Pitfalls of Common Estimators for Continuous Dynamic Treatments*

Lukas Delgado-Prieto† Jan Stuhler‡

May 26, 2026

Abstract

We examine commonly used estimators for continuous treatments that vary over time. The classical static panel regression that relates outcomes to current treatment intensity does not accurately capture lagged effects. The most commonly used dynamic difference-in-differences regression instead uses a time-constant measure of treatment to estimate period-specific coefficients, capturing dynamic effects more accurately. We show, however, when higher-exposed units accumulate their treatment earlier than less-exposed units – a pattern we document for immigration and labor demand shocks – the estimated coefficients become harder to interpret causally. A panel regression including leads and lags of the time-varying treatment provides an alternative that is robust to these concerns, although it requires sufficient time-series variation in the treatment. We characterize when each estimator fails to recover the treatment effects, propose diagnostic tests, and discuss additional complications arising from heterogeneous treatment effects.

Keywords: Continuous Treatments, Dynamic Effects, Two-Way Fixed Effects, Panel Data, Proportional Exposure.
JEL Codes: C23, C52.

*We are grateful to Martin E. Andresen, Maxwell Kellogg, and Mael Astruc–Le Souder for their comments and suggestions. We thank participants at the LACEA 2025 Annual Meeting and the Oslo Applied Lunch Seminar for helpful discussions.

†Department of Economics, University of Oslo, Norway. Email: laprieto@econ.uio.no.

‡Department of Economics, Universidad Carlos III de Madrid, Spain. Email: jstuhler@eco.uc3m.es.

1 Introduction

Researchers routinely use regressions with time and unit fixed effects (Two-Way Fixed Effects, TWFE) to assess the impact of continuous treatments, such as migration, trade, or minimum wage shocks.[1] However, different researchers choose different specifications. Traditionally, the most common approach has been to include a time-varying measure of treatment intensity in a *static panel* model. An increasingly popular alternative is to use a time-invariant measure of treatment intensity interacted with event time, in what can be termed a *dynamic difference-in-differences* (DiD) model. However, opinions on these approaches vary: while many researchers adopt dynamic DiD models as their default or preferred specification – see the latest Handbook Chapters of Labor Economics on migration (Dustmann and Schönberg, 2025), trade (Autor et al., 2025), and minimum wages (Dube and Lindner, 2024) for prominent examples – others remain skeptical.[2]

In this paper, we compare the strengths and weaknesses of the static panel and dynamic DiD approaches, highlighting the settings in which each is most appropriate. We also consider implementing a third method — *panel regression with dynamic treatment leads and lags* — that aims to combine the advantages of the other two approaches (Schmidheiny and Siegloch 2023; Freyaldenhoven et al. 2021). We illustrate the trade-offs between these methods both analytically and through simulations, using the classic question of how immigration affects labor markets as a running example. Drawing on this evidence, we argue that dynamic DiD and leads/lags models are generally more robust to continuous dynamic treatments, as both provide transparent evidence on whether treatment timing aligns with its supposed effects, which many researchers consider essential for credible causal inference. The choice between these two, in turn, hinges on whether units' *relative* exposure to the treatment over time is constant across units.

Static Panel. "Static panel" linear regressions, which relate current outcomes to current treatment, have long been a popular choice for analyzing continuous treatments. For example, Carrington and De Lima (1996) regresses wages in period t on a measure of regional exposure to immigration in that same period, together with time and region fixed effects,

[1] De Chaisemartin and d'Haultfoeuille (2024) review the 100 most-cited papers in the American Economic Review from 2015 to 2019 to find that 26 estimate a TWFE regression, but only six involve a binary treatment. Of the remaining 20, eleven use static panel regressions.

[2] The motivation for this paper stems from negative referee feedback both authors received on their empirical specifications in immigration studies. In a submission to *Econometrica*, a referee expressed confusion about why outcomes were long differenced from a baseline year, noting that the dependent and independent variables were not "aligned temporally" and recommending a static panel specification instead. In a paper submitted to *Labour Economics*, a referee was concerned about using a time-invariant independent variable for a dynamic treatment, which made the results "a lot harder" to interpret.

to estimate a single static treatment effect. This approach is straightforward to implement, accurately measures time-varying treatments, absorbs time-constant endogeneity, and requires at least two periods of data. It has long been popular in the immigration literature and is still commonly used as the baseline specification (e.g., Caruso et al., 2021; Aksu et al., 2022; Groeger et al., 2024), or as a complement to summarize treatment effects (Fuest et al., 2018).

However, the approach is subject to important limitations. The key issue is that the static panel approach offers no insight into the timing of effects. First, it offers no evidence on pre-trends — a critical shortcoming, as aligning the timing of treatment and its supposed effects is now widely used for assessing the credibility of the research design (Roth et al., 2023; Baker et al., 2025). [3] Second, by summarizing marginal treatment effects with a single coefficient, the approach implicitly assumes a static relationship and ignores potential dynamics in how outcomes evolve in response to treatment. However, in many economic settings, such dynamics are expected: for example, canonical models of immigration imply negative short-run wage effects that dissipate as other input factors adjust (Borjas, 1999). When such dynamics exist, the panel regression is misspecified, and estimates become sensitive, for instance, to the time periods included in the regression. Even if the marginal effects are immediate, the panel estimate, on its own, reveals nothing about how the *overall* treatment effect unfolds over time.

Dynamic Difference-in-Differences. An increasingly common alternative to address the limitations of static panel regressions is the dynamic DiD approach, which uses a time-constant measure of exposure to estimate separate coefficients for each time or event period. For instance, Dustmann et al. (2017) exploits regional variation in a commuting shock over a two-year window to estimate separate regressions of the change in outcome relative to a baseline year for each pre- and post-treatment period. The post-treatment coefficients are informative about how the *overall* treatment effect evolves over time. Since effects are estimated separately for each period, this approach is robust to dynamic treatment effects and variations in the number of included treatment periods. And since researchers only need to measure the shock in a single period, the method is especially useful when treatment intensity is more accurately measured at some points in time than at others.[4]

[3] It has become rare for a DiD-type design to appear in a top-five journal without an accompanying event study plot (Currie et al., 2020). Roth (2022) report that among the papers published in three leading economics journals between 2014 and 2018, 70 included an event-study plot. Importantly, such evidence on pre-trends is informative not just about the validity of the causal design, but also about detecting issues with statistical inference.

[4] For example, Delgado-Prieto (2024a) relies on the single census enumeration of Venezuelan immigrants, thereby avoiding the measurement bias in counting migrants with surveys (Aydemir and Borjas, 2011).

Moreover, the pre-treatment coefficients in a dynamic DiD framework serve as built-in, transparent placebo tests to probe the key identifying assumption of parallel trends. The choice between static panel and dynamic DiD goes thus beyond the accuracy of impact estimates, but also influences the credibility of the research design. Crucially, pre- and post-periods enter symmetrically in a DiD design. A frequent issue in empirical research is that researchers implement a *separate* specification to test for pre-trends, possibly estimated on a different sample, which can then make it difficult to compare the results with the main estimates. Moreover, tests for pre-trends are frequently underpowered, such that formal significance tests may offer little insight.[5] The symmetric structure of dynamic DiD enables more meaningful comparisons that go beyond the narrow question of whether pre-treatment coefficients are statistically different from zero.

However, the DiD approach also has potential drawbacks, which we illustrate below. While well-suited for settings with clearly defined pre-treatment periods – where treatment has not yet occurred or its intensity has remained stable – the approach is less suitable when such a pre-period is absent. Moreover, the dynamic DiD coefficients recover the true treatment effects only if the constant measure of treatment intensity – typically taken at the time the shock has fully materialized – is *perfectly correlated* with treatment intensity in other post-treatment periods. Specifically, if the ratio between period-specific treatment and the fully materialized one is constant across units (i.e., *proportional exposure*), the dynamic coefficients represent a combination of relative shock realizations and treatment effects, which retain a clear causal interpretation and can be rescaled as needed for obtaining specific parameters of interest (e.g., the marginal treatment effects). But if the treatment builds up at varying speeds across units (i.e., with *non-proportional exposure*), the dynamic DiD estimates become difficult to interpret and no longer recover a well-defined average treatment effect. For example, if units with greater eventual exposure also experience earlier exposure to the shock, a common scenario in immigration studies, the estimated effects for the earlier periods will be upward-biased.

Panel with Treatment Leads and Lags. A third specification combines the advantages of static panels and dynamic DiD by incorporating leads and lags of the treatment, explicitly capturing *both* the dynamics of the treatment and the dynamics of its effects (Schmidheiny and Siegloch, 2023; Freyaldenhoven et al., 2021). Although less common than the other two

[5] See Kahn-Lang and Lang (2020) and Roth (2022) for discussions of low power and related issues in pre-trend testing, and Rambachan and Roth (2023) for a partial identification approach that allows for some violations of parallel trends based on estimated pre-trends. Our point here is that these concerns are compounded when researchers implement separate specifications for main and pre-trend analysis, making coefficient estimates less comparable.

approaches, it is sometimes used in applied research; for example, Autor (2003) employs it in their study of temporary help services in the United States, while Mello (2022) uses it to study the effects of Brazil's centralized admission system and affirmative action expansion on student enrollment. As we show, this leads-and-lags approach is robust to the main concerns we highlight regarding the static panel and dynamic DiD.

The puzzle, then, is why this "more systematic" approach is not more widely adopted in applied research. We discuss three potential explanations. First, the approach is data-intensive, requiring longer time series of outcomes and treatments, and increases computation time, especially for individual-level analysis. Second, estimates from this approach are more difficult to communicate: unlike dynamic DiD, they reflect marginal rather than total effects and lack a straightforward graphical representation. To address this, we propose a simple method for transforming leads/lags panel estimates into an event-study–style graph that presents results in a format familiar from standard DiD applications and, by treating pre- and post-treatment periods symmetrically, provides a transparent basis for evaluating the validity of the research design. Third, the approach performs poorly in settings where treatments are highly serially correlated over time: if *relative* treatment intensities remain stable across periods, it becomes difficult or impossible to disentangle the contemporaneous impact of current from the lagged effect of past treatment.[6]

This last point implies a fundamental distinction between the two dynamic approaches: the leads/lags approach identifies treatment effects precisely by exploiting variation in how the treatment builds up across units – that is, it requires non-proportional exposure. The dynamic DiD, by contrast, is well suited to sharply defined treatments that evolve with proportional exposure, in which a single reference measure reliably captures relative exposure. The two approaches are therefore complements, not substitutes, and the choice between them hinges on the nature of the treatment process.

To illustrate the trade-offs between static panel, dynamic DiD, and leads/lags panel methods, we revisit the classic questions of how demand and supply shocks affect labor markets. Using simulations and empirical applications based on Yagan (2019) and Delgado-Prieto (2024a), we illustrate how each method performs under different types of dynamic effects and shock patterns, identify scenarios in which bias occurs, and propose simple diagnostic tests to detect its direction.

[6] For example, Jaeger et al. (2019) shows that immigration shocks across US regions were strongly correlated with shocks in later decades, especially when using the standard shift-share instrument to predict immigration shocks.

Heterogeneous Treatment Effects and Related Literature. Importantly, the concerns we highlight arise even under the assumption of homogeneous treatment effects across units. Allowing for heterogeneity in treatment effects, for example, through nonlinear treatment effects, introduces additional complications that have been emphasized in the recent DiD literature for binary treatments (Roth et al., 2023). Such heterogeneity may invalidate the approaches we study, though not in design-based specifications (Borusyak and Hull, 2024), and therefore represents an additional challenge beyond the mechanisms we emphasize. While these issues have received substantial attention in the binary treatment setting, they remain an open problem for continuous treatments that evolve over time. Recent contributions, including Callaway et al. (2024) and De Chaisemartin and d'Haultfoeuille (2024), propose estimators designed for continuous treatments that accommodate dynamic and heterogeneous effects. Callaway et al. (2024) focuses on staggered continuous treatments in which the treatment dose remains constant after adoption, whereas De Chaisemartin and d'Haultfoeuille (2024) allows for treatments whose intensity evolves over time. These approaches differ in how they construct and aggregate causal parameters and in the versions of the parallel trends assumption they rely on.

Rather than comparing these new estimators directly, our focus is on empirical settings with continuous dynamic treatments in which all units are simultaneously exposed to some treatment intensity, such as immigration or trade shocks, and researchers rely on more conventional regression approaches that can be combined with instrumental variables. We add to this discussion in the final section, examining the conditions under which treatment effect heterogeneity can create identification challenges in these settings and suggesting a flexible extension of the leads/lags panel model, following the extended TWFE framework of Wooldridge (2021), that can potentially address heterogeneous effects.

The remainder of the paper is structured as follows. Section 2 introduces common empirical strategies for analyzing continuous dynamic treatments and presents the potential outcome framework that guides the analysis. Section 3 discusses the role of heterogeneous treatment effects and relates our results to recent developments in the DiD literature. Section 4 concludes.

2 Empirical Strategies

This section outlines common empirical strategies for studying continuous treatments that evolve over time. We first formalize the setting in a potential outcome framework and state the identifying assumptions. We then show how three common specifications map into this framework, clarifying what each recovers.

2.1 Potential Outcome Framework

We consider a setting in which units i (which may represent individuals or groups) are treated *simultaneously* with a continuous, non-negative treatment $m_{it} \geq 0$ that may either amplify or dissipate over time t.[7] Let $Y_{it}(\mathbf{m}_i)$ denote the potential outcome for unit i at time t under treatment path $\mathbf{m}_i = \{m_{i1}, \ldots, m_{iT}\}$. We impose two assumptions, no-anticipation and parallel trends, that are standard in the difference-in-differences literature (Roth et al. 2023):[8]

Assumption 1 (No Anticipation)**.** *For all units and treatment paths, potential outcomes do not depend on future treatments:*

$$Y_{it}(m_{i1}, \ldots, m_{iT}) = Y_{it}(m_{i1}, \ldots, m_{it}).$$

Assumption 2 (Parallel Trends)**.** *Let $Y_{it}(0)$ denote the potential outcome for unit i in period t with zero treatment path. For all treatment paths $\boldsymbol{m}_i \neq \boldsymbol{m}_j$:*

$$\mathbb{E}\left[Y_{it}(0) - Y_{it-1}(0) \mid \boldsymbol{m}_i\right] = \mathbb{E}\left[Y_{jt}(0) - Y_{jt-1}(0) \mid \boldsymbol{m}_j\right].$$

Assumption 1 rules out behavioral responses to future treatments. Given that all units are untreated in the initial period, this ensures valid baseline outcomes. Assumption 2 states that, in the absence of treatment, the average change in outcomes for treated units receiving $\mathbf{m}_i$ would have followed the same trend as any other treatment unit receiving $\mathbf{m}_j$ for all $\mathbf{m}_i \neq \mathbf{m}_j$.

Under these assumptions, we define the potential outcomes as a function of both current and lagged treatment effects, with marginal treatment effects $b_i(m_{it})$ that could vary flexibly across units, time, and treatment intensity. To make the model tractable, we first assume homogeneous linear treatment effects, such that $b_i(m_{it}) \equiv b$, which are constant across i and linear on m_{it}, while in Appendix C we relax this assumption. In addition, we assume the effects of current and past treatments are additive and time-

[7] If treatment is staggered across units, additional complications arise, such as heterogeneous treatment effects across treatment cohorts (Sun and Abraham, 2021). In such cases, recent econometric advances provide some tools for addressing issues with both binary and continuous treatments (see De Chaisemartin and d'Haultfoeuille (2023) for an overview).

[8] By maintaining these assumptions throughout, we abstract from endogeneity concerns that are central in practice – such as selection into treatment or confounding trends – but orthogonal to our focus. Our goal is to understand how different estimators perform when "clean" causal variation is available in principle.

invariant. Under these restrictions, the potential outcome of Y_{it} is equal to:

$$Y_{it}(\mathbf{m}_i) = Y_{it}(0) + \sum_{l=0}^{L} b_l \, m_{it-l}. \tag{2.1}$$

Here, l represents the number of lags that affect the outcome up to lag L. The parameter b inside the summation captures the marginal treatment effect (i.e., the rate of change in the outcome per dose of treatment), and m_{it-l} is the treatment level. This formulation reflects the dynamic accumulation of treatment effects over time.

Particularly, when analyzing the dynamic DiD estimator, we will be interested in comparing the target estimand, termed as "true scaled effect", based on the period-specific treatment in T,

$$\beta_{t,true} = \mathbb{E}\left[\frac{Y_{it}(\mathbf{m}_i) - Y_{it}(0)}{m_{iT}}\right] = \sum_{l=0}^{L} b_l \, \mathbb{E}\left[\frac{m_{it-l}}{m_{iT}}\right], \tag{2.2}$$

with the one obtained from that regression framework.

If, as in the static panel regression, we further assume that treatment effects are only contemporaneous, i.e., $Y_{it}(\mathbf{m}_i) \equiv Y_{it}(m_{it})$, the potential outcome simplifies to:

$$Y_{it}(m_{it}) = Y_{it}(0) + b \, m_{it}, \tag{2.3}$$

with the causal estimand being $\mathbb{E}\left[\frac{Y_{it}(m_{it}) - Y_{it}(0)}{m_{it}}\right] = b$. These formulations provide a unified framework for the outcome models used in the simulations below (see Table 1) and for the interpretation of the effects delivered by each estimator.

2.2 Static Panels

To analyze the impact of a continuous dynamic treatment, such as immigration or trade shocks, researchers often use a panel regression that relates an outcome y_{it} – such as wages or employment at the regional level – to the continuous treatment m_{it}, while controlling for unobserved time-invariant unit heterogeneity and common time trends. The basic form of this model, also referred to as a static TWFE regression, is:

$$y_{it} = \beta \cdot m_{it} + \gamma_i + \gamma_t + \epsilon_{it} \tag{M1}$$

Here, γ_i captures unit fixed effects, γ_t captures time fixed effects, m_{it} represents the time-varying treatment shock, and ϵ_{it} the error term. By regressing the outcome on only the contemporaneous treatment, this specification implicitly assumes that treatment affects the outcome immediately and fully within the same period. While accounting for dynamics in the shock itself, the coefficient β summarizes its impact as a single, "static" parameter.[9]

This static panel approach has a long tradition in empirical research. For example, Carrington and De Lima (1996) pools district-level data across 18 Portuguese districts over time and regresses log wages on a measure of the district-level immigration rate, together with year and district fixed effects, to estimate the contemporaneous effect of immigration on wages. The approach has broad applicability. It absorbs time-constant unobserved heterogeneity, a key source of endogeneity, and imposes minimal requirements on the timing of the treatment process.[10] Although dynamic specifications have gained popularity in recent years, the static panel remains widely used, often as the baseline specification or as a complement to summarize post-treatment effects.[11] It continues to be a common choice in studies of immigration and minimum wages, as well as in other settings involving continuous treatments.[12]

While still widely used, static panels are no longer the default choice, owing to two key limitations.[13] First, they provide no built-in evidence of the presence or absence of pre-treatment trends, making it difficult to determine whether post-shock differences reflect treatment effects or the continuation of a pre-treatment trajectory. More generally, aligning the timing of the shock with the observed effects is now widely regarded as essential for credible causal inference (Roth et al., 2023; Baker et al., 2025). The absence of such evidence is a major limitation, especially in settings with sudden shocks, where a clearly defined pre-treatment period is present.

Second, by overlooking the possibility of dynamic responses over time, static panels can introduce bias that amplifies, weakens, or flips the sign of the estimated effects. Before

[9] A first-difference specification, $\Delta y_{it} = \beta \cdot \Delta m_{it} + \Delta \epsilon_{it}$, similarly relates changes in outcomes to contemporaneous changes in treatment, implicitly ruling out lagged effects.

[10] Moreover, it can be estimated with as few as two periods of data. In this case, first-difference regressions of the form of Altonji and Card (1991), Card (1992), and Autor et al. (2013) are equivalent to static panel *and* dynamic DiD estimates, but offer no direct evidence on pre-trends.

[11] Fuest et al. (2018) uses a static approach to summarize the effects of corporate taxes on wages.

[12] In the immigration literature, see, among others, Bratsberg and Raaum (2012) on immigrant inflows in Norway, Aksu et al. (2022) on Syrian arrivals in Turkey, Caruso et al. (2021), Bonilla-Mejía et al. (2024), and Lebow (2022) on Venezuelan migration to Colombia, and Groeger et al. (2024) on Peru. In the minimum wage literature, several papers reviewed in Dube and Lindner (2024) use this specification, including Meer and West (2016) and Neumark et al. (2014).

[13] A related important concern with TWFE regressions is the implicit weighting in β under heterogeneous treatment effects, which can assign negative weights to the treatment effects of certain (i, t) groups (De Chaisemartin and d'Haultfoeuille, 2023). This is particularly pronounced in settings with continuous, gradually increasing treatments, where units with below-average treatment levels or those in early treatment periods tend to receive negative weights, even when all units receive the treatment simultaneously. For instance, De Chaisemartin and d'Haultfoeuille (2023) decompose the TWFE coefficient as $\beta_{static} = \sum_{i,t} w_{it} TE_{it}$, where the weights w_{it} are proportional to the residuals of the regression $m_{it} = \alpha + \gamma_i + \gamma_t + \epsilon_{it}$. In comparison, we focus on issues that arise even if treatment effects are homogeneous across units.

formalizing this issue, we illustrate it with a simulation. We consider a treatment that amplifies over time, such as immigration shocks, under three effect patterns (see table notes): (i) *static*, where a one-unit increase in m_{it} reduces the outcome, potentially wages, by 0.3% and the effect persists permanently; (ii) *gradual*, where the effect builds up at -0.1% per period over three periods, accumulating to -0.3%, and then persists; and (iii) *U-shaped*, which follows the same build-up but then fades out symmetrically.

Table 1 illustrates what the static panel estimator captures under these different outcome models. Under static effects, the panel estimator recovers the true value, regardless of which periods are included in the estimation. But if treatment effects unfold dynamically – as is plausible in many settings – the estimates fluctuate widely. One might hope that they capture a weighted average of short- and long-run responses, but this is generally not the case. Under the gradual effects assumption, the estimate based on early and late post-periods (-0.51) overstates the true long-run effect (-0.3). Under U-shaped effects, the estimate switches sign entirely. The intuition is that the static coefficient captures not only the contemporaneous effect of treatment, but also the omitted lagged effect of earlier treatments – which may either reinforce (gradual effects) or offset (U-shaped effects) the current response. This is a classic misspecification problem: by assuming that only the current treatment affects the outcome, the model conflates contemporaneous and lagged responses.

Table 1: Simulated Effects with Static Panel

Effect	Outcome Model	Static Panel Estimator		
		All periods	Pre- & early post-periods	Early & late post-periods
Static	$y_{it} = -0.3\, m_{it} + \epsilon_{it}$	-0.30	-0.30	-0.31
Gradual	$y_{it} = \sum_{l=0}^{2} -0.1\, m_{it-l} + \epsilon_{it}$	-0.27	-0.23	-0.51
U-shaped	$y_{it} = \sum_{l=0}^{2} -0.1\, m_{it-l} + \sum_{l=3}^{5} 0.1\, m_{it-l} + \epsilon_{it}$	-0.14	-0.19	0.14
N (units x periods)		14,000	11,000	5,000

Notes: Simulation based on 1,000 units and 14 periods, divided into 6 pre- and 8 post-treatment periods. A continuous dynamic treatment begins in period 0, amplifies through period 4, and reaches its full scale by that point. The full shock is drawn from $m_{iT} \sim U(0,1) * 0.1$, and the random component is modeled as $\epsilon_{it} = 5 + U(0,1) * 0.01$. "Pre- and early post-periods" include all pre-treatment and post-treatment periods 0 to 4, during which the shock is increasing over time. "Early and late post-periods" refer to post-treatment periods 3 to 7, during which the full shock materializes.

2.2.1 How The Static Coefficient Aggregates Dynamic Effects?

To provide clarity on how the static panel coefficient aggregates dynamic effects, we derive it in terms of potential outcome parameters from Eq. (2.1). The OLS estimand from

Eq. (M1), putting time and unit fixed effects inside the error term, is given by:

$$\beta_{static} = \frac{\text{Cov}(m_{it}, y_{it})}{\text{Var}(m_{it})} = \frac{\text{Cov}(m_{it}, \sum_{l=0}^{L} b_l\, m_{it-l} + \varepsilon_{it})}{\text{Var}(m_{it})}. \tag{2.4}$$

Under the Parallel Trends Assumption 2, m_{it} is uncorrelated with ε_{it}, so in the case when treatment effects are static ($b_l = 0, \forall\, l \geq 1$), the panel regression recovers the contemporaneous marginal effect $\beta_{static} = b_0$. However, when treatment effects are dynamic, then by linearity of covariances β_{static} can be decomposed as a weighted sum of the contemporaneous marginal effect b_0 and the lagged marginal effects $b_1, \ldots, b_L$, where the weights on lagged effects are generally different from one whenever treatment varies over time. The following proposition makes this precise.

Proposition 1 (Static Panel Weighting of Dynamic Effects)**.** *The static panel coefficient can be written as a weighted sum of contemporaneous and lagged marginal effects:*

$$\beta_{static} = b_0 + \sum_{l=1}^{L} b_l \underbrace{\frac{\text{Cov}(m_{it}, m_{it-l})}{\text{Var}(m_{it})}}_{\omega_l}, \tag{2.5}$$

where $\omega_l \gtrless 1$. These weights determine the magnitude of β_{static} relative to the cumulative effect $\sum_{l=0}^{L} b_l$ as follows: i) if treatment is stable between periods, $\omega_l = 1$; ii) if higher-treatment units in t have larger treatment differences between t and $t-l$, $\omega_l \in (0,1)$; and iii) if higher-treatment units in t have smaller treatment differences between t and $t-l$, $\omega_l > 1$.[14]

Hence, with dynamic effects *and* continuous dynamic treatments, the weights will differ from one and $\hat{\beta}_{\text{static}}$ will not capture accurately the lagged marginal effects. In our simulations, higher treatment units increase more over time, so $\omega_l \in (0,1)$ and $\beta_{static} > \sum_{l=0}^{L} b_l$. Specifically, for the gradual effects case in Table 1, $\omega_1 \approx 0.9$ and $\omega_2 \approx 0.8$, implying $\hat{\beta}_{\text{static}} = -0.1 - 0.1 \times 0.8 - 0.1 \times 0.9 \approx -0.27$, against a true cumulative effect of -0.3.

Notably, these weights are influenced by the *number* and *type* of periods included in the regression. Including more post-treatment periods during which the treatment is fully realized will assign higher weights to the lagged effects, whereas adding more pre-treatment periods will assign lower weights to those effects. Overall, this shows

[14]The last two cases can be seen more clearly after rewriting the weights as $\omega_l = 1 - \text{Cov}(m_{it}, \Delta m_{it}) / \text{Var}(m_{it})$. A further case arises when treatments decrease over time, in which $\omega_l < 0$ and the sign of β_{static} can differ from that of $\sum_{l=0}^{L} b_l$. With unit and time fixed effects, ω_l is instead the regression coefficient of residualized $\tilde{m}_{it-l}$ on residualized $\tilde{m}_{it}$, where $\tilde{m}_{it}$ denotes m_{it} after removing unit and time fixed effects, making the weights less clear along with the direction of β_{static} relative to $\sum_{l=0}^{L} b_l$.

the potential misspecification issues with static regressions that aggregate lagged effects in sensible ways that depend, for instance, on the number of periods included in the regression.[15] Appendix A presents a similar derivation for a first differences specification, and shows that this static coefficient has similar aggregation issues of treatment effects.

A second limitation of the static specification is that it does not allow for testing for pre-treatment dynamics of the outcome. Suppose the Parallel Trends Assumption 2 fails, the static coefficient can be re-expressed as: $\beta_{\text{static}} = b_0 + b_1\, w_{t-1} + b_2\, w_{t-2} + \lambda_{trend} \frac{\text{Cov}(m_{it},\, y_{it}(0) - y_{i0}(0))}{\text{Var}(m_{it})}$, where λ_{trend} captures the association of the shock with the counterfactual outcome trend, since $y_{it}(0)$ is the potential outcome in the absence of treatment. Whenever the covariance term is non-zero, the static coefficient is confounded with differential trends in the outcome. Given the lack of counterfactual post-treatment trends in the data, researchers often assess pre-treatment trends to support the parallel trends assumption. This is why it is increasingly relevant to provide suggestive tests on this issue, where the static panel regression provides no assessment.

2.3 Dynamic DiD Regression – Dynamic Effects

An alternative approach to estimating the effects of continuous treatments is the "dynamic" or "generalized" DiD regression, which allows for a more flexible analysis of dynamic responses. Moreover, it incorporates pre-treatment trends into the same regression equation, in a symmetric manner, making it a powerful tool for assessing the plausibility of the parallel trends assumption (Dustmann et al., 2017; Roth et al., 2023).

The dynamic or generalized DiD regression examines the outcome response relative to a baseline period before the treatment starts, t_{pre}, and can be implemented either in levels:

$$y_{it} = \sum_{\tau \in \mathcal{T}, \tau \neq -1} \beta_\tau D_\tau \cdot m_{iT} + \gamma_i + \gamma_t + \epsilon_{it}, \tag{M2}$$

or in long differences:

$$y_{it} - y_{i,t_{pre}} = \beta_\tau \cdot m_{iT} + \gamma_t + \epsilon_{it} - \epsilon_{i,t_{pre}}, \quad \text{for each } \tau \in \mathcal{T}. \tag{M2LD}$$

Compared to the static panel in (M1), the key differences are that the shock m_{iT} is measured at a fixed reference period T – typically once the shock is fully realized or when

[15] Appendix C shows a similar derivation with heterogeneous effects across units. For the static coefficient to accurately capture average treatment effects, even with static treatment effects, the heterogeneity must be unrelated to the treatment exposure.

reliable post-treatment data is available – and that it is interacted with event time.[16] In the levels specification (M2), the unit fixed effects γ_i capture outcome differences in the baseline period t_{pre}, while the interactions of m_{iT} with the event time dummies D_τ capture the differential changes over time relative to the omitted baseline period, which can be defined more explicitly as $\mathbf{1}\{t - t_{pre} = \tau\}$. The long-difference specification (M2LD) nets out the unit fixed effects and yields the same β_τ coefficients (depending on the error structure, not equal standard errors) but avoids the need for interactions with event time by estimating a separate regression for each pre- and post-treatment period.[17] Either specification yields a full set of dynamic β_τ coefficients for the event-time window $\mathcal{T}$, defined over a set of leads and lags relative to the baseline period, which researchers typically plot in a "dynamic DiD" or "event study" graph.

These coefficients have a useful interpretation: they trace the *overall* impact of the shock over time, reflecting not only the dynamics of its marginal effects but also the shock's evolution. For example, we may observe small coefficients β_τ in the early post-treatment periods either because the shock has not yet fully materialized or because the marginal effects operate with a lag – or both. The static panel "resolves" this ambiguity by assuming that the marginal effects are immediate and constant over time. In contrast, the dynamic DiD is more agnostic and does not attempt to disentangle these two sources of dynamics. As a useful side effect, dynamic DiD is robust to the number of included periods, as reflected by Eq. (M2LD) being estimated separately for each period t. The approach provides a unified view of both the shock's overall dynamic impact and whether the timing of the estimated effects supports a causal interpretation. In the case of immigration, for example, researchers can assess whether differentially exposed regions followed comparable economic trends before the migration event, whether the onset of estimated effects aligns with the timing of the migration shock, and whether these effects eventually stabilize or dissipate over time.

Given these advantages, the dynamic DiD approach has gained increasing popularity across fields; for example, several chapters in the most recent Handbook of Labor Economics adopt dynamic DiD to study the effects of migration, trade, and minimum wages (Dube and Lindner, 2024; Dustmann and Schönberg, 2025; Autor et al., 2025). The approach has become common in studies on the labor market impacts of immigration.

[16] In practice, this shock may be constructed as a change relative to a pre-treatment baseline, $m_{iT} \equiv \Delta m_{iT}$, which accommodates settings without a strictly zero pre-treatment period while preserving the interpretation of the shock as a cumulative change.

[17] Although some researchers include contemporaneous covariates X_{it}, the logic of the dynamic DiD specification is more consistent with measuring covariates at the baseline period and interacting them with event time, thereby assuming some form of conditional parallel trends. See Section 4.3 of Baker et al. (2025) for a detailed discussion.

Using a specification in long differences, Dustmann et al. (2017) study the inflow of Czech commuters into Germany, Edo (2020) studies the arrival of Algerian repatriates in France, Monras (2020) studies Mexican inflows into the United States, and Muñoz (2024) investigates the impacts of posted workers in France.[18] The approach has also been widely adopted in other literatures studying continuous local shocks.[19] Yet even within the immigration literature, the approach is not universally accepted (see footnote 2). In the context of Venezuelan immigration to Colombia, most studies rely on static specifications (Bonilla-Mejía et al., 2024; Caruso et al., 2021; Lebow, 2022), and the dynamic specification used by Delgado-Prieto (2024a) has been subject to criticism: Lebow (2022) argues that relying on a single-year measure of the treatment variable is a limitation and recommends using a time-varying variable within a static regression.

To address such concerns, we next formalize the conditions under which the dynamic DiD approach is suitable and discuss how its coefficients can be interpreted. Settings with a relatively sharp shock and a clearly defined pre-treatment period are natural candidates. However, a crucial question is whether the shock measured at time T provides a useful approximation for *relative* shock exposure in other periods. If exposure evolves in constant proportions, so that the relative exposure of different units remains stable, the dynamic DiD coefficients retain a straightforward causal interpretation. Problems arise when this is not the case.

2.3.1 When Dynamic DiD Just Works: Proportional Exposure

We first express the DiD coefficients in terms of potential outcome parameters from Eq. (2.1). Considering long differences (relative to a baseline pre-period) to remove all constant terms, the DiD estimands from Eq. (M2LD) are given by:

$$\beta_\tau^{DiD} = \frac{\text{Cov}\left(m_{iT}, y_{it} - y_{i,t_{pre}}\right)}{\text{Var}\left(m_{iT}\right)} = \frac{\text{Cov}\left(m_{iT}, \sum_{l=0}^{L} b_l \; m_{it-l} + (\varepsilon_{it} - \varepsilon_{i,t_{pre}})\right)}{\text{Var}\left(m_{iT}\right)}. \tag{2.6}$$

Under the Parallel Trends Assumption 2:

$$\beta_\tau^{DiD} = \sum_{l=0}^{L} b_l \frac{\text{Cov}\left(m_{iT}, m_{it-l}\right)}{\text{Var}\left(m_{iT}\right)}, \tag{2.7}$$

[18] In Denmark, Foged and Peri (2016) study the arrival of refugees using a panel approach, and then employ a dynamic DiD regression with individual-level controls.

[19] Examples include the impacts of the Great Recession (Yagan, 2019; Redondo, 2022), import shocks (Autor et al., 2014, 2021), and minimum wages (Derenoncourt et al., 2025; Luduvice et al., 2024).

corresponding to a weighted sum of the dynamic marginal effects b_l, weighted by the lagged covariance structure of the shock itself.

The interpretation of DiD coefficients thus depends on how each unit's treatment evolves relative to its period-T exposure, as captured by $\delta_{it} \equiv m_{it}/m_{iT}$.

If exposure evolves proportionally ($\delta_{it} = \delta_t$ for all i), so that relative exposure remains stable across units ($m_{it}/m_{jt} = m_{iT}/m_{jT}$ for all i, j), the weights in equation (2.7) simplify to $\text{Cov}\left(m_{iT}/m_{it-l}\right)\text{Var}\left(m_{iT}\right) = \mathbb{E}\left[\frac{m_{i,t-l}}{m_{iT}}\right]$ and the dynamic DiD coefficients retain a straightforward causal interpretation: they recover the true scaled effects implied by the potential outcome model as stated in Eq. (2.2),

$$\beta_{\tau}^{DiD} = \sum_{l=0}^{L} b_l \mathbb{E}\left[\frac{m_{i,t-l}}{m_{iT}}\right] = \beta_{t,true}. \tag{2.8}$$

Assuming that future treatments do not affect pre-treatment outcomes, they also recover that the true scaled effects are zero in the pre-treatment period.

This "proportional exposure" condition is trivially satisfied when the shock is immediate and sharp. For example, Dustmann et al. (2017) studies a commuting shock that builds up sharply within just two years. In such settings, the dynamic DiD coefficients capture the dynamic impacts of a one-time, sharp shock, simplifying their interpretation. In such settings, the dynamic DiD provides strictly more information about the effects than the static panel does, with few downsides. While period-specific DiD estimates may be noisy, this is informative in itself – and researchers can always aggregate across periods, either informally (visually) or by complementing the dynamic specification with a static DiD estimate.

It is less obvious whether a time-T shock measure remains appropriate in settings where treatment builds up more gradually over time – such as Venezuelan immigration to Colombia, where inflows accumulated over several years. It is useful to distinguish between two scenarios. In Scenario 1, the *proportional exposure* condition still holds, as the shock increases proportionally across units, so that *relative* exposure remains constant over time. The dynamic DiD works well in this case and retains a clear causal interpretation, as shown in eq. (2.8). In Scenario 2, the units' relative exposure changes over time – for example, because some regions are exposed earlier than others. Here, interpreting the dynamic DiD coefficients becomes more complicated. Whether one views them as "biased" depends on the target: they no longer recover the true scaled effect, but they continue to capture how outcomes in differentially exposed units diverge over time.[20]

We provide simulations to illustrate these two scenarios before studying actual applications in the next subsection. Scenario 1 in Figure 1a illustrates the *proportional exposure* case, in which the shock evolves in constant proportions across units. In contrast to static panel estimates, the dynamic DiD retains a simple and useful interpretation: its coefficients accurately capture the *true scaled effect*, corresponding to the treatment effect divided by the average shock $\bar{m}_{iT}$ (see Eq. 2.2). Intuitively, small early coefficients indicate that the treatment is still unfolding – as becomes apparent when comparing these estimates with descriptive evidence on the build-up of the aggregate shock, which researchers commonly provide. Moreover, they reflect that the treatment has dynamic (in our simulation, gradual) rather than immediate effects.

While constant proportions may seem like a knife-edge assumption, it is a good approximation in relevant settings, especially when using instrumental variable methods. In immigration studies, immigrants' location choices may remain relatively stable even as overall inflows grow. Moreover, most papers instrument immigrants' location choices, and the predicted inflows will satisfy the proportional-exposure condition by construction whenever the instrument is based on a common baseline measure across all years — as is often the case. For example, in the Colombian case, the instruments are typically constructed using distance measures to Venezuela or historical settlements of Venezuelans. In these cases, the predicted exposure evolves proportionally.

2.3.2 Dynamic DiD Under Non-Proportional Exposure

The interpretation of dynamic DiD estimates becomes tricky when units' relative treatment exposure is unstable over time – that is, when the shock builds up earlier or more rapidly in some units than in others (*non-proportional exposure*, Scenario 2). While the dynamic DiD coefficients no longer recover the true scaled effect, they retain a useful – although less general – interpretation even in such settings.[21]

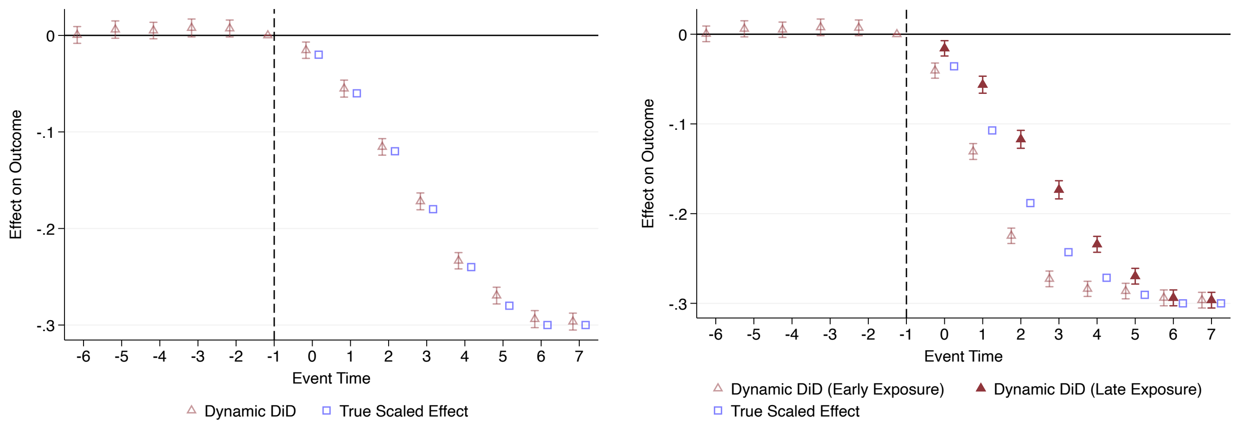
Under non-proportional exposure, the weights in equation (2.7) do not simplify as $\text{Cov}\left(m_{iT}/m_{it-l}\right)\text{Var}\left(m_{iT}\right) \neq \mathbb{E}\left[\frac{m_{i,t-l}}{m_{iT}}\right]$. Figure 1b illustrates two cases. In the *early-exposure* case, units with higher exposure in T experience earlier amplification of the

[20] A simplifying practice to help interpretation could be to binarize the continuous treatments. Appendix B shows that the resulting coefficient naturally captures the outcome difference between above- and below-threshold units, but the binarization discards within-group variation in treatment intensity and introduces complications when units that are above some threshold in T experience differential evolution of their treatment over time.

[21] Appendix C discusses other concerns when allowing for heterogeneous treatment effects across units. For the DiD coefficient to be interpreted causally, even in the proportional exposure case, the heterogeneity must be unrelated to treatment exposure in T.

Figure 1: Simulated Effects with Dynamic DiD

(a) Shock Evolving Proportional (b) Shock Evolving Non-Proportional



Notes: We focus on gradual effects of Table 1 that follow the pattern: $y_{it} = \sum_{k=0}^{2} -0.1 * m_{it-k} + \epsilon_{it}$. The true scaled effect represents the average effect per unit of the shock in period T, as described in Eq. 2.2. The analysis is based on a continuous dynamic shock that begins in period 0, amplifies either in proportional exposure across units (a) or not (b) until period 4, and fully reaches its scale by that point. The full shock is drawn from $m_{iT} \sim U(0,1) * 0.1$, and the random component is modeled as $\epsilon_{it} = 5 + U(0,1) * 0.01$. The simulation is conducted with 1,000 units and spans 14 periods, divided into 6 pre- and 8 post-treatment periods. We use 95% confidence intervals.

shock, such that $\frac{\text{Cov}(m_{iT}, m_{i,t-l})}{\text{Var}(m_{iT})} > \mathbb{E}\left[\frac{m_{i,t-l}}{m_{iT}}\right]$; in the *late-exposure* case, these units are instead exposed later, implying $\frac{\text{Cov}(m_{iT}, m_{i,t-l})}{\text{Var}(m_{iT})} < \mathbb{E}\left[\frac{m_{i,t-l}}{m_{iT}}\right]$ (see Appendix Section E for details on the construction of the shock dynamics). Under early exposure, the dynamic DiD estimates overstate the true scaled effects; under late exposure, they understate them.[22]

By decomposing the unit-specific relative exposure measures $\delta_{it} = m_{it}/m_{iT}$ into a period-specific mean and its individual deviations, $\delta_{it} = \bar{\delta}_t + \tilde{\delta}_{it}$ where $\bar{\delta}_t = \bar{m}_t/\bar{m}_T$, we can characterize this bias as follows:

Proposition 2 (DiD Interpretation With Dynamic Shocks)**.** *The dynamic DiD estimands can be written as:*

$$\beta_\tau^{DiD} = \sum_{l=0}^{L} b_l \frac{\text{Cov}(m_{iT}, \delta_{it-l}\, m_{iT})}{\text{Var}(m_{iT})} = \sum_{l=0}^{L} b_l \left(\bar{\delta}_{t-l} + \underbrace{\frac{\text{Cov}(m_{iT}, \tilde{\delta}_{it-l}\, m_{iT})}{\text{Var}(m_{iT})}}_{\text{Bias}[\tilde{\delta}_{t-l}]} \right). \tag{2.9}$$

This bias term:

(i) Equals zero if $\delta_{it} = \delta_t$ for all i (proportional exposure).

[22]Appendix Figures D.2a and D.2b show the distribution of estimates in period 3 across 500 simulations. The densities are clearly shifted away from the true effect, with minimal overlap, confirming that the discrepancy is systematic.

(ii) Is positive if units with higher m_{iT} have higher δ_{it-l} (early exposure for high-exposure in T).

(iii) Is negative if units with higher m_{iT} have lower δ_{it-l} (late exposure for high-exposure in T).

This expression implies that the direction of bias depends on whether higher-exposed units have systematically higher or lower proportions of their treatments, resulting in an attenuated (overstated) β_τ^{DiD} when the bias parameters are negative (positive).[23] Moreover, it clarifies that even during the period T when the treatment is *accurately* measured, there is bias, as it inherits bias from lagged treatment effects.[24]

Different exposure paths across units can therefore cause the dynamic DiD estimand to diverge from, or even have the opposite sign of the true scaled effect. Even in such scenarios, the coefficients retain their interpretation as capturing how outcomes in units more versus less exposed in period T diverge over time. However, without additional information on relative exposure paths, it is difficult to make sense of their dynamic pattern. Figure D.1 provides a simple numerical illustration of two treated units with early- and late-exposure cases. In the late-exposure example of Figure D.1, the DiD coefficient is positive in period 1 despite all marginal effects being negative. The explanation is straightforward – in that period, region B is more exposed than region A, whereas in the reference period their relative ranking is reversed – but this reversal cannot be inferred from the dynamic impact and mean exposure graphs typically provided in dynamic DiD studies. In such cases, researchers should either provide additional evidence on how relative exposure evolves across units or use the leads-and-lags approach described in the next section, which aims to explicitly disentangle the dynamics of the treatment from those of its effects.

2.3.3 Empirical Diagnostic of Treatment Proportions

Based on insights from previous derivations, we provide guidance for detecting issues arising from using a constant measure for a dynamic shock. If researchers observe the treatment measure in every period, they can run a regression of m_{it} on m_{iT} for each period t. If the slope equals the average proportion $\bar{\delta}_t$, the constant-proportion assumption holds. If the shock is not observed in every period, the researcher should focus on the period

[23]For a shift-share instrument $z_i = \sum_k s_{ik} g_k$ that interacts baseline shares s and growth rates g for all subgroups k. The IV-DiD estimands are $\beta_\tau^{DiD-IV} = \sum_{l=0}^{L} b_l \left(\bar{\delta}_{t-l} + \text{Cov}(z_i, \tilde{\delta}_{it-l} m_{iT}) / \text{Cov}(z_i, m_{iT})\right)$. The additional covariance term vanishes only if z_i is uncorrelated with $\tilde{\delta}_{it-l} m_{iT}$. However, because static shift-share instruments predict treatment levels with proportional exposure, they do not eliminate this concern.

[24]This is also useful for understanding DiD coefficients when $t > T$. If treatment stabilizes, they capture marginal treatment effects. However, if treatment continues to grow or dissipate in *unit-varying* proportions, the DiD coefficients might not retain a clear interpretation, which is relevant because these longer-term coefficients are often used to indicate whether effects persist or stabilize over time (Autor et al., 2021).

where the shock is measured or argue why the shock might evolve in constant proportions across units and explicitly assume some form of mean independence assumption as $\mathbb{E}[\delta_{it}|m_{iT}] = \mathbb{E}[\delta_{it}]$.

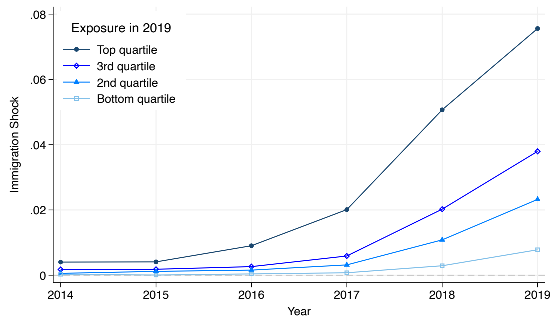
We now examine this issue in two studies. First, using evidence from an immigration shock in Colombia. Figure 2a shows how shock exposure – defined as Venezuelan-born migrants over total population – builds up over time across quartiles of the exposure distribution in $T = 2019$. Figure 2b normalizes the year-specific exposure by period-T exposure, showing that the shock amplifies more rapidly in departments with higher immigrant inflows in T (labeled as the early exposure case).[25] In this case, the most affected departments are close to the Venezuelan border, but as time passes, the shock gradually extends to more distant departments. This pattern is intuitive and likely applies to other settings involving immigration shocks. As a result, dynamic coefficients measured before the shock is realized will be overstated.

In a second context, we analyze labor demand shocks during the Great Recession using Yagan (2019), which estimates a dynamic DiD model with individual-level U.S. data. To start, we plot in Figure 2c the time series of the unemployment shock, and in Figure 2d the evolution of the unit-proportions of the shock. The latter suggests that the shock first amplifies in commuting zones that were more affected in 2009, as in the early-exposure case, and that these zones then recover more rapidly.

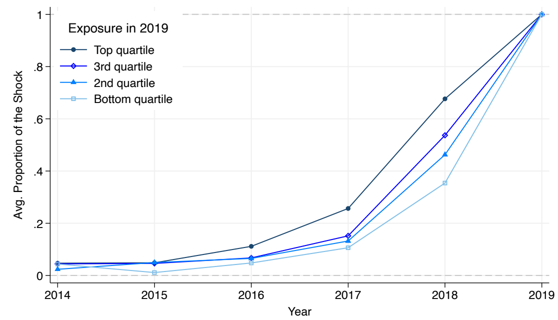
[25] Running the regression for post-treatment periods before the full realization of the shock reveals positive coefficients, yet the sample size is small ($N = 24$).

Figure 2: Shocks and Treatment Proportions by Exposure

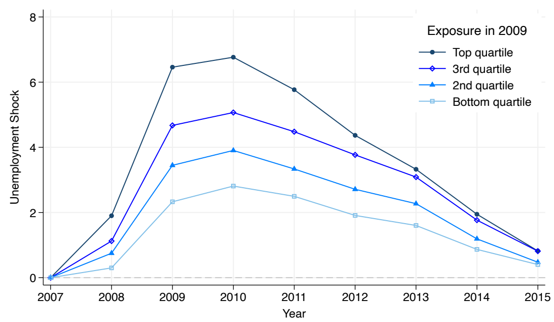
(a) Average Immigration Shock, Colombia



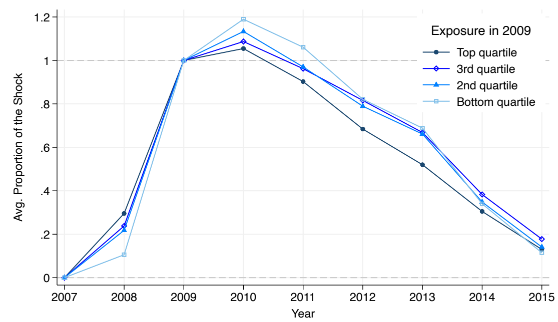
(b) Average Treatment Proportions, Colombia



(c) Average Unemployment Shock, US



(d) Average Treatment Proportions, US



Notes: For Colombia, we focus on the employed population aged 18–64 as in Delgado-Prieto (2024a). The numerator denotes Venezuelan-born migrants and the denominator the total population. We use departmental survey weights to aggregate microdata across 24 departments. Treatment proportions are based on exposure quartiles using the shock exposure in 2019. Source: Labor Force Survey (GEIH), 2013–2019. For the US, we use commuting-zone-level unemployment data from Yagan (2019). The full shock is measured as the percentage-point change in the unemployment rate between 2007 and 2009, and treatment proportions are defined relative to the observed shock over 2008–2015. The averages across quartiles are weighted by their 2007 population sizes.

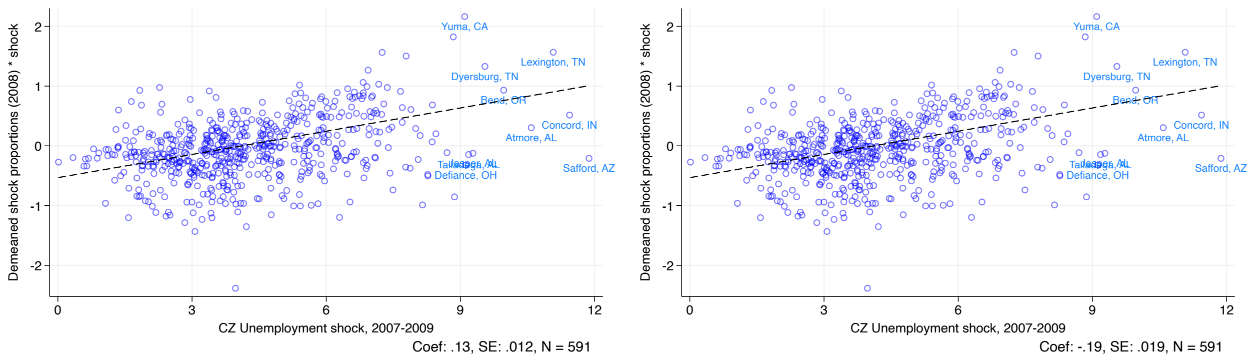
Figures 3a and 3b visually show the biases. The shock, measured between 2007 and 2009, enables us to assess treatments in selected years, such as 2008 and 2010 (the years preceding and following the shock). Figure 3a shows that in 2008, commuting zones more affected by the recession experienced a faster initial accumulation of their unemployment shock, implying the dynamic DiD coefficients for 2008 would be overstated. In contrast, after 2010, unemployment recovered more strongly in commuting zones that were more severely affected by the 2007–2009 shock, suggesting that the dynamic coefficients for 2010 and later periods are likely attenuated.[26] This intuition also matters for studies that use a continuous measure of exposure based on pre-reform wages relative to a future minimum wage level (Derenoncourt et al., 2025; Luduvice et al., 2024). In those settings, firms or local labor markets with higher baseline exposure see their gap to the future minimum

[26]For more details on the measurement of the shocks, see Delgado-Prieto (2024a) and Yagan (2019).

wage close more quickly, suggesting that the estimated post-treatment coefficients before the new minimum wage takes effect may be attenuated.

Figure 3: Illustration of Potential Bias for Unemployment Shocks

(a) Positive Bias for 2008 Coefficient (b) Negative Bias for 2010 Coefficient



Notes: We use the unemployment information at the commuting zone level from Yagan (2019). We refer the reader to that paper for more detailed data. The full shock is measured as the percentage-point difference in the unemployment rate between 2007 and 2009, and we then compare this shock with the demeaned shock ($\tilde{\delta}_{it}$) for 2008 and 2010. We weight the regressions by the 2007 population in the commuting zone.

An interesting application from this result is that we could construct the bias of the estimated DiD coefficients for each period with the general formula in Eq. (2.9), and directly account for it in our simulations: $\beta_\tau - \text{Bias}[\delta]$. While this could be a general solution, it relies on the assumption that the outcome model accurately reflects the true data-generating process. Since this is unverifiable, we propose an alternative solution below using a different regression model.

2.4 Panel Regression with Treatment Leads and Lags – Dynamic Shocks and Effects

An alternative, more robust approach to estimating the dynamic effects of continuous treatments is to extend panel regression by incorporating treatment leads and lags. This approach combines the benefits of static panel and dynamic DiD models, accurately measuring the dynamic shock and allowing for dynamic effects. This helps address the issues we raise in each approach individually. More formally, the estimating equation is the following:

$$y_{it} = \sum_{l=-L}^{-1} \beta_l \cdot m_{i,t+l} + \beta_0 \cdot m_{it} + \sum_{f=1}^{F} \beta_f \cdot m_{i,t+f} + \gamma_i + \gamma_t + \epsilon_{it}. \tag{M3}$$

This specification extends the panel regression in Eq. (M1) by including up to L leads and F lags of the dynamic shock m_{it}, thereby estimating lagged, current, and future treatment effects. This framework provides a closer approximation to the true outcome model, potentially giving more accurate estimates of the treatment effects than the dynamic DiD approach. This modeling strategy has been used in the applied economics literature (e.g., Autor, 2003; Fuest et al., 2018; Mello, 2022; Matsuzawa et al., 2025). In fact, Freyaldenhoven et al. (2021) proposes a similar specification, labeling it a linear panel model with dynamic policy effects, and recommends visualizing its coefficients in an "event-study" plot. Likewise, this specification resembles the distributed-lag models discussed in Schmidheiny and Siegloch (2023).[27]

Abstracting from the specification challenges of modeling treatment leads or not and determining how many leads and lags to include, which implicitly is some form of binning (Schmidheiny and Siegloch, 2023), this regression produces coefficients that lack an intuitive, visually appealing interpretation, unlike the dynamic DiD framework. To illustrate this, we estimate the panel regression in Eq. (M3) with treatment leads and lags, using a predefined selection of three leads and lags. Figure 4a displays the coefficients of this regression. These coefficients are not easily interpretable because they reflect the treatment's contemporaneous and lagged impacts. Based on the potential outcome Eq. (2.1), these coefficients would correspond to $\beta_0 = b_0$, $\beta_{-1} = b_1$, and $\beta_{-2} = b_2$. Hence, they do not accumulate lagged effects as in the dynamic DiD approach, which complicates interpretation. For instance, if we focus on the coefficient of the third lag, where the true effects are expected to be negative, the lagged coefficient transitions to zero with gradual treatment effects, or even becomes positive, with U-shaped treatment effects.

To address this, we apply a simple transformation to the leads/lags coefficients to produce an event-study plot. We treat the pre- and post-treatment periods symmetrically by using the post-treatment shock evolution to normalize both the post- and pre-treatment estimates, rather than using pre-treatment measures of the shock, which have little variation, to test the design validity. We focus on the average treated unit and use its treatment evolution from the baseline period $t = 0$ to the furthest post-treatment period we analyze $t = 7$.[28] We choose the same number of periods as in the Dynamic DiD for comparison, but this is a flexible choice. More formally, the average post-treatment effect in t is de-

[27] They demonstrate the equivalence of the coefficients between such models and event-study specifications in settings with staggered treatment timing that we do not address. Importantly, they discuss the importance of "binning" treatment effects before or after an event time to recover them in settings without never-treated units.

[28] This representation differs from the scaled effects obtained in Freyaldenhoven et al. (2021) that are based on a regression using dynamic first differences of the treatment variable.

fined as $\beta_t^{Lags,post} = \sum_{l=0}^{\min(L,t)} \beta_l * \overline{m}_{t-l}$, where $\overline{m}_t$ is the average continuous treatment in t. Intuitively, we sum the lags of the panel coefficients, scaled by the average treatment, to obtain an average effect. For instance, the average effect in $t = 0$ is $\beta_0 * \overline{m}_{t=0}$, in $t = 1$ it is $\beta_0 * \overline{m}_{t=1} + \beta_{-1} * \overline{m}_{t=0}$ and in $t = 2$ it is $\beta_0 * \overline{m}_{t=2} + \beta_{-1} * \overline{m}_{t=1} + \beta_{-2} * \overline{m}_{t=0}$.

Conversely, we construct each pre-treatment period t average effect as $\beta_t^{\text{Leads,pre}} = \sum_{f=1}^{\min(F,-t)} \beta_f \, \overline{m}_{t+f}$, and no pre-period needs to be omitted as in the dynamic DiD.[29] As an example, in $t = -1$ it is $\beta_1 * \overline{m}_{t=0}$, and in $t = -2$ it is $\beta_1 * \overline{m}_{t=1} + \beta_2 * \overline{m}_{t=0}$. This transformation then allows for a direct comparison with the true average effects. We focus on the continuous treatment evolving in non-proportional exposure (late exposure), where the dynamic DiD attenuates the estimated impacts, and show in Figure 4b that the panel treatment leads-and-lags regression yields coefficients that closely recover the true average effect.

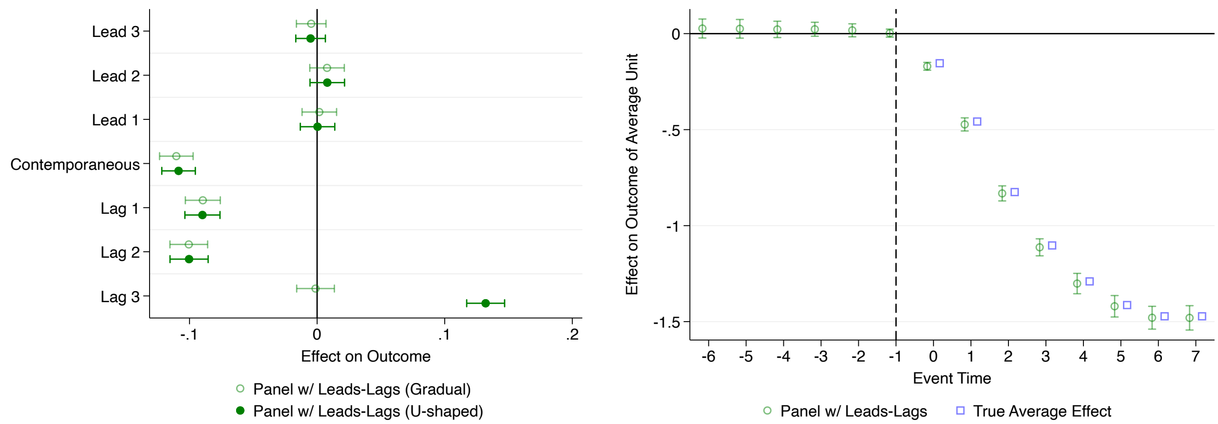
However, the selection of leads and lags depends on assumptions about the underlying outcome model, which is a potential drawback of this approach. If the specification fails to capture the remaining dynamic responses, the resulting estimates do not accurately weight the lagged effects, as we show for the static panel coefficient.[30] Furthermore, when treatments evolve in unit constant proportions, the leads and lags of m_{it} are linear combinations of one another because $m_{it-l} = \left(\frac{\delta_{t-l}}{\delta_t}\right) m_{it}$. Identifying their effects, therefore, requires time variation in δ_t, often with reduced precision in the estimated effects, whereas dynamic DiD performs well in this setting.

[29] We use tools like `xlincom` or similar commands in `Stata` to transform the coefficients.

[30] For instance, in the case of U-shaped effects of Table 1, the assumption of three lags would aggregate the dynamic effects of future periods into one in the third lag coefficient, so you would need to include an additional lag to capture the entire path of dynamic effects accurately.

Figure 4: Simulated Effects with Treatment Leads and Lags

(a) Marginal Effects (b) Average Effects



Notes: We focus on gradual and U-shaped effects in (a), while in (b) we only focus on gradual effects that follow the pattern: $y_{it} = \sum_{k=0}^{2} -0.1 * m_{it-k} + \epsilon_{it}$. In both panels, we focus on shocks that evolve with non-proportional exposure (late exposure), where units with higher exposure at T experience a later amplification of their shocks. We construct average effects in (b) based on the transformation described in Table 2. The analysis is for a continuous dynamic treatment that begins in period 0, evolves with non-proportional exposure across units through period 4, and reaches its full scale by that point. The full shock is drawn from $m_{iT} \sim U(0,1) * 0.1$, and the random component is modeled as $\epsilon_{it} = 5 + U(0,1) * 0.01$. The simulation is conducted with 1,000 units and spans 14 periods, divided into 6 pre- and 8 post-treatment periods. We use 95% confidence intervals.

2.5 Comparison of Approaches

We present a figure comparing results from our three main regressions—static panel, dynamic DiD, and panel with treatment leads and lags—highlighting the effect for the average unit. To ensure comparability, we multiply the regression coefficients by the average shock, as described in Table 2.

Table 2: Coefficient Transformation

Approach	Effects on the Average Unit
Static Panel (M1)	$\beta * \overline{m}_t$
Dynamic DiD (M2)	$\beta_t * \overline{m}_T$
Panel with Lags (M3)	$\sum_{l=0}^{\min(L,t)} \beta_l \, \overline{m}_{t-l}$
True Average Effect	$\sum_{k=0}^{2} -0.1 * \overline{m}_{t-k}$

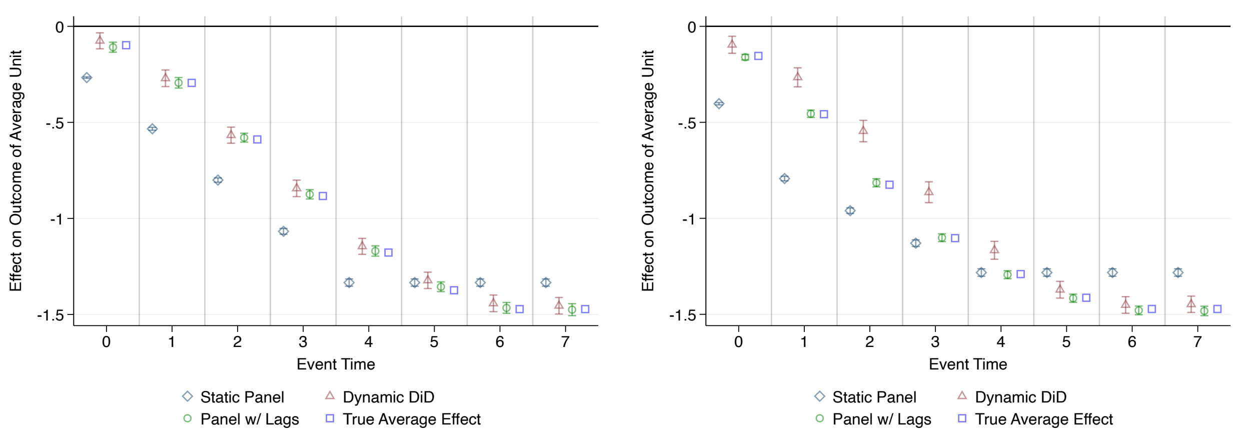
Notes: The coefficients are scaled using the average treatment at time t, $\overline{m}_t$, or the full treatment value $\overline{m}_T$, depending on the model specification. The panel with leads incorporates lagged average treatments, with l lags. The true average effect is constructed based on the average treated unit. All post-treatment coefficients are transformed into these effects for Figures 5a and 5b.

With this transformation, Figure 5a presents estimates from different approaches in a standardized manner. First, we find that the static panel approach fails to capture dynamic effects, as it overstates the effect in the initial periods and subsequently underestimates it in later periods. Second, when we use a shock evolving at constant proportions, both the dynamic DiD and the panel with treatment leads-and-lags regressions yield estimates that closely approximate the true average effect, indicating that the dynamic DiD performs well in these cases. In contrast, Figure 5b shows that when the shock evolves with non-proportional exposure, and units with higher exposure in T face later exposure to their shock, the dynamic DiD regression attenuates the effects, whereas the panel with leads and lags recovers the true average effect. These results summarize our main points, indicating the importance of analyzing the treatment paths when specifying dynamic DiD models, while the panel with lags serves as an alternative when this does not hold.[31]

Figure 5: Simulated Effect on the Outcome of Average Unit by Approach

(a) Shock Evolving Proportional

(b) Shock Evolving Non-Proportional



Notes: A detailed explanation of the transformation of the coefficients is presented in Table 2. We focus on gradual effects in these figures that follow the pattern: $y_{it} = \sum_{k=0}^{2} -0.1 * m_{it-k} + \epsilon_{it}$, and on the shock with proportional exposure in (a) and non-proportional exposure (late exposure) in (b). The analysis is based on a continuous dynamic shock that begins in period 0, amplifies through period 4, and reaches its full scale by that point. The full shock is drawn from $m_{iT} \sim U(0,1) * 0.1$, and the random component is modeled as $\epsilon_{it} = 5 + U(0,1) * 0.01$. The simulation is conducted with 1,000 units and spans 14 periods, divided into 6 pre- and 8 post-treatment periods.

[31] Including a time-varying treatment to the dynamic DiD model in Eq. (M2LD), instead of the constant realized shock, does not resolve the points we raise. It omits pre-trend testing and, while capturing marginal effects, it overlooks the timing and scale of the dynamic shocks, thereby complicating the interpretation of post-treatment trends.

3 Heterogeneous Effects and Modern DiD Estimators

So far, we have focused on complications that arise even under the assumption of homogeneous treatment effects across units, a common assumption in related models (Freyaldenhoven et al., 2021). We now turn to the role of heterogeneous treatment effects, a central concern in the recent DiD literature. Initial concerns about TWFE regressions emerged in settings with binary treatments and staggered timing, where the estimator can rely on "forbidden comparisons" between early- and late-treated groups (Roth et al., 2023; Baker et al., 2025). Subsequent work has shown that related issues also arise in continuous treatments, in which the static TWFE estimator can assign negative weights to the treatment effects of particular *group-time* cells (De Chaisemartin and d'Haultfoeuille, 2023). These concerns become particularly relevant when treatment effects are heterogeneous or nonlinear, as the TWFE estimate may fail to satisfy the "no-sign reversal" property, meaning that negative weights on groups with negative treatment effects could in principle lead to a positive overall estimate, although this appears to be rare in practice (Chiu et al., 2023) and not relevant for "design-based" specifications (Borusyak and Hull, 2024).

Therefore, we now analyze how the approaches we study perform under heterogeneous effects across units. This is relevant because the existing DiD literature surveyed in De Chaisemartin and d'Haultfoeuille (2023) has not yet fully addressed the case of continuous treatments that phase in gradually, particularly when all units receive different treatment intensities simultaneously. The most recent papers have proposed robust estimators for continuous treatments when units with no treatment dose (or *stayers*) are present. Specifically, Callaway et al. (2024) and De Chaisemartin and d'Haultfoeuille (2024) have introduced estimators that accommodate continuous treatments and are robust to dynamic and heterogeneous effects.[32] We do not compare these estimators, as they rely on different parallel trends assumptions and use various technical approaches for identification and aggregation of treatment effects. Instead, we use the one closest to our setting to illustrate, more concretely, how heterogeneity in TEs may bias the panel with leads and lags.

We first formally characterize the sources of bias arising from heterogeneous TEs in Appendix C. The derivations show that both the static panel and dynamic DiD estimators can be expressed as the sum of an average treatment effect $\bar{b}$ and an additional term arising from mean deviations in the marginal effects $\tilde{b}_i$. In particular, each contemporaneous and lagged effect now includes a "bias" term of the form $\frac{\mathrm{Cov}(m_{it},\tilde{b}_{il}m_{it-l})}{\mathrm{Var}(m_{it})}$ in the static panel,

[32]These estimators propose different frameworks for estimating causal effects and are suited to different treatment types. Callaway et al. (2024) focuses on staggered treatment settings in which the treatment dose remains constant after the initial exposure, whereas De Chaisemartin and d'Haultfoeuille (2024) allows for continuous treatments that may increase or decrease over time.

whereas for dynamic DiD, the same expression applies after replacing m_{it} with m_{iT}. This highlights that bias arises when heterogeneous TEs are correlated with treatment exposure. In practice, nonlinear treatment effects satisfy this condition, for instance, increasing immigrant exposure from 1 to 2 percent may have little impact, while moving from 9 to 10 percent could generate a larger effect.[33]

We adapt our simulations to these conditions and then illustrate how heterogeneity can pose a challenge to the recovery of average treatment effects under our potential outcome framework. For this analysis, we compare the results with the estimator proposed by De Chaisemartin and d'Haultfoeuille (2024), which is robust to these concerns and accommodates the type of continuous dynamic shocks we analyze. We need to include units in our simulations that never receive any treatment dose to use that estimator.[34]

Figure 6a shows that when treatment effect heterogeneity is uncorrelated with treatment exposure, the panel regression with leads and lags approach remains unbiased and recovers the average treatment effect on the treated across all periods. However, once the heterogeneity becomes positively correlated with exposure, the panel with lead estimates becomes biased, while the estimator of De Chaisemartin and d'Haultfoeuille (2024) continues to recover the true average effect (see Figure 6b). This indicates that not all forms of heterogeneity pose problems for the panel regression with treatment leads and lags, and that the modern DiD estimator remains robust in such cases.

A potential way to address this is to extend the model in Eq. (M3) to allow for some form of group- or covariate-specific heterogeneity based on final treatment exposure, following the spirit of extended TWFE estimators in Wooldridge (2021), which would then yield unbiased estimates of the treatment effects. Specifically, the model could allow treatment effects to vary across exposure groups g_i by estimating separate regression of the form: $y_{it} = \sum_{l=-L}^{-1} \beta_{lg} m_{i,t+l} + \beta_{0g} m_{it} + \sum_{f=1}^{F} \beta_{fg} m_{i,t+f} + \gamma_i + \gamma_t + \epsilon_{it}$, or by interacting

[33]A similar conclusion arises when discussing heterogeneous effects of the minimum wage (Dube and Lindner, 2024, p. 29).

[34]More formally, the non-normalized estimator compares the evolution of outcomes from the pre-treatment period τ to $\tau + \ell$ for treated units i against that of units j that have not changed their baseline treatment (or did not receive any treatment). Specifically,

$$\text{DID}_{i,\ell} = Y_{i,\tau+\ell} - Y_{i,\tau} - \frac{1}{\mathcal{N}_{\tau+\ell}} \sum_{j \in \mathcal{N}_{\tau+\ell}} \left(Y_{j,\tau+\ell} - Y_{j,\tau} \right),$$

where $\mathcal{N}_{\tau+\ell}$ denotes the number of control units in each ℓ. Finally, the aggregated estimate at horizon ℓ is computed as

$$\text{DID}_{\ell} = \frac{1}{N_\ell} \sum_{i:\tau+\ell \leq T_i} S_i \cdot \text{DID}_{i,\ell},$$

where S_i is equal to one (-1) for treatment that increases (decreases), and N_ℓ is the number of units for which the effect can be estimated up to period T_i.

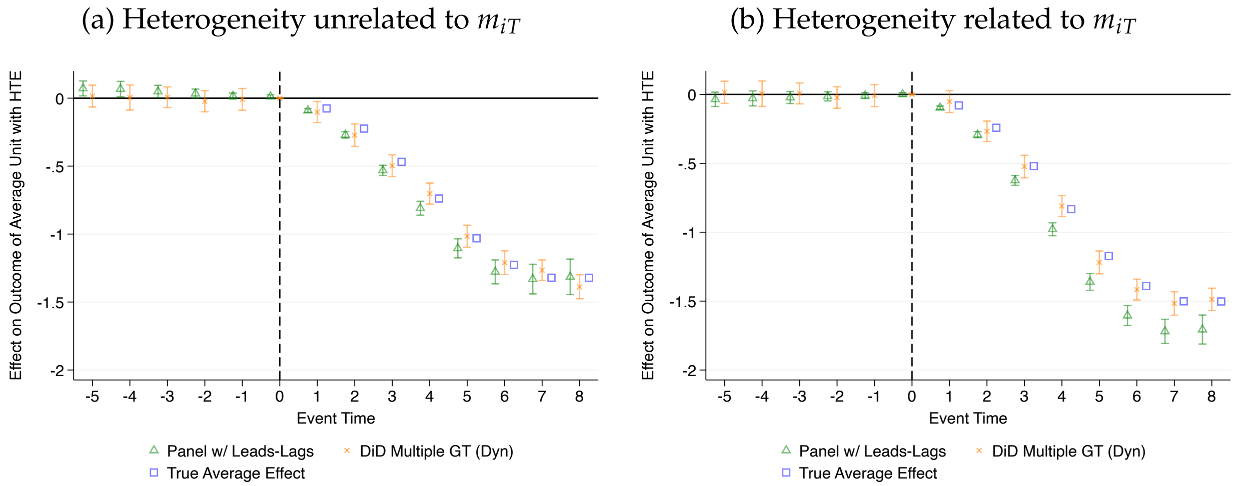
the leads and lags with the exposure indicators g_i.[35] Alternatively, de Chaisemartin et al. (2025) develops a test for homogeneous linear treatment effects in TWFE regressions that accommodates continuous shocks, which provides a validation exercise for the panel with treatment leads and lags.

With this in mind, we emphasize that our setting involves shocks, such as immigration or trade shocks, which potentially affect all units, limiting the use of modern DiD estimators.[36] Moreover, many papers using this type of shocks use it in conjunction with instrumental variables, which limits the applicability of similar estimators. In these settings, a panel regression with treatment leads and lags can flexibly accommodate group heterogeneity, thereby addressing additional concerns about heterogeneous effects of dynamic shocks. Dynamic DiD, in contrast, tends to perform better with sudden shocks or when treatment evolves in constant proportions, whereas both approaches are more robust than static panel regression when treatment effects are dynamic.

[35] Gardner et al. (2025) proposes an alternative way to address heterogeneity. Instead of modifying the baseline specification, one can estimate the standard model after transforming the outcome variable. This involves first-stage regression of the outcome on group and unit fixed effects using only the control units, then predicting the outcomes for all units, subtracting them from the observed outcomes, and using the adjusted outcomes in the main estimation. In this case, one would need groups with both treatment and control units, unlike for immigration shocks, where all units receive some treatment dose.

[36] An alternative estimator is the one proposed by de Chaisemartin et al. (2025), which considers designs where no unit remains untreated and constructs quasi-untreated units using methods inspired by regression discontinuity designs.

Figure 6: Simulated Effects Under Heterogeneous Effects Across Units



Notes: DiD Multiple GT (Dyn) refers to the non-normalized estimator proposed by De Chaisemartin and d'Haultfoeuille (2024). In these figures, we focus on the shock evolving with non-proportional exposure (late exposure) and gradual effects that follow the pattern: $y_{it} = \sum_{k=0}^{2} -b_i * m_{it-k} + \epsilon_{it}$. The parameter b_i is drawn from the distribution $b_i \sim |N(\mu, 0.2^2)| * 0.1$, where $\mu \in \{0.6, 0.8, 1, 1.2\}$ is randomly assigned across units. In the right panel, μ increases with exposure: units in the highest exposure quartile (Q_4) are assigned $\mu = 1.2$, while those in the lowest (Q_1) receive $\mu = 0.6$. The analysis is based on a continuous dynamic shock that begins in period 0, amplifies through period 4, and reaches its full scale by that point. The full shock is drawn from $m_{iT} \sim U(0,1) * 0.1$, and the random component is modeled as $\epsilon_{it} = 5 + U(0,1) * 0.01$. The simulation is conducted with 1,000 units and spans 14 periods, divided into 6 pre- and 8 post-treatment periods.

4 Conclusion

In this paper, we examine common approaches to quantify the effects of continuous dynamic treatments and highlight potential pitfalls. We first show that the traditional panel regression, widely used across studies, provides no suggestive evidence on the validity of the research design. Moreover, it does not accurately aggregate the dynamic responses, leading to a biased estimate. We then assess a regression that incorporates dynamic responses, typically defined as dynamic DiD. We argue that this approach more accurately captures dynamic responses. However, using a single constant measure of a dynamic shock implicitly assumes a proportional exposure across units. In practice, the shock amplification may vary across units. In two examples, immigration and labor demand shocks, we show that units more exposed during the period when the shock is measured commonly experience an earlier exposure to the shock than units less exposed. In those cases, the dynamic DiD coefficients become harder to interpret and might not fully recover the underlying treatment effects.

To validate these concerns, we simulate a range of dynamic effects and identify cases where bias arises. We then propose an alternative estimator that includes treatment leads and lags in the panel regression and find that it accounts for the dynamics of the effects and the shocks. Importantly, we highlight that these issues arise even when TEs are homogeneous across units. Yet we also examine the potential implications of relaxing this assumption. We compare the panel regression with leads and lags to estimates from a modern DiD estimator, which is robust to heterogeneity after including non-treated control units in our simulations, and identify conditions under which this can be problematic and propose solutions for heterogeneity in the leads-and-lags panel regression. Overall, our work contributes to the empirical estimation of continuous dynamic treatments, a common setting in the applied economics literature.

References

AKSU, E., R. ERZAN, AND M. G. KIRDAR (2022): "The impact of mass migration of Syrians on the Turkish labor market," *Labour Economics*, 76, 102183. (Cited on p. 2, 8, and 41)

ALTONJI, J. G. AND D. CARD (1991): "The Effects of Immigration on the Labor Market Outcomes of Less-skilled Natives," in *Immigration, Trade, and the Labor Market*, University of Chicago Press, 201–234. (Cited on p. 8)

AUTOR, D., D. DORN, G. H. HANSON, M. R. JONES, AND B. SETZLER (2025): "Places versus people: the ins and outs of labor market adjustment to globalization," Tech. rep., National Bureau of Economic Research. (Cited on p. 1 and 12)

AUTOR, D. H. (2003): "Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing," *Journal of labor economics*, 21, 1–42. (Cited on p. 4 and 21)

AUTOR, D. H., D. DORN, AND G. H. HANSON (2013): "The China syndrome: Local labor market effects of import competition in the United States," *American economic review*, 103, 2121–2168. (Cited on p. 8)

——— (2021): "On the Persistence of the China Shock," *Brookings Papers on Economic Activity*, 2021, 381–476. (Cited on p. 13 and 17)

AUTOR, D. H., D. DORN, G. H. HANSON, AND J. SONG (2014): "Trade adjustment: Worker-level evidence," *The Quarterly Journal of Economics*, 129, 1799–1860. (Cited on p. 13)

AYDEMIR, A. AND G. J. BORJAS (2011): "Attenuation bias in measuring the wage impact of immigration," *Journal of Labor Economics*, 29, 69–112. (Cited on p. 2)

BAKER, A., B. CALLAWAY, S. CUNNINGHAM, A. GOODMAN-BACON, AND P. H. SANT'ANNA (2025): "Difference-in-differences designs: A practitioner's guide," *arXiv preprint arXiv:2503.13323*. (Cited on p. 2, 8, 12, and 25)

BONILLA-MEJÍA, L., L. F. MORALES, D. HERMIDA, AND L. A. FLÓREZ (2024): "The labor market effect of south-to-south migration: Evidence from the venezuelan crisis," *International Migration Review*, 58, 764–799. (Cited on p. 8, 13, and 41)

BORJAS, G. J. (1999): "The economic analysis of immigration," *Handbook of labor economics*, 3, 1697–1760. (Cited on p. 2)

BORUSYAK, K. AND P. HULL (2024): "Negative weights are no concern in design-based specifications," in *AEA Papers and Proceedings*, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, vol. 114, 597–600. (Cited on p. 5 and 25)

BRATSBERG, B. AND O. RAAUM (2012): "Immigration and wages: Evidence from construction," *The economic journal*, 122, 1177–1205. (Cited on p. 8 and 41)

CALLAWAY, B., A. GOODMAN-BACON, AND P. H. SANT'ANNA (2024): "Difference-in-differences with a continuous treatment," Tech. rep., National Bureau of Economic Research. (Cited on p. 5 and 25)

CARD, D. (1992): "Using regional variation in wages to measure the effects of the federal minimum wage," *Ilr Review*, 46, 22–37. (Cited on p. 8)

CARRINGTON, W. J. AND P. J. DE LIMA (1996): "The impact of 1970s repatriates from Africa on the Portuguese labor market," *Ilr Review*, 49, 330–347. (Cited on p. 1 and 8)

CARUSO, G., C. G. CANON, AND V. MUELLER (2021): "Spillover effects of the Venezuelan crisis: migration impacts in Colombia," *Oxford Economic Papers*, 73, 771–795. (Cited on p. 2, 8, 13, and 41)

CHIU, A., X. LAN, Z. LIU, AND Y. XU (2023): "Causal panel analysis under parallel trends: lessons from a large reanalysis study," *American Political Science Review*, 1–22. (Cited on p. 25)

CURRIE, J., H. KLEVEN, AND E. ZWIERS (2020): "Technology and big data are changing economics: Mining text to track methods," in *AEA Papers and Proceedings*, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, vol. 110, 42–48. (Cited on p. 2)

DE CHAISEMARTIN, C., D. CICCIA, X. D'HAULTFOEUILLE, AND F. KNAU (2025): "Difference-in-Differences Estimators When No Unit Remains Untreated," SSRN Working Paper, available at SSRN. (Cited on p. 27)

DE CHAISEMARTIN, C. AND X. D'HAULTFOEUILLE (2024): "Difference-in-differences estimators of intertemporal treatment effects," *Review of Economics and Statistics*, 1–45. (Cited on p. 1, 5, 25, 26, and 28)

DE CHAISEMARTIN, C. AND X. D'HAULTFOEUILLE (2023): "Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey," *The econometrics journal*, 26, C1–C30. (Cited on p. 6, 8, and 25)

DELGADO-PRIETO, L. (2024a): "Immigration, wages, and employment under informal labor markets," *Journal of Population Economics*, 37, 55. (Cited on p. 2, 4, 13, 19, and 41)

——— (2024b): "Worker Responses to Immigration across Firms: Evidence from Colombia," Tech. rep. (Cited on p. 41)

DERENONCOURT, E., F. GERARD, L. LAGOS, AND C. MONTIALOUX (2025): "Minimum Wages and Informality," Tech. rep., National Bureau of Economic Research. (Cited on p. 13 and 19)

DUBE, A. AND A. LINDNER (2024): "Minimum wages in the 21st century," *Handbook of Labor Economics*, 5, 261–383. (Cited on p. 1, 8, 12, and 26)

DUSTMANN, C. AND U. SCHÖNBERG (2025): "Linking Empirical Evidence to Theory: A Framework for Understanding Immigration's Labor Market Effects," . (Cited on p. 1 and 12)

DUSTMANN, C., U. SCHÖNBERG, AND J. STUHLER (2017): "Labor supply shocks, native wages, and the adjustment of local employment," *The Quarterly Journal of Economics*, 132, 435–483. (Cited on p. 2, 11, 13, 14, and 41)

EDO, A. (2020): "The Impact of immigration on wage dynamics: Evidence from the Algerian independence war," *Journal of the European Economic Association*, 18, 3210–3260. (Cited on p. 13 and 41)

FOGED, M. AND G. PERI (2016): "Immigrants' effect on native workers: New analysis on longitudinal data," *American Economic Journal: Applied Economics*, 8, 1–34. (Cited on p. 13 and 41)

FREYALDENHOVEN, S., C. HANSEN, J. P. PÉREZ, AND J. M. SHAPIRO (2021): "Visualization, identification, and estimation in the linear panel event-study design," Tech. rep., National Bureau of Economic Research. (Cited on p. 1, 3, 21, and 25)

FUEST, C., A. PEICHL, AND S. SIEGLOCH (2018): "Do higher corporate taxes reduce wages? Micro evidence from Germany," *American Economic Review*, 108, 393–418. (Cited on p. 2, 8, and 21)

GARDNER, J., N. THAKRAL, L. T. TÔ, AND L. YAP (2025): "Two-Stage Differences in Differences," *Mimeo*. (Cited on p. 27)

GROEGER, A., G. LEÓN-CILIOTTA, AND S. STILLMAN (2024): "Immigration, labor markets and discrimination: Evidence from the Venezuelan Exodus in Perú," *World Development*, 174, 106437. (Cited on p. 2, 8, and 41)

JAEGER, D. A., J. RUIST, AND J. STUHLER (2019): "Shift-share Instruments and Dynamic Adjustments: The Case of Immigration," Tech. rep., mimeo. (Cited on p. 4)

KAHN-LANG, A. AND K. LANG (2020): "The Promise and Pitfalls of Differences-in-Differences: Reflections on 16 and Pregnant and Other Applications," *Journal of Business & Economic Statistics*, 38, 613–620. (Cited on p. 3)

LEBOW, J. (2022): "The labor market effects of Venezuelan migration to Colombia: reconciling conflicting results," *IZA Journal of Development and Migration*, 13, 1–49. (Cited on p. 8, 13, and 41)

LUDUVICE, A. V. D., T. R. MARTINEZ, AND A. B. SOLLACI (2024): "Minimum wage, business dynamism, and the life cycle of firms," Tech. rep., IDB Working Paper Series. (Cited on p. 13 and 19)

MATSUZAWA, K., D. I. REES, J. J. SABIA, AND R. MARGOLIT (2025): "Minimum Wages and Teenage Childbearing in the United States," *Journal of Applied Econometrics*, 40, 471–484. (Cited on p. 21)

MEER, J. AND J. WEST (2016): "Effects of the minimum wage on employment dynamics," *Journal of Human Resources*, 51, 500–522. (Cited on p. 8)

MELLO, U. (2022): "Centralized Admissions, Affirmative Action and Access of Low-income Students to Higher Education," *American Economic Journal: Economic Policy*, 14, 166–197. (Cited on p. 4 and 21)

MONRAS, J. (2020): "Immigration and wage dynamics: Evidence from the mexican peso crisis," *Journal of Political Economy*, 128, 3017–3089. (Cited on p. 13 and 41)

MUÑOZ, M. (2024): "Trading nontradables: The implications of europe's job-posting policy," *The Quarterly Journal of Economics*, 139, 235–304. (Cited on p. 13 and 41)

NEUMARK, D., J. I. SALAS, AND W. WASCHER (2014): "Revisiting the minimum wage—Employment debate: Throwing out the baby with the bathwater?" *Ilr Review*, 67, 608–648. (Cited on p. 8)

RAMBACHAN, A. AND J. ROTH (2023): "A more credible approach to parallel trends," *Review of Economic Studies*, 90, 2555–2591. (Cited on p. 3)

REDONDO, H. (2022): "From bricklayers to waiters: Reallocation in a deep recession," . (Cited on p. 13)

ROTH, J. (2022): "Pretest with caution: Event-study estimates after testing for parallel trends," *American Economic Review: Insights*, 4, 305–322. (Cited on p. 2 and 3)

ROTH, J., P. H. SANT'ANNA, A. BILINSKI, AND J. POE (2023): "What's trending in difference-in-differences? A synthesis of the recent econometrics literature," *Journal of Econometrics*, 235, 2218–2244. (Cited on p. 2, 5, 6, 8, 11, and 25)

SCHMIDHEINY, K. AND S. SIEGLOCH (2023): "On event studies and distributed-lags in two-way fixed effects models: Identification, equivalence, and generalization," *Journal of Applied Econometrics*, 38, 695–713. (Cited on p. 1, 3, and 21)

SUN, L. AND S. ABRAHAM (2021): "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," *Journal of econometrics*, 225, 175–199. (Cited on p. 6)

WOOLDRIDGE, J. M. (2021): "Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators," *Available at SSRN 3906345*. (Cited on p. 5 and 26)

YAGAN, D. (2019): "Employment hysteresis from the great recession," *Journal of Political Economy*, 127, 2505–2558. (Cited on p. 4, 13, 18, 19, and 20)

Appendix

A First Differences

Since first differences are also used to study continuous treatments, we show that they face similar aggregation issues to those in the static panel when treatment effects are dynamic.[37] Starting from the first-difference estimand and substituting the potential outcome model in Eq. (2.1), we obtain:

$$\beta_{FD} = \frac{\mathrm{Cov}\Big(\Delta m_{it}, \sum_{l=0}^{L} b_l \, \Delta m_{i,t-l} + \Delta\varepsilon_{it}\Big)}{\mathrm{Var}(\Delta m_{it})} = \sum_{l=0}^{L} b_l \, \omega_l^{\Delta}, \tag{A.1}$$

where $\omega_l^{\Delta} = \mathrm{Cov}(\Delta m_{it}, \Delta m_{i,t-l}) / \mathrm{Var}(\Delta m_{it})$ follows the same logic as ω_l in Proposition 1. Because first differences focus on period-to-period changes rather than levels, the weights ω_l^{Δ} can take a wider range of values than ω_l, and can even turn negative with treatments that increase over time, making β_{FD} potentially more sensitive to the specific pattern of treatment dynamics.

B Dynamic DiD Coefficients Under a Binary Treatment

A common practice in applied research is to binarize a continuous treatment by splitting it into two groups, for instance, using the median exposure. Define $D_{iT} = \mathbf{1}(m_{iT} > \mathrm{med}(m_{iT}))$. The dynamic DiD estimands from regressing $y_{it} - y_{i,t_{pre}}$ on D_{iT} are:

$$\beta_{\tau}^{B} = \frac{\mathrm{Cov}\left(D_{iT}, y_{it} - y_{i,t_{pre}}\right)}{\mathrm{Var}(D_{iT})}. \tag{B.1}$$

Under Assumptions 1–2 and the potential outcome model in Eq. (2.1):

$$\beta_{\tau}^{B} = \sum_{l=0}^{L} b_l \frac{\mathrm{Cov}(D_{iT}, m_{it-l})}{\mathrm{Var}(D_{iT})} = \sum_{l=0}^{L} b_l \Big(\mathbb{E}[m_{it-l} \mid D_{iT} = 1] - \mathbb{E}[m_{it-l} \mid D_{iT} = 0]\Big) \equiv \sum_{l=0}^{L} b_l \, \Delta_{t-l}, \tag{B.2}$$

where the second equality uses the standard property of regressing on a binary variable. The coefficient β_{τ}^{B} therefore has a natural interpretation under constant treatment proportions: it captures the outcome difference between above- and below-median units, with each marginal effect b_l weighted by Δ_{t-l}, the difference in mean treatment at lag l between the two groups.

[37] It is well known that static panel and first-difference regressions yield the same coefficient when $T = 2$.

While this group-contrast interpretation is transparent, the binarization introduces complications. For instance, by collapsing exposure into two groups, the specification discards within-group variation in treatment intensity. But more fundamentally, when treatment exposure profiles are heterogeneous across units, the composition of the above- and below-median groups may differ in their relative exposure over time, making the coefficient harder to interpret. In this case, Δ_{t-l} reflects not only the treatment contrast between high- and low-exposure units in T, but also the differential treatment evolution within groups over time. This introduces a source of bias due to heterogeneous exposure profiles, similarly to the term $\text{Bias}[\tilde{\delta}_{t-l}]$ in Eq. (2.9), and hence in the estimated dynamic treatment effects.

Moreover, the magnitude of Δ_{t-l} depends on the distribution of treatment.[38] For instance, in settings where the treatment distribution has heavy tails, the above-median group may have a substantially higher conditional mean, so that the coefficients β_τ^B are bigger, while the marginal treatment effects b remain unchanged.

C Introducing Heterogeneous Treatment Effects Across Units

C.1 Static Panel

We now allow for marginal treatment effects that vary across units and over time. Extending the potential outcome model in Eq. (2.1) with heterogeneity, the OLS estimand from Eq. (M1) is given by:

$$\beta_{static} = \frac{\text{Cov}(m_{it}, y_{it})}{\text{Var}(m_{it})} = \frac{\text{Cov}\left(m_{it}, \sum_{l=0}^{L} b_{il}\, m_{it-l} + \varepsilon_{it}\right)}{\text{Var}(m_{it})}. \tag{C.1}$$

Under the Parallel Trends Assumption 2, m_{it} is uncorrelated with the error term. Writing the marginal effects as $b_{il} = \bar{b}_l + \tilde{b}_{il}$, where $\bar{b}_l$ is the average effect at lag l and $\tilde{b}_{il}$ is the unit-specific deviation, we obtain:

$$\beta_{static} = \sum_{l=0}^{L} \bar{b}_l \frac{\text{Cov}(m_{it}, m_{it-l})}{\text{Var}(m_{it})} + \sum_{l=0}^{L} \frac{\text{Cov}(m_{it}, \tilde{b}_{il} m_{it-l})}{\text{Var}(m_{it})}. \tag{C.2}$$

Defining the weights $\omega_l = \frac{\text{Cov}(m_{it}, m_{it-l})}{\text{Var}(m_{it})}$, the static panel coefficient can be written

[38] Under a uniform distribution for m_{iT}, Δ_{t-l} equals half the support range, so the gap between groups is symmetric and determined entirely by the bounds of the distribution. Under a normal distribution, Δ_{t-l} scales with the standard deviation of treatment. Under a log-normal distribution, the conditional mean of the above-median group is pulled upward by the right tail, making Δ_{t-l} larger and more sensitive to extreme values, so that β_τ^B is disproportionately driven by heavily treated units.

compactly as:

$$\beta_{static} = \sum_{l=0}^{L} \bar{b}_l \, \omega_l + \sum_{l=0}^{L} \frac{\text{Cov}(m_{it}, \tilde{b}_{il} m_{it-l})}{\text{Var}(m_{it})}. \tag{C.3}$$

Thus, with heterogeneous treatment effects across units, the static coefficient captures not only a weighted average of contemporaneous and lagged average marginal effects, but also an additional term that depends on the covariance between treatment exposure and deviations from those average effects. When heterogeneity is correlated with treatment exposure, the static coefficient no longer recovers the average treatment effect. The same complication would also apply to the panel with treatment leads and lags, unless the model is extended to allow for systematic heterogeneity in treatment effects.

C.2 Dynamic DiD

We now analyze heterogeneous marginal treatment effects under the dynamic DiD framework. Using Eq. (M2LD), the estimands in long differences are:

$$\beta_\tau^{DiD} = \frac{\text{Cov}\left(m_{iT}, \sum_{l=0}^{L} b_{il} \, m_{it-l} + (\varepsilon_{it} - \varepsilon_{i,t_{pre}})\right)}{\text{Var}(m_{iT})}. \tag{C.4}$$

Under the Parallel Trends Assumption 2, m_{iT} is uncorrelated with $(\varepsilon_{it} - \varepsilon_{i,t_{pre}})$. Again writing $b_{il} = \bar{b}_l + \tilde{b}_{il}$, we have:

$$\beta_\tau^{DiD} = \sum_{l=0}^{L} \bar{b}_l \frac{\text{Cov}(m_{iT}, m_{it-l})}{\text{Var}(m_{iT})} + \sum_{l=0}^{L} \frac{\text{Cov}(m_{iT}, \tilde{b}_{il} m_{it-l})}{\text{Var}(m_{iT})}. \tag{C.5}$$

Consider first the case in which treatment builds up in constant proportions across units, so that $m_{it} = \delta_t m_{iT}$. Then:

$$\beta_\tau^{DiD} = \sum_{l=0}^{L} \bar{b}_l \, \delta_{t-l} + \sum_{l=0}^{L} \delta_{t-l} \frac{\text{Cov}(m_{iT}, \tilde{b}_{il} m_{iT})}{\text{Var}(m_{iT})}. \tag{C.6}$$

This shows that even under constant proportions, the dynamic DiD coefficient recovers the true scaled effect only if heterogeneity in marginal treatment effects is orthogonal to treatment exposure in period T.

Next, suppose treatment proportions vary across units, so that $m_{it} = \delta_{it} m_{iT}$, with

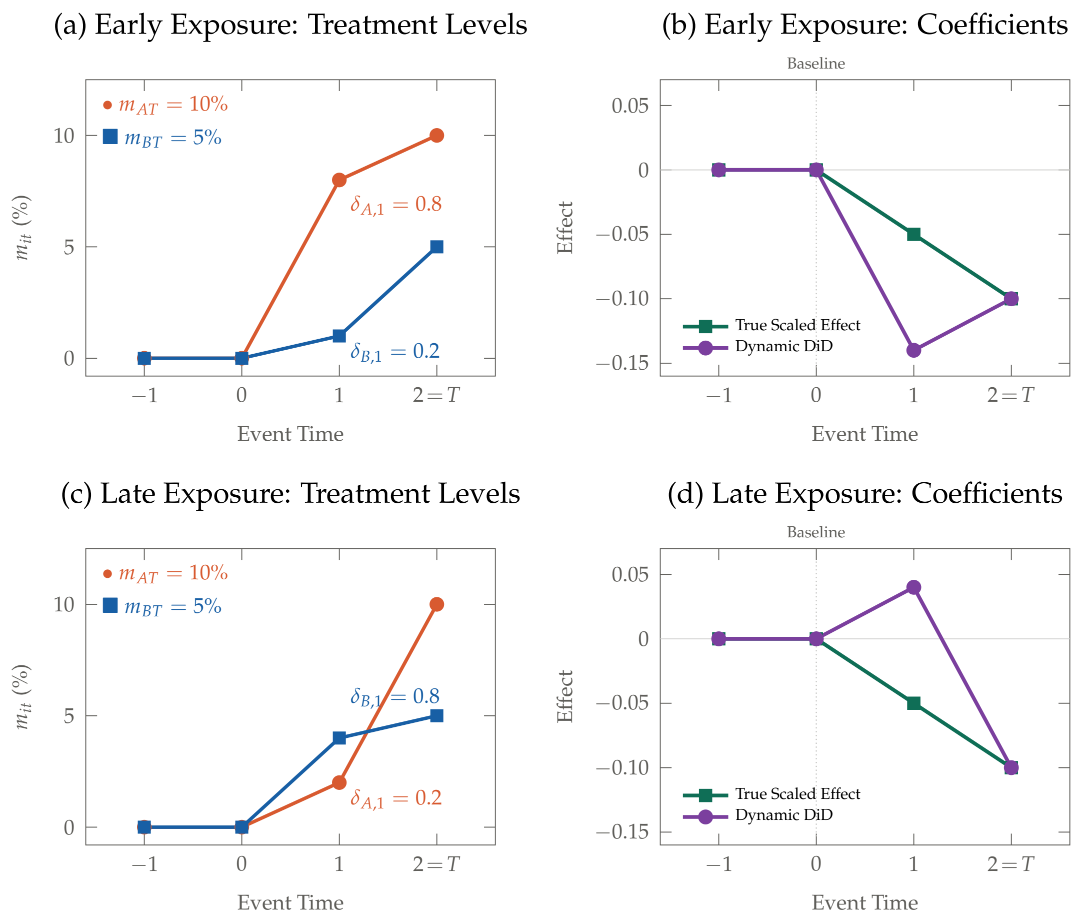
$\delta_{it} = \overline{\delta}_t + \tilde{\delta}_{it}$. In that case:

$$\beta_\tau^{DiD} = \sum_{l=0}^{L} \overline{b}_l \left(\overline{\delta}_{t-l} + \frac{\text{Cov}(m_{iT}, \tilde{\delta}_{it-l} m_{iT})}{\text{Var}(m_{iT})} \right) + \sum_{l=0}^{L} \frac{\text{Cov}(m_{iT}, \tilde{b}_{il} \delta_{it-l} m_{iT})}{\text{Var}(m_{iT})}. \quad \text{(C.7)}$$

This expression makes clear that, with heterogeneous treatment effects, the dynamic DiD coefficient can be more challenging to interpret causally due to two additional complications: variation in treatment proportions across units, and correlation between treatment exposure and heterogeneous marginal effects. The first source remains observable and can be studied using the treatment data. The second is more difficult to address, since treatment effect heterogeneity is typically unobserved.

D Supplementary Results

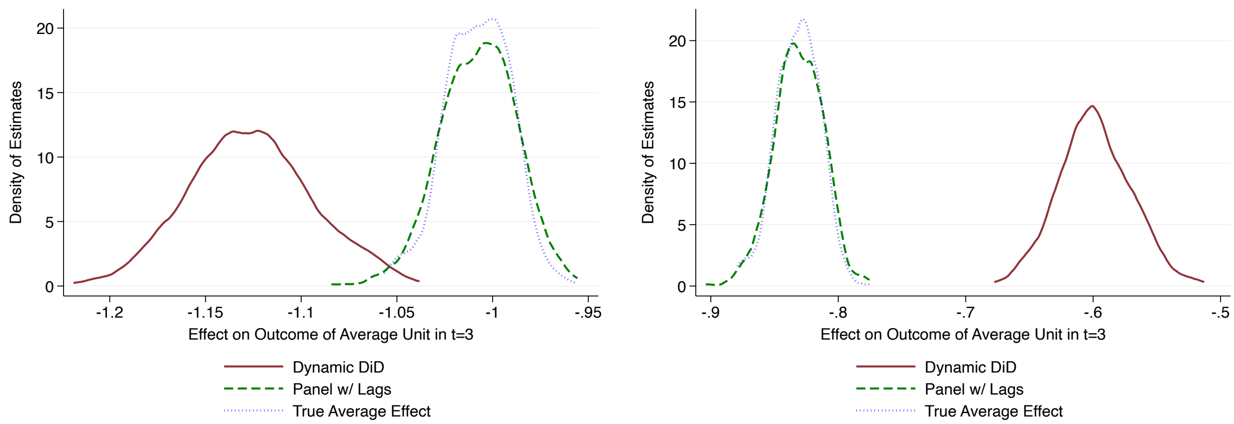
Figure D.1: Numerical Illustration of Dynamic DiD with Non-Proportional Exposure



Notes: The figure uses two units A and B with final treatment exposures $m_{AT} = 10\%$ and $m_{BT} = 5\%$, both realized at the reference period $T = 2$. Treatment is zero in pre-treatment periods. We assume static treatment effects, $b_0 = -0.1$ for a one-unit increase in exposure, where one unit corresponds to moving from 0 to 100%. The figure reports scaled effects, so the true scaled effect at period t is $b_0 \mathbb{E}[m_{it}/m_{iT}]$, while the dynamic DiD coefficient is $b_0 \operatorname{Cov}(m_{iT}, m_{it}) / \operatorname{Var}(m_{iT})$. Panels (a)–(b) show the early-exposure case, where by event time 1 unit A has received 80% of its final exposure and unit B has received 20%. Since $\operatorname{Cov}(m_{iT}, m_{i1}) / \operatorname{Var}(m_{iT}) = 1.4$ while $\mathbb{E}[m_{i1}/m_{iT}] = 1/2$, the DiD coefficient overstates the true scaled effect. Panels (c)–(d) show the late-exposure case, where the proportions are reversed. In this case, the covariance term equals -0.4, so the DiD coefficient has the wrong sign. At T, both the treatments and the effects coincide because the treatment is fully realized.

Figure D.2: Density of Estimates under Shocks Evolving with Non-Proportional Exposure By Approach

(a) Dynamic Shocks – Early Exposure (b) Dynamic Shocks – Late Exposure



Notes: The distribution of estimates comes from 500 simulations, and we use post-treatment period 3 as the reference. We focus on the average effects in each approach as described in Table 2. The analysis is based on a continuous dynamic shock that begins in period 0, amplifies through period 4, and reaches its full scale by that point. The full shock is drawn from $m_{iT} \sim U(0,1) \cdot 0.1$, and the random component is modeled as $\epsilon_{it} = 5 + U(0,1) \cdot 0.01$. Each simulation is conducted with 1,000 units and spans 14 periods, divided into 6 pre- and 8 post-treatment periods. We focus on gradual treatment effects of Table 1 in these figures that follow the pattern: $y_{it} = \sum_{k=0}^{2} -0.1 \cdot m_{it-k} + \epsilon_{it}$. The shock evolves with non-proportional exposure, in which units with higher exposure in T experience either early or late exposure.

Table D.1: Summary of Recent Studies on Immigration Impacts with the Spatial Approach

	Country	Empirical Strategy
Static Panel		
Bratsberg and Raaum (2012)	Norway	Panel regression with individual-level data
Aksu et al. (2022)	Turkey	Panel regression with individual-level data
Caruso et al. (2021) & Bonilla-Mejía et al. (2024)	Colombia	Panel regression with individual-level data
Groeger et al. (2024)	Peru	Panel regression with individual-level data
Lebow (2022)	Colombia	Panel regression with regional-level data
Dynamic DiD		
Foged and Peri (2016)	Denmark	Panel regression and dynamic DiD in levels
Dustmann et al. (2017)	Germany	Dynamic DiD in long differences with regional-level data
Edo (2020)	France	Dynamic DiD in long differences with regional-level data
Monras (2020)	United States	Dynamic DiD in long differences with regional-level data
Delgado-Prieto (2024b)	Colombia	Dynamic DiD in long differences with individual-level data
Delgado-Prieto (2024a)	Colombia	Dynamic DiD in levels with regional-level data
Muñoz (2024)	France	Dynamic DiD in levels with regional-level data

Notes: This table summarizes certain studies that investigate the labor market impacts of immigration published or posted as working papers in recent years.

E Dynamic Shock Construction in the Simulations

The continuous shock for unit i evolves from zero at $t = 0$ to a full value $m_{iT} \sim U(0,1) \cdot 0.1$ at $T = 4$ and stays constant after that. We define two types of dynamic shock evolution.

1. **Shock Evolving with Proportional Exposure** (homogeneous proportions δ_t over time for every unit):

$$m_{it} = m_{iT} * \overbrace{\sum_{t=0}^{T} (1+t) * 0.2}^{\delta_t}, \qquad t = 0, \ldots, 4.$$

2. **Shock Evolving with Non-Proportional Exposure** (heterogeneous proportions δ_{it} over time for every unit):

$$m_{it} = m_{iT} \cdot \overbrace{\frac{s_{it}}{\sum_{t=0}^{T} s_{it}}}^{\delta_{it}}, \qquad t = 0, \ldots, 4.$$

The proportions δ_{it} are derived from the factors s_{it}, which depend on the unit's ventile of exposure $Q_i \in \{1, \ldots, 20\}$, assigned based on its treatment level m_{iT}:

$$s_{it} = \left(\tfrac{Q_i}{20} \cdot \mathbf{1}(t < 3) + 0.1 \cdot \mathbf{1}(t \geq 3) \right) \cdot U(0,1).$$

In the *early exposure* case, we use Q_i directly, so that units with higher long-run exposure accumulate their shock earlier. In the *late exposure* case, we replace Q_i with $21 - Q_i$, reversing the ranking so that the same units accumulate their shock later. Because this is a relabelling of the same set of integers, the cross-sectional average $\bar{\delta}_t$ is identical in both cases at every t.