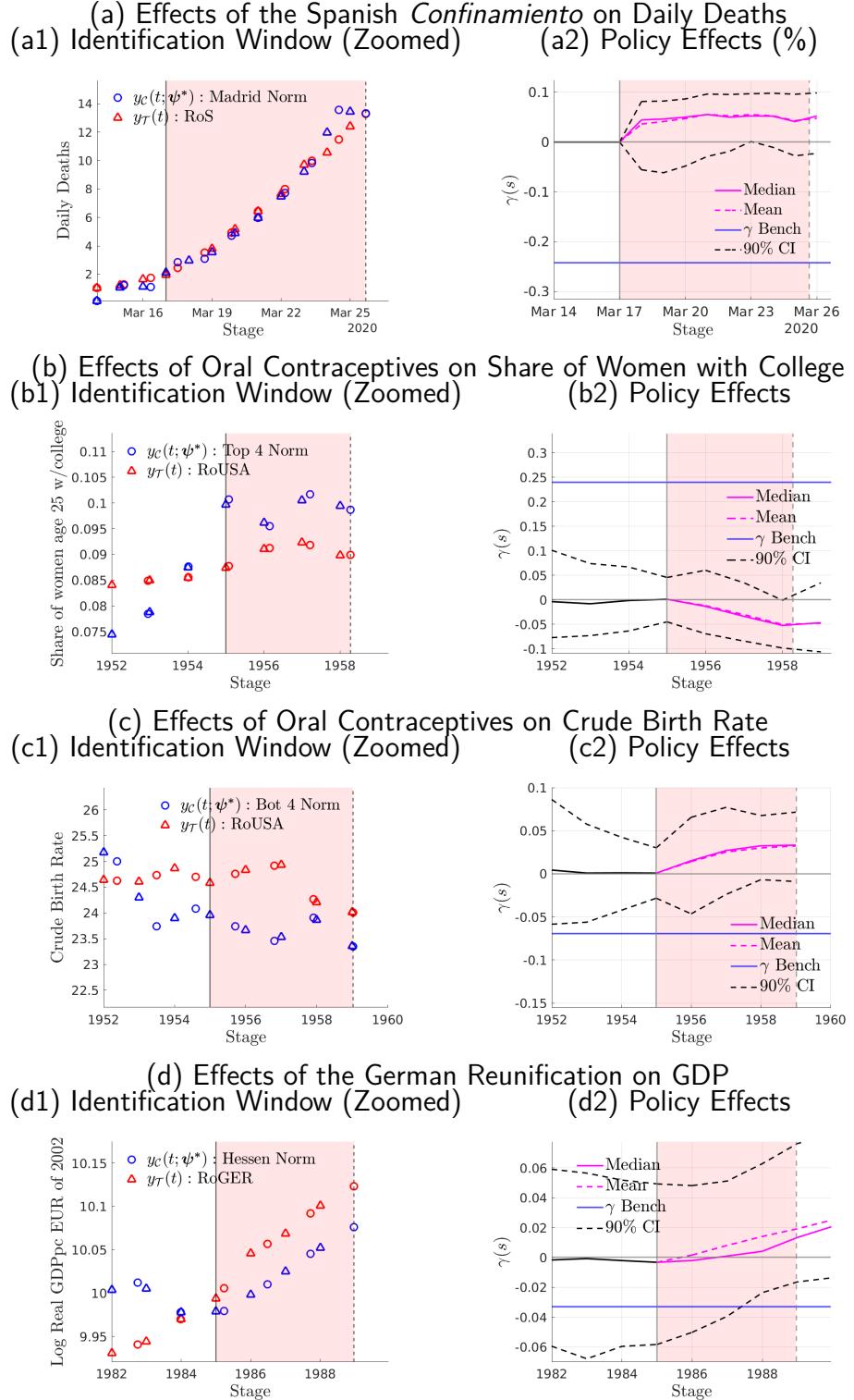
Here we show the policy effects that emerge from our method if we conduct our normalization at a some period before the actual policy takes place. In our placebo, we assume that the stay-home policy was imposed on March 10 2020; the pill was introduced 5 year before its actual market release; and that German Reunification occured 5 years before the actual reunification date. Our method will survive this test if it identify policy effects that are not significantly different from zero. We find that this is the case for each and all our applications; see Figure 33.

Figure 32: Identified Policy Effects Without the Trend-Extraction Step



Notes: We show the policy effects identified without the implementation of the smoothing step as described in Section 3.4.2.

Figure 33: Stage-Based Identification of Policy Effects: Placebo Effects



Notes: In our placebo, we assume that the stay-home policy was imposed on March 10 2020; the pill was introduced 5 year before its actual market release; and that German Reunification occurred 5 years before the actual reunification date.

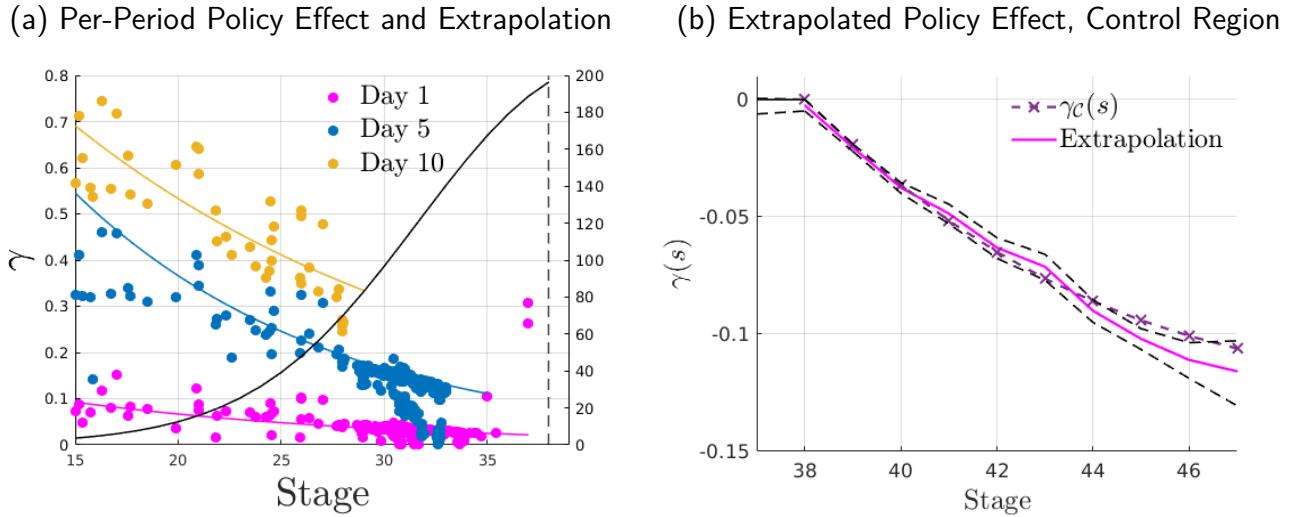
F Extrapolating to Find the Policy Effects for the Control Region

One may be tempted to use the heterogeneous policy effects in Section 5 in order to extrapolate seeking for the policy effects of the control region. In particular, given that the policy effects depend on the size of the identification window one could extrapolate the per-period policy effects forward. That is, one could fit a curve to the well-identified per-period policy effects (within the identification window) across stages and use the fitted curves to find the effects for the control region via extrapolation. We do this using a log-linear fit (top panels) and a quadratic fit (bottom panels) using our benchmark econ-epi model in Section 3.1.1; see, respectively, panel (a) and panel (c) of Figure 34. The policy effect of the control region is constructed using the value of the extrapolated per-period policy effect (after policy implementation). Precisely, we find where the extrapolated per-period policy effect hits the period when the policy effectively enters the control region.

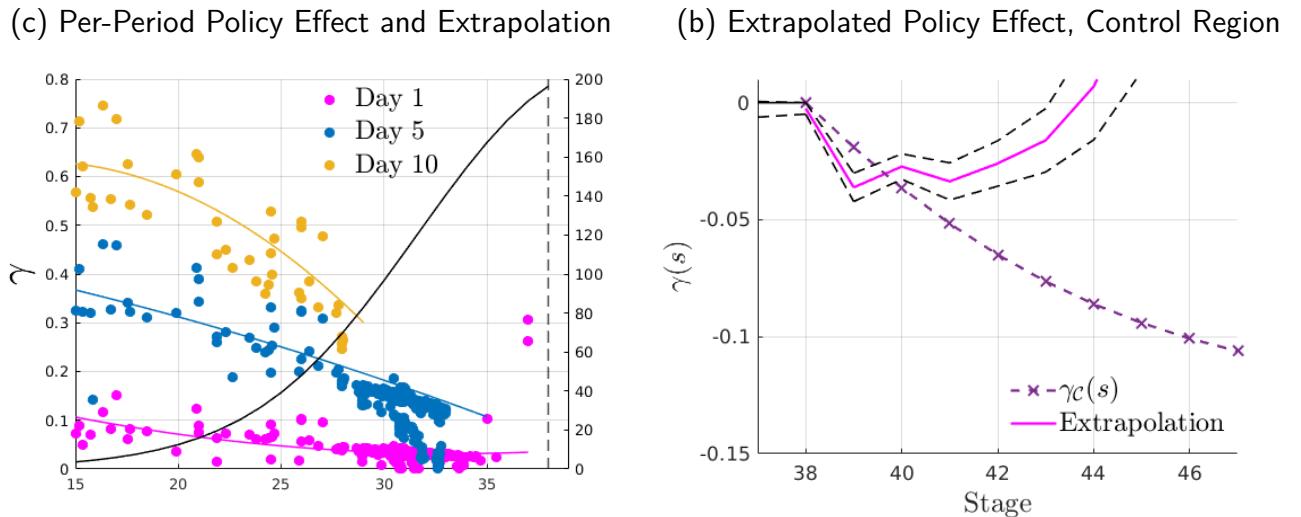
Overall, we find that conducting extrapolations in order to pin down the policy effects on the control region is not necessarily a good idea. In this specific example, a extrapolations from a log linear fit recovers the true per-period policy effect on the control region; see panel (b) of Figure 34. However, this is not the case with a quadratic fit; see panel (d) of Figure 34. One can imagine situations in which the extrapolation works are those in which the per-period policy effects are well approximated by the fitted curve, at least, locally around the last stages where the policy effects are available (hence, closer to when the control enters policy). However, there is no reason to think that is generally the case. In Figure 35, we show the extrapolated policy effects in the control region, Madrid, emerging from extrapolating the heterogenous effects by stage in our empirical application of the stay-home policy implemented in Spain during Covid-19.

Figure 34: Extrapolation to Find the Policy Effects for the Control Region: Model

Log-Linear Fit



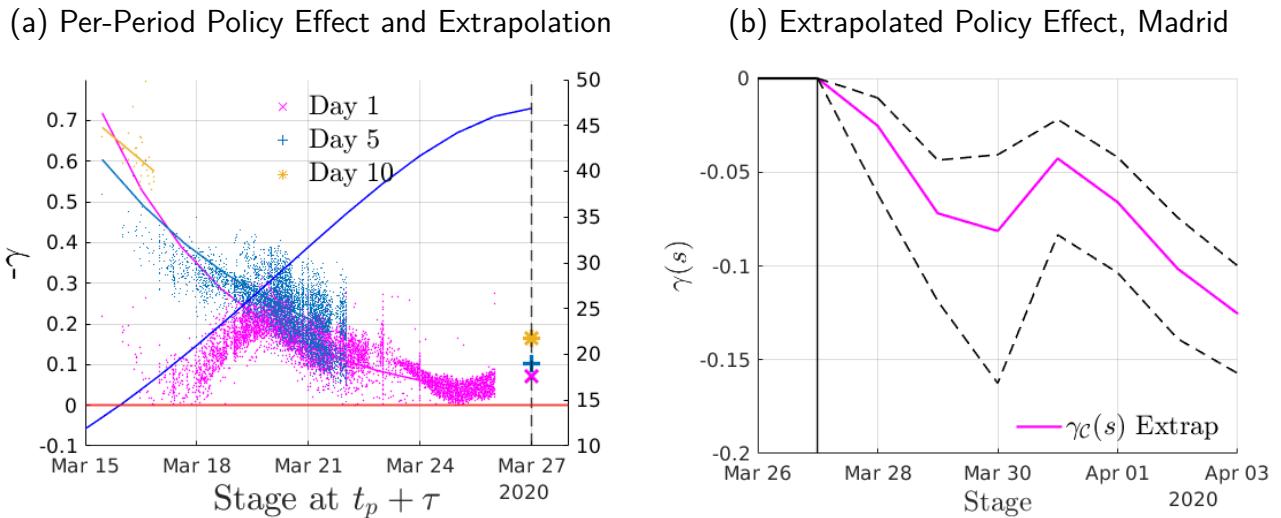
Quadratic Fit



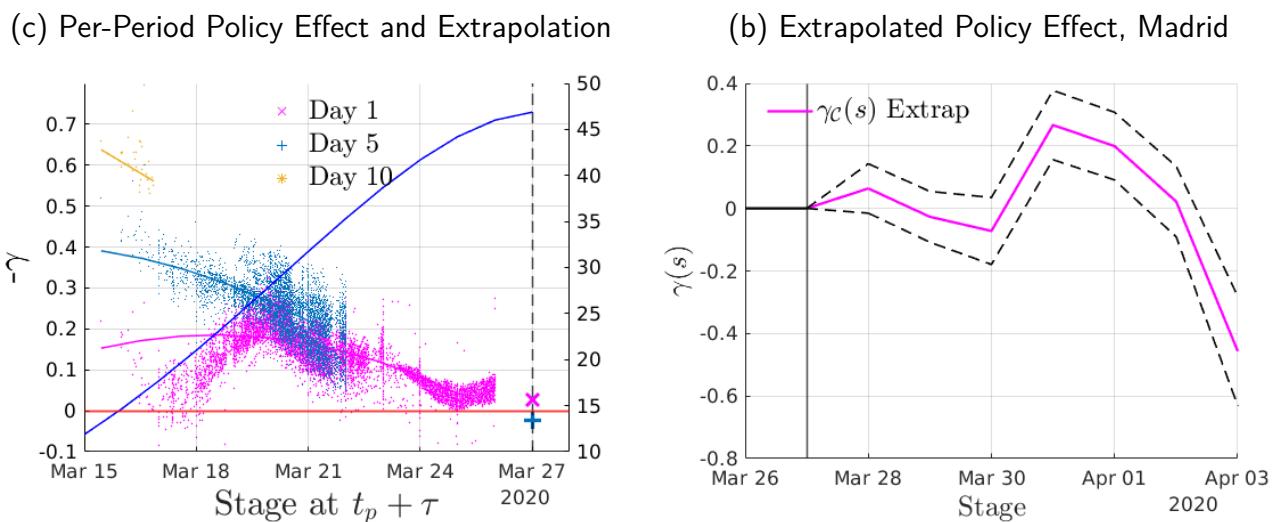
Notes:

Figure 35: Extrapolation to Find the Policy Effects for the Control Region: Madrid

Log-Linear Fit



Quadratic Fit



Notes:

G More on Policy Effects with Spillovers

Here we show a case in which we can separately identify the effects of policy and the spillover effects building on our benchmark econ-epi model; see Section 6.2. In particular, consider an environment with three regions that we show in Figure 36. In this environment there is a leading region \mathcal{C} (blue lines) that implements a stay-home policy at time t_p ; a non-leading region \mathcal{T}_1 (gray lines) that implements the same stay-home policy at time t_p and also affected by spillovers from the leading region throughout the entire epidemic path following (33); and a non-leading region \mathcal{T}_2 (red lines) that does not implement the stay-home policy but is affected by spillovers from the leading region also through (33). In the case of \mathcal{T}_2 our method recovers the spillover effects which are partly affected by the policy implemented in \mathcal{C} , whereas in the case of \mathcal{T}_1 our method recovers the effects of policy in the presence of spillovers—including the effects that the policy in region \mathcal{C} has on the spillovers. Hence, the difference between the effects across these two regions captures the policy effects (net of spillovers) of implementing the policy in region \mathcal{T}_1 in the presence of spillovers which are intern affected by policy in region \mathcal{C} . In Figure 37, we show an analogous case with these three regions with the sole differnece that region \mathcal{C} is assumed not to implement the policy. In this case, when focusing on \mathcal{T}_2 our method recovers the spillover effects, whereas focusiong on \mathcal{T}_1 our method recovers the effects of policy in the presence of spillovers. Then, the difference between the effects across these two regions captures the policy effects (net of spillovers) of implementing the policy in region \mathcal{T}_1 in the presence of spillovers from region \mathcal{C} .

Figure 36: More on Spillovers (1): With Policy in \mathcal{C}

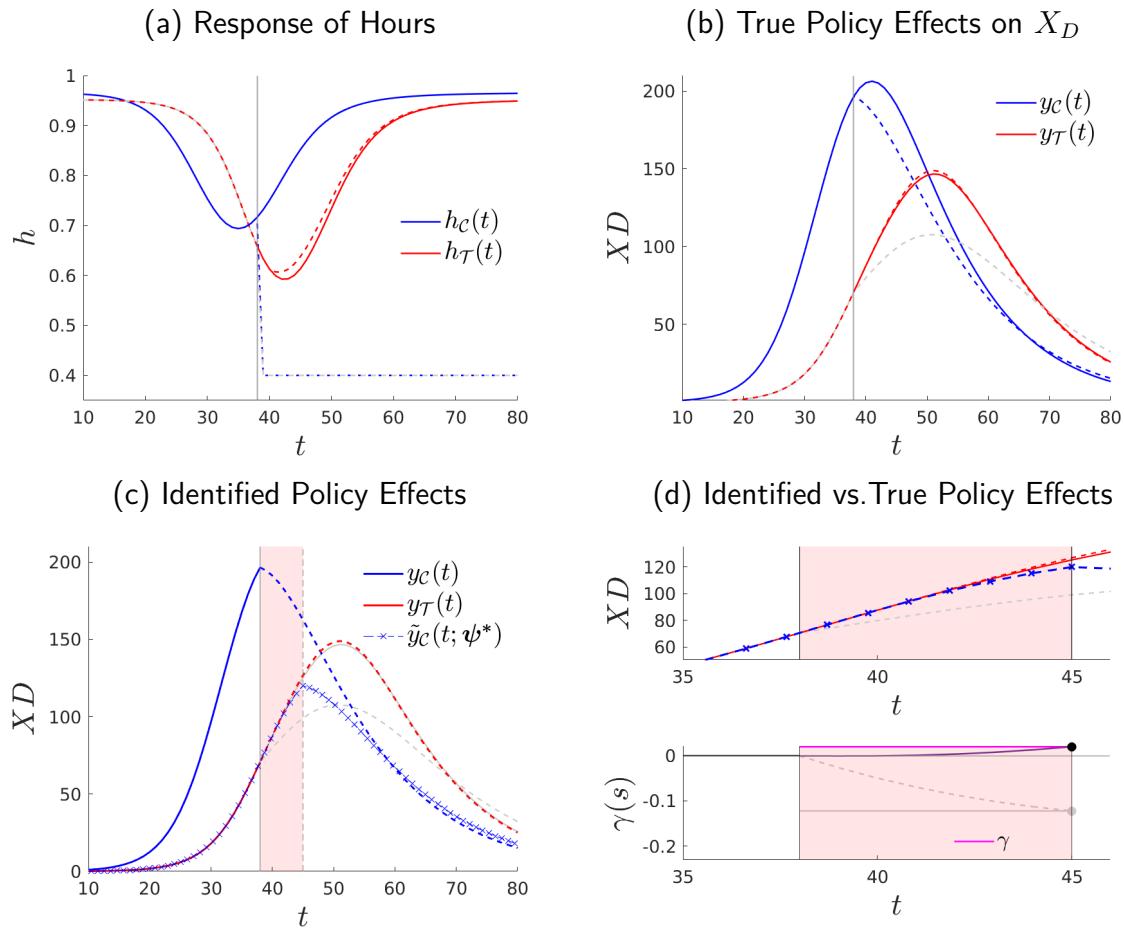
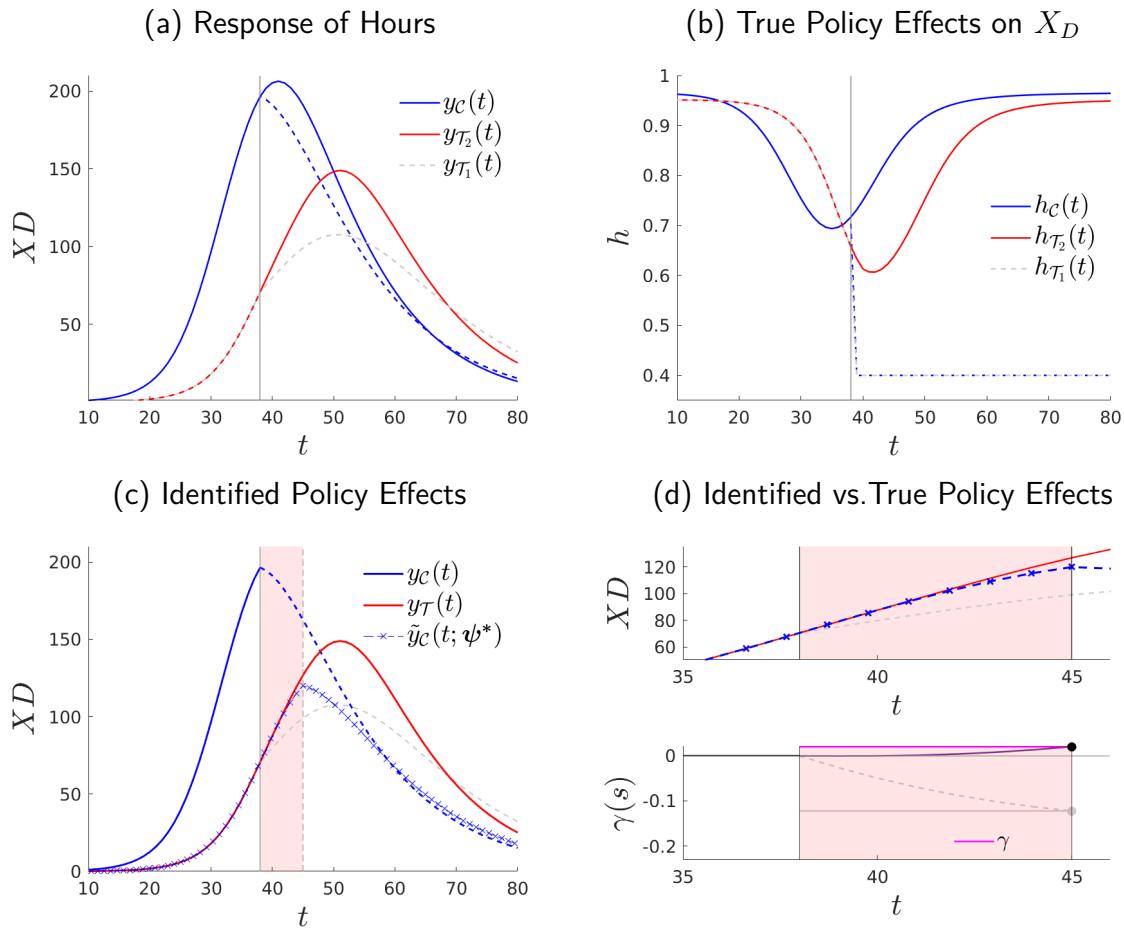


Figure 37: More on Spillovers (1): With Policy in \mathcal{C}



Notes: Where $\bar{h} = 0.4$, $t_p = 38$, $t_f = 250$, $\tilde{\xi}_{\mathcal{P},r} = 0.2$, $\sigma = 30$, $\gamma = 1.97\%$, $\gamma_{true} = 0.51\%$, $\epsilon(\gamma) = 283.95\%$.

Figure 38: More on Spillovers (2): Without Policy in \mathcal{C}

