

Insper Instituto de Ensino e Pesquisa
Doutorado em Economia dos Negócios

Pedro Picchetti

Essays in Instrumental Variables

São Paulo - SP

2025

Pedro Picchetti

Essays in Instrumental Variables

Tese apresentada ao programa de Doutorado em
Econômica dos Negócios como requisito parcial
para obtenção do rótulo de Doutor em Econo-
mia.

Inspere Instituto de Ensino e Pesquisa
Doutorado em Economia dos Negócios

Orientador: Profa. Cristine Campos de Xavier Pinto

São Paulo - SP

2025

Pedro Picchetti

Essays in Instrumental Variables/ Pedro Picchetti. - São Paulo - SP: 2025

137 p.

: Insper Instituto de Ensino e Pesquisa

Doutorado em Economia dos Negócios , 2025.

Orientador: Profa. Cristine Campos de Xavier Pinto

1. Variáveis Instrumentais. 2. Inferência Causal. 3. Econometria. I. Pedro Picchetti.

II. Essays in Instrumental Variables.

Pedro Picchetti

Essays in Instrumental Variables

Tese apresentada ao programa de Doutorado em
Econômica dos Negócios como requisito parcial
para obtenção do título de Doutor em Econo-
mia.

Banca examinadora

Profa. Cristine Campos de Xavier Pinto
Orientador(a)

Prof. Sergio Firpo
Examinador(a)

Prof. Peter Hull
Examinador(a)

Prof. Jonathan Roth
Examinador(a)

São Paulo - SP
2025

Agradecimentos

Agradeço profundamente minha orientadora Cristine Pinto, pelo apoio e paciência ao longo dos anos e por confiar no meu trabalho. Também sou grato aos demais membros da banca: Sergio Firpo, Vitor Possebom, Jonathan Roth e Peter Hull, cujo apoio foi fundamental ao longo do doutorado e durante o mercado de trabalho acadêmico.

O doutorado teria sido uma empreitada solitária se não fosse pelos meus colegas, que com sua camaradagem e companheirismo fizeram dos momentos difíceis mais tranquilos. Um agradecimento especial para André Mancha, Bruno Cavani, Caio Bloise, Guilherme Gallego, Ivan Prenhaca, Newton Oliveira e Tomás Nobrega, que além de colegas se tornaram grandes amigos. Mas de forma geral, sou profundamente grato por todos colegas que a Sala Fábio Chaddad me proporcionou.

A vida é mais fácil com amigos. Faço aqui um agradecimento especial para os meus velhos amigos de colégio, a Família do Mercado, e meus amigos de faculdade, o FEA7 e a turma do QG, por me lembrarem da beleza da vida fora da pesquisa.

Me faltam palavras para agradecer minha família. Aos meus pais, Paulo e Fabiana, e meus irmãos, João Paulo e Fernando, minha gratidão eterna pelo apoio.

Por fim, agradeço à minha esposa, Aline. Os meses que passamos separados durante o sanduíche foram tristes, mas superar eles foi a prova de que você era o amor da minha vida, e que é um prazer imenso dividir ela com você. Obrigado, nada disso seria possível sem seu apoio, e nada disso teria o mesmo valor se não pudesse compartilhar com você.

Resumo

Essa tese consiste em três ensaios em Econometria. Mais especificamente, com em métodos de Variáveis Instrumentais (IV) para Inferência Causal.

O primeiro capítulo é coautorado com Cristine Pinto, e trata da identificação de Efeitos Marginais (MTE) de Tratamento em designs de Diferenças-em-Diferenças (DiD). Muitas aplicações de DiD se assemelham ao métodos de variáveis instrumentais no sentido de que a exposição ao choque que afeta a escolha do tratamento não determina quem é tratado e quem permanece no controle. Nesses designs, chamados de Fuzzy DiD, efeitos de tratamento são identificados apenas para *compliers*. Neste capítulo, mostro como MTEs podem ser identificados em designs de Fuzzy DiD substituindo a hipótese de independência do instrumento pela hipótese de tendências paralelas.

O segundo capítulo propõe um novo método de identificação para efeitos causas dinâmicos usando IVs em painel. O ponto de partida da investigação é que 2SLS não identifica efeitos causais com interpretação clara diante de efeitos dinâmicos e instrumentos serialmente correlacionados no tempo. Diante desse resultado, proponho um novo método para identificar efeitos causais associados à sequências alternativas de tratamentos ao longo do tempo sob restrições de dinâmica no primeiro estágio.

Por fim, o terceiro capítulo propõe uma nova abordagem para Análise de Sensibilidade em métodos de IV onde a hipótese de independência não é crível. Focando no caso de outcomes binários, eu derivo resultados de identificação parcial de efeitos de tratamento sob violação de independência, e identifico os valores máximos de violação para qual é possível sustenar conclusões sobre os efeitos verdadeiros.

Palavras-chave: Variáveis Instrumentais, Inferência Causal, Econometria.

Abstract

This chapter consists of three essays in Econometrics. More specifically, three essays in Instrumental Variables (IV) methods for Causal Inference.

The first chapter is coauthored with Cristine Pinto, and proposes identification results for Marginal Treatment Effects (MTE) in Difference-in-Difference (DiD) settings. Many DiD applications resemble IV methods in the sense that there is imperfect compliance towards the shock affecting selection into treatment. In such called Fuzzy DiD designs, treatment effect parameters are only identified for *compliers*. In this chapter, I show how MTEs can be identified in Fuzzy DiD settings by replacing the instrument independence assumption with a parallel trends assumption for potential outcomes and potential treatments.

The second chapter proposes a new identification method for dynamic causal effects in IV settings with panel data. The starting point of this work is that 2SLS parameters do not hold a clear causal interpretation in the presence of dynamic treatment effects and serial correlation of the instrument. Under this result, I propose a new identification method for causal effects associated to alternative sequences of treatment taken through time under restrictions of dynamics in the first-stage.

Finally, the third chapter proposes a new Sensitivity Analysis approach for IV methods where the instrument independence assumption is not credible. With a focus on the case of binary outcomes, I derive partial identification results for treatment effects under violations of independence and identify breakdown values, which are the largest violations of independence under which a particular conclusion holds.

Keywords: Instrumental Variables, Causal Inference, Econometrics.

Sumário

1	MARGINAL TREATMENT EFFECTS IN DIFFERENCE-IN-DIFFERENCES	9
1.1	Introduction	9
1.2	General Framework	12
1.2.1	Framework	12
1.2.2	Identification	14
1.3	Estimation and Inference	16
1.3.1	Partially Linear Model	16
1.3.2	Sieve Estimator	17
1.4	Doubly-Robust estimation of LATE	19
1.5	Monte Carlo Simulations	22
1.6	Empirical illustration	24
1.6.1	LATE	25
1.6.2	MTE Curve	26
1.7	Conclusion	27
2	INSTRUMENTAL VARIABLES WITH MULTIPLE TIME PERIODS	28
2.1	Introduction	28
2.2	Identification with $T = 2$	32
2.2.1	Interpretation of the Static Wald Estimand	33
2.2.2	Separate Identification Approach	35
2.2.3	The Role of the Time Structure in Identification	38
2.3	General Framework and Identification	39
2.3.1	Assumptions	39
2.3.2	Target Parameters	42
2.3.3	Interpretation of the Static Wald Estimator	43
2.3.4	A General Expression for the Multiple Differences Estimand	44
2.4	Estimation and Design-Based Inference	45
2.4.1	Correcting the Wald Estimator	45
2.4.2	Nonparametric Estimator for Dynamic Causal Effects	47
2.5	Extensions	48
2.5.1	Covariates	48
2.5.2	Intermediary Realizations	49
2.6	Monte Carlo Simulations	50
2.7	Application: Law Enforcement and Deforestation	53
2.7.1	AGR (2023)	53

2.7.2	Serial Correlation of the Instrument	54
2.7.3	Dynamic Effects of Law Enforcement	54
2.8	Conclusion	57
3	BREAKDOWN ANALYSIS FOR INSTRUMENTAL VARIABLES WITH BINARY OUTCOMES	59
3.1	Introduction	59
3.2	Framework	61
3.3	Partial Identification with Binary Outcomes	63
3.3.1	First-Stage	63
3.3.2	Reduced Form	65
3.3.3	Local Average Treatment Effect	66
3.4	Breakdown Analysis	67
3.5	Numerical Illustration	69
3.6	Estimation and Inference	71
3.6.1	Bootstrap Inference	75
3.7	Partial Identification under joint <i>c-dependence</i>	76
3.8	Monte Carlo Simulations	78
3.9	Empirical Application: Family Size and Employment	79
3.10	Conclusion	80
	REFERÊNCIAS	82
4	APPENDIX TO CHAPTER 1	89
4.1	Appendix A	89
4.2	Appendix B	97
4.3	Appendix C: Identification with Panel Data	99
5	APPENDIX TO CHAPTER 2	102
5.1	Appendix A	102
5.2	Appendix B	109
5.3	Appendix C	113
5.4	Appendix D	114
6	APPENDIX TO CHAPTER 3	116
6.1	Appendix A	116
6.2	Appendix B	125

1 Marginal Treatment Effects in Difference-in-Differences

Difference-in-Differences (DiD) is a popular method that uses variations in treatment status resulting from exposure to a shock—typically a policy change—to estimate treatment effects. The usual DiD estimand fails to recover relevant causal parameters when there is imperfect compliance towards the shock. This article introduces an identification strategy for DiD settings with imperfect compliance, focusing on Marginal Treatment Effects (MTE). We demonstrate how to modify and combine standard instrumental variables (IV) and DiD assumptions to identify treatment effects in cases where individuals enter treatment with at least partial knowledge of their unobserved gains. We propose a semiparametric estimator for the MTE that is consistent under partially linear potential outcomes and prove its asymptotic normality. Additionally, we derive an estimator for the local average treatment effect (LATE) that is robust to the misspecification of the MTE model. We assert the desirable finite-sample properties of the estimators through simulation studies of a linear MTE model. Finally, we use our results to investigate heterogeneity in the returns to primary education attendance in Indonesia.

1.1 Introduction

Difference-in-differences (DiD) is a popular method for identifying treatment effects using observational data. The method achieves identification by exploiting variation in treatment status that comes from a binary shock, usually in the form of a policy change, that only affects a sub-group of the population. In that sense, exposure to the shock defines treatment and control groups in the setting. Its simplest version has a treatment and a control group, a pre-shock and a post-shock period. In the pre-shock period, no individual is treated, and after the shock, the treatment group becomes fully treated while the control group remains fully untreated. Under the parallel trends assumption, the DiD estimand identifies the average treatment effect on the treated (ATT).

However, it is common to observe imperfect compliance toward the shock in many applications of the DiD method. The control group does not remain fully untreated, at the same time that the treatment group does not become fully treated after the shock. A canonical example is (DUFLO, 2001), which measures the return to education by exploiting the variation in education supply from a major primary school expansion program in Indonesia. In this example, we have a cohort of men exposed to the reform and another old cohort who finished primary school when the program was launched. Some men were exposed to the expansion of primary schools and did not attend primary school, and some men in the old cohort attended primary

school. In this case, compliance is not perfect.

Despite its widespread use, identification in DiD settings with imperfect compliance has received little formal attention. (HUDSON; HULL; JACK, 2017) study the properties of the two-stage DiD estimand in an approach called "Instrumented"DiD. The authors outline high-level assumptions that ensure the validity of their approach while remaining agnostic about the structural functions governing potential outcomes and selection mechanisms in the given setting. (CHAISEMARTIN; D'HAULTEFOEUILLE, 2018) adopt a different approach within the same setting, referring to it as the "Fuzzy"DiD framework. While they provide a choice model for the framework, they remain agnostic about the structural function underlying potential outcomes and do not discuss exclusion restrictions explicitly.

This paper shows how to identify the marginal treatment effect (MTE) in DiD settings with imperfect compliance. The MTE is a causal inference parameter with several desirable properties: it has a straightforward economic interpretation as the willingness to pay for treatment for individuals indifferent between taking treatment or not, and all other average treatment effects can be expressed as weighted averages of the MTEs. We show that the MTE can be identified in DiD settings with imperfect compliance under the standard Instrumental Variables (IV) assumptions combined with the parallel trends assumption. Our approach allows researchers to navigate between the 'instrumented' DiD approach and the 'Fuzzy' DiD approach with transparency and rigor.

We interpret the DiD setting with imperfect compliance as an IV setting in which there is no variation in the instrument in the first period, a shock to the instrument in the second period, and heterogeneous responses across all individuals.

Our identification approach is based on a modification of the Local Instrumental Variables (LIV) estimand (HECKMAN; VYTLACIL, 1999; HECKMAN; VYTLACIL, 2001; HECKMAN; VYTLACIL, 2005) to account for the time structure in the DiD setting. We show that the estimand identifies the MTE under a parallel trends assumption for potential outcomes and the choice model, which is weaker than the usual independence of the instrument assumption from IV settings.

We propose a semiparametric estimator that follows the control function approach. We impose a functional form that relates the unobserved heterogeneity with the propensity score and use this function to control for endogeneity as an omitted variable problem. The estimator is consistent and asymptotically normal under additional assumptions further discussed in detail. Monte Carlo simulation studies illustrate its desirable finite-sample performance.

We revisit (DUFLO, 2001) and document substantial heterogeneity in the returns to primary education. First, we estimate the local average treatment (LATE) using the two methods proposed in the literature: two-stage least squares applied to DiD and the fuzzy DiD approach proposed by (CHAISEMARTIN; D'HAULTEFOEUILLE, 2018). Our more robust approach

suggested an increase of around 19% in wages caused by attending primary education. This effect is smaller than the ones estimated using the other approaches. However, this effect has an average of positive and negative returns on education. We document a pattern of reverse selection on gains that comes as an upward-sloping MTE curve. Individuals with a higher distaste for education have a positive return to primary education, while individuals with a lower distaste may have a negative return. This reverse selection was found in (CORNELISSEN et al., 2018) that investigates the returns of early child care in Germany.

Our work relates to three strands of the causal inference literature.

First, the results are intimately related to other papers in the MTE literature. The marginal treatment effect was first introduced by (BJÖRKLUND; MOFFITT, 1987) as the gain from treatment for individuals who are shifted into treatment by a marginal change in its cost. (HECKMAN; VYTLACIL, 1999; HECKMAN; VYTLACIL, 2001; HECKMAN; VYTLACIL, 2005; HECKMAN; VYTLACIL, 2007) further defined the MTE as the gain from treatment for individuals shifted into treatment by a marginal change in the propensity score, the predicted probability of treatment (see (CORNELISSEN et al., 2016) for a comprehensive review of the literature). Our approach is closely related to recent advances in the MTE literature (BRINCH; MOGSTAD; WISWALL, 2017; MOGSTAD; SANTOS; TORGOVITSKY, 2018). We are the first to define, identify, and estimate marginal treatment effects in DiD settings.

Second, our results are directly related to the recent advances in the DiD literature. (BONHOMME; SAUDER, 2011) consider a DiD model that allows for heterogeneous effects of time, but (CHAISEMARTIN; D’HAULTEFOEUILLE, 2018) were the first to explore the identification of heterogeneous treatment effects in DiD settings with imperfect compliance. We build on the framework proposed by the authors and show how the MTE relates to the LATE parameter. This paper is also related to the recent advances in the choice-theoretical foundations of the DiD setting (MARX; TAMER; TANG, 2023; GHANEM; SANT’ANNA; WÜTHRICH, 2023).

Third, our work relates to the literature on doubly robust estimators. More specifically, to the papers that provide estimands that are robust to the misspecification of the MTE (SASAKI; URA, 2023; ESCANCIANO; PEREZ-IZQUIERDO, 2023) and the reshaped inverse probability weighting (RIPW) estimators (ARKHANGELSKY et al., 2021).

Organization of the paper: In the next section, we define the DiD setting with imperfect compliance in a Roy model and present the main identification results. In section 3, we present estimation and inference results in a partially linear model. In Section 4 we introduce an estimator for the LATE that is robust to the misspecification of the MTE functional form. We examine the properties of the estimators using a Monte Carlo simulation study in Section 5 and provide an empirical illustration in Section 6. Section 7 concludes.

1.2 General Framework

1.2.1 Framework

We present the framework through a model best suited for repeated cross-sections or single cross-sections where the birth cohort plays the role of time (individuals born in a cohort are affected by the shock, whereas individuals from a different cohort are not). We are interested in the effect of a binary treatment D on a scalar outcome Y .

We denote by T the random variable that indicates which individuals belong to a specific cohort observed after the shock. For example, $T_i = 1 \{t = 1\}$ equals 1 if the individual belongs to cohort 1 after the shock. For a random variable R we write R_t as the expectation of R_i conditional on the cohort t ($R_t = \mathbb{E}[R_i|T_i = t]$). We observe an instrument Z_i , which is assumed to be constant for cohort 0 and varies across individuals in cohort 1. Thus, we normalize the instrument to have value zero in cohort 0¹. In the (DUFLO, 2001) example, for instance, the instrument is a shock in the supply of public primary education per capita at the district level. This school expansion program has different effects on primary education availability across different districts, characterizing a continuous instrument². In (FERRARA; CHONG; DURYEYEA, 2012), which analyzes the effects of soap opera content on female fertility in Brazil, the instrument is the presence of Globo (the primary soap opera producer in the country) coverage in a Brazilian municipality.

Let $\mathcal{S}(Z)$ denote the support of the instrument. For each $d \in \{0, 1\}$ and $z \in \mathcal{S}(z)$ let $Y_i(d, z)$ denote the potential outcome for individuals i that he/she would have if their treatment and instrument have been set to d and z . Similarly, let $D_i(z)$ denote the potential treatment choice of individual i from cohort t if the instrument were z . The observed and potential variables are related through

$$Y_i = \sum_{d \in \{0,1\}} \sum_{z \in \mathcal{S}(Z)} 1 \{D_i = d, Z_i = z\} Y_i(d, z), \quad D_i = \sum_{z \in \mathcal{S}(Z)} 1 \{Z_i = z\} D_i(z)$$

Since $Z_i = 0$ for all individuals in cohort 0, observed quantities for individuals in this cohort are always associated with $z = 0$. Observed outcomes are equal to $Y_i = \sum_{d \in \{0,1\}} Y_i(d, 0)$ and observed treatment statuses are equal to $D_i = D_i(0)$.

In the 'sharp' 2×2 design, treatment status is a deterministic function of the exposure to the shock. Z_i is binary and $D_i = Z_i$. In 'fuzzy' DiD settings, however, such an equality does not

¹ We assume time-invariant covariates X_i are also observed. Still, they are suppressed in this section for the sake of the exposition. Covariates are explicitly considered in Section 3.

² DiD settings usually exploit variation in treatment that comes from a binary shock. This Section considers a general continuous instrument as the shock driving selection. Section 3 presents the functional form assumptions under which the results are valid for a binary instrument.

hold. We start with a simple choice-theoretic framework that shows how exposure to the shock affects the selection into treatment.

Selection into treatment follows a single-crossing threshold model vylta:

$$D_i = 1 \{ \mu(Z_i) \geq U_{iD} \}$$

One can think of the function $\mu(\cdot)$ as the perceived net gain from treatment and the random variable U_{iD} as an unobservable "resistance" to treatment. Individuals take treatment if their perceived gain of treatment is greater than their resistance towards it.

The marginal distribution of U_{iD} can be normalized to a standard uniform distribution $V_i \equiv F_{U_{iD}}(U_{iD})$, which can be interpreted as the quantile of resistance to treatment. Similarly, applying the CDF of U_{iD} to the function $\mu(\cdot)$ yields a propensity score interpretation: $\pi(Z_i) \equiv \mathbb{P}(D_i = 1 | Z_i) = F_{U_{iD}}(\mu(Z_i))$. Hence, the choice model can be represented simply by $D_i = 1 \{ \pi(Z_i) \geq V_i \}$.

The unobservable heterogeneity driving selection into treatment, U_{iD} is assumed to vary across cohorts. Specifically, we assume that U_{iD} has the following mixture distribution:

$$\mathbb{P}(U_{iD} \leq u, T_i = t) = t \cdot \mathbb{P}(T_i = 1) \mathbb{P}(U_{iD,1} \leq u | T_i = 1) + (1-t) \cdot \mathbb{P}(T_i = 0) \mathbb{P}(U_{iD,0} \leq u | T_i = 0)$$

It follows that individuals from cohort 0 take treatment if $\pi(Z_i) \leq V_0$ and individuals from cohort 1 take treatment if $\pi(Z_i) \leq V_1$. Under the choice model above, the "sharp" DiD design could be interpreted as a model in which there is a binary instrument and the function $\pi(Z_i)$ is such that $\pi(1) = 1$ and $\pi(0) = 0$ for all individuals and thus, $D_{i0} = 0$ for all i , and $D_{i1} = 1$ for individuals with $Z_i = 1$ and $D_{i1} = 0$ for individuals with $Z_i = 0$.

Note that for all individuals from cohort 0, the instrument has a value of zero. Thus, there is no exogenous variation for the choice model for these individuals, and selection is purely endogenous. Hence, variations in the instrument can only be used to identify treatment effects for individuals from cohort 1. The assumption below imposes a time trend on selection into treatment:

Assumption 1 (Trend in Selection) : For all $z \in \mathcal{S}(Z)$ all i , $\mathbb{P}(V_1 \leq \pi(z)) \geq \mathbb{P}(V_0 \leq \pi(z))$.

Assumption 1 states that the distribution of the distaste for treatment from individuals from cohort 1 is first-order stochastically dominated by the distaste from individuals from cohort 0. Assumption 1 implies that keeping Z_i fixed, there is no individual that would choose to take treatment in cohort 0 that would not have chosen to take treatment if he was from cohort 1. The assumption can be interpreted as a time trend in selection. In the (DUFLO, 2001) setting, it translates as the well-documented secular trend of increasing enrollment in primary education observed worldwide last century.

We also consider the following assumption, which is standard in IV settings:

Assumption 2 (Exclusion Restriction): For all $z \in \mathcal{S}(Z)$, all i and $d \in \{0, 1\}$, $Y_i(z, d) = Y_i(d)$.

Assumption 2 states that the instrument does not affect potential outcomes directly, but only to the extent that it affects the choice of taking treatment or not. In the applied example, it means that the school expansion program does not affect potential future earnings directly, but it affects the choice to enroll in primary education, which will impact future earnings. Given the exclusion restriction, the observed outcomes for individual i is simply $Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$.

Another standard assumption in IV settings is that the instrument is independent of potential outcomes and the choice model (Assumption A-2 from (HECKMAN; VYTLACIL, 2005)). In fuzzy DiD settings, the independence assumption can be relaxed and replaced by a modified version of the parallel trends assumption:

Assumption 3 (Parallel Trends): For all $z \in \mathcal{S}(Z)$, $\mathbb{E}[Y_1(D_1(z)) - Y_0(D_0(0)) | Z_i = z] \perp Z_i$ and $\mathbb{P}(V_1 \leq \pi(z) < V_0 | Z_i = z) \perp Z_i$.

Assumption 3 states the evolution of potential outcomes and treatment rates across cohorts is independent of the actual instrument assignment. One way to interpret the parallel trends assumption is that instrument assignment was poorly randomized (e.g., school expansion was greater in districts where potential earnings were greater), but data from a time period/cohort before the assignment can be used to correct this poor randomization. An equivalent representation for the parallel trends selection in the choice model is that $\mathbb{P}(D_{i1}(z) | Z_i = z) - \mathbb{P}(D_{i0}(0) | Z_i = z)$ is independent of Z_i . That is, the growth in the treatment rate across cohorts is independent of treatment assignment.

The next assumptions are technical assumptions which are standard in the MTE literature.

Assumption 4: The function $\mu(Z_i)$ is a nondegenerate random variable.

Assumption 5: The distribution of U_{iD} is absolutely continuous with respect to the Lebesgue Measure.

Assumption 6: The values of $\mathbb{E}[|Y_i(1)|]$ and $\mathbb{E}[|Y_i(0)|]$ are finite.

Assumption 4 imposes that there is a variable that determines the selection into treatment. Assumption 5 is made for expositional convenience, and Assumption 6 guarantees that the treatment effects are well-defined.

1.2.2 Identification

The goal of this paper is to derive identification results for the marginal treatment effect in the cohort after the instrument is assigned. Formally, we define this object as

$$MTE(\pi_1(z)) = \mathbb{E} [Y_1(1) - Y_1(0) | V_1 = \pi_1(z) < V_0]$$

The MTE is the average treatment effect for individuals in cohort 1 who are indifferent between taking treatment or not and would not take treatment if they were from cohort 0.

Identification of the MTE in “fuzzy” DiD settings follows closely from the LIV estimand (HECKMAN; VYTLACIL, 2001). However, the estimand is modified to account for the fact that the instrument is not independent of the vector of potential quantities. The result below shows that under Assumptions 1-6, $MTE(\pi_1(z))$ is identified. See Section 1 of the Appendix for the proof.

Theorem 1. *Under Assumptions 1-6,*

$$\frac{\partial \mathbb{E} \left[\frac{Y_i T_i}{\mathbb{E}[T_i]} - \frac{Y_i(1-T_i)}{(1-\mathbb{E}[T_i])} \middle| Z_i = z \right] / \partial z}{\partial \mathbb{E} \left[\frac{D_i T_i}{\mathbb{E}[T_i]} - \frac{D_i(1-T_i)}{(1-\mathbb{E}[T_i])} \middle| Z_i = z \right] / \partial z} = MTE(\pi_1(z))$$

The identification result in Theorem 1 can be interpreted as a correction for the standard LIV estimand to account for parallel trends instead of independence. The MTE is identified for individuals in cohort 1 with $Z_i = z$, $V_{i1} = \pi_1(z)$. Thus, the parameter can be interpreted as the average treatment effect for individuals with assignment z , who are indifferent between taking treatment or not. Another interpretation is the one of willingness to pay for treatment for individuals with assignments equal to z .

The LIV estimand as defined by (HECKMAN; VYTLACIL, 1999) is given by

$$\frac{\partial \mathbb{E} [Y_i | Z_i = z] / \partial z}{\partial \mathbb{E} [D_i | Z_i = z] / \partial z}$$

Consider the estimand for the reduced form. (HECKMAN; VYTLACIL, 2001) show that

$$\begin{aligned} \mathbb{E} [Y_i | Z_i = z] &= \mathbb{E} [Y_i(0) | Z_i = z] + \mathbb{E} [Y_i(1) - Y_i(0) | D_i = 1, Z_i = z] \mathbb{P}(D_i = 1 | Z_i = z) \\ &= \mathbb{E} [Y_i(0) | Z_i = z] + \int_0^{\pi(z)} \mathbb{E} [Y_i(1) - Y_i(0) | D_i = 1, Z_i = V_i = v] dv \end{aligned}$$

Under the independence assumption, we have $\mathbb{E} [Y_i(0) | Z_i = z] = \mathbb{E} [Y_i(0)]$, from which it follows that the limit form of the reduced form estimand identifies $\mathbb{E} [Y_i(1) - Y_i(0) | V_i = \pi(z)] \frac{\partial \pi(z)}{\partial z}$, and the standard MTE parameter is identified by taking the ratio with respect to the limit form of the first-stage estimand.

In the absence of independence, but under parallel trends, we exploit variations across cohorts to leverage the fact that $\mathbb{E} [Y_1(D_1(z)) - Y_0(D_0(0)) | Z_i = z] \perp Z_i$ and the fact that

$\mathbb{P}(V_1 \leq \pi(z) < V_0 | Z_i = z) \perp Z_i$ in order to identify treatment effects for individuals in cohort 1 who are indifferent between taking treatment or not.

The modified LIV estimand is a limited version of the Wald-DiD estimand (CHAISEMARTIN; D’HAULTEFOEUILLE, 2018). Consider the estimand associated with a shift in the instrument from zero to z :

$$\frac{\mathbb{E} \left[\frac{Y_i T_i}{\mathbb{E}[T_i]} - \frac{Y_i(1-T_i)}{(1-\mathbb{E}[T_i])} \middle| Z_i = z \right] - \mathbb{E} \left[\frac{Y_i T_i}{\mathbb{E}[T_i]} - \frac{Y_i(1-T_i)}{(1-\mathbb{E}[T_i])} \middle| Z_i = 0 \right]}{\mathbb{E} \left[\frac{D_i T_i}{\mathbb{E}[T_i]} - \frac{D_i(1-T_i)}{(1-\mathbb{E}[T_i])} \middle| Z_i = z \right] - \mathbb{E} \left[\frac{D_i T_i}{\mathbb{E}[T_i]} - \frac{D_i(1-T_i)}{(1-\mathbb{E}[T_i])} \middle| Z_i = 0 \right]}$$

Under assumptions 1-6, the estimand identifies the average treatment effect for compliers from cohort 1 that would not take treatment if they were from cohort 0, defined as treatment group switchers by (CHAISEMARTIN; D’HAULTEFOEUILLE, 2018) in the case of a binary instrument. Under the choice-theoretical framework from Section 2.1, the parameter can be expressed as

$$\begin{aligned} & \mathbb{E}[Y_1(1) - Y_1(0) | D_1(z) > D_1(0), D_0(z) = D_0(0) = 0] \\ & = \mathbb{E}[Y_1(1) - Y_1(0) | \pi(0) < V_1 \leq \pi(z) < V_0] \end{aligned}$$

The discrete shift in the instrument can be narrowed into a partial derivative by making the shift arbitrarily small:

$$\lim_{z \rightarrow 0} \frac{\mathbb{E} \left[\frac{Y_i T_i}{\mathbb{E}[T_i]} - \frac{Y_i(1-T_i)}{(1-\mathbb{E}[T_i])} \middle| Z_i = z \right] - \mathbb{E} \left[\frac{Y_i T_i}{\mathbb{E}[T_i]} - \frac{Y_i(1-T_i)}{(1-\mathbb{E}[T_i])} \middle| Z_i = 0 \right]}{\mathbb{E} \left[\frac{D_i T_i}{\mathbb{E}[T_i]} - \frac{D_i(1-T_i)}{(1-\mathbb{E}[T_i])} \middle| Z_i = z \right] - \mathbb{E} \left[\frac{D_i T_i}{\mathbb{E}[T_i]} - \frac{D_i(1-T_i)}{(1-\mathbb{E}[T_i])} \middle| Z_i = 0 \right]} = \frac{\partial \mathbb{E} \left[\frac{Y_i T_i}{\mathbb{E}[T_i]} - \frac{Y_i(1-T_i)}{(1-\mathbb{E}[T_i])} \middle| Z_i = z \right] / \partial z}{\partial \mathbb{E} \left[\frac{D_i T_i}{\mathbb{E}[T_i]} - \frac{D_i(1-T_i)}{(1-\mathbb{E}[T_i])} \middle| Z_i = z \right] / \partial z}$$

Hence, it follows that the LIV estimand identifies the average treatment effect for the marginal complier in cohort 1, which would be a never-taker in cohort 0. The modified LIV estimand above is suited for the case of repeated cross-sections, but a similar estimand can be obtained when Panel data is available. See Appendix C for the results.

1.3 Estimation and Inference

1.3.1 Partially Linear Model

In this section, we propose a semiparametric estimator for the MTE function. From now on, time-invariant covariates X_i are considered explicitly in the exposition. In the absence of additional assumptions, researchers are likely to suffer from the curse of dimensionality when accounting nonparametrically for covariates. Below, we impose a functional form assumption which is pervasive in the MTE literature:

Assumption 7 (Additive Separability): For all i , $Y_i(j) = X_i' \beta_j + U_{ij}$, $j \in \{0, 1\}$.

Assumption 7 states that potential outcomes are additively separable into linear functions of covariates and an error term. This assumption is commonly used in the literature, particularly when the instrument is not continuous, as it enables the identification of MTEs even with a binary instrument, provided additional functional form assumptions are met (BRINCH; MOGSTAD; WISWALL, 2017).

Both the coefficients associated to the covariates and the error term are allowed to vary across treatment status and cohorts. Under the separable model, endogeneity comes from the correlation of U_{ij} and V_i , which is commonly called essential heterogeneity (HECKMAN; URZUA; VYTLACIL, 2006).

Define the cohort-specific propensity score as $\mathbb{P}(D_{it} = 1 | X_i = x, Z_i = z) := \pi_t(x, z)$. Under the separability assumption, the conditional expectation of Y_{it} given $X_i = x$ and $Z_i = z$ can be written as

$$\begin{aligned} \mathbb{E}[Y_t | X_i = x, Z_i = z] &= x' \beta_{0t} + x' (\beta_{1t} - \beta_{0t}) \pi_t(z, x) + \mathbb{E}[U_t | Z_i = z] \\ &= x' \beta_{0t} + x' (\beta_{1t} - \beta_{0t}) \pi_t(z, x) + K(\pi_t(z, x)) \end{aligned}$$

where $U_t = \mathbb{E}[U_i | T_i = t]$, with $U_i = D_i U_{i1} + (1 - D_i) U_{i0}$ and the coefficients β are allowed to vary across cohorts.

The corollary below provides the identification result for MTEs under Assumption 7. See Section 2 of the Appendix for the proof.

Corollary 1. *Under Assumptions 1-7,*

$$\frac{\partial \mathbb{E} \left[\frac{Y_i T_i}{\mathbb{E}[T_i]} - \frac{Y_i (1 - T_i)}{(1 - \mathbb{E}[T_i])} \middle| Z_i = z \right] / \partial z}{\partial \mathbb{E} \left[\frac{D_i T_i}{\mathbb{E}[T_i]} - \frac{D_i (1 - T_i)}{(1 - \mathbb{E}[T_i])} \middle| Z_i = z \right] / \partial z} = x' (\beta_{11} - \beta_{01}) + \frac{\partial K(\pi_1(z, x))}{\partial z}$$

1.3.2 Sieve Estimator

Assume a partially linear functional form for the choice model, in which covariates enter linearly and the functional form associated to the instrument is unspecified and approximated by a vector of basis functions:

$$\begin{aligned} \mathbb{E}[D_t | X_i = x, Z_i = z] &:= \pi_t(z, x) = x' \beta_{Dt} + g_D(z) \\ &\approx x' \beta_{Dt} + P_D^K(z)' \gamma_{Dt} \end{aligned}$$

From which it follows that the sieve estimator for the limit of the first-stage is equal to

$$\partial \widehat{\mathbb{E}} \left[\frac{D_i T_i}{\mathbb{E}[T_i]} - \frac{D_i(1-T_i)}{(1-\mathbb{E}[T_i])} \middle| X_i = x, Z_i = z \right] / \partial z = \nabla P_D^K(Z_i)' (\widehat{\gamma_{D1}} - \widehat{\gamma_{D0}}) \quad (1.1)$$

where

$$(\widehat{\gamma_{D1}} - \widehat{\gamma_{D0}}) = \left(P_D^K(Z_i)' W P_D^K(Z_i) \right)^{-1} P_D^K(Z_i)' W \left(Y_i - X_i' \widehat{\beta}_D \right)$$

with $W = \text{diag} \left(\frac{D_i T_i}{\mathbb{E}[T_i]} - \frac{D_i(1-T_i)}{(1-\mathbb{E}[T_i])} \right)$ and $\widehat{\beta}_D$ is a weighted version of the Donald & Newey (1994) estimator for the linear component.

For the reduced form, we build on Assumption 7 and assume that the nonparametric component of the conditional mean can be approximated by a vector of basis functions $P_Y^K(\cdot)$:

$$\begin{aligned} \mathbb{E}[Y_{it} | X_i = x, Z_i = z] &= x' \beta_{0t} + x' (\beta_{1t} - \beta_{0t}) \pi_t(z, x) + K(\pi_t(z, x)) \\ &\approx x' \beta_{0t} + x' (\beta_{1t} - \beta_{0t}) \left(x' \beta_{Dt} + P_D^K(z)' \gamma_{Dt} \right) + P_Y^K(z)' \gamma_{Yt} \end{aligned}$$

The estimand for the reduced form is thus approximated by

$$\mathbb{E} \left[\frac{Y_i T_i}{\mathbb{E}[T_i]} - \frac{Y_i(1-T_i)}{(1-\mathbb{E}[T_i])} \middle| X_i = x, Z_i = z \right] \approx x' \alpha + x' \delta \left(x' \beta_D + P_D^K(z)' \gamma_D \right) + P_Y^K(z)' \gamma_Y$$

The vector of coefficients $(\alpha, \delta, \gamma_Y)$ is estimated by

$$\begin{pmatrix} \widehat{\alpha} \\ \widehat{\delta} \\ \widehat{\gamma}_Y \end{pmatrix} = \left(R' W R \right)^{-1} R' W Y$$

where $R = \left(X, X \left(X' \widehat{\beta}_D + P_D^K(Z)' \widehat{\gamma}_D \right), P_Y^K(Z) \right)$

and thus the limit version of the estimand is

$$\partial \widehat{\mathbb{E}} \left[\frac{Y_i T_i}{\mathbb{E}[T_i]} - \frac{Y_i(1-T_i)}{(1-\mathbb{E}[T_i])} \middle| X_i = x, Z_i = z \right] / \partial z = x' \widehat{\delta} \nabla P_D^K(z)' \widehat{\gamma}_D + \nabla P_Y^K(z)' \widehat{\gamma}_Y \quad (1.2)$$

The sieve estimator is the ratio of the first-stage estimator and the reduced-form estimator. Theorem 2 shows that the sieve estimator is \sqrt{N} -consistent and asymptotically normal. The functional form of the variance is defined in Section 3 of the Appendix. However, since the functional form can take complicated expressions, using the bootstrap might be convenient for inference. Theorem 2 also shows that that bootstrap confidence intervals are asymptotically valid for the estimator.

Theorem 2. *Let Assumptions 1-7 and Assumptions 8-10 from Section 3 of the appendix hold. Then,*

$$\sqrt{N} \left(\widehat{MTE}(\pi_1(z, x)) - MTE(\pi_1(z, x)) \right) \xrightarrow{d} N(0, V_{MTE(\pi_1(z, x))})$$

where $V_{MTE(\pi_1(z, x))}$ is defined in Section 3 of the Appendix.

Moreover, the bootstrap is consistent for $\widehat{MTE}(\pi_1(z, x))$.

The estimation procedure outline in this section can be summarized as follows:

1. Estimation of the modified first-stage through a weighted sieve estimator of the propensity score (equation 1.1).
2. Estimation of the modified reduced through a weighted sieve estimator of the propensity score (equation 1.2).
3. Estimates for the MTE are obtained by taking the ratio of these quantities.
4. The MTE curve is estimated by evaluating the ration along the support of the instrument.
5. Estimates for the Confidence Interval are obtained through bootstrap.

1.4 Doubly-Robust estimation of LATE

In the 2×2 setting with a binary instrument, the series estimator is equivalent to a TWFE estimator augmented by a nonparametric control function. Assume the following working model for potential outcomes:

$$Y_i(d) = Z_i + T_i + X_i' \beta_d + U_{id}, \quad d \in \{0, 1\}$$

and define

$$K_1(Z_i, T_i) = \mathbb{E}[U_{i1} | Z_i, T_i]$$

$$K_0(Z_i, T_i) = \mathbb{E}[U_{i0} | Z_i, T_i]$$

as control functions (OLSEN, 1980; HECKMAN; ROBB, 1985).

Assuming linearity in the regressors, we can write the switching regression model as the following TWFE version of the regression model from (HECKMAN; URZUA; VYTLACIL, 2006):

$$Y_i = Z_i + T_i + X_i' \alpha + D_i X_i' \tau + D_i K_1(Z_i, T_i) + (1 - D_i) K_0(Z_i, T_i) + \varepsilon_i$$

with $\alpha = \beta_0$ and $\tau = \beta_1 - \beta_0$. The coefficient τ captures the observed gains from treatment, while the unobservable gains from treatment are obtained by differentiating and rearranging the control functions (BRINCH; MOGSTAD; WISWALL, 2017). In the regression model, the parameter τ represents the LATE.

In practice, it is not uncommon for researchers to use parametric approximations for the control function (CORNELISSEN et al., 2016; CORNELISSEN et al., 2018). The misspecification of the control function and/or the propensity score will lead to inconsistent estimates of treatment effects.

We propose a doubly robust estimator for the LATE, which augments the TWFE regression specification with a control function and unit-specific weights derived from the propensity score and a reshaped distribution of D .

The estimator is a function of two objects: the estimated propensity score \hat{P} and the outcome model $M_d(Z, T, X, p) = \mu_0(X, T) + K_d(Z, T, p)$, $d \in \{0, 1\}$. In that sense, it is an adaptation of the Reshaped Inverse Probability Weighting (RIPW) estimator for panel data proposed by (ARKHANGELSKY et al., 2021). We modify it to account for essential heterogeneity by including the control function in the outcome model and the case of repeated cross-sections.

Given an estimate \hat{P} for the propensity score and estimates for $M_d(G, T, X, p)$ we consider the following RIPW estimator:

$$\hat{\tau}(\pi) = \underset{\mu, \tau}{\operatorname{argmin}} \sum_{i=1}^N ((Y_i - M_{id}(Z_i, T_i, X_i, p_i)) - \mu - D_i\tau)^2 \frac{\pi(D_i; X_i)}{\hat{P}}$$

where $\pi(D_i; X_i)$ is a density for D_i conditional on the covariates X_i ³.

The following theorem states the solution for the RIPW minimization problem.

Theorem 3. *Define*

$$\Phi_i = \frac{\pi(D_i; X_i)}{\hat{P}_D(Z_i, T_i, X_i)}$$

and

$$\tilde{Y}_i = Y_i - M_{ij}(Z_i, T_i, X_i, \hat{P}_D(Z_i, T_i, X_i))$$

Furthermore, let $\Gamma_\Phi = \frac{1}{N} \sum_{i=1}^N \Phi_i$, $\Gamma_{DD} = \frac{1}{N} \sum_{i=1}^N \Phi_i D_i^2$, $\Gamma_{DY} = \frac{1}{N} \sum_{i=1}^N \Phi_i D_i \tilde{Y}_i$, $\Gamma_D = \frac{1}{N} \sum_{i=1}^N \Phi_i D_i$ and $\Gamma_Y = \frac{1}{N} \sum_{i=1}^N \Phi_i \tilde{Y}_i$.

Under Assumptions 1-7,

³ (ARKHANGELSKY et al., 2021) show that the density can be arbitrary in the 2×2 DiD setting, as long as positive probabilities are assigned for both treatment status.

$$\widehat{\tau}(\pi) = \frac{\Gamma_{DY} - \Gamma_{\Phi}^{-1}\Gamma_D\Gamma_Y}{\Gamma_{DD} - \Gamma_{\Phi}^{-1}\Gamma_D^2}$$

See Section 4 of the Appendix for the proof.

To state the double-robust property of the RIPW estimator, we invoke another assumption regarding the convergence of the potential outcomes model and the model for the control function. First, we define the maximal conditional correlation between $\mathcal{W}_i = (Y_i(1), Y_i(0), Z_i, T_i, X_i, D_i)$ and $\mathcal{W}_j = (Y_j(1), Y_j(0), Z_j, T_j, X_j, D_j)$ as

$$\rho_{ij} = \sup_{f,g} \text{Corr}(f(\mathcal{W}_i), g(\mathcal{W}_j) | Z, T, X)$$

We write ρ_i when evaluating the correlation of the measurable functions f, g at i .

Assumption 11: There exists $q \in (0, 1]$ such that

$$\frac{1}{N^2} \rho_i \left\{ \mathbb{E} \left[\left| \tilde{Y}_i(1) \right|^2 \right] + \mathbb{E} \left[\left| \tilde{Y}_i(0) \right|^2 \right] + 1 \right\} = O(N^{-q})$$

and

$$\frac{1}{N} \left\{ \mathbb{E} \left[\left| \tilde{Y}_i(1) \right|^2 \right] + \mathbb{E} \left[\left| \tilde{Y}_i(0) \right|^2 \right] \right\} = O(1)$$

Theorem 4. Under Assumptions 1-7 and 11, $\widehat{\tau}$ is a consistent estimator for τ if either

- (1) The outcome model $\widehat{M}_d(G, T, X, P)$ is correctly specified for $d \in \{0, 1\}$ or
- (2) The propensity score \widehat{P} is correctly specified.

See Section 5 of the Appendix for the proof and Section 6 for the derivation of a Wald-type confidence interval for $\widehat{\tau}$.

Remark: Although the double-robustness property can be proved, it is hard to think of applications in which the control function does not depend on the propensity score, which means double-robustness is possible but is unlikely to hold in practice under the misspecification of the propensity score. Hence, we recommend that the applied researcher take this estimator as a parametric estimator that is robust to the misspecification of the control function. For a locally robust estimator that is robust to small deviations from the true model, see (ESCANCIANO; PEREZ-IZQUIERDO, 2023).

1.5 Monte Carlo Simulations

In this section, we conduct a series of Monte Carlo exercises to evaluate the finite-sample properties of our proposed estimators. First, we compare the modified semiparametric LIV estimator to the standard LIV semiparametric estimator, as outlined in Appendix B of (CORNELISSEN et al., 2016), across different values of the MTE. Next, we assess various estimators for the LATE, including the standard 2SLS estimator, the doubly robust DiD estimator (DR-DID) proposed by (SANT'ANNA; ZHAO, 2020), the Wald-DID estimator used in (DUFLO, 2001), and the Time-Corrected WALD-DID estimator (TC-WALD) suggested by (CHAISEMARTIN; D'HAULTEFOEUILLE, 2018). These estimators are compared to an integral of the semiparametric estimator (CF-DID) from Section 3 and the RIPW estimator from Section 4.

In the simulation exercises, we consider a linear probability model for selection into treatment and a linear working model for the evolution of the outcomes. The choice model is

$$D_i = 1 \{ \alpha_1 X_i + \alpha_2 Z_i + \alpha_3 T_i \geq V_i \}$$

where V_i follows a standard uniform distribution, the coefficients are specified so that the left-hand side of the inequality is also bounded within the unit interval.

We specify the potential outcomes as

$$\begin{aligned} Y_i(0) &= \mu_0(Z_i, T_i, X_i) + U_{i0} \\ &= c_i + \beta T_i + \phi_0 X_i + U_{i0} \end{aligned}$$

and

$$\begin{aligned} Y_i(1) &= \mu_1(Z_i, T_i, X_i) + U_{i1} \\ &= c_i + \beta T_i + \phi_1 X_i + U_{i1} \end{aligned}$$

where c_i is a fixed effect with mean Z_i . We specify a linear correlation of U_{i1} , U_{i0} and V_i , so that the MTE function is linear.

Table 1 shows the results of the finite sample properties of the semiparametric MTE-DiD estimator for three values of $V_{i1} = \pi_1(z, x)$.

In terms of bias, the performance of the pointwise estimator is stable across different values of the curve of distaste for treatment. However, the confidence intervals for the more extreme values are considerably larger than for $V_{i1} = 0.5$.

Figure 1 (a) compares the true MTE curve to the one estimated by the semiparametric procedure, evaluated at the mean value of the covariates. The blue line represents the true MTE

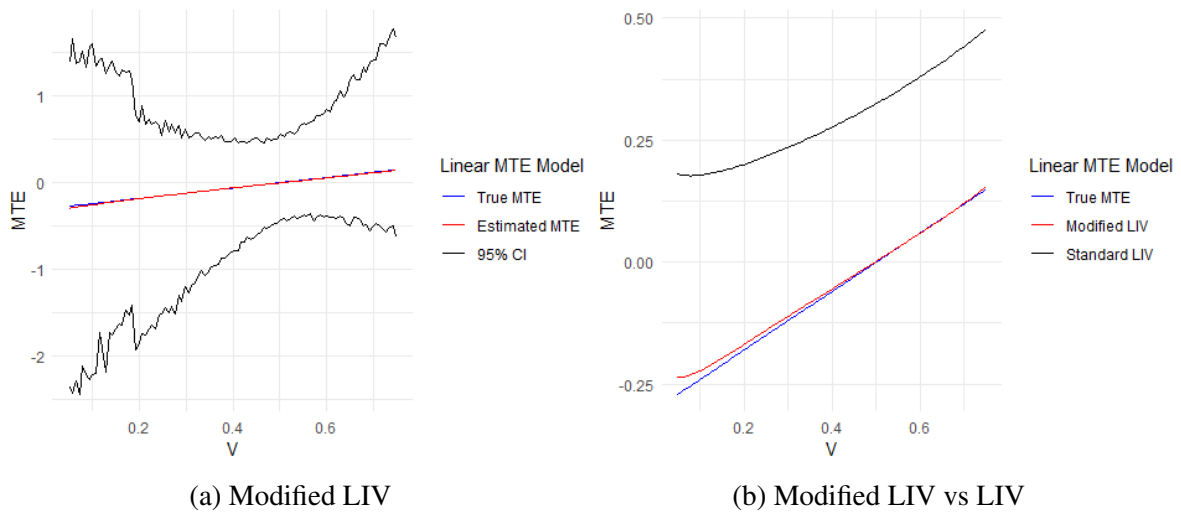
Table 1 - Monte Carlo Simulations: Modified LIV Estimator

$MTE(\pi_1(z, x))$	Av. Bias	Med. Bias	RMSE	Cover	CIL
$\pi_1(z, x) = 0.5$	-0.001	-0.001	0.203	0.942	1.049
$\pi_1(z, x) = 0.7$	0.002	-0.002	0.628	0.943	3.615
$\pi_1(z, x) = 0.3$	0.001	0.004	0.594	0.932	4.346

Note: Simulations based on 10,000 Monte Carlo experiments. ‘‘Av. Bias’’, ‘‘Med. Bias’’, ‘‘RMSE’’, ‘‘Cover’’ and ‘‘CIL’’, stand for the average simulated bias, median simulated bias, simulated root-mean-squared errors, 95% coverage probability, and 95% confidence interval length, respectively.

curve, the red line shows the estimated curve, and the black lines show the bounds for the 95% bootstrapped confidence interval.

Figure 1 - MTE Curve



Note: This figure displays the average MTE estimates based on the DGP from Section 5. The MTE is evaluated at the mean value of the covariates. The 95% percent confidence interval is computed from a bootstrap with 500 replications. The y -axis measures the value of the MTE from the DGP, whereas the x -axis measures the unobserved distaste for treatment.

The estimator successfully recovers the MTE curve for areas where the empirical common support is verified. However, there is substantial variation in the length of the confidence interval of pointwise estimates across the curve. For more extreme values of $V_{i1} = p$ close to 0.2 or 0.8, for instance, the confidence intervals are considerably larger than the intervals for values around the $[0.4, 0.7]$ interval.

Figure 1 (b) compares the proposed semiparametric estimator with the standard LIV estimator and the true MTE curve. The standard LIV estimator does not recover the true MTE values at any point of the curve, its curve is shifted upwards from the true curve because the estimator captures the fixed effect c_i from the DGP which violates the instrument independence assumption under which the estimator is consistent.

Table 2 shows the results of the finite sample properties of the standard semiparametric MTE estimator for three values of $V_{i1} = \pi_1(z, x)$. The estimator is severely biased in different points of the MTE curve.

Table 2 - Monte Carlo Simulations: Standard LIV Estimator

$MTE(\pi_1(z, x))$	Av. Bias	Med. Bias	RMSE	Cover	CIL
$\pi_1(z, x) = 0.5$	0.338	0.327	0.362	0.647	1.030
$\pi_1(z, x) = 0.7$	0.378	0.337	0.689	0.743	3.576
$\pi_1(z, x) = 0.3$	0.375	0.339	0.664	0.623	2.772

Note: Simulations based on 10,000 Monte Carlo experiments. “Av. Bias”, “Med. Bias”, “RMSE”, “Cover” and “CIL”, stand for the average simulated bias, median simulated bias, simulated root-mean-squared errors, 95% coverage probability, and 95% confidence interval length, respectively.

Table 3 - Control Function and Propensity Score are correct.

Estimator	Av. Bias	Med. Bias	RMSE	Cover	CIL
2SLS	-1.613	-1.614	1.615	0.000	0.208
DR-DID	1.611	1.612	1.615	0.327	2.632
WALD-DID	2.971	2.863	2.988	0.657	4.542
TC-WALD	-0.450	-0.439	0.449	0.768	1.822
CF-DID	-0.010	-0.011	0.209	0.953	1.088
RIPW	-0.012	-0.008	0.209	0.947	1.327

Note: Simulations based on 10,000 Monte Carlo experiments. TWFE is the two-way fixed effects estimator, DR-DID is the Doubly-Robust DiD estimator as proposed in Sant’anna and Zhao (2020), WALD-DID is the Wald-DiD estimator as used in Duflo (2001), TC-WALD is the Time-Corrected Wald Ratio proposed by De Chaisemartin and D’Haultfoeuille (2018), CF-DID and RIPW are our proposed estimator. “Av. Bias”, “Med. Bias”, “RMSE”, “Cover” and “CIL”, stand for the average simulated bias, median simulated bias, simulated root-mean-squared errors, 95% coverage probability, and 95% confidence interval length, respectively.

Table 3 displays the simulation results regarding the estimation of LATE. The existing estimators from the literature fail to recover the LATE in the presence of imperfect compliance. The semiparametric series estimator (CF-DID) and the RIPW estimator have similar finite-sample performance in terms of bias and coverage.

1.6 Empirical illustration

We illustrate the use of our estimators by revisiting (DUFLO, 2001), which uses data from the 1995 intercensal survey to analyze the returns to schooling in Indonesia by exploiting a major government school construction program as a natural experiment in a DiD design.

In 1973, the Indonesian government launched the INPRES program, a major primary school construction program. In the setting, the year of birth plays the role of time. Men born between 1957 and 1962 are defined as cohort 0, as they should have finished primary school by the time the program was launched. Men born between 1968 and 1972 are described as cohort 1 since they had the age to enroll in primary education after the program. Treatment and control groups are defined according to the number of primary schools per capita in each district. The author regresses the number of primary schools constructed on the number of school-age children in each district and defines treatment districts as those with a positive residual in that regression.

Table 4 - Returns to primary school using the groups from Dulfo (2001)

	WALD-DID	TC-WALD	CF-DID	RIPW
Returns to education	0.345	0.317	0.318	0.186
95% CI	[-0.075, 1.432]	[-0.062, 1.363]	[0.041, 0.584]	[0.116, 0.354]

Note: Sample size: 10.805 observations. Confidence Intervals are based on nonparametric bootstrapped standard errors.

The outcome of interest is the logarithm of wages, and the treatment variable is the individual's schooling years. Since our method is suited for binary treatments, we recategorize the treatment variable to take value 1 for individuals who attended primary school and 0 for individuals who did not. Individuals with greater educational attainment were excluded from the sample. Our sample consists of 10.805 observations. We include the district's school enrollment rate in 1971, the presence of a water and sanitation program, the number of families in the individual's household, and the individual's family size as covariates in the regressions.

Nonparametric estimators are severely under-powered in this setting (see (CHAISEMARTIN; D'HAULTEFOEUILLE, 2018)), so we estimate the LATE using the polynomial parametric estimator and the RIPW and compare the results obtained using the WALD-DID and the TC-WALD. We also use the estimates from the parametric control function to derive an estimate of the MTE curve to recover the heterogeneity in returns to primary education across individuals as a function of their probability of attending primary school. We estimate the propensity score through a logistic regression of treatment status on the covariates above and the number of primary schools per capita per district in the post-expansion period as the instrument.

1.6.1 LATE

Table 4 displays the value of the estimates obtained using the WALD-DID, the TC-WALD, the sieve control function estimator(CF-DID), and the RIPW. Standard errors for the first three estimators were obtained using 999 bootstrap replications of a nonparametric bootstrap. In contrast, the standard error for the RIPW was obtained through the plug-in estimator from Section 5 of Appendix A. We test several different basis functions for the control function estimator. We display the results for the linear specification, which is the one that shows the smallest mean-squared error in the "leave-one-out"cross-validation exercise.

Returns to primary school attendance are large. However, the estimate that is statistically different from 0 is the one that comes from the control function, and the RIPW estimators, which are the parametric estimators ⁴.

The RIPW estimator is flexible enough to allow for misspecification of the control function or the propensity score. Therefore, the more credible estimates suggest that primary school attendance increases wages by around 19%.

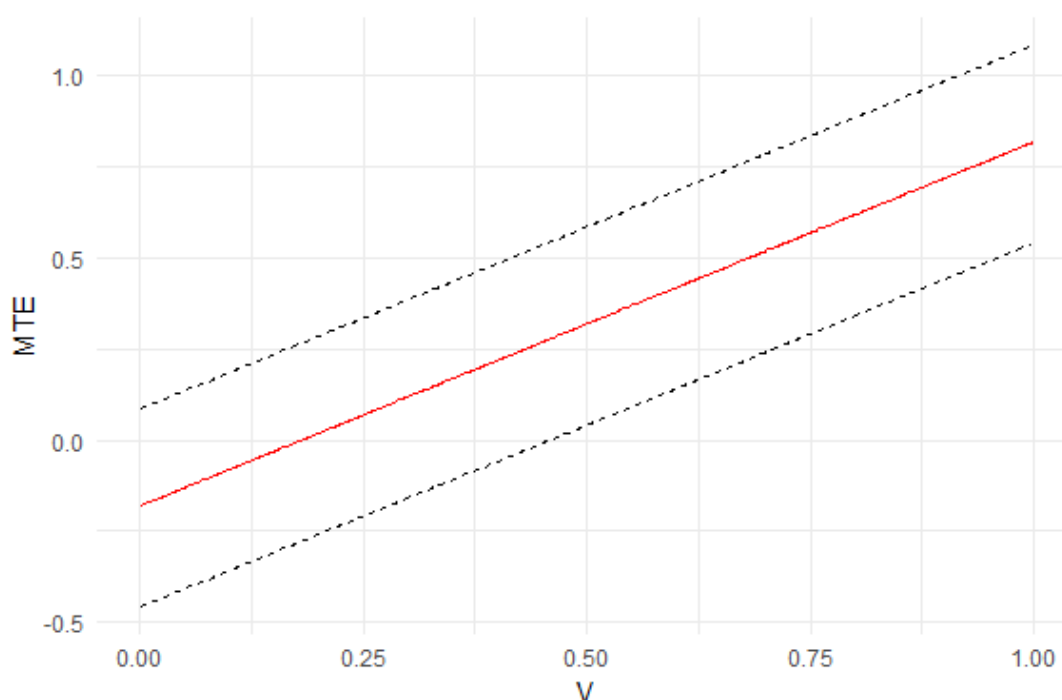
⁴ (CHAISEMARTIN; D'HAULTEFOEUILLE, 2018) also find positive but not statistically significant returns to education, see Table 1.

1.6.2 MTE Curve

Figure 2 shows that the returns to primary school attendance depend on V_1 , with 95% confidence intervals constructed from a nonparametric bootstrap. The MTEs are evaluated at the mean value of the covariates.

Following (CARNEIRO; HECKMAN; VYTLACIL, 2011), we drop observations for which the estimated propensity score is below the 0.01 percentile or above the 0.99 percentile in the distribution of the first-stage.

Figure 2 - MTE curve for the returns to primary education



Note: Figure 2 depicts the MTE curve for returns to primary education estimated by the parametric polynomial method. The 95% confidence interval is based on nonparametric bootstrapped standard errors.

The estimates suggest that the returns to primary school vary substantially in magnitude and sign. More specifically, individuals with V_1 greater than 0.17 have positive returns to primary school, while individuals with V_1 smaller than 0.17 experience negative returns to education.

This pattern of reverse selection in gains appears in other settings that measure early education returns, for example, in (CORNELISSEN et al., 2018), which analyzes the returns of early childcare attendance in Germany.

The results in Table 4 exhibit the positive returns of primary school attendance for the post-expansion period but mask substantial heterogeneity in the returns.

1.7 Conclusion

The Marginal Treatment Effect provides a choice-theoretic foundation that unifies the econometric literature on causal inference. It has a clear economic interpretation and summarizes all other conventional treatment parameters.

In this paper, we show how the difference-in-differences method can identify the marginal treatment effect (MTE) under a functional framework that accounts for treatment heterogeneity. This approach considers unobservable characteristics that influence selection into treatment, using a modified version of the local instrumental variable (LIV) estimand.

We propose a semiparametric estimator for the marginal treatment effect (MTE) that remains consistent within a partially linear model. This model assumes linearity in the regressors, while leaving the unobserved terms unspecified, and approximating them using a vector of basis functions.

We also demonstrate that reweighting the objective function of our parametric estimator at the unit level enhances the robustness of the resulting estimator for the local average treatment effect (LATE). This approach offers a less demanding procedure for applied researchers to estimate the LATE in "fuzzy" difference-in-differences (DiD) settings.

We derive the large-sample properties of our MTE estimators and illustrate their desirable finite-sample properties via simulation exercises. We recommend using bootstrap for inference, the standard practice for semiparametric MTE estimators.

The empirical application illustrates the economic insights that can come from the estimation of the MTE curve. Applying the semiparametric control function DiD estimator in the setting of (DUFLO, 2001), we find substantial heterogeneity in the effects of primary school attendance on future earnings and a pattern of reverse selection on gains which could not be recovered by any other conventional policy evaluation parameter.

This work is only the first effort to build a complete theory for the MTE framework in DiD designs. A natural direction to advance this agenda is to analyze the conditions under which dynamic MTEs can be identified in DiD settings with staggered adoption and dynamic treatment effects, following recent advances in the literature (CALLAWAY; SANT'ANNA, 2021; GOODMAN-BACON, 2021; BORUSYAK; JARAVEL; SPIESS, 2021).

Another important step in future research is to consider the 2×2 setting and the cases with multiple periods where the treatment is multi-valued or continuous (CALLAWAY; GOODMAN-BACON; SANT'ANNA, 2021).

2 Instrumental Variables with Multiple Time Periods

The instrumental variables (IV) method has been widely studied in cross-sectional settings. However, many practical applications involve panel data, in settings where a unit's treatment status may turn on or off over time. I show that in the presence of dynamic treatment effects, i.e., if past treatments affect current potential outcomes, standard methods are no longer valid if the instruments are serially correlated. This paper shows the nonparametric identification of dynamic causal effects in a potential outcomes framework in which potential outcomes depend on the treatment path taken by a unit through time but the first stage is static, in the sense that at each period the IV only instruments its contemporary treatment. I provide a nonparametric estimator that is consistent over the randomization distribution and derive its finite-population limiting distribution as the sample size increases. Monte Carlo Simulations illustrate the desirable finite-sample properties of the estimators. An application of the estimator show that law enforcement curbs illegal deforestation until two years after the actual detection of illegal practice, but effects fade out afterwards.

2.1 Introduction

In a seminal paper, (IMBENS; ANGRIST, 1994) showed that the instrumental variables (IV) estimand in cross-sectional settings can be interpreted as the Local Average Treatment Effect (LATE), defined as the average treatment effect for the subpopulation that has its treatment status shifted by an excluded instrument, the so-called compliers. However, applications of the IV method using panel data are also pervasive in the literature. In such cases, it is common to find the following static two-stage least squares (2SLS) specification estimated across time periods:

$$\begin{aligned} Y_{i,t} &= \alpha + \beta D_{i,t} + c_t + \varepsilon_{i,t} \\ D_{i,t} &= \kappa + \theta Z_{i,t} + c_t + \eta_{i,t} \end{aligned}$$

where $Y_{i,t}$ is the observed outcome for unit i in period t , α and κ are constants, $D_{i,t}$ is a binary non-absorbing treatment¹ of interest in period t , $Z_{i,t}$ is the instrument in period t , c_t are time fixed effects and $\varepsilon_{i,t}$ and $\eta_{i,t}$ are the 2SLS error terms.

As a leading case, consider this stylized example inspired by the setting of (ASSUNÇÃO; GANDOUR; ROCHA, 2023) (hereafter referred to as AGR), in which the authors investigate

¹ Absorbing treatments are treatments that are not "forgotten" by units and remains constant in post-treatment periods (e.g a policy change, a change in the minimum wage, etc...).

the effects of a satellite law enforcement program, the DETER program, on deforestation in the Brazilian Amazon. In this example, $Y_{i,t}$ is the deforested area of municipality i in the year t , $D_{i,t}$ is an indicator for “intense” law enforcement in municipality i in year t , and $Z_{i,t}$ is an indicator for “intense” cloud coverage in municipality i and year t .

Such specifications are common in applied research. To document this practice, I searched the Web of Science database for articles published between 2000 and 2023 containing the words “instrumental variable” and “panel data” in the abstract, title, or topic words. The restricted search for six leading journals² found 33 articles matching the criteria. Some prominent examples are in the development literature (ACEMOGLU et al., 2019), inequality (AGHION et al., 2018), education (JACKSON; JOHNSON; PERSICO, 2015), trade (BLANCHARD; MATSCHKE, 2015) and behavioral economics (STANGO; ZINMAN, 2022).

The static estimand β can be decomposed as a weighted average of period-specific Wald estimands (KOLESÁR, 2013). The well-know results from (IMBENS; ANGRIST, 1994) imply that the static 2SLS specification yields a LATE in the absence of dynamics, that is, if $Z_{i,t}$ only affects $D_{i,t}$ and $Y_{i,t}$, but not $D_{i,t'}$ and $Y_{i,t'}$ for $t \neq t'$. In practice we might often expect dynamic effects of treatment. In my motivating example, law enforcement in period t might affect deforestation in period $t + 1$. By decomposing the period-specific Wald estimand, I show that the static specification generally does not identify well defined causal effects if past treatments affect the outcome when there is serial correlation of the instrument³. Similar results have been derived in (CHAISEMARTIN; LEI, 2023) and (SHEN; CHOI; SEONG, 2024) in related, but different dynamic IV settings⁴.

In this paper, I propose a novel identification approach for dynamic causal effects using instrumental variables and panel data when potential outcomes that are defined in terms of the sequence of treatments taken by a unit. As in (ROBINS, 1986; ROBINS, 1987), potential outcomes for unit i in period t depend on the path of treatments taken until period t , $Y_{i,t}(D_{i,1}, \dots, D_{i,t})$. I then consider the effect of changing the path of treatments. In particular, I focus on what (BOJINOV; RAMBACHAN; SHEPHARD, 2021) called the lag- p dynamic causal effect, defined as the difference between the outcomes from following different treatment paths from period $t - p$ to t , fixing the assignments for unit i to follow the observed path up to time $t - p - 1$.

The key restriction for identification is the absence of dynamics in the first stage, that

² *Econometrica*, *Journal of Political Economy*, *Quarterly Journal of Economics*, *American Economic Review*, *Review of Economic Studies* and *Review of Economics and Statistics*

³ An intuitive alternative would be a multivariate 2SLS specification with the sequence of endogenous treatments until period t being instrumented by the sequence of instruments until period t . I show that this approach identifies a weighted average of different dynamic causal effects in which weights can be negative. Much like in the multiple treatments IV literature (KIRKEBOEN; LEUVEN; MOGSTAD, 2016; MOUNTJOY, 2022; BHULLER; SIGSTAD, 2024), multivariate 2SLS estimates only hold a clear causal interpretation in the absence of treatment effect heterogeneity.

⁴ (CHAISEMARTIN; LEI, 2023) analyzes the causal decomposition of a first-difference IV estimand under static potential outcomes and (SHEN; CHOI; SEONG, 2024) analyzes the causal decomposition of 2SLS estimates when treatment follows an AR(1) process.

is, $D_{i,t}(Z_{i,1}, \dots, Z_{i,T}) = D_{i,t}(Z_{i,t})$ which is plausible in the AGR setting, in which the authors argue that cloud coverage in period t only affects satellite visibility, and thus law enforcement, in period t . The static first-stage assumption is plausible in many other settings. For instance, multi-period experiments evaluating the effects of advertisements on purchase behavior using advertisement display as the instrument or criminal justice settings with repeated offenders using judge leniency as the instrument.

The strategy consists in identifying the potential outcomes associated to different treatment paths separately, by exploiting all possible variations in the assignment path from period $t - p$ to t under a Wald-like estimand. That is, potential outcomes associated to contemporary treatments ($p = 0$) are identified by exploiting the variation in assignment in period t holding the path until $t-1$ fixed, which takes the form of a simple difference (compare units with $Z_{i,t} = 1$ versus units with $Z_{i,t} = 0$). Potential outcomes associated to a path of treatments from $t - 1$ to t ($p = 1$) are identified exploiting variations in the path of assignment in a Wald “difference-in-differences” format.

The approach takes a purely design-based approach to uncertainty and is fully agnostic about functional forms. I provide nonparametric estimators that are consistent over the randomization distribution and asymptotically normal as the finite population grows to infinity. The limiting distribution can be used to perform conservative tests on weak null hypotheses of no average dynamic causal effects. I propose the construction of a conservative dynamic version of the (BLOOM, 1984) confidence interval for hypothesis testing. Monte Carlo Simulation studies illustrate the desirable finite-sample properties of the proposed estimators.

Finally, I use the proposed estimators to revisit AGR and analyze the dynamic effects of law enforcement on illegal deforestation in the Brazilian Amazon. The results suggest there are dynamic causal effect of law enforcement, which persist through time until a certain point when they fade out. The estimates show that law enforcement curbs illegal deforestation until two years after the actual detection of illegal practice, but effects fade out afterwards.

Related Literature: This paper relates to several strands of the causal inference literature. There is an extensive literature on dynamic treatment settings, specifically from the biostatistics literature ((ROBINS, 1986), (MURPHY; LAAN; ROBINS, 2001), (HERNAN; ROBINS, 2023)). There is also a more recent interest from econometricians in such settings ((BOJINOV; RAMBACHAN; SHEPHARD, 2021); (ARKHANGELSKY; IMBENS, 2022)). The aforementioned works focus on identification under the assumption that treatment assignment is sequentially randomized and that compliance is perfect. By contrast, I consider settings where there is an instrument and imperfect compliance.

There is also a large literature on the identification of dynamic treatment effects using instrumental variables, dating back to (ARELLANO; BOND, 1991). (ABBRING; HECKMAN, 2007), (HECKMAN; NAVARRO, 2007) and (HECKMAN; HUMPHRIES; VERAMENDI, 2016), focus on the identification of time-to-treatment effects in settings with absorbing tre-

atments. This paper differs from the ones mentioned above by focusing on non-absorbing treatments and taking a fully nonparametric approach to identification without relying on identification-at-infinity arguments or parametric factor models.

When it comes to sequential treatment settings, (PHAM; CHEN, 2017) propose a non-parametric identification approach that relies on the knowledge of the compliance type from each individual in the sample. (HAN, 2021) and (HAN, 2023) focus on how the support of the instrument and covariates allows the identification of average causal effects in sequential regimes. (SOTRA; SYRGKANIS, 2024) focuses on identification of dynamic causal effects in dynamic IV settings with one-sided noncompliance, and (SHEN; CHOI; SEONG, 2024) studies Panel IV regressions when the instrument follows a first-order autoregressive (AR-1) model. This paper also focus on sequential treatment regimes, but relies on different assumptions for identification. Namely, by restricting how the path of assignments affects potential treatments, I provide a new identification approach that is fully non-parametric that avoids two-way exclusion restrictions and restrictions on noncompliance.

In a certain way, the sequential treatment regime is a setting with multiple treatments, multiple potential outcomes and multiple instruments. In that sense, this paper is related to the recent advances in the literature regarding multiple treatments in IV settings. ((KIRKEBOEN; LEUVEN; MOGSTAD, 2016); (MOUNTJOY, 2022)). The panel setting gives additional exclusion restrictions which can be exploited in order to overcome the well documented identification challenges in the multiple treatments and multiple IVs cross-sectional literature.

Finally, the paper relates to the literature on design-based inference in IV settings ((IMBENS; RUBIN, 2015); (KANG; PECK; KEELE, 2018); (RAMBACHAN; ROTH, 2025), (BORUSYAK; HULL, 2024)). While these papers focus solely on estimators for the case where treatment effects are static, I consider design-based inference for IV settings with dynamic causal effects.

Outline of the paper: Section 2 provides the interpretation of the static 2SLS specification and the proposed causal estimand in a framework with only two-periods. Section 3 defines the dynamic potential outcomes framework, states the identification assumptions, define the target causal parameters and provides the general estimand for the identification of dynamic causal effects. In Section 4 I provide a nonparametric estimator for the dynamic causal effects and derive its finite-population asymptotic distribution. Section 5 shows how the randomization distribution can be modified to incorporate common features from applied work using panel data. Section 6 presents the Monte Carlo simulations and Section 7 the empirical application. Section 8 concludes.

Notation: I use the notation from (BOJINOV; RAMBACHAN; SHEPHARD, 2021). For an integer $t \geq 1$ and a random variable R_t , we write $R_{1:t} = (R_1, \dots, R_t)$. Sets of observation units and time periods are respectively defined in a compact form as $[N] := \{1, \dots, N\}$ and $[T] := \{1, \dots, T\}$.

2.2 Identification with $T = 2$

As a leading case, consider a two-period model in which we observe at each period a binary treatment status ($D_{i,t}$), a binary instrument ($Z_{i,t}$) and a scalar outcome ($Y_{i,t}$). In the motivating application, $Y_{i,t}$ is deforestation, $D_{i,t}$ is law enforcement and $Z_{i,t}$ is cloud coverage. Then deforestation in period 2 might depend on law enforcement and cloud coverage from period 1.

Suppose that in the first period, the standard (IMBENS; ANGRIST, 1994) assumptions for cross-sectional settings hold. Thus, there are two potential treatments ($D_{i,1}(1)$ and $D_{i,1}(0)$) and the exclusion restriction implies that there are two potential outcomes ($Y_{i,1}(1)$ and $Y_{i,1}(0)$). Under the standard monotonicity assumption, individuals can be divided into three groups ($G_{i,1}$) according to their response to $Z_{i,1}$: always-takers (AT_1), never-takers (NT_1) and compliers (C_1). Note that in period 1, dynamics play no role in potential treatments nor potential outcomes. Hence, using data only from the first period, the static Wald estimand identifies the LATE for period 1:

$$\frac{\mathbb{E}[Y_{i,1}|Z_{i,1} = 1] - \mathbb{E}[Y_{i,1}|Z_{i,1} = 0]}{\mathbb{E}[D_{i,1}|Z_{i,1} = 1] - \mathbb{E}[D_{i,1}|Z_{i,1} = 0]} = \mathbb{E}[Y_{i,1}(1) - Y_{i,1}(0)|G_{i,1} = C_1]$$

In the second period, potential outcomes are functions of the sequence of treatments taken in periods 1 and 2. Thus, there are four potential treatments, and under a dynamic version of the exclusion restriction assumption ($Y_{i,2}(Z_{i,1:2}, D_{i,1:2}(Z_{i,1:2})) = Y_{i,2}(D_{i,1:2}(Z_{i,1:2}))$), there are four potential outcomes.

Assumption PO: $Y_{i,2} = Y_{i,2}(D_{i,1:2}(Z_{i,1:2}))$.

Suppose that potential outcomes are functions only of their contemporary treatments ($D_{i,1}(Z_{i,1:2}) = D_{i,1}(Z_{i,1})$ and $D_{i,2}(Z_{i,1:2}) = D_{i,2}(Z_{i,2})$). Furthermore, suppose that the (IMBENS; ANGRIST, 1994) assumptions regarding the first-stage hold for each period. Then there are three response types in period 1 and three response types in period 2, which means individuals can be divided into nine groups with respect to their compliance towards the assignment path corresponding to the nine possible combinations of always-taker, never-taker and complier at each period. The individuals that comply in both periods ($C_{1:2}$) are called path-compliers.

Assumption IA: $D_{i,1}(z_{1:2}) = D_{i,1}(z_1)$ and $D_{i,2}(z_{1:2}) = D_{i,2}(z_2)$. Moreover, $D_{i,1}(1) \geq D_{i,1}(0)$ and $D_{i,2}(1) \geq D_{i,2}(0)$.

The vector of instruments $Z_{i,1:2}$ is assumed to be independent from the vector of potential treatments and outcomes ($D_{i,1:2}(z_{1:2}), Y_{i,1}(D_{i,1}(z_1)), Y_{i,2}(D_{i,1:2}(z_{1:2}))$).

Assumption SX: $(D_{i,1:2}(z_{1:2}), Y_{i,2}(D_{i,1:2}(z_{1:2}))) \perp Z_{i,1:2}$.

Below, I discuss the causal interpretation of the static Wald estimand under these potential outcomes model.

2.2.1 Interpretation of the Static Wald Estimand

In most applications of IV methods, researchers usually rely on a static specification of the Wald estimator, that is, there is an outcome equation in which the outcome in period t is specified as a function of an endogenous treatment taken in period t , instrumented by an excluded variable also realized in period t . I will refer to such specification as the Time- t static Wald estimator. In $T = 2$, the estimand amounts to

$$\beta_2^{Wald} = \frac{\beta_2^{RF}}{\beta_2^{FS}} = \frac{\mathbb{E}[Y_{i,2}|Z_{i,2} = 1] - \mathbb{E}[Y_{i,2}|Z_{i,2} = 0]}{\mathbb{E}[D_{i,2}|Z_{i,2} = 1] - \mathbb{E}[D_{i,2}|Z_{i,2} = 0]}$$

The static Wald estimand is the building block for the static 2SLS specification stacked across time periods, which is pervasive in the applied literature.

Next, I provide a causal decomposition of the static Wald estimand under the model described above. Proposition 1 shows that the estimand does not hold a straightforward causal interpretation without further assumptions:

Proposition 1. *Suppose that assumptions **PO**, **IA** and **SX** hold. Then,*

$$\begin{aligned} \beta_2^{Wald} &= \sum_{z \in \{0,1\}} \omega_z(1) \mathbb{E}[Y_{i,2}(D_{i,1}(\mathbf{z}), 1) - Y_{i,2}(D_{i,1}(\mathbf{z}), 0) | G_{i,2} = C_2] \\ &+ (\omega_1(1) - \omega_1(0)) \mathbb{E}[Y_{i,2}(1, D_{i,2}(0)) - Y_{i,2}(0, D_{i,2}(0)) | G_{i,1} = C_1] \frac{\mathbb{P}(G_{i,1} = C_1)}{\mathbb{P}(G_{i,2} = C_2)} \end{aligned}$$

where $\omega_z(1) = \mathbb{P}(Z_{i,1} = z | Z_{i,2} = 1)$ and $\omega_z(0) = \mathbb{P}(Z_{i,1} = z | Z_{i,2} = 0)$.

Proposition 1 shows that in general the static Wald estimand does not identify well defined causal effects. The result shows the estimand can be decomposed into a weighted average of the contemporary treatment in period 2 holding fixed the potential treatment in period 1 for period 2 compliers, with weights that are always positive, and the dynamic treatment effect holding the potential treatment in period 2 at $D_{i,2}(0)$ for period 1 compliers. However, this term can enter negatively into the expression if $\omega_1(0) > \omega_1(1)$.

Thus, the static Wald estimand can be interpreted as a weighted average of different dynamic causal effects for different groups, with potentially negative weights. As the expression shows, the weights associated to the causal effect of treatment in period 2 are always positive. The sign of the weights associated to the causal effects of receiving treatment in period, however, depend on the joint distribution of the instrument across time periods.

If $\omega_1(1) > \omega_1(0)$, then the causal effects of treatment in period 1 are incorporated with positive weights and the static Wald estimand is what (BLANDHOL et al., 2022) defined as a *weakly causal* estimand, which is an estimand that preserves the sign of the causal effects if all the causal effects have the same sign. Mapping back to the AGR setting, if cloud coverage is positively correlated over time, and all dynamic effects of law enforcement have the same

sign (eg. law enforcement always curbs deforestation), then the static Wald estimate will be negative as well. If $\omega_1(1) < \omega_1(0)$, then β_2^{Wald} reflects not only the causal effect of treatment 2 for compliers in period 2, but also negatively weighted causal effects of treatment in period 1 for compliers in period 1⁵.

To put it simply, Proposition 1 shows that the static Wald estimand does not hold a straightforward causal interpretation when there is negative serial correlation of the instrument.

Note that in the case where $Z_{i,2}$ and $Z_{i,1}$ are independent, the weights associated to the causal effect of treatment in period 1 collapse to zero, and thus the static Wald estimand hold a straightforward causal interpretation as a weighted average of contemporary treatment effects.

Lemma 1. *Suppose that assumptions **PO**, **IA** and **SX** hold. Furthermore, assume that $Z_{i,1}$ is independent from $Z_{i,2}$. Then,*

$$\beta_2^{Wald} = \sum_{z \in \{0,1\}} \mathbb{P}(Z_{i,1} = z) \mathbb{E}[Y_{i,2}(D_{i,1}(z), 1) - Y_{i,2}(D_{i,1}(z), 0) | G_{i,2} = C_2]$$

Lemma 1 shows that when the instrument in one period is independent from the instrument in the other period, then the β_2^{Wald} is a weighted average of contemporary treatments effects holding the potential treatment in period 1 fixed, with weights being given by the probability of observing the realized assignment in period 1.

(Linear Toy Model): For the sake of concreteness, suppose the outcome is generated by a linear causal model where potential outcomes are assumed to be homogeneous. The mean potential outcome for unit i in period 2 is

$$\mathbb{E}[Y_{i,2}(D_{i,1:2})] = \beta_0 + \beta_1 D_{i,1} + \beta_2 D_{i,2}$$

Here, β_0 is mean the fully untreated potential outcome of unit i , $\beta_0 + \beta_1$ is the potential outcome associated to taking treatment in the first period only, $\beta_0 + \beta_2$ is the potential outcome associated to taking treatment in period 2, and $\beta_0 + \beta_1 + \beta_2$ is the potential outcome associated to the path of full exposure to treatment.

Suppose that in the first period the share of compliers is α_1 , in the second period the share is α_2 , and that the joint first stage is the product of first-stages so that the share of path-compliers is simply $\alpha_1 \alpha_2$:

⁵ An alternative interpretation to Proposition 1 would be an Omitted Variable Bias (OVB) result. Under the presence of dynamic treatment effects, the correctly specified model includes not only $D_{i,2}$ but $D_{i,1}$ as well. However, $D_{i,1}$ is omitted from the static specification and might be correlated with $Z_{i,2}$ due to the serial correlation of $Z_{i,2}$ and $Z_{i,1}$.

$$\begin{aligned}\mathbb{P}(D_{i,1}(Z_{i,1}) = 1) &= \alpha_1^0 + \alpha_1^1 Z_{i,1} \\ \mathbb{P}(D_{i,2}(Z_{i,2}) = 1) &= \alpha_2^0 + \alpha_2^1 Z_{i,2} \\ \mathbb{P}(D_{i,1}(Z_{i,1}) = 1, D_{i,2}(Z_{i,2}) = 1) &= (\alpha_1^0 + \alpha_1^1 Z_{i,1})(\alpha_2^0 + \alpha_2^1 Z_{i,2})\end{aligned}$$

The expected value of the Time-t static Wald estimator under assumptions **PO**, **IA** and **SX** amounts to

$$\frac{\beta_2^{RF}}{\beta_2^{FS}} = \beta_2 + \frac{\beta_1 \alpha_1^1 (\mathbb{E}[Z_{i,1} = 1 | Z_{i,2} = 1] - \mathbb{E}[Z_{i,1} = 1 | Z_{i,2} = 0])}{\alpha_2^2}$$

It is clear that the ratio of means equals β_2 only if there is no dynamic effect of treatment ($\beta_1 = 0$), if there are no compliers in previous periods ($\alpha_1 = 0$), or the assignment of the instrument in period 2 is independent from the instrument in period 1 ($\mathbb{E}[Z_{i,1} = 1 | Z_{i,2} = 1] = \mathbb{E}[Z_{i,1} = 1 | Z_{i,2} = 0]$).

Translating to the AGR setting, the estimates for the effect of present law enforcement on illegal deforestation from the static 2SLS specification are biased if there are dynamic effects of law enforcement if the probability of a municipality experiencing "intense" cloud coverage depends on its history of past cloud coverage.

What does Proposition 1 imply for the usual approach of taking static 2SLS estimates and stacking them across time periods? The Proposition implies that the stacked estimator can be decomposed in terms of causal effects of contemporary treatment ($D_{i,2}$ on $Y_{i,2}$ and $D_{i,1}$ on $Y_{i,1}$) and dynamic effects of treatment ($D_{i,1}$ on $Y_{i,2}$), where the dynamic effects might enter the expression with negative weights, depending on the serial correlation of the instrument.

2.2.2 Separate Identification Approach

To overcome the limitations of the standard static Wald estimand and the multivariate 2SLS, I develop an alternative dynamic IV approach that separately identifies causal effects of different treatment paths for individuals that are compliers for the path of interest.

I propose a separate identification approach that isolates mean potential outcomes for path compliers. Taking the difference between two of these separate components generates a dynamic causal effect for path-compliers.

I illustrate the procedure by showing how the proposed estimand identifies the mean causal effect of full exposure for path-compliers, that is, $\mathbb{E}[Y_{i,2}(1, 1) - Y_{i,2}(0, 0) | G_{i,1:2} = C_{1:2}]$. The estimand is nonparametric and relies on the propensity score for the path of the instrument, defined as $\pi_i(\mathbf{z}) = \mathbb{P}(Z_{i,1:2} = \mathbf{z})$. Throughout the paper, it is assumed that there is a consistent estimator for $\pi_i(\mathbf{z})$.

Table 1- Separate Identification Approach - Modified First-stage ($D_{i,1:2} = (1, 1)$)

Potential Path	Groups	Δ^1	Δ^2
$\mathbf{1}\{D_{i,1:2}(1, 1) = (1, 1)\}$	$(AT_{1:2}), (C_1, AT_2), (AT_1, C_2), (C_{1:2})$		
$\mathbf{1}\{D_{i,1:2}(1, 0) = (1, 1)\}$	$(AT_{1:2}), (C_1, AT_2)$	$(AT_1, C_2), (C_{1:2})$	$(C_{1:2})$
$\mathbf{1}\{D_{i,1:2}(0, 1) = (1, 1)\}$	$(AT_{1:2}), (AT_1, C_2)$		
$\mathbf{1}\{D_{i,1:2}(0, 0) = (1, 1)\}$	$(AT_{1:2})$	(AT_1, C_2)	

Note: Table 1 shows what groups take the fully treated path ($D_{i,1:2} = (1, 1)$) under the four possible paths of assignment.

In the case of $T = 2$, there are four potential outcomes and four potential treatments that can be arranged in order to identify the mean potential outcomes. Below, I show that a “difference-in-differences” Wald-like causal estimand exploiting the variations in the path of assignment recovers the mean potential outcomes for path-compliers.

I begin by discussing the identification of the potential outcome associated to full exposure, which I define as $m_2(1, 1) := \mathbb{E}[Y_{i,2}(1, 1)|G_{i,1:2} = C_{1:2}]$. Table 1 shows how the fully treated path appears for different groups of compliers under different assignment paths. Always-takers for both periods ($AT_{1:2}$) will take the fully treated sequence under the four possible assignments, compliers for the two periods only take the fully treated path under assignment to treatment for the full path. The third column shows what is identified under the first-difference, referred to as Δ^1 , of the fully treated path. That is, the quantity that is identified by exploiting the variation in assignment in period 2, keeping assignment in period 1 fixed. The fourth column shows the “difference-in-differences”, that is, the difference in Δ^1 across different assignments in period 1, referred to as Δ^2 .

Definition: For a random variable R , the “difference-in-differences” of path assignments is defined as

$$\begin{aligned} \Delta^2 (\mathbb{E}[R_i \mathbf{1}\{Z_{i,1:2} = (z_{1:2})\}]) &= \mathbb{E}[R_i \mathbf{1}\{Z_{i,1:2} = (1, 1)\}] - \mathbb{E}[R_i \mathbf{1}\{Z_{i,1:2} = (1, 0)\}] \\ &- (\mathbb{E}[R_i \mathbf{1}\{Z_{i,1:2} = (0, 1)\}] - \mathbb{E}[R_i \mathbf{1}\{Z_{i,1:2} = (0, 0)\}]) \end{aligned}$$

Table 1 shows that exploiting variations in $Z_{i,2}$ keeping $Z_{i,1}$ fixed (Δ^1) cancel out noncompliers in period 2, identifying compliers for that period. Exploiting the variation in $Z_{i,1}$ within the group of compliers in period 2 (Δ^2) cancels out noncompliers from period 1, identifying compliers for periods 1 and 2. Therefore, it follows that a “difference-in-differences” two-stage estimand exploiting the variation in the path of instruments for the treatment path of full exposure identifies $m_2(1, 1)$.

Now, consider the potential outcome associated to full control, defined as $m_2(0, 0) := \mathbb{E}[Y_{i,2}(0, 0)|G_{i,1:2} = C_{1:2}]$. Table 2 shows how the fully untreated path appears for different groups of compliers under different assignment paths. Never-takers for both periods ($NT_{1:2}$) will take the fully untreated sequence under the four possible assignments, compliers for the two

Table 2 - Separate Identification Approach - Modified First-stage ($D_{i,1:2} = (0, 0)$)

Potential Path	Groups	Δ^1	Δ^2
$\mathbf{1}\{D_{i,1:2}(1, 1) = (0, 0)\}$	$(NT_{1:2})$		
$\mathbf{1}\{D_{i,1:2}(1, 0) = (0, 0)\}$	$(NT_{1:2}), (NT_1, C_2)$	(NT_1, C_2)	
$\mathbf{1}\{D_{i,1:2}(0, 1) = (0, 0)\}$	$(NT_{1:2}), (C_1, NT_2)$		$-(C_{1:2})$
$\mathbf{1}\{D_{i,1:2}(0, 0) = (0, 0)\}$	$(NT_{1:2}), (C_1, NT_2), (NT_1, C_2), (C_1, C_2)$	$(NT_1, C_2), (C_{1:2})$	

Note: Table 2 shows what groups take the fully untreated path ($D_{i,1:2} = (0, 0)$) under the four possible paths of assignment.

periods only take the fully untreated path under assignment to control for the full path.

Following the same reasoning for the identification of $m_2(1, 1)$, Table 2 shows that a “difference-in-differences” modified two-stage estimand identifies $m_2(0, 0)$. The proposition below shows that the generalized propensity score for the instrument $\pi_i(\mathbf{z})$ can be used to build Inverse Probability Weighting (IPW) estimands that identify mean potential outcomes for path-compliers:

Proposition 2. *Suppose that Assumptions PO, IA and SX hold. Then,*

$$\frac{\Delta^2 \left(\mathbb{E} \left[Y_{i,2} \frac{D_{i,1} D_{i,2} \mathbf{1}\{Z_{i,1:2}=(z_{1:2})\}}{\pi_i(z_{1:2})} \right] \right)}{\Delta^2 \left(\mathbb{E} \left[\frac{D_{i,1} D_{i,2} \mathbf{1}\{Z_{i,1:2}=(z_{1:2})\}}{\pi_i(z_{1:2})} \right] \right)} = m_2(1, 1)$$

and

$$\frac{\Delta^2 \left(\mathbb{E} \left[Y_{i,2} \frac{(1-D_{i,1})(1-D_{i,2}) \mathbf{1}\{Z_{i,1:2}=(z_{1:2})\}}{\pi_i(z_{1:2})} \right] \right)}{\Delta^2 \left(\mathbb{E} \left[\frac{(1-D_{i,1})(1-D_{i,2}) \mathbf{1}\{Z_{i,1:2}=(z_{1:2})\}}{\pi_i(z_{1:2})} \right] \right)} = m_2(0, 0)$$

See Section 3 of Appendix A for the proof.

Linear Toy Model (Continued). In the linear toy model introduced above, the dynamic causal effects associated to full exposure is $\beta_1 + \beta_2$. The lemma below shows that the separate identification successfully identifies $\beta_1 + \beta_2$.

Lemma 2. *Suppose that the linear toy model hold, and that for each period that assumptions from (IMBENS; ANGRIST, 1994) for the first-stage hold. Then,*

$$\frac{\Delta^2 \left(\mathbb{E} \left[Y_{i,2} \frac{D_{i,1} D_{i,2} \mathbf{1}\{Z_{i,1:2}=(z_{1:2})\}}{\pi_i(z_{1:2})} \right] \right)}{\Delta^2 \left(\mathbb{E} \left[\frac{D_{i,1} D_{i,2} \mathbf{1}\{Z_{i,1:2}=(z_{1:2})\}}{\pi_i(z_{1:2})} \right] \right)} - \frac{\Delta^2 \left(\mathbb{E} \left[Y_{i,2} \frac{(1-D_{i,1})(1-D_{i,2}) \mathbf{1}\{Z_{i,1:2}=(z_{1:2})\}}{\pi_i(z_{1:2})} \right] \right)}{\Delta^2 \left(\mathbb{E} \left[\frac{(1-D_{i,1})(1-D_{i,2}) \mathbf{1}\{Z_{i,1:2}=(z_{1:2})\}}{\pi_i(z_{1:2})} \right] \right)} = \beta_1 + \beta_2$$

In the next subsection, I discuss the intuition behind the identification approach.

2.2.3 The Role of the Time Structure in Identification

Dynamic IV settings and multiple treatment IV settings share some common features. Namely, there are multiple potential outcomes and hence multiple treatment effects, from which it follows that standard IV methods (e.g Wald estimands, Multivariate 2SLS) fail to recover relevant causal effects in the presence of heterogeneity. However, the identification approach is fundamentally different in these settings. One fundamental difference, is that the multiple differences Wald-like estimand identifies dynamic causal effects without any additional assumptions, which is not the case in multiple treatments settings, which must rely on additional restrictions, such as assuming irrelevant alternatives (KIRKEBOEN; LEUVEN; MOGSTAD, 2016), or comparable compliers (MOUNTJOY, 2022).

What makes identification in the dynamic setting fundamentally different from the multiple treatment setting? The answer lies in the sequential nature of choice and compliance.

In the multiple treatments settings, there are mutually exclusive and exhaustive treatments, and shifts in the instruments for a treatment affect compliers coming from different treatments. In (MOUNTJOY, 2022), for example, shifts in the distance of a community college make compliers that would not enroll in any superior education and compliers that would enroll in a 4-year college move to the community college option. A shift in an instrument affects multiple margins.

In the dynamic setting, treatment paths are mutually exclusive and exhaustive, but period-specific treatment choices are not, and the instrument is assigned at the period-specific treatment level, that is, there is a path of instruments, not instruments for the path. Thus, the time structure in the setting allows the identification of compliers for a sequence of periods under the multiple differences estimands.

Under no-anticipation, shifts in $Z_{i,t}$ do not affect treatment choices in periods 1 up to $t - 1$. Keeping the path of instruments $z_{1:t-1}$ fixed, variations in $Z_{i,t}$ identify the compliers in period t . Compliers in period t mimic their assignment in period t . Hence, variation in instruments from previous periods do not affect their decision in period t . Thus, keeping $z_{1:t-2}$ fixed, the joint variation of $Z_{i,t-1}$ and $Z_{i,t}$ in the difference-in-differences format identifies the share of individuals that comply in both $t-1$ and t . Taking a triple difference would identify compliers from $t-2$ to t , so on and so forth.

In the next Section, I formalize the potential outcomes model and the assumptions so that the proposed identification strategy can be generalized to be used in a panel with T periods.

2.3 General Framework and Identification

2.3.1 Assumptions

Consider a balanced panel in which N units are observed over T periods of time. For each unit $i \in [N]$ and time period $t \in [T]$, we observe a binary instrumental variable $Z_{i,t} \in \{0, 1\}$, a binary treatment status $D_{i,t} \in \{0, 1\}$ and a real-valued scalar outcome $Y_{i,t}$.

Without further restrictions, potential outcomes of unit i in period t are a function of the full panel of treatments and assignments, $Y_{i,t}(d_{1:N,1:T}, z_{1:N,1:T})$, and potential treatments are a function of the full panel of assignments, $D_{i,t}(z_{1:N,1:T})$. The next assumptions are invoked for identification.

Assumption 1 (No Spillovers and No Anticipation) For all $i \in [N]$, $t \in [N]$,

$$Y_{i,t} = \sum_{(d_{1:t}, z_{1:t}) \in \{0,1\}^{t \times t}} \mathbf{1}\{D_{i,1:t} = d_{1:t}, Z_{i,1:t} = z_{1:t}\} Y_{i,t}(d_{1:t}, z_{1:t})$$

$$D_{i,t} = \sum_{z_{1:t} \in \{0,1\}^t} \mathbf{1}\{Z_{i,1:t} = z_{1:t}\} D_{i,t}(z_{1:t})$$

Assumption 1 imposes that the potential outcome of unit i in period t depends only on the treatment and assignment paths of unit i until period t , ruling out the possibility of spillover of both treatments and assignments across units, as well as future treatments affecting past potential outcomes. It also imposes that the potential treatment from unit i in period t depends only on the assignment paths of unit i until period t . To put it shortly, Assumption 1 imposes both SUTVA and no-anticipation.

In the AGR setting, Assumption 1 implies that deforestation in a municipality in a certain year depends only on the path of law enforcement for the that municipality until that year, thus ruling out spillovers in deterrence effects and the possibility that expectations of future law enforcement affect current deforestation decisions. It also implies that future cloud coverage does not affect the present capacity of detection from the satellite monitoring deforestation.

One of the fundamental assumptions of IV settings is the exclusion restriction. Its dynamic version is stated below, alongside an additional exclusion restriction for the first stage.

Assumption 2 (Exclusion Restrictions): (i) $Y_{i,t}(d_{i,1:t}, z_{i,1:t}) = Y_{i,t}(d_{i,1:t})$ and (ii) $D_{i,t}(z_{1:t}) = D_{i,t}(z_t)$, for all $i \in [N]$ and $t \in [T]$.

The first part of Assumption 2 is the standard exclusion restriction for dynamic IV settings. It states that the path of assignments does not affect potential outcomes directly. Assignments only affect potential outcomes to the extent that they affect treatment choices. Mapping it to the application, it implies that the history of cloud coverage over a municipality affects illegal deforestation in that municipality only through its effect on law enforcement capabilities.

The second part of Assumption 2 imposes that potential treatments in period t depend

only on the instruments in period t . Although it might seem strong in a dynamic setting, it is justifiable in many experimental and observational settings. In AGR, the capacity of the satellite to detect illegal deforestation at a given period depends on the cloud coverage below the satellite at that period. Previous coverage affects the previous law enforcement capacity but does not affect the current capacity to detect deforestation.

When considering the validity of Assumption 2 in other contexts, one can think, for example, of a multi-period experiment with imperfect compliance in which past assignments do not provide any “encouragement” to present treatment due to institutional aspects of the experiment. A canonical case would be the leading example in (PHAM; CHEN, 2017), which is an experiment evaluating the effect of a sequence of advertisements (ads) on Facebook users’ purchasing behaviors over time, using the display of the advertisements user’s Facebook homepage as the instrument. The display of the advertisement on the homepage only affects the probability of viewing the ad while being displayed. Now, consider an example for observational settings. In a criminal justice setting with repeated offenders, sentences in a period depend only on the leniency of the judge assigned to trial in that period, and not on previous judges characteristics⁶.

The fundamental behavioral assumption in IV settings is the monotonicity assumption, which is provided below.

Assumption 3 (Monotonicity): For all $i \in [N]$ and $t \in [T]$, $D_{i,t}(1) \geq D_{i,t}(0)$.

Monotonicity states that at each period units assigned to treatment ($Z_{i,t} = 1$) are at least as likely to take treatment as units assigned to control ($Z_{i,t} = 0$). Under Assumption 3, units can be divide into three groups at each period of time defined by how units treatment choice in period t relates to treatment assignment in period t : Always-takers (AT_t), Never-takes (NT_t) and Compliers (C_t). Note that an individual can be part of a group in a given period, it does not need to be in the same group through the whole path of assignments. In the application, Assumption 3 implies that municipalities with intense cloud coverage at a certain period are less likely to have intense law enforcement in that period than the ones with little cloud coverage⁷.

Assumption 4 generalizes Assumption **SX** from Section 2, I am going to use it for the decomposition of the static Wald estimand under the general framework:

Assumption 4 (Strict Exogeneity): For all $i \in [N]$, $t \in [T]$,

$$(D_{i,1:T}(\cdot), Y_{i,1:T}(\cdot)) \perp Z_{i,1:T}$$

Assumption 4 is the independence of the instrument assumption that is usually invoked in dynamic IV settings (HAN, 2021; HAN, 2023) and states the panel of instrument is independent

⁶ One can think that the lenience of past judges do not affect current sentences, but past sentences might. This is not a violation of Assumption 2, and cases such as this are discussed in Section 5.2

⁷ For the application, it is more adequate to state Assumption 3 with the inverted inequality, $D_{i,t}(1) \leq D_{i,t}(0)$, which does not affect the validity of the results exposed in this paper.

from the panel of potential outcomes and potential treatments. For identification, I use a weaker assumption for the independence of the instrument, alongside an assumption regarding the estimation of the generalized propensity score.

As shown in Section 2, the generalized propensity score for assignment paths is a fundamental objection for identification. Define the generalized propensity score for an assignment path from period $t - p$ to p as $\pi_{i,t-p}(\mathbf{z}) = \mathbb{P}(Z_{i,t-p:t} = \mathbf{z} | Z_{i,1:t-p-1} = z_{1:t-p-1})$. Let $\mathcal{F}_{1:N,1:T}$ denote the filtration generated by $Z_{1:N,1:T}$. Assumption 5 states that the instrument is sequentially randomizes and that there is a consistent estimator for the generalized propensity score.

Assumption 5 (Independence of the Instrument and Estimator for the Propensity Score): For all $i \in [N]$, $t \in [T]$,

$$(D_{i,1:t}(\cdot), Y_{i,1:t}(\cdot)) \perp Z_{i,t} | Z_{i,1:t-1}$$

Moreover, there is an estimator $\hat{\pi}_{i,t-p:t}(\mathbf{z})$ which is \sqrt{N} -Consistent for $\pi_{i,t-p:t}(\mathbf{z})$.

Assumption 4' states that treatment assignment in period t is ‘‘as-good-as-random’’ conditional on the path of assignments from period 1 to period $t - 1$ ⁸. In our applied setting, it implies that cloud coverage in a certain period is independent from the panel of potential law enforcement and potential deforestation conditional on the path of cloud coverage until that period. Moreover, it is assumed that the generalized propensity score can be consistently estimated at the \sqrt{N} rate.

In Section 2 I defined the ‘‘difference-in-differences’’ as Δ^2 . Below, I define an expression for the general multiple difference in means across assignment paths, which I will refer to hereafter as the $p + 1$ difference in means (Δ^{p+1}).

Definition (Δ^{p+1}): Define the $p + 1$ -th difference across assignment paths from period $t-p$ onwards of conditional means of a random variable R as

$$\Delta^{p+1}(\mathbb{E}[R_{i,t} \mathbf{1}\{Z_{i,t-p:t} = z_{t-p:t}\}]) = \sum_{z_{t-p:t} \in \{0,1\}^{p+1}} (-1)^{\prod_{k \in \{t-p:t\}} (z_k + 1)} \mathbb{E}[R_{i,t} \mathbf{1}\{Z_{i,t-p:t} = z_{t-p:t}\}]$$

It is easy to see that when $p = 0$, the expression takes form as a difference in means for units with $Z_{i,t} = 1$ versus units with $Z_{i,t} = 0$, holding the assignment from period 1 to period $t - 1$ fixed. For the case $p = 1$, the expression takes the form of a difference-in-differences across assignment paths from period $t - 1$ to period t , keeping the path until period $t - 2$ fixed. For $p = 2$, it takes the form a triple difference. Thus, the $p + 1$ difference can be used to exploit all possible variations in assignment paths until period t .

Before defining the causal estimand, one more assumption is required for the proposed object to be well-defined:

Assumption 6 (Relevance): For all $i \in [N]$, $t \in [T]$ and $z_{t-p:t} \in \{0, 1\}^{p+1}$,

⁸ Modified versions of Assumption 4' are discussed in Section 5.

$$\Delta^{p+1} (\mathbb{E} [\mathbf{1} \{D_{i,t-p:t} = \mathbf{d}\} \mathbf{1} \{Z_{i,t-p:t} = z_{t-p:t}\} | Z_{i,1:t-p-1}]) \neq 0$$

Assumption 6 is a dynamic version of the standard relevance assumption from cross-sectional IV settings. In the case of $p = 0$, it amounts to $\mathbb{E} [D_{i,t} | Z_{i,t} = 1] - \mathbb{E} [D_{i,t} | Z_{i,t} = 0] \neq 0$, which is the standard cross-sectional relevance assumption.

Finally, it is assumed that there is common support for the generalized propensity score at all time periods

Assumption 7 (Common Support): There exists $C^L < C^U \in (0, 1)$ such that for all $i \in [N], t \in [T], C^L < \pi_{i,t-p}(\mathbf{z}) < C^U$ for all $\mathbf{z} \in \{0, 1\}^{p+1}$.

In the next subsection, I define the target parameters that can be identified under the aforementioned assumptions.

2.3.2 Target Parameters

We are interested in the identification of dynamic causal effects, which are parameters that compare potential outcomes for unit i at period t along different treatment paths. Define the dynamic causal effect of a treatment path versus an alternative treatment path in period t as $\tau_{i,t}(d_{i,1:t}, \tilde{d}_{i,1:t}) = Y_{i,t}(d_{i,1:t}) - Y_{i,t}(\tilde{d}_{i,1:t})$.

The number of potential outcomes grows exponentially with the periods of time. For the sake of tractability, it is common to focus on lag- p dynamic causal effect as defined in (BOJINOV; RAMBACHAN; SHEPHARD, 2021). For $0 \leq p \leq t$, and $\mathbf{d}, \tilde{\mathbf{d}} \in \{0, 1\}^{p+1}$, the lag- p dynamic causal effect is defined as

$$\tau_{i,t}(\mathbf{d}, \tilde{\mathbf{d}}; p) := Y_{i,t}(d_{i,1:t-p-1}^{obs}, \mathbf{d}) - Y_{i,t}(d_{i,1:t-p-1}^{obs}, \tilde{\mathbf{d}})$$

which can be interpreted as the causal effect of taking a treatment path from period $t - p$ to p versus an alternative path, keeping the path until period $t - p - 1$ fixed.

The lag- p dynamic causal effect can be used to construct generalized impulse response functions, which are weighted averages of the lag- p dynamic causal effects, defined as

$$\tau_{i,t}^\dagger(\mathbf{d}, \tilde{\mathbf{d}}; p) = \sum_{\mathbf{v} \in \{0,1\}^p} a_{\mathbf{v}} \left\{ Y_{i,t}(d_{1:t-p-1}^{obs}, (\mathbf{d}, \mathbf{v})) - Y_{i,t}(d_{1:t-p-1}^{obs}, (\tilde{\mathbf{d}}, \mathbf{v})) \right\}$$

where $a_{\mathbf{v}}$ are non-stochastic weights chosen by the researcher. By varying the values of the choice for the lag p , the generalized impulse response function can be used to construct event-study plots. In that sense, the generalized impulse response function parameter holds a similar interpretation to the event study parameters estimated in Difference-in-Differences settings (GOODMAN-BACON, 2021; CALLAWAY; SANT'ANNA, 2021) as the effect of the length of exposure to treatment.

There are different ways to aggregate such parameters in order to highlight different aspects of heterogeneity in dynamic causal effects. As in cross-sectional IV settings, the average causal effects are only identifiable for the group of compliers. I define two target parameters for the path of compliers. The time- t lag- p local dynamic causal effect and the total lag- p local dynamic causal effect are defined respectively as

$$\begin{aligned}\bar{\tau}_{C_{t-p:t},t}(\mathbf{d}, \tilde{\mathbf{d}}; p) &= \mathbb{E} \left[\tau_{i,t}(\mathbf{d}, \tilde{\mathbf{d}}; p) | G_{i,t-p:t} = C_{t-p:t} \right] \\ \bar{\tau}_{C_{t-p:t}}(\mathbf{d}, \tilde{\mathbf{d}}; p) &= \frac{1}{T-p} \sum_{t=p+1}^T \mathbb{E} \left[\tau_{i,t}(\mathbf{d}, \tilde{\mathbf{d}}; p) | G_{i,t-p:t} = C_{t-p:t} \right]\end{aligned}$$

These estimands extend to the generalized impulse response function by analogously defining $\bar{\tau}_{C_{t-p:t},t}^\dagger$ and $\bar{\tau}_{C_{t-p:t}}^\dagger$.

Next, I show how the dynamic IV can be used to identify dynamic causal effects for units that are contemporary compliers in the periods of interest. The identification of the generalized impulse response function follows directly.

2.3.3 Interpretation of the Static Wald Estimator

In this section, I generalize the decomposition presented in Section 2 for the case with a general T . First, note that the static specification stacked across time periods can be decomposed into a weighted average of period-specific Wald estimands. Using (KOLESÁR, 2013):

$$\beta = \sum_{t=1}^T \left[\frac{\lambda_t \mu_t \rho_t (1 - \rho_t)}{\sum_j \lambda_j \mu_j \rho_j (1 - \rho_j)} \right] \beta_t^{Wald}$$

where $\lambda_t = \mathbb{P}(T_t = 1)$ with T_t being an indicator of period t , $\mu_t = \mathbb{E}[D_{i,t} | Z_{i,t} = 1] - \mathbb{E}[D_{i,t} | Z_{i,t} = 0]$ is the period specific first-stage and $\rho_t = \mathbb{P}(Z_{i,t} = 1)$.

I analyze the period-specific Wald estimand, which is the building block in the static approach. The estimand for period t is

$$\beta_t^{Wald} = \frac{\mathbb{E}[Y_{i,t} | Z_{i,t} = 1] - \mathbb{E}[Y_{i,t} | Z_{i,t} = 0]}{\mathbb{E}[D_{i,t} | Z_{i,t} = 1] - \mathbb{E}[D_{i,t} | Z_{i,t} = 0]}$$

The next proposition shows that β_t^{Wald} generally does not hold a straightforward causal interpretation.

Proposition 3. *Suppose Assumptions 1-4 and 6 hold. Then,*

$$\begin{aligned}\beta_t^{Wald} &= \sum_{z \in \{0,1\}^{t-1}} \omega_z(1) \mathbb{E}[Y_{i,t}(D_{i,1:t-1}(z), 1) - Y_{i,t}(D_{i,1:t-1}(z), 0) | G_{i,t} = C_t] \\ &+ \sum_{j=1}^{t-1} \left\{ (\omega_{z_j}(1) - \omega_{z_j}(0)) \mathbb{E}[Y_{i,t}(D_{i,1:j-1}(\mathbf{0}), 1, D_{i,j+1:t}(\mathbf{0})) - Y_{i,t}(D_{i,1:j-1}(\mathbf{0}), 0, D_{i,j+1:t}(\mathbf{0})) | G_{i,j} = C_j] \frac{\mathbb{P}(G_{i,j} = C_j)}{\mathbb{P}(G_{i,t} = C_t)} \right\}\end{aligned}$$

where $\omega_z(1) = \mathbb{P}(Z_{i,1:t-1} = \mathbf{z} | Z_{i,t} = 1)$, $\omega_{z_j}(z_t) = \mathbb{P}(Z_{i,j} = 1 | Z_{i,t} = z_t)$ and $\mathbf{0}$ is a conformable vector of zeros.

Proposition 3 holds a similar interpretation to Proposition 1 in that sense that it shows that the static Wald estimand does not hold a straightforward causal interpretation in the presence of dynamics and serial correlation of the instrument. It additionally shows that the weighting scheme for dynamic causal effects becomes increasingly more complex as the number of time periods in the panel grows. That translates to the interpretation of the static estimand stacked across time periods, which is a weighted average of the effect of the contemporary treatment and dynamic treatments, where the dynamic treatments have potentially negative weights. In order for the estimator to hold a *weakly causal* interpretation, it must be the case that for all $t \in [T]$, the instrument $Z_{i,t}$ is positively correlated with its lagged values. For the proof of the decomposition, see Section 5 of Appendix A. In the next section, I define the estimand that identifies mean potential outcomes for path-compliers in the general case.

2.3.4 A General Expression for the Multiple Differences Estimand

Here, I focus on the identification of the local lag- p response functions, which are mean potential outcomes for compliers from period $t - p$ to period t , along the observed treatment path until period $t - p - 1$. I define them as

$$m_t(\mathbf{d}) = \mathbb{E} \left[Y_{i,t}(d_{1:t-p-1}^{obs}, \mathbf{d}) | G_{i,t-p:t} = C_{t-p:t}, \mathcal{F}_{i,1:t-p-1} \right]$$

where \mathbf{d} is the treatment path of interest from $t - p$ to t . Following the intuition from Proposition 2, the main identification result shows that mean potential outcomes associated to a treatment path from $t - p$ to t can be identified exploiting variations in the path of assignments from $t - p$ to p .

Theorem 1. *Under Assumptions 1-3, and 5-7,*

$$\frac{\Delta^{p+1} \left(\mathbb{E} \left[\frac{Y_{i,t} \mathbf{1}\{D_{i,t=p:t}=\mathbf{d}\} \mathbf{1}\{Z_{i,t-pt:t}=z_{t-p:t}\}}{\pi_{i,t-p:t}(z_{t-p:t})} | \mathcal{F}_{i,t-p-1} \right] \right)}{\Delta^{p+1} \left(\mathbb{E} \left[\frac{\mathbf{1}\{D_{i,t=p:t}=\mathbf{d}\} \mathbf{1}\{Z_{i,t-pt:t}=z_{t-p:t}\}}{\pi_{i,t-p:t}(z_{t-p:t})} | \mathcal{F}_{i,t-p-1} \right] \right)} = m_t(\mathbf{d})$$

Theorem 1 provides the main identification result, see Section 6 of Appendix A for the proof. Time- t local lag- p dynamic causal effects are identified by taking the difference between two local lag- p response functions identified separately.

2.4 Estimation and Design-Based Inference

In this section, I study the asymptotic properties of the estimator corresponding to the estimand introduced in Section 3.4. Instead of relying on superpopulation arguments, I take a purely design-based approach towards uncertainty.

Potential outcomes and potential treatments for a fixed size population are taken as given, and uncertainty arises from the stochastic nature of the path of instruments. Such an approach allows the applied researcher to remain agnostic about functional forms and spatial correlations, which is particularly desirable for the motivating application. The results are closely related to the hypothesis testing methods proposed by (BOJINOV; SHEPHARD, 2019) and (BOJINOV; RAMBACHAN; SHEPHARD, 2021).

The design-based analogues for the target parameter and the local lag- p response functions are, respectively,

$$\bar{\tau}_{C_{t-p:t},t}(\mathbf{d}, \tilde{\mathbf{d}}; p) = \sum_{i \in C_{t-p:t}} \tau_{i,t}(\mathbf{d}, \tilde{\mathbf{d}}; p) \frac{1}{|C_{t-p:t}|}$$

and

$$m_t(\mathbf{d}) = \frac{1}{|C_{t-p:t}|} \sum_{i \in C_{t-p:t}} Y_{i,t}(d_{1:t-p-1}^{obs}, \mathbf{d})$$

I first derive the asymptotic properties of the lag-0 dynamic causal effect over the randomization distribution and then proceed with the asymptotic properties of the estimator for general lag- p dynamic causal effects.

2.4.1 Correcting the Wald Estimator

Proposition 1 shows that the static "difference-in-means" Wald estimator fails to recover relevant causal effects in the presence of dynamics. Thus, any estimator for such specification will be biased. However, the Wald estimator can be easily modified to account for the path of instruments and hence recover local lag-0 dynamic causal effects.

I propose a simple two-stage Horvitz-Thompson (HT) estimator, which is built using an estimate of the generalized propensity score for the path of the instrument, $\hat{p}_{i,t-p:t}^l(\mathbf{z})$. The nonparametric estimator for $\tau_{i,t}(1, 0; 0)$ is

$$\hat{\tau}_{i,t}(1, 0; 0) = \frac{\hat{\tau}_{i,t}^{RF}(1, 0; 0)}{\hat{\tau}_{i,t}^{FS}(1, 0; 0)}$$

where

$$\begin{aligned} \hat{\tau}_{i,t}^{RF}(1, 0; 0) &= \frac{Z_{i,t}Y_{i,t}}{\hat{\pi}_{i,t}(1)} - \frac{(1 - Z_{i,t})Y_{i,t}}{\hat{\pi}_{i,t}(0)} \\ \hat{\tau}_{i,t}^{FS}(1, 0; 0) &= \frac{Z_{i,t}D_{i,t}}{\hat{\pi}_{i,t}(1)} - \frac{(1 - Z_{i,t})D_{i,t}}{\hat{\pi}_{i,t}(0)} \end{aligned}$$

Both the time-t lag-0 dynamic causal effect and the total lag-0 dynamic causal effect can be estimated by plugging in $\widehat{\tau}_{i,t}(1, 0)(0)$:

$$\begin{aligned}\widehat{\tau}_{C_t,t}(1, 0; 0) &= \frac{\frac{1}{N} \sum_{i=1}^N \widehat{\tau}_{i,t}^{RF}(1, 0; 0)}{\frac{1}{N} \sum_{i=1}^N \widehat{\tau}_{i,t}^{FS}(1, 0; 0)} \\ \widehat{\tau}_{C_t}(1, 0; 0) &= \frac{\frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N \widehat{\tau}_{i,t}^{RF}(1, 0; 0)}{\frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N \widehat{\tau}_{i,t}^{FS}(1, 0; 0)}\end{aligned}$$

Theorem 2 shows that the estimators are consistent over the randomization distribution, and asymptotically normal as the population size grows larger.

Theorem 2. *Suppose that potential outcomes are bounded. Under Assumptions 1-3 and 5-7,*

$$\frac{\sqrt{N} \left\{ \widehat{\tau}_{C_t,t}(1, 0; 0) - \bar{\tau}_{C_t,t}(1, 0; 0) \right\}}{\sigma_t(1, 0; 0)} \xrightarrow{d} \mathcal{N}(0, 1), \text{ as } N \rightarrow \infty$$

$$\text{and} \quad \frac{\sqrt{NT} \left\{ \widehat{\tau}_{C_t}(1, 0; 0) - \bar{\tau}_{C_t}(1, 0; 0) \right\}}{\sigma(1, 0; 0)} \xrightarrow{d} \mathcal{N}(0, 1), \text{ as } NT \rightarrow \infty$$

where $\sigma_t(1, 0; 0)$ and $\sigma(1, 0; 0)$ are defined in Section 7 of Appendix A.

The variances of $\widehat{\tau}_{C_t,t}(1, 0; 0)$ and $\bar{\tau}_{C_t}(1, 0; 0)$ are the appropriate averages of the variance of $\widehat{\tau}_{C_t,t}(1, 0; 0)$, which are generally not estimable as they depends on individual potential outcomes and potential treatments under both treatment and counterfactual. However, Lemma 8 in Appendix B shows that the variance of the reduced form and the first stage are bounded from above by a term that is estimable.

For hypothesis testing, I propose the estimation of a conservative (BLOOM, 1984) confidence interval, built with estimates of the upper bound of the variance of the reduced form, and the square of the estimate of the first stage.

The upper bound for the variance of the estimator of the reduced form $\widehat{\tau}_{i,t}^{RF}(1, 0; 0)$ is

$$\left(\gamma_{i,t}^{RF}(1, 0; 0) \right)^2 = \frac{Y_{i,t}(d_{1:t-1}^{obs}, D_{i,t}(1))^2}{\pi_{i,t}(1)} + \frac{Y_{i,t}(d_{1:t-1}^{obs}, D_{i,t}(0))^2}{\pi_{i,t}(0)}$$

The resulting $1 - \alpha$ confidence intervals for the time-t lag-0 dynamic causal effect and the total lag-0 dynamic causal effect are, respectively,

$$\begin{aligned}\widehat{\tau}_{C_t,t}(1, 0; 0) \pm z_{1-\alpha/2} \sqrt{\left\{ \frac{\frac{1}{N} \sum_{i=1}^N \left(\gamma_{i,t}^{RF}(1, 0; 0) \right)^2}{\left(\frac{1}{N} \sum_{i=1}^N \widehat{\tau}_{i,t}^{FS}(1, 0; 0) \right)^2} \right\}} \\ \widehat{\tau}_{C_t}(1, 0; 0) \pm z_{1-\alpha/2} \sqrt{\left\{ \frac{\frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N \left(\gamma_{i,t}^{RF}(1, 0; 0) \right)^2}{\left(\frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N \widehat{\tau}_{i,t}^{FS}(1, 0; 0) \right)^2} \right\}}\end{aligned}$$

(BLOOM, 1984) intervals exhibit good performance in terms of coverage rates for compliance rates greater than 10%. See (KANG; PECK; KEELE, 2018) for a thorough discussion about inference using IV in cross-sectional settings.

2.4.2 Nonparametric Estimator for Dynamic Causal Effects

For the general lag- p dynamic causal effect, I propose the separate estimation of lag- p dynamic response functions through a multiple-differences Horvitz-Thompson type of estimator. That is, $\widehat{\tau}_{i,t}(\mathbf{d}, \tilde{\mathbf{d}}; p) = \widehat{m}_{i,t}(\mathbf{d}) - \widehat{m}_{i,t}(\tilde{\mathbf{d}})$, where $\widehat{m}_{i,t}(\mathbf{d}) = \frac{\widehat{m}_{i,t}^{RF}(\mathbf{d})}{\widehat{m}_{i,t}^{FS}(\mathbf{d})}$ with

$$\begin{aligned}\widehat{m}_{i,t}^{RF}(\mathbf{d}) &= \Delta^{p+1} \left(\frac{\mathbf{1}\{Z_{i,t-p:t} = z_{t-p:t}\} \mathbf{1}\{D_{i,t-p:t} = \mathbf{d}\} Y_{i,t}}{\widehat{\pi}_{i,t-p:t}(z_{t-p:t})} \right) \\ \widehat{m}_{i,t}^{FS}(\mathbf{d}) &= \Delta^{p+1} \left(\frac{\mathbf{1}\{Z_{i,t-p:t} = z_{t-p:t}\} \mathbf{1}\{D_{i,t-p:t} = \mathbf{d}\}}{\widehat{\pi}_{i,t-p:t}(z_{t-p:t})} \right)\end{aligned}$$

Plugging the unit i , period t estimates for the lag- p local response function as shown in Section 5.1 leads to estimates of the time- t lag- p local response function and the total lag- p response function. The local lag- p response functions are estimated, respectively, by

$$\begin{aligned}\widehat{m}_t(\mathbf{d}) &= \frac{\frac{1}{N} \sum_{i=1}^N \widehat{m}_{i,t}^{RF}(\mathbf{d})}{\frac{1}{N} \sum_{i=1}^N \widehat{m}_{i,t}^{FS}(\mathbf{d})} \\ \widehat{m}(\mathbf{d}) &= \frac{\frac{1}{N(T-p)} \sum_{t=p+1}^T \sum_{i=1}^N \widehat{m}_{i,t}^{RF}(\mathbf{d})}{\frac{1}{N(T-p)} \sum_{t=p+1}^T \sum_{i=1}^N \widehat{m}_{i,t}^{FS}(\mathbf{d})}\end{aligned}$$

The appropriate lag- p dynamic causal effects are generated by the difference of the estimates for the response functions. Theorem 3 shows that the estimators are consistent over the randomization distribution, and asymptotically normal as the population size grows larger.

Theorem 3. *Suppose that potential outcomes are bounded. Under Assumptions 1-3 and 5-7,*

$$\frac{\sqrt{N} \left\{ \widehat{\tau}_{C_{t-p:t},t}(\mathbf{d}, \tilde{\mathbf{d}}; p) - \bar{\tau}_{C_{t-p:t},t}(\mathbf{d}, \tilde{\mathbf{d}}; p) \right\}}{\sigma_t(\mathbf{d}, \tilde{\mathbf{d}}; p)} \xrightarrow{d} \mathcal{N}(0, 1), \text{ as } N \rightarrow \infty$$

and

$$\frac{\sqrt{N(T-p)} \left\{ \widehat{\tau}_{C_{t-p:t}}(\mathbf{d}, \tilde{\mathbf{d}}; p) - \bar{\tau}_{C_{t-p:t}}(\mathbf{d}, \tilde{\mathbf{d}}; p) \right\}}{\sigma(\mathbf{d}, \tilde{\mathbf{d}}; p)} \xrightarrow{d} \mathcal{N}(0, 1), \text{ as } NT \rightarrow \infty$$

where $\sigma_t(\mathbf{d}, \tilde{\mathbf{d}}; p)$ and $\sigma(\mathbf{d}, \tilde{\mathbf{d}}; p)$ are defined in Section 8 of Appendix A.

For hypothesis testing, conservative (BLOOM, 1984) confidence intervals are constructed using the upper bound for the variance of the reduced forms for the lag- p response functions and estimates of the first stage for the lag- p response functions.

The upper bound for the variance of the estimator of the reduced form is

$$\begin{aligned} & (\gamma_{i,t}^{RF}(\mathbf{d}))^2 + (\gamma_{i,t}^{RF}(\tilde{\mathbf{d}}))^2 \\ = & \sum_{\mathbf{z} \in \{0,1\}^{p+1}} \frac{(Y_{i,t}(d_{1:t-p-1}^{obs}, \mathbf{d}) \mathbf{1}\{D_{i,t-p:t}(\mathbf{z}) = \mathbf{d}\})^2}{\pi_{i,t-p:t}(\mathbf{z})} + \sum_{\mathbf{z} \in \{0,1\}^{p+1}} \frac{(Y_{i,t}(d_{1:t-p-1}^{obs}, \tilde{\mathbf{d}}) \mathbf{1}\{D_{i,t-p:t}(\mathbf{z}) = \tilde{\mathbf{d}}\})^2}{\pi_{i,t-p:t}(\mathbf{z})} \end{aligned}$$

The resulting $1 - \alpha$ confidence intervals for the time- t lag- p dynamic causal effect and the total lag- p dynamic causal effect are, respectively,

$$\begin{aligned} \widehat{\tau}_{C_{t-p:t},t}(\mathbf{d}, \tilde{\mathbf{d}}; p) \pm z_{1-\alpha/2} & \sqrt{\left\{ \frac{\frac{1}{N} \sum_{i=1}^N (\gamma_{i,t}^{RF}(\mathbf{d}))^2 + (\gamma_{i,t}^{RF}(\tilde{\mathbf{d}}))^2}{\frac{1}{N} \sum_{i=1}^N (\widehat{m}_{i,t}^{FS}(\mathbf{d}))^2 + (\widehat{m}_{i,t}^{FS}(\tilde{\mathbf{d}}))^2} \right\}} \\ \widehat{\tau}_{C_{t-p:t}}(\mathbf{d}, \tilde{\mathbf{d}}; p) \pm z_{1-\alpha/2} & \sqrt{\left\{ \frac{\frac{1}{N(T-p)} \sum_{t=p+1}^T \sum_{i=1}^N (\gamma_{i,t}^{RF}(\mathbf{d}))^2 + (\gamma_{i,t}^{RF}(\tilde{\mathbf{d}}))^2}{\frac{1}{N(T-p)} \sum_{t=p+1}^T \sum_{i=1}^N (\widehat{m}_{i,t}^{FS}(\mathbf{d}))^2 + (\widehat{m}_{i,t}^{FS}(\tilde{\mathbf{d}}))^2} \right\}} \end{aligned}$$

2.5 Extensions

In this section, I briefly discuss some extensions and how to modify the estimand and estimators in order to correctly account for the modifications in the identifying assumptions.

2.5.1 Covariates

Most applications of IV methods with panel rely on conditional versions of the identifying assumptions, whether they are baseline or time-varying. For example, it is common in multi-period experiments to assume that assignment to treatment is randomized within strata of baseline covariates. In that case, the sequential randomization assumption is modified to take the following form. Let X denote a vector of baseline covariates (e.g. gender, race). Then, sequential randomization holds conditional on covariates if for all $i \in [N]$, $t \in [T]$, $(D_{i,1:t}(z_{1:t}), Y_{i,1:t}(D_{i,1:t}(z_{1:t}))) \perp Z_{i,t} | (Z_{i,1:t-1}, X_i)$.

When covariates are time-varying, they can be incorporated into the sequential randomization assumption in two ways. The first is to assume that sequential randomization holds conditional on contemporary covariates. In such cases, the randomization distribution can be expressed as $(D_{i,1:t}(z_{1:t}), Y_{i,1:t}(D_{i,1:t}(z_{1:t}))) \perp Z_{i,t} | (Z_{i,1:t-1}, X_{i,t})$, for all $i \in [N]$, $t \in [T]$.

If sequential randomization holds conditional on the path of covariates until period t , then the distribution can be expressed as $(D_{i,1:t}(z_{1:t}), Y_{i,1:t}(D_{i,1:t}(z_{1:t}))) \perp Z_{i,t} | (Z_{i,1:t-1}, X_{i,1:t})$, for all $i \in [N]$, $t \in [T]$.

In all of three cases, the estimand can be easily modified by adequately incorporating the covariates in the adapted propensity score $\pi_{t-p}(\mathbf{z})$. Such procedure can be interpreted as a nonparametric analogue of the so-called 'g-formula' (see (HERNAN; ROBINS, 2023) for a review).

2.5.2 Intermediary Realizations

Another common specification in dynamic treatment settings is to assume that sequential randomization holds conditional on the path of all realizations, that is, $Z_{i,t}$ is independent from the panel of potential outcomes and treatments conditional on the path of instruments, treatments, and outcomes until period $t - 1$:

$$(D_{i,1:t}(z_{1:t}), Y_{i,1:t}(D_{i,1:t}(z_{1:t}))) \perp Z_{i,t} | (Z_{i,1:t-1}, Y_{i,1:t-1}, D_{i,1:t-1})$$

for all $i \in [N]$, $t \in [T]$.

In such cases, the identification of the first stage of treatment paths becomes more challenging, as the identification of the joint distribution of treatments as proposed in Sections 3 and 4 is no longer valid if there are treatment effects from previous periods. However, the procedure can be modified, and the path of compliers can be identified using the general product rule.

Consider the leading case of $T = 2$. Note that there is no history affecting the first stage in period 1. Hence, potential treatments are readily identified by $\frac{1}{N} \sum_{i=1}^N D_{i,t}(z_1)$ and can readily be estimated using the Horvitz-Thompson estimator $\frac{1}{N} \sum_{i=1}^N \frac{D_{i,1} \mathbf{1}\{Z_{i,1}=z_1\}}{\pi_{i,1}(z_1)}$. Potential treatments in period 2 are identified conditional on potential treatments from period 1:

$$\begin{aligned} \mathbb{P}(D_{i,2}(z_2) = 1 | D_{i,1}(z_1) = 1) &= \int_{y_1} \mathbb{P}(D_{i,2}(z_2) = 1 | D_{i,1}(z_1) = 1, Y_{i,1} = y_1) dF_{Y_{i,1}}(y_1) \\ &= \int_{y_1} \mathbb{P}(D_{i,2} = 1 | D_{i,1} = 1, Z_{i,1:2} = (z_1, z_2), Y_{i,1} = y_1) dF_{Y_{i,1}}(y_1) \\ &= \int_{y_1} \mathbb{E}[D_{i,2} | D_{i,1} = 1, Z_{i,1:2} = (z_1, z_2), Y_{i,1} = y_1] dF_{Y_{i,1}}(y_1) \end{aligned}$$

Hence, the potential treatment path $D_{i,1}(z_1) = 1, D_{i,2}(z_2) = 1$ is identified using a design-based analogue of the general product rule:

$$\mathbb{P}(D_{i,1}(z_1) = 1, D_{i,2}(z_2) = 1) = \mathbb{P}(D_{i,2}(z_2) = 1 | D_{i,1}(z_1) = 1) \mathbb{P}(D_{i,1}(z_1) = 1)$$

Once, these quantity is identified, identification of compliers for the path follows the same "difference-in-differences" format as presented in Section 2.

When it comes to the reduced form, note that

$$\begin{aligned}
& \mathbb{E} [D_{i,2}Y_{i,2}|D_{i,1} = 1, Z_{i,1:2} = (z_1, z_2), Y_{i,1} = y_1] \\
&= \mathbb{E} [Y_{i,2}(1, 1)|D_{i,1} = 1, D_{i,2} = 1, Z_{i,1:2} = (z_1, z_2), Y_{i,1} = y_1] \mathbb{P} (D_{i,2} = 1|D_{i,1} = 1, Z_{i,1:2} = (z_1, z_2), Y_{i,1} = y_1) \\
&= \mathbb{E} [Y_{i,2}(1, 1)|D_{i,1}(z_1) = 1, D_{i,2}(z_2) = 1, Y_{i,1} = y_1] \mathbb{P} (D_{i,2}(z_2) = 1|D_{i,1}(z_1) = 1, Y_{i,1} = y_1)
\end{aligned}$$

And, therefore, the reduced form of the potential outcome associated to full exposure for individuals with assignment path (z_1, z_2) is identified by

$$\begin{aligned}
& \int_{y_1} \mathbb{E} [D_{i,2}Y_{i,2}|D_{i,1} = 1, Z_{i,1:2} = (z_1, z_2), Y_{i,1} = y_1] dF_{Y_{i,1}}(y_1) \mathbb{E} [D_{i,1}|Z_{i,1} = z_1] \\
&= \mathbb{E} [Y_{i,2}(1, 1)|D_{i,1}(z_1) = 1, D_{i,2}(z_2) = 1] \mathbb{P} (D_{i,1}(z_1) = 1, D_{i,2}(z_2) = 1)
\end{aligned}$$

And one can proceed with Wald-like modified "difference-in-differences" to identify the local lag- p response function.

Hence, the Horvitz-Thompson estimator can be easily modified to account for the alternative sequential randomization assumption. Consider, once again, the estimation of the local lag- p response function associated to full exposure, $Y_{i,2}(1, 1)$. This time, we work with two different propensity scores instead of the generalized propensity score. Let $\pi_{i,1}(z_1) = \mathbb{P} (Z_{i,1} = z_1)$ and $\pi_{i,2}(z_2) = \mathbb{P} (Z_{i,2} = z_2 | Y_{i,1} = y_1, D_{i,1}(z_1) = 1)$. Then, the first stage is consistently estimated by

$$\hat{m}_t^{FS}(1, 1) = \frac{1}{N} \sum_{i=1}^N \Delta^2 \left(\frac{D_{i,2} \mathbf{1} \{Z_{i,2} = z_2\}}{\pi_{i,2}(z_2)} \frac{D_{i,1} \mathbf{1} \{Z_{i,1} = z_1\}}{\pi_{i,1}(z_1)} \right)$$

and the reduced form is estimated by

$$\hat{m}_t^{RF}(1, 1) = \frac{1}{N} \sum_{i=1}^N \Delta^2 \left(\frac{D_{i,1} D_{i,2} Y_{i,2} \mathbf{1} \{Z_{i,1:2} = (z_1, z_2)\}}{\pi_{i,2}(z_2) \pi_{i,1}(z_1)} \right)$$

Therefore, estimation and inference can be easily adapted to different versions of the sequential randomization assumption.

2.6 Monte Carlo Simulations

In this section, I show the desirable finite-sample properties of the proposed nonparametric estimator for the local dynamic causal effects.

I consider a balanced panel setting with $N = 1000$ and $T = 2$. The simulation focuses on the lag- p dynamic causal effects with $p = t$, that is, the dynamic causal effects associated to the whole treatment path in the setting. I simulate a continuous covariate that has a Gaussian

Table 3 - Simulation results for the Lag-0 dynamic causal effect

Time-t effect	t=1		t=2		Total	
	HT	2SLS	HT	2SLS	HT	2SLS
Av. Bias	-0.003	0.003	0.027	0.368	0.015	0.156
Med. Bias	-0.002	0.003	0.034	0.392	0.019	0.163
RMSE	0.097	0.081	0.119	0.484	0.237	0.169
Cover	0.944	0.948	0.952	0.612	0.955	0.784
CIL	0.356	0.223	0.508	0.285	0.441	0.258

Note: Simulations based on 10.000 Monte Carlo experiments with sample size $N = 1.000$ and $T = 2$. CIs for the 2SLS estimator were built using the standard estimated variance using the Delta Method.

distribution and a binary treatment instrument that is sequentially randomized following a Bernoulli distribution:

$$\mathbb{P}(Z_{i,t} = z_t | Z_{i,1:t-1} = z_{1:t-1}, Y_{i,1:t-1} = y_{1:t-1}, X_i = x) \propto \prod_{i \in [N]} p_{i,t}^{z_{i,t}} (1 - p_{i,t})^{1 - z_{i,t}}$$

I set $p_{i,t} = p_i = 0.6$. For the choice model. Outcomes in period t are specified to have the following linear working model:

$$Y_{i,t} = \delta'_t Y_{i,1:t-1} + \beta'_{1:t} D_{i,1:t} + \theta X_i + U_{i,t}(D_{i,1:t})$$

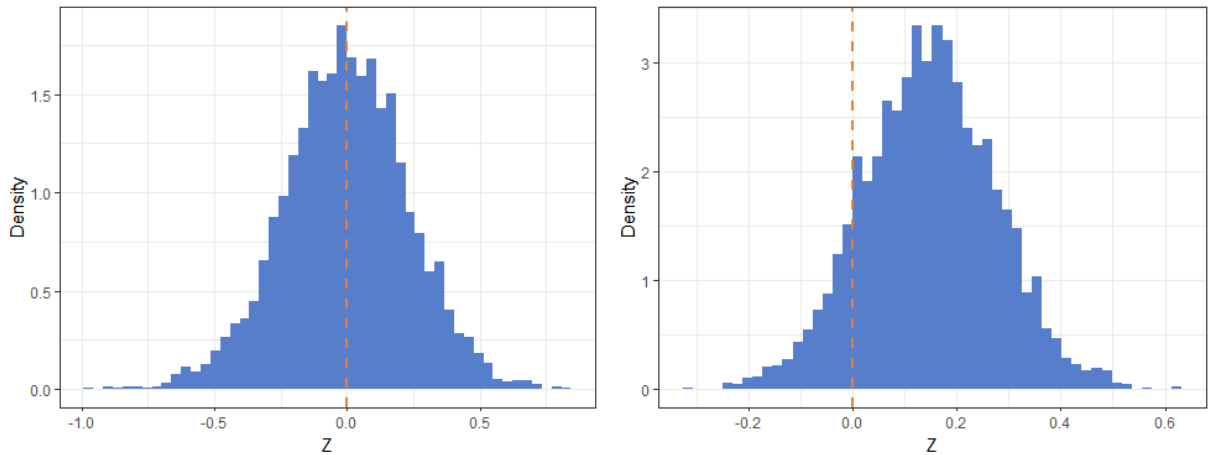
I set $\beta_t = 1$ and $\beta_{1:t-1} = 0.5$.

When analyzing the competing estimators, I study their finite-sample properties in terms of the average bias (Av. Bias), median bias (Med. Bias), root mean-squared error (RMSE), coverage of the Confidence Interval (Cover) and the Confidence Interval length (CIL).

In Table 3, I compare the performance of the proposed Horvitz-Thompson estimator (HT) with the performance of the static 2SLS specification (2SLS), taking the local lag-0 dynamic causal effect as the target parameter.

The first two columns show the results for the first period. When $t = 1$, dynamics play no role in the model. Hence, both the nonparametric estimator and the static 2SLS show little to none Monte Carlo bias. Moreover, the coverage is close to the desired 95%, with the 2SLS estimator showing a tighter Confidence Interval on average. When it comes to the second period, the Horvitz-Thompson estimator remains consistent in the Monte Carlo exercise, as shown by the third column. The static 2SLS estimator is severely biased, with coverage far from the desired 95%. The last two columns stack the lag-0 estimates across the two time periods. Thus, the results can be interpreted as a weighted average of the time- t results, which explains why the performance of the 2SLS estimator is better than in the second period alone. As the number of periods grows larger, however, one should expect the 2SLS estimator to perform increasingly worse with the number of time periods.

Figure 1 - Simulated Randomization Distributions



Note: The figure on the left shows the simulated randomization distribution of the Horvitz-Thompson estimator for the total local lag-0 dynamic causal effect. The figure on the right shows the simulated randomization distribution of the static 2SLS estimator for the total local lag-0 dynamic causal effect. Simulations based on 10,000 Monte Carlo experiments with sample size $N = 1,000$ and $T = 2$. CIs for the 2SLS estimator were built using the standard estimated variance using the Delta Method.

In Table 4, I present the simulation results for different time- t local lag-1 dynamic causal effects. The lag-1 effects can only be estimated for period 2. I present results for the “difference-in-differences” modified Horvitz-Thompson estimator (HT) and a multivariate 2SLS estimator (MV2SLS) such as the ones specified in Section 2. I consider the performance of the estimators with respect to effects of full exposure in the first two columns, exposure in the second period in the third and fourth columns, and exposure in the first period in the last two columns.

Unsurprisingly, the multivariate 2SLS specification yields substantially biased estimates for the three different target parameters. Coverage is closer to the desired 95% than in the simulations for the static 2SLS. However, it is never greater than 81.5%.

Figure 1 plots the randomization distribution for total lag-0 dynamic effect estimated by the HT estimator (left) and the 2SLS estimator (right). Both estimators converge to a normal distribution under the Monte Carlo exercise. However, the distribution of the 2SLS estimator is clearly not centered around the true parameter.

The proposed nonparametric estimator exhibits great Monte Carlo performance. The average bias, median bias and root mean-squared error for the causal effects are small, and the coverage of the conservative confidence interval is close to the desired 95% coverage. Confidence interval lengths are fairly stable across the considered treatment paths.

Overall, the Monte Carlo Simulations assert the desirable finite-sample performance of the proposed estimators for dynamic causal effects over the randomization, while bringing evidence of the inadequacy of the standard 2SLS methods in the presence of time-varying heterogeneity.

Table 4 - Simulation results for the Lag-1 dynamic causal effects

Lag-1 effect	((1,1),(0,0))		((0,1),(0,0))		((1,0),(0,0))	
Estimator	HT	MV2SLS	HT	MV2SLS	HT	MV2SLS
Av. Bias	-0.0008	0.1968	-0.0039	0.1306	-0.0051	0.1641
Med. Bias	-0.0012	0.1970	0.0002	0.1312	-0.0046	0.1646
RMSE	0.0772	0.1975	0.0762	0.1404	0.0845	0.1937
Cover	0.938	0.813	0.956	0.798	0.928	0.815
CIL	0.8721	0.3019	0.8023	0.3145	0.8674	0.3222

Note: Simulations based on 10.000 Monte Carlo experiments with sample size $N = 1.000$ and $T = 2$. CIs for the 2SLS estimator were built using the standard estimated variance using the Delta Method.

2.7 Application: Law Enforcement and Deforestation

2.7.1 AGR (2023)

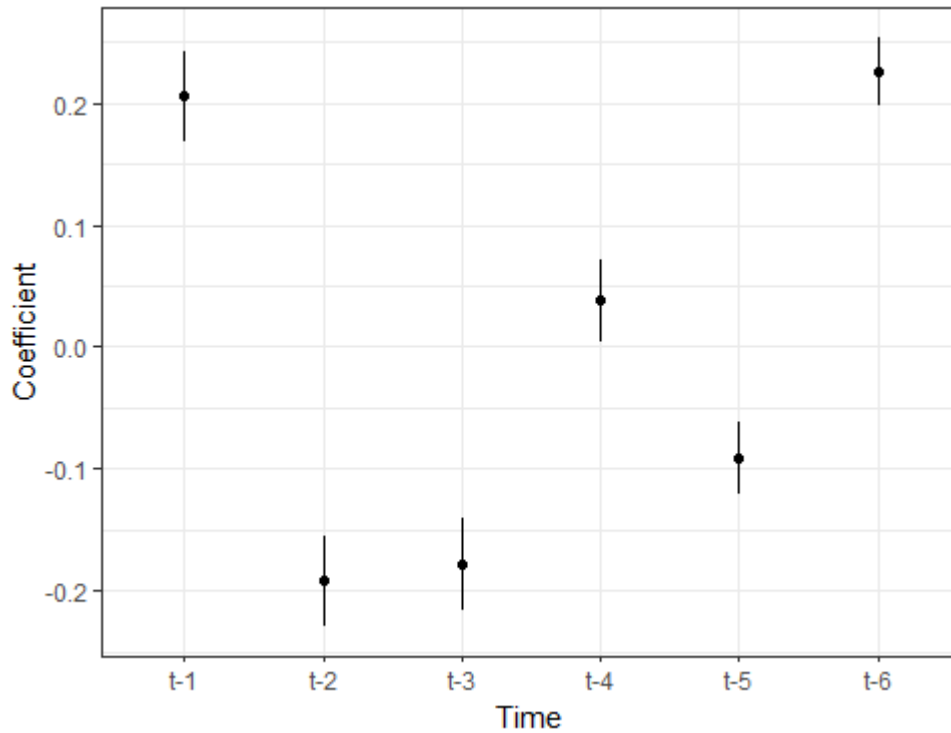
In this section, I revisit the work of (ASSUNÇÃO; GANDOUR; ROCHA, 2023) (hereafter referred to as AGR, 2023). The authors investigate the effects of a satellite law enforcement program, the DETER program, on deforestation in the Brazilian Amazon.

In 2004, the Brazilian government launched the Action Plan for the Prevention and Control of Deforestation in the Legal Amazon (PPCDAm). In the early 2000s, Brazil was the country that cleared the most tropical forest area in both absolute and relative terms (HANSEN et al., 2008). The program was a response to the growing concerns regarding the dangers of deforestation of tropical forests.

One of the main innovations of the PPCDAm was the development of a satellite-based system that regularly collected and processed georeferenced imagery on Amazon land cover to detect forest loss, the Real-Time System for Detection of Deforestation (DETER). Deforestation alerts that came from DETER imagery became the cornerstone for law enforcement. Officers visited alert sites and applied sanctions if evidence of illegal deforestation was found. Law enforcement officers would typically issue a fine for every environmental infraction they detected, but other forms of penalty, such as the seizure and destruction of products and equipment, are also contemplated by law.

AGR (2023) estimate the effects of law enforcement, as a function of the number of fines issued by the DETER satellite, on deforestation using panel data from 2006 to 2016. To address the issue of reverse causality between deforestation and law enforcement, the authors use an IV approach, leveraging variation in cloud coverage as exogenous variation on law enforcement. The intuition behind the strategy is that cloud coverage blocks visibility in satellite imagery and thereby limits DETER's capacity to detect changes in land cover patterns and therefore law enforcement.

Figure 2 - Serial Correlation of Cloud Coverage



Note: Figure 2 plots the results of the regression of cloud coverage in period t vs cloud coverage in previous periods and municipal fixed effects.

2.7.2 Serial Correlation of the Instrument

The 2SLS regressions from AGR (2023) can be mapped to the static Wald estimand as presented in Section 2.1, with the introduction of covariates. From proposition 1, it follows that the estimates are consistent as long as cloud coverage in a given period is independent from cloud coverage at previous periods.

Figure 2 plots the estimates of the coefficients and the confidence intervals associated to serial correlation of the instrument in a regression where cloud coverage in period t is the dependent variable, and lags of cloud coverage and municipal fixed effects are the regressors. The coefficients are statistically significant up to the 6-th lag, with exception of the 4-th lag. The correlation with the first and sixth lags are positive, whereas the remaining significant ones are negative. Looking at the magnitude of the coefficients, the serial correlation plot suggests a pattern of mean-reversion for cloud coverage, which, as shown in Proposition 3, affects the causal interpretation of the static estimates.

2.7.3 Dynamic Effects of Law Enforcement

To address the issue of serial correlation of the instrument, I use the multiple-differences Horvitz-Thompson estimator presented in Section 4.2 to estimate the impulse response functions in order to analyze the dynamics effects of law enforcement. Since the procedure is suited for binary instruments and binary treatments, I binarize the treatment creating an indicator of

Table 5 - First-Stage Results

Cloud Coverage:	t	$t - 1$	$t - 2$	$t - 3$
Law Enforcement t	-0.497	-0.598	-0.237	-0.009
	(0.125)	(0.415)	(0.186)	(0.109)
Observations	5210	4689	4168	3647
Municipalities	521	521	521	521

Note: The dependent variable is the indicator of intense law enforcement. The set of time-varying covariates contains precipitation and temperature (weather), PRODES cloud coverage and other nonobservable areas (satellite visibility), and agricultural commodity prices.

“intense” law enforcement for municipalities in which the number of fines in a given year is greater than that year’s average, and a binary instrument for “intense” cloud coverage under the same reasoning. The binarized results can be interpreted under the light of (SCHULER; LEE; HUBBARD, 2024)⁹.

The key assumption for the valid implementation of the estimator is that the first-stage is static. In Table 5 I provide suggestive evidence that the assumption is valid.¹⁰

Table 5 shows the effects of the contemporary and legged instruments in an OLS approach for the G-estimation (ROBINS; MARK; NEWHEY, 1992) of the first-stage. The coefficient associated to the contemporary instrument, presented in Column 1, is statistically significant, while none of the lagged instruments is statistically different from zero.

The generalized propensity score $\pi_{i,t-p}(\mathbf{z})$ was estimated following the factorization shown in Appendix B.1 from (BOJINOV; RAMBACHAN; SHEPHARD, 2021):

$$\pi_{i,t-p}(\mathbf{z}) = \mathbb{P}(Z_{i,t-p} = z_{t-p} | Z_{i,1:t-p-1}, X_{i,1:t-p}) \\ \times \prod_{s=1}^p \mathbb{P}(Z_{i,t-p+s} = z_{t-p+s} | Z_{i,1:t-p-1}, Z_{i,t-p:t-p+s+1}, X_{i,1:t-p+s})$$

where each of the period-specific propensity scores was estimated following a Probit model. The estimator includes the same covariates used in the main specification of AGR (2023), which are covariates controlling for the weather (rain and temperature), other complementary measures of cloud coverage and agricultural commodity prices.

Table 6 presents the results. The dependent variable is the ratio between a municipality’s deforested area in a year and its total area. The mean of the ratio for the whole panel is about 0.007. The first column of Table 1 present the results for a binarized version of the pooled static 2SLS estimate from AGR (2023). The estimate should be interpreted as a “naive” estimate for the lag-0 dynamic causal effect. The second column shows the estimated lag-0 dynamic causal effect using the HT estimator. The estimated causal effect is almost half of the one from AGR

⁹ As a robustness check, I use the estimator under different binarization procedures. See Appendix D for the results.

¹⁰ Table 1 from AGR (2023) provides a similar result, but for the case of the continuous instrument and treatment.

(2023). The difference in the results, however, cannot be interpreted solely as the correction for the bias term since the authors work with a linear specification for first stage and reduced form while my procedure is fully nonparametric. Nevertheless, the results in column 2 corroborate the initial finding that law enforcement successfully curbs deforestation, even when accounting for serial correlation in the instrument.

Table 6 - Dynamic Effects of Law Enforcement: Impulse Response Functions

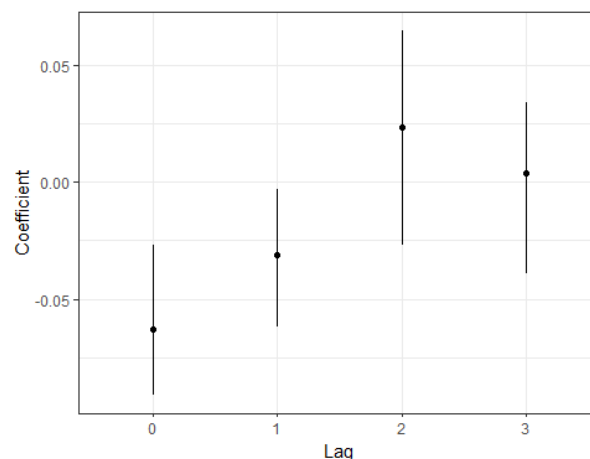
	AGR(2023)	lag-p			
		0	1	2	3
Point estimate	-0.107	-0.0627	-0.0311	0.0234	0.0036
95% CI	(-0.137, -0.076)	(-0.091, -0.003)	(-0.062, -0.003)	(-0.027, 0.065)	(-0.039, 0.034)
Baseline	0.0071	0.0071	0.0071	0.0071	0.0071
Observations	5210	5210	4689	4168	3647
Municipalities	521	521	521	521	521

Note: The dependent variable is the ratio between deforested area in a year and the municipality area. The set of control variables contains precipitation and temperature (weather), PRODES cloud coverage and other nonobservable areas (satellite visibility), and agricultural commodity prices.

Column 3 presents the estimate for the lag-1 impulse response function. The estimated coefficient shows that lagged law enforcement also successfully curbs deforestation. Thus, it can be interpreted as evidence of dynamic effects of law enforcement. Columns 4 and 5 present the results of the impulse response functions for the lag-2 and lag-3 dynamic causal effects. The estimates, however, are not statistically significant.

How are the AGR (2023) estimates to be interpreted under Proposition 2? Figure 2 shows what seems to be a mean-reversion pattern for cloud coverage. Correlation is positive for some lags and negative for others. Thus, the estimated effects are outside the convex hull of dynamic causal effects. Therefore, even if the lag- p estimates are not significant for $p \geq 2$, the estimates from AGR (2023) do not hold a straightforward causal interpretation.

Figure 3 - Event-Study: Dynamic Effects of Law Enforcement



Note: Figure 3 plots the results displayed in the columns 2-5 from Table 6.

A potential violation of the maintained assumptions would be law enforcement in the past affecting directly present law enforcement. However, since law enforcement depends on the alerts that are issued by the DETER satellite, and those alerts are issued based only on contemporary satellite images, there does not seem to be a path for past law enforcement to affect present law enforcement directly, or at least such spillovers across periods are second-order concerns.

The results in Table 6 suggest that there are dynamic effects of enforcement in the short-run but they don't seem to carry over to the long-run. The mechanisms driving the results are not clear, however. One could interpret the evidence of dynamic effects of law enforcement as evidence of persistence of deterrence effects through time. Law enforcement curbs deforestation by causing potential offenders to update their beliefs about the probability of getting caught and thus their expected costs from engaging in the illegal activity. However, the effect fades out, which could indicate that law enforcement does not provide a permanent shock on the perception of potential offenders.

On the other hand, law enforcement could have lead to to the loss of capital goods and machinery used in deforestation, which may have reduced potential offenders' ability to commit future offenses. If that's the case, the observed pattern of the dynamic effects of law enforcement could be suggesting that law enforcement affects the means to engage in illegal deforestation, but with time potential offenders reequip themselves with the necessary machinery and go back to their illegal deforestation activities. While the pattern of dynamic causal effects documented in Table 5 is consistent with loss of machinery driving the results (recovering stock of capital takes time), the empirical strategy does not allow us to disentangle the underlying mechanisms driving the results and that should be an effort for future research.

Still, the results presented contribute to the literature on the enforcement of environmental regulation in developing countries. While (GREENSTONE; HANNA, 2014) and AGR (2023) study such phenomenon, to the extent of my knowledge this is the first paper analyzing the dynamics effects of environmental regulation and disentangling the dynamic effects of law enforcement between a contemporary and a dynamic component. Dynamics are particularly important in the context of environmental regulation, as understanding how the effects of regulation vary with time is fundamental for improving the efforts against illegal deforestation. Yet, further inquiries are necessary in order to understand how to improve the efforts against deforestation in the Brazilian Amazon, since the mechanisms driving the effects are not clear.

2.8 Conclusion

This paper develops a finite-population framework for the identification of dynamic causal effects using instrumental variables with panel data.

First, I show that the usual approach to IV settings with panel data, which I called the

static Wald estimator, does not generally estimate well defined causal parameters unless the instrument is fully randomized at each period of time and discuss the conditions for it to hold a *weakly causal* interpretation.

Second, I present a modification of the Wald estimator that identifies contemporary treatment effects under sequential randomization, and provide a general framework for the identification of dynamic causal effects in which mean potential outcomes are identified separately. I introduce a nonparametric estimator, and derive its finite population asymptotic distribution.

I then show how covariates and intermediary outcomes and treatments can be introduced in the framework by accounting for them in the randomization distribution.

Monte Carlo simulation studies assert the desirable finite-sample property of the estimators for both the local and the marginal dynamic causal effects.

Finally, I use the estimators to analyze the dynamic effects of law enforcement on illegal deforestation in the Brazilian Amazon. Results suggest that there is substantial time-varying heterogeneity in the effects of law-enforcement and that while the effects of law enforcement on illegal deforestation persist for some time, they are not permanent.

3 Breakdown Analysis for Instrumental Variables with Binary Outcomes

This paper studies the partial identification of treatment effects in Instrumental Variables (IV) settings with binary outcomes under violations of independence. I derive the identified sets for the treatment parameters of interest in the setting, as well as breakdown values for conclusions regarding the true treatment effects. I derive \sqrt{N} -consistent nonparametric estimators for the bounds of treatment effects and for breakdown values. These results can be used to assess the robustness of empirical conclusions obtained under the assumption that the instrument is independent from potential quantities, which is a pervasive concern in studies that use IV methods with observational data. In the empirical application, I show that the conclusions regarding the effects of family size on female unemployment using same-sex sibling as the instrument are highly sensitive to violations of independence.

3.1 Introduction

Instrumental variables (IV) techniques are among the most widely used empirical tools in social sciences. In the canonical IV setting, the causal effect of a binary treatment is identified by exploiting variations in a binary instrument in the form of the (WALD, 1940) estimand. Point identification is achieved if the instrument satisfies a set of assumptions. For instance, the instrumental variable must be independent from potential treatments and potential outcomes.

Although in certain cases the independence assumption is readily justified (experimental studies), it is often unverifiable and must be defended by appealing to context specific knowledge, specially in observational studies. In this paper I study what can be learned about treatment effects in IV settings under weaker versions of the instrument independence assumption.

I focus in the case where the outcome is binary. I derive bounds for the first-stage and reduced form parameters, as well as bounds for the Local Average Treatment Effect (LATE) under a bounded dependence assumption called *c-dependence* (MASTEN; POIRIER, 2018), which bounds the distance between the probability of receiving the instrument given observed covariates and unobserved potential quantities and the probability of being treated given just the observed covariates.

The first-stage parameter, the share of compliers, is partially identified as function of the difference between the probability of assignment given covariates and the probability of assignment given covariates and potential treatments. The reduced form parameter, the intention-to-treat effect, is partially identified as function of the difference between the probability of

assignment given covariates and the probability of assignment given covariates and potential outcomes. The LATE is partially identified as a function of both sensitivity parameters.

I identify breakdown values for the first-stage and reduced form, as well as the breakdown frontier for the LATE. Breakdown values are the violations of the independence assumptions under which a particular conclusion no longer holds. For instance, one could be interested to learn under which violations of independence the conclusion that the treatment effect has a particular sign holds. If a researcher is concerned about the external validity of the study, the breakdown analysis of the first-stage is useful to understand under which violations of independence the share of compliers is above a certain share of the study population.

I propose nonparametric estimators for the bounds of causal effects and breakdown values, and derive their asymptotic properties using convergence results for *Hadamard directional differentiable* functions (FANG; SANTOS, 2018).

The identified sets for the LATE are not sharp. I derive sharp bounds for the LATE under a joint *c-dependence* assumption for potential outcomes and potential treatments. The bounds of the set can be used to derive the breakdown point for conclusions regarding the LATE.

Monte Carlo simulations show the desirable finite sample properties of the proposed estimators.

For the empirical application, I revisit (ANGRIST; EVANS, 1998), which studies the effects of family size on female employment using same-sex siblings as the instrument, and estimates the identified sets for the share of compliers, the ITT and the LATE under different relaxations of independence.

Related Literature: This paper relates broadly to three strands of the causal inference literature. First, to the literature on partial identification and sensitivity analysis in IV settings. While most results on the literature focus on partial identification under violations of the exclusion restriction (CONLEY; HANSEN; ROSSI, 2012; WANG et al., 2018; MASTEN; POIRIER, 2021; CINELLI; HAZLETT, 2025), this paper focuses solely on violations of independence. In that sense, it is similar to (KLINE; MASTEN, 2025) and (RAMBACHAN; ROTH, 2025), but differs from the former by allowing two-sided noncompliance and from the latter by choosing a different sensitivity parameter.

Second, this paper relates to the literature on the identification of breakdown values, introduced by (HOROWITZ; MANSKI, 1995). My approach to inference follows closely the one introduced in (MASTEN; POIRIER, 2020). While most of the work in this literature focuses on missing data settings (KLINE; SANTOS, 2013) and selection on observables (MASTEN; POIRIER, 2020), this is the first paper studying inference for breakdown values in settings with non-compliance.

Finally, this paper is related to the literature on IV settings with binary outcomes, which dates back to the seminal work of (HECKMAN, 1978). While most prominent work on this

literature focuses on the identification of the average structural functions (VYTLACIL; YILDIZ, 2007; SHAIKH; VYTLACIL, 2011) or partial identification of Average Treatment Effects (MACHADO; SHAIKH; VYTLACIL, 2019), this paper is more closely related to (CHESHER; ROSEN, 2013) as it builds on the LATE framework for identification and sensitivity analysis.

Outline of the paper: The rest of the paper is organized as follows: Section 2 describes the framework and target parameters in the setting. Section 3 provides the partial identification results for the case of binary outcomes and in Section 4 I show the identification of the breakdown values. In Section 5 I perform a numerical illustration of the method. Section 6 introduces the estimators and their asymptotic properties. Section 7 presents the partial identification results for the case of joint *c-dependence*. Section 8 presents the Monte Carlo simulations, Section 9 presents the empirical application and Section 10 concludes.

3.2 Framework

Let $Z \in \{0, 1\}$ denote a binary variable that indicates whether an individual was assigned to treatment ($Z = 1$) or control ($Z = 0$). In the setting, non-compliance is allowed, which means that not all individuals assigned to treatment will actually take the treatment and not all individuals assigned to control will remain untreated. Rather than determining treatment status, the assignment represents an encouragement towards treatment.

Let $D \in \{0, 1\}$ denote the actual treatment status. Define the potential treatment associated to assignment z as $D(z)$. We observe the treatment status

$$D = ZD(1) + (1 - Z)D(0)$$

Let $Y \in \{0, 1\}$ denote the observed binary outcome. The potential outcome associated to assignment z is defined as $Y(D(z), z)$. At first, I allow potential outcomes depend arbitrarily on treatment and assignment. Observed and potential outcomes are related by

$$Y = ZY(D(1), 1) + (1 - Z)Y(D(0), 0)$$

Let $X \in \mathcal{S}(X)$ be a vector of observed covariates and $p_{z|x} = \mathbb{P}(Z = z|X = x)$ be the observed propensity score for assignment. I maintain the following assumption regarding the joint distribution of $(D(z), Y(D(z), z), Z, X)$ throughout the paper:

Assumption 1: For each $z, z' \in \{0, 1\}$ and $x \in \mathcal{S}(X)$:

1. $\mathbb{P}(D(z) = 1|Z = z', X = x) \in (0, 1)$
2. $\mathbb{P}(Y(D(z), z) = 1|Z = z', X = x) \in (0, 1)$

3. $p_{z|x} > 0$

Assumptions 1.1 and 1.2 state that the support of potential quantities does not depend on the assignment. Assumption 1.3 states that all individuals can be assigned to treatment and control with probability greater than zero, and is usually referred to as the common support, or overlap assumption.

The fundamental behavioral assumption for identification in IV settings restricts how individuals respond to assignment, and is formalized below:

Assumption 2: For all $x \in \mathcal{S}(X)$, we have $D(1) \geq D(0)$ conditional on $X = x$.

Assumption 2 is referred to as the monotonicity condition imbensangrist and it states that there are no individuals that would take treatment if assigned to control and would not take treatment in the presence of the encouragement. Under assumption 2 individuals can be divided into three groups regarding their response to assignment: Always-takers (individuals that take treatment regardless of the assignment), Compliers (individuals that mimic their assignment) and Never-takers (individuals that don't take treatment regardless of their assignment).

Another necessary assumption for identification is the exclusion restriction:

Assumption 3: For all $x \in \mathcal{S}(X)$ and $z \in \{0, 1\}$, $Y(D(z), z) = Y(D(z))$.

The exclusion restriction states that the outcome is only affected directly by the actual uptake of the treatment. Hence, assignment only affects outcomes to the extent that it affects the choice of treatment. Under the exclusion restriction, the observed outcome relates to potential outcomes simply by $Y = ZY(D(1)) + (1 - Z)Y(D(0))$.

Point identification in IV settings usually relies on two additional assumptions, which are stated below:

Assumption 4: For all $x \in \mathcal{S}(X)$, $\mathbb{E}[D|Z = 1, X = x] \neq \mathbb{E}[D|Z = 0, X = x]$

Assumption 4 is a technical assumption often referred to as relevance.

Assumption 5: For all $x \in \mathcal{S}(X)$, $(Y(D(z)), D(z)) \perp Z|X = x$.

Assumption 5 states that the assignment of treatment is independent of the potential quantities. Although it is usually justified in experimental settings, it is hard to justify and verify in observational settings.

Under these five assumptions, it is well known that the average treatment effect for compliers (LATE) conditional on $X_i = x$ is identified by the conditional Wald estimand:

$$\begin{aligned} \mathbb{E}[Y(1) - Y(0)|D(1) > D(0), X = x] &= \frac{\mathbb{E}[Y|Z = 1, X = x] - \mathbb{E}[Y|Z = 0, X = x]}{\mathbb{E}[D|Z = 1, X = x] - \mathbb{E}[D|Z = 0, X = x]} \\ &= \frac{\mathbb{E}[Y(D(1)) - Y(D(0))|X = x]}{\mathbb{E}[D(1) - D(0)|X = x]} \equiv \frac{\tau_Y(x)}{\tau_D(x)} \equiv \tau(x) \end{aligned}$$

The unconditional LATE is identified by integrating $\tau_Y(x)$ and $\tau_D(x)$ over the distribution of covariates. In this paper, I study the partial identification of the LATE in settings where the independence assumption is violated. The approach consists in finding bounds for the first-stage ($\tau_D(x)$) and the reduced form ($\tau_Y(x)$) as functions of the magnitude of the dependence of treatment assignment on potential quantities.

The partial identification results are used to conduct sensitivity analysis and identifying breakdown frontiers, that is, the boundary between the set of assumptions which lead to a specific conclusion and those which do not. For instance, one might be interested in the values of dependence which change the conclusion that the LATE is greater than zero.

3.3 Partial Identification with Binary Outcomes

I consider the partial identification as a function of violations of independence in a setting where the outcome Y is binary. In that case, the conditional LATE can be interpreted as the increase in probability of observing $Y = 1$ due to the treatment for compliers with covariates $X = x$:

$$\tau(x) = \mathbb{E}(Y(1) - Y(0) | D(1) > D(0), X = x)$$

I begin with the partial identification of the share of compliers.

3.3.1 First-Stage

We begin with the partial identification of $\tau_D(x)$. For that purpose, write $\tau_D(x) = \tau_{D(1)}(x) - \tau_{D(0)}(x)$, where $\tau_{D(z)}(x) = \mathbb{E}[D(z) | X = x]$. The parameter $\tau_D(x)$ can be interpreted as the share of compliers with $X = x$: $\mathbb{P}(D(1) > D(0) | X = x)$. In the case where Assumptions 1-5 and independence hold, $\tau_D(x)$ is identified by

$$\tau_D(x) = \mathbb{E}[D | Z = 1, X = x] - \mathbb{E}[D | Z = 0, X = x]$$

which is usually referred to as the first-stage in the Wald estimand.

We replace the independence assumption by a bounded dependence assumption, called *c-dependence* (Masten and Poirier, 2018):

Definition: Let $x \in \mathcal{S}(X)$. Let c_1 be a scalar between 0 and 1. We say that Z is *c₁-dependent* with potential treatment $D(z)$ given $X = x$ if

$$\sup_{d \in \{0,1\}} |\mathbb{P}(Z = 1 | D(z) = d, X = x) - \mathbb{P}(Z = 1 | X = x)| \leq c_1$$

Under c_1 -dependence, the unobserved propensity score is allowed to deviate c_1 probability units away from the observed propensity score $p_{1|x}$. For $c_1 = 0$, the assignment is independent from potential treatments (Assumption 5 holds). Throughout the paper, I assume c_1 -dependence:

Assumption 5A: Z is c_1 -dependent with $D(1)$ given X and with $D(0)$ given X .

Let $p_{D|z,x} = \mathbb{P}(D = 1|Z = z, X = x)$. Lemma 1 provides sharp identified sets for potential treatments and the share of compliers:

Lemma 1. *Suppose Assumptions 1-3 and 5A hold. Then, the sharp identified set for potential treatment associated to assignment z is $\tau_{D(z)}(x) \in \left[\tau_{D(z)}^{LB}(c_1, x), \tau_{D(z)}^{UB}(c_1, x) \right]$ is, where*

$$\tau_{D(z)}^{LB}(c_1, x) = \frac{p_{D|z,x} p_{z|x}}{\min\{p_{z|x} + c_1, 1\}}$$

and

$$\tau_{D(z)}^{UB}(c_1, x) = \min \left\{ \frac{p_{D|z,x} p_{z|x}}{p_{z|x} - c_1} \mathbf{1}(p_{z|x} > c_1) + \mathbf{1}(p_{z|x} \leq c_1), p_{D|z,x} p_{z|x} + (1 - p_{z|x}) \right\}$$

Consequently, the sharp identified set for the share of compliers is $\tau_D(x) \left[\tau_D^{LB}(c_1, x), \tau_D^{UB}(c_1, x) \right]$, where

$$\tau_D^{LB}(c_1, x) = \max(0, \tau_{D(1)}^{LB}(c_1, x) - \tau_{D(0)}^{UB}(c_1, x))$$

and

$$\tau_D^{UB}(c_1, x) = \tau_{D(1)}^{UB}(c_1, x) - \tau_{D(0)}^{LB}(c_1, x)$$

Lemma 1 follows directly from Proposition 5 of (MASTEN; POIRIER, 2018). The upper bound for the first-stage is identified by the difference between the upper bound of $\tau_{D(1)}(c_1, x)$ and the lower bound of $\tau_{D(0)}(c_1, x)$. These are both quantities between zero and one, and the monotonicity assumptions ensures that the difference between these quantities is positive.

The lower bound is identified by the difference between the lower bound of $\tau_{D(1)}(c_1, x)$ and the upper bound of $\tau_{D(0)}(c_1, x)$. There is no guarantee that these difference is greater than zero. Since the first-stage identifies a share between zero and one, the lower bound for is restricted to be at least as great as zero.

The unconditional bounds for the first-stage are obtained by integrating the conditional bounds over the distribution of covariates:

$$\begin{aligned}\tau_D^{LB}(c_1) &= \max \left(0, \int \tau_{D(1)}^{LB}(c_1, x) dF_X(x) - \int \tau_{D(0)}^{UB} dF_X(x)(c_1, x) \right) \\ \tau_D^{UB}(c_1) &= \int \tau_{D(1)}^{UB}(c_1, x) dF_X(x) - \int \tau_{D(0)}^{LB}(c_1, x) dF_X(x)\end{aligned}$$

3.3.2 Reduced Form

Now focus on the identification of $\tau_Y(x)$, which we write as $\tau_Y(x) = \tau_{Y(D(1))}(x) - \tau_{Y(D(0))}(x)$, where $\tau_{Y(D(z))}(x) = \mathbb{E}[Y(D(z))|X = x]$. The parameter $\tau_Y(x)$ captures the effect of assignment on potential outcomes, which is often referred to in the literature as the Intention-to-Treat effect (ITT). In the case where Assumptions 1-5 and independence hold, $\tau_Y(x)$ is identified by

$$\tau_Y(x) = \mathbb{E}[Y|Z = 1, X = x] - \mathbb{E}[Y|Z = 0, X = x]$$

which is referred to as the reduced form. As in the first-stage, I relax the independence assumption replace it by a *c-dependence* assumption of Z_i with the potential outcomes.

Definition: Let $x \in \mathcal{S}(X)$. Let c_2 be a scalar between 0 and 1. We say that Z is *c₂-dependent* with potential outcome $Y(D(z))$ given $X = x$ if

$$\sup_{y \in \{0,1\}} |\mathbb{P}(Z = 1|Y(D(z)) = y, X = x) - \mathbb{P}(Z = 1|X = x)| \leq c_2$$

For $c_2 = 0$, the assignment is independent from potential outcomes (Assumption 5 holds). Throughout the paper, I assume *c₂-dependence*:

Assumption 5B: Z is *c₂-dependent* with $Y(D(1))$ given X and with $Y(D(0))$ given X .

Let $p_{Y|z,x} = \mathbb{P}(Y = 1|Z = z, X = x)$. I use the results from (MASTEN; POIRIER, 2018) to derive sharp identified sets for potential outcomes and the ITT:

Lemma 2. *Suppose Assumptions 1-3 and 5B hold. Then, the sharp identified set for potential outcome associated to assignment z is $\tau_{Y(D(z))}(x) \in \left[\tau_{Y(D(z))}^{LB}(c_2, x), \tau_{Y(D(z))}^{UB}(c_2, x) \right]$ is, where*

$$\tau_{Y(D(z))}^{LB}(c_2, x) = \frac{p_{Y|z,x} p_{z|x}}{\min \{ p_{z|x} + c_2, 1 \}}$$

and

$$\tau_{Y(D(z))}^{UB}(c_2, x) = \min \left\{ \frac{p_{Y|z,x} p_{z|x}}{p_{z|x} - c_2} \mathbf{1}(p_{z|x} > c_2) + \mathbf{1}(p_{z|x} \leq c_2), p_{Y|z,x} p_{z|x} + (1 - p_{z|x}) \right\}$$

Consequently, the sharp identified set for the share of compliers is $\tau_Y(x) [\tau_Y^{LB}(c_2, x), \tau_Y^{UB}(c_2, x)]$, where

$$\tau_Y^{LB}(c_2, x) = \max(0, \tau_{Y(D(1))}^{LB}(c_2, x) - \tau_{Y(D(0))}^{UB}(c_2, x))$$

and

$$\tau_Y^{UB}(c_2, x) = \tau_{Y(D(1))}^{UB}(c_2, x) - \tau_{Y(D(0))}^{LB}(c_2, x)$$

The bounds are similar to those obtained for the first-stage. Note that the ITT is not bounded by definition between zero and one, and thus, there is no need to constraint the lower bound to be at least as great as zero.

The unconditional bounds for the ITT are identified by integrating the conditional bounds over the distribution of covariates:

$$\begin{aligned} \tau_Y^{LB}(c_2) &= \int \tau_{Y(D(1))}^{LB}(c_2, x) dF_X(x) - \int \tau_{Y(D(0))}^{UB}(c_2, x) dF_X(x) \\ \tau_Y^{UB}(c_2) &= \int \tau_{Y(D(1))}^{UB}(c_2, x) dF_X(x) - \int \tau_{Y(D(0))}^{LB}(c_2, x) dF_X(x) \end{aligned}$$

3.3.3 Local Average Treatment Effect

The Local Average Treatment Effect (LATE), the average treatment effect for the subgroup of compliers, under the standard IV assumptions, is point-identified by the ratio of the reduced form and the first-stage. Replacing Assumption 5 with Assumptions 5A and 5B, we obtain the following bounds for the LATE. Putting the pieces from Sections 3.1 and 3.2 together, we find that $\tau(x) \in [\tau^{LB}(c_1, c_2, x), \tau^{UB}(c_1, c_2, x)]$, with

$$\begin{aligned} \tau^{LB}(c_1, c_2, x) &= \max\left(\frac{\tau_Y^{LB}(c_2, x)}{\tau_D^{UB}(c_1, x)}, -1\right) \\ \tau^{UB}(c_1, c_2, x) &= \min\left(\frac{\tau_Y^{UB}(c_2, x)}{\tau_D^{LB}(c_1, x)}, 1\right) \end{aligned}$$

In the case of a binary outcome, treatment effects are not greater than 1 nor smaller than -1 . Hence, the lower bound is the greatest value between the ratio of the lower bound of the ITT and the upper bound of the first-stage, and -1 . The upper bound is the smallest value between the ratio of the upper bound of the ITT and the lower bound of the first-stage, and 1. The unconditional bounds for the LATE are identified by integrating the conditional LATE bounds over the distribution of covariates:

$$\tau^{LB}(c_1, c_2) = \max\left(\frac{\int \tau_Y^{LB}(c_2, x) dF_X(x)}{\int \tau_D^{UB}(c_1, x) dF_X(x)}, -1\right)$$

and

$$\tau^{UB}(c_1, c_2) = \min \left(\frac{\int \tau_Y^{UB}(c_2, x) dF_X(x)}{\int \tau_D^{LB}(c_1, x) dF_X(x)}, 1 \right)$$

The bounds for the LATE are functions of violations of instrument independence with respect to both potential treatments and potential outcomes. In that sense, it is similar to the partial identification result presented in Section 5.3 of (RAMBACHAN; ROTH, 2025), which partially identifies the LATE as a function of the finite-population covariance between the assignment probabilities and the potential outcomes and treatments. In the next section, I show how researchers can identify the set of violations of independence under which a conclusion is invalidated.

3.4 Breakdown Analysis

The fundamental interest here is to understand under which violations of independence a particular conclusion still hold. In this section I define breakdown points for conclusions regarding the first-stage and the reduced form separately, and a breakdown frontier for conclusions regarding the LATE.

First, I define the breakdown point for conclusions of the first-stage. That is, what is the largest value of c_1 under which we can conclude that $\mathbb{P}(D(1) > D(0)) \geq \mu$? First, define the robust region for the conclusion as the set of values of c_1 under which the conclusion holds:

$$RR_{FS}(\mu) = \{c_1 \in [0, 1] : \tau_D^{LB}(c_1) \geq \mu\}$$

The robust region for the first-stage is the set of values of c_1 for which the identified set for the share of compliers is above μ . The breakdown point is the value of c_1 on the boundary of the robust region. Formally, the breakdown point is c_1^* such that $\tau_D^{LB}(c_1^*) = \mu$. In the case where $\mu \geq 0$, solving for c_1 in the equation yields

$$bf_{FS}(\mu) = \frac{\int p_{D|1,x} p_{1|x} + p_{D|0,x} p_{0|x} - \mu(p_{1|x} - p_{0|x}) \pm \sqrt{h_{FS}(x, \mu)} dF_X(x)}{2\mu}$$

where

$$h_{FS}(x, \mu) = \left\{ p_{D|1,x} p_{1|x} + p_{D|0,x} p_{0|x} - \mu(p_{1|x} - p_{0|x}) \right\}^2 - 4p_{1|x} p_{0|x} \mu (p_{D|1,x} - p_{D|0,x} - \mu)$$

In the case where $\mu = 0$, the expression simplifies to

$$bf_{FS}(\mu) = \frac{\int p_{1|x} p_{0|x} (p_{D|1,x} - p_{D|0,x}) dF_X(x)}{\int p_{D|1,x} p_{1|x} + p_{D|0,x} p_{0|x} dF_X(x)}$$

Therefore, we obtain the following expression for the breakdown point:

$$c_1^* = \min \{ \max \{ bf_{FS}(\mu), 0 \}, 1 \}$$

Similarly, when it comes to the reduced form, the breakdown point is the largest value of c_2 under which one can conclude that $\mathbb{E} [Y(D(1)) - Y(D(0))] \geq \mu$. It is defined implicitly by c_2^* such that $\tau_Y^{LB}(c_2) \geq \mu$. In the case where $\mu \neq 0$, Solving for c_2 yields

$$bf_{RF}(\mu) = \frac{\int p_{Y|1,x}p_{1|x} + p_{Y|0,x}p_{0|x} - \mu(p_{1|x} - p_{0|x}) \pm \sqrt{h_{RF}(x, \mu)} dF_X(x)}{2\mu}$$

where

$$h_{RF}(x, \mu) = \{ p_{Y|1,x}p_{1|x} + p_{Y|0,x}p_{0|x} - \mu(p_{1|x} - p_{0|x}) \}^2 - 4p_{1|x}p_{0|x}\mu(p_{Y|1,x} - p_{Y|0,x} - \mu)$$

In the case where $\mu = 0$, the expression simplifies to

$$bf_{RF}(\mu) = \frac{\int p_{1|x}p_{0|x} (p_{Y|1,x} - p_{Y|0,x}) dF_X(x)}{\int p_{Y|1,x}p_{1|x} + p_{Y|0,x}p_{0|x} dF_X(x)}$$

Therefore, we obtain the following expression for the breakdown point:

$$c_2^* = \min \{ \max \{ bf_{RF}(\mu), 0 \}, 1 \}$$

When it comes to the LATE, the parameter is partially identified as function of both sensitivity parameters c_1, c_2 . In that case, the robust region of identification is the set of values of c_1 and c_2 under which the conclusion holds. It is defined as

$$RR(\mu) = \{ (c_1, c_2) \in [0, 1]^2 : \tau^{LB}(c_1, c_2) \geq \mu \}$$

The breakdown frontier is the set of values of c_1 and c_2 on the boundary of the robust region. Specifically, the breakdown frontier is defined as

$$BF(\mu) = \{ (c_1, c_2) \in [0, 1]^2 : \tau^{LB}(c_1, c_2) = \mu \}$$

Solving in the equation $\tau^{LB}(c_1, c_2) = \mu$ for c_2 yields

$$bf(c_1, \mu) = \frac{\int p_{Y|1,x}p_{1|x} + p_{Y|0,x}p_{0|x} - \mu\tau_D^{UB}(c_1, x)(p_{1|x} - p_{0|x}) \pm \sqrt{h(x, c_1, \mu)} dF_X(x)}{2\mu \int \tau_D^{UB}(c_1, x) dF_X(x)}$$

where

$$h(x, c_1, \mu) = \{p_{Y|1,x}p_{1|x} + p_{Y|0,x}p_{0|x} - \mu\tau_D^{UB}(c_1, x)(p_{1|x} - p_{0|x})\}^2 \\ - 4p_{1|x}p_{0|x}(p_{Y|1,x} - p_{Y|0,x} - \mu\tau_D^{UB}(c_1, x))\mu\tau_D^{UB}(c_1, x)$$

Therefore, we obtain the following expression for the breakdown frontier as a function of c_1 :

$$BF(c_1, \mu) = \min \{ \max \{ bf(c_1, \mu), 0 \}, 1 \}$$

The shape of the breakdown frontier provides insights on the tradeoffs between these two types of relaxations of independence when researchers are drawing conclusions. The relaxations c_1 and c_2 are measured in the same unit, which helps the interpretation of the breakdown analysis.

3.5 Numerical Illustration

I illustrate the breakdown approach with a simple numerical illustration. Let X be a binary covariate that follows a Bernoulli distribution with parameter $p_x = 0.5$. Let Z denote the instrument which also follows a Bernoulli distribution with parameter p_z . Let $p_{z|x} = 0.5$, for the sake of simplicity. Selection into treatment follows a threshold-crossing model as in (VYTLACIL, 2002):

$$D = \mathbf{1} \{ \pi_z Z + \pi_x X \geq V \}$$

where V has a standard uniform distribution. The parameter π_z is the share of compliers, and is set to be equal to 0.5.

The binary outcome also follows a threshold model, as in Heckman (1978):

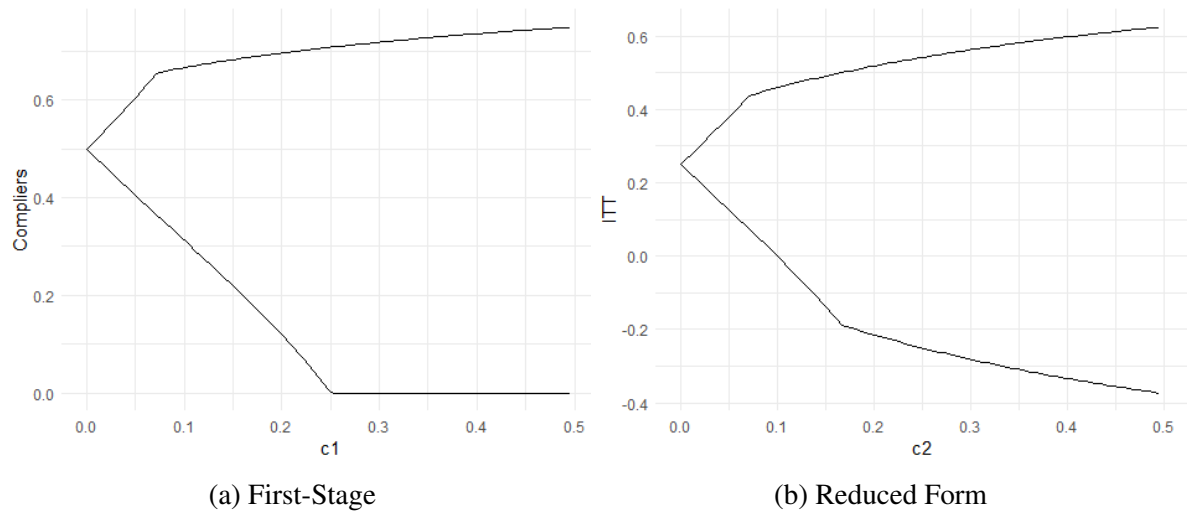
$$Y = \mathbf{1} \{ \beta_d D + \beta_x X \geq U \}$$

where U is also uniformly distributed. The random variables U and V are linearly correlated as in (OLSEN, 1980), which generates the selection problem. The parameter β_d is the LATE in this DGP, which is set to be equal to 0.5. Hence, it follows that the ITT is equal to 0.25.

Figure 1 (a) shows the identified set for the first-stage as a function of c_1 . The breakdown point for the conclusion that the share of compliers is greater than zero is $c_1^* = 0.25$. Figure 1 (b) show the identified set of the ITT as a function of c_2 . The breakdown point for the conclusion that the ITT, and hence the LATE, is greater than zero is $c_1^* = 0.1$.

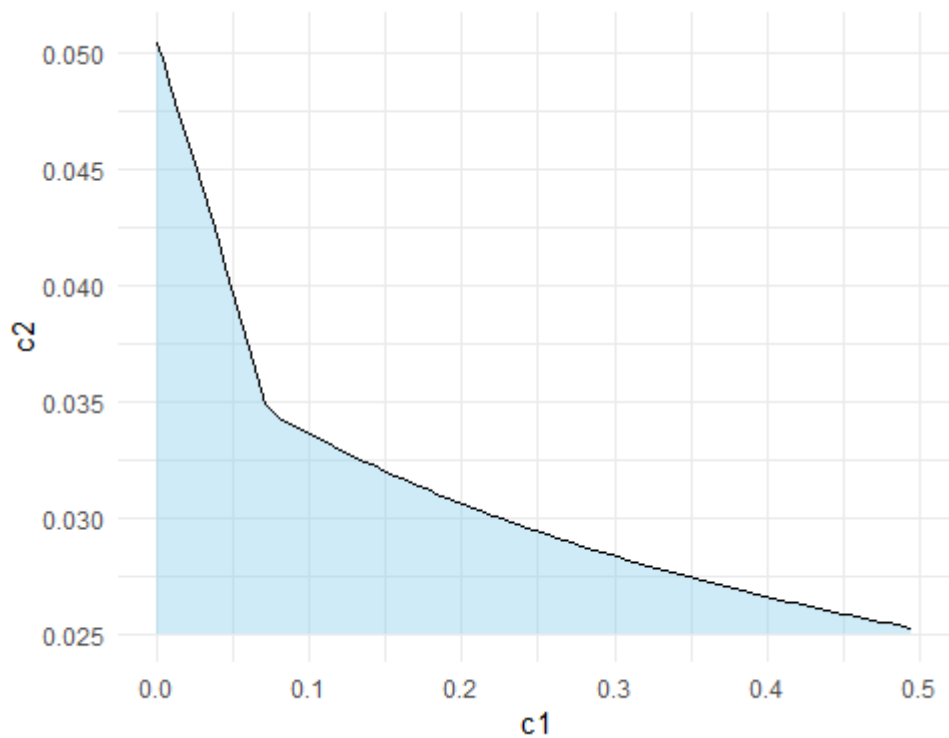
Figure 2 shows the breakdown frontier for the LATE. I specify the breakdown frontier for the conclusion that the LATE is greater than 0.25, which is half of its true value. The blue

Figure 1 - Identified sets for the first-stage and reduced form.



Note: Figure (a) show the identified set for the share of compliers under the dgp described in Section 6 and different values of c_1 . Figure (b) show the identified set for the ITT under the dgp described in Section 6 and different values of c_2

Figure 2 - Breakdown Frontier for the LATE



Note: Figure 2 plots the breakdown frontier for the conclusion that $\tau \geq 0.25$.

area represents the robust region, that is, this is the set of values for violations of independence under which the conclusion still holds.

3.6 Estimation and Inference

In this section I study estimation and inference on the identified sets and breakdown values defined above. The estimands for the bounds of assignment effects, the LATE and the breakdown values are functionals of the conditional cdf of outcomes and treatment given assignment and covariates, the probability of assignment given covariates, and the marginal distribution of the covariates. I propose nonparametric sample analog estimators and derive asymptotic distributional results using a delta method for directionally differentiable functionals. First, I assume we observe a random sample of data.

Assumption 6: The random variables $\{(Y_i, D_i, Z_i, X_i)\}_{i=1}^n$ are independently and identically distributed according to the distribution of (Y, D, Z, X) .

Furthermore, I assume the support of covariates is discrete.

Assumption 7: The support of X is discrete and finite. Let $\mathcal{S}(X) = \{x_1, \dots, x_K\}$ up to a finite K .

All target parameters are functionals of the underlying parameters $p_{Y|z,x}$, $p_{D|z,x}$, $p_{z|x}$ and $q_x = \mathbb{P}(X = x)$. Their parametric estimators are, respectively,

$$\widehat{p}_{Y|z,x} = \frac{\frac{1}{N} \sum_{i=1}^N \mathbf{1}(Y_i = 1) \mathbf{1}(Z_i = z, X_i = x)}{\frac{1}{N} \sum_{i=1}^N \mathbf{1}(Z_i = z, X_i = x)},$$

$$\widehat{p}_{D|z,x} = \frac{\frac{1}{N} \sum_{i=1}^N \mathbf{1}(D_i = 1) \mathbf{1}(Z_i = z, X_i = x)}{\frac{1}{N} \sum_{i=1}^N \mathbf{1}(Z_i = z, X_i = x)},$$

$$\widehat{p}_{z|x} = \frac{\frac{1}{N} \sum_{i=1}^N \mathbf{1}(Z_i = z, X_i = x)}{\frac{1}{N} \sum_{i=1}^N \mathbf{1}(X_i = x)}$$

and

$$\widehat{q}_x = \frac{1}{N} \sum_{i=1}^N \mathbf{1}(X_i = x)$$

These quantities converge uniformly to a Gaussian process at a \sqrt{N} -rate; see Lemma B1 in Appendix B. Next, consider the bounds for potential treatments and outcomes. I estimate these bounds by a plug-in estimator of the quantities above. First, I introduce an additional assumption:

Assumption 8: For all $x \in \mathcal{S}(X)$, we have $\max\{c_1, c_2\} < \min\{p_{1|x}, p_{0|x}\}$.

Assumption 8 is a technical assumption which simplifies the bounds for potential treatments to

$$\begin{aligned}\tau_{D(z)}^{LB}(c_1, x) &= \frac{p_{D|z,x}p_{z|x}}{\min\{p_{z|x} + c_1, 1\}} = \max\left\{\frac{p_{D|z,x}p_{z|x}}{p_{z|x} + c_1}, p_{D|z,x}p_{z|x}\right\} \\ \tau_{D(z)}^{UB}(c_1, x) &= \min\left\{\frac{p_{D|z,x}p_{z|x}}{p_{z|x} - c_1}, p_{D|z,x}p_{z|x} + (1 - p_{z|x})\right\}\end{aligned}$$

and the bounds for potential outcomes to

$$\begin{aligned}\tau_{Y(D(z))}^{LB}(c_2, x) &= \frac{p_{Y|z,x}p_{z|x}}{\min\{p_{z|x} + c_2, 1\}} = \max\left\{\frac{p_{Y|z,x}p_{z|x}}{p_{z|x} + c_2}, p_{Y|z,x}p_{z|x}\right\} \\ \tau_{Y(D(z))}^{UB}(c_2, x) &= \min\left\{\frac{p_{Y|z,x}p_{z|x}}{p_{z|x} - c_2}, p_{Y|z,x}p_{z|x} + (1 - p_{z|x})\right\}\end{aligned}$$

This simplification is particularly important to guarantee *Hadamard directional differentiability* of the estimators. The bounds for potential quantities are estimated by replacing the population quantities in the expressions above by its sample analogues. In Lemma B2 of Appendix B I show that the estimators for potential quantities converge in distribution to a nonstandard distribution given by a continuous transformation of Gaussian processes. This result is the building block for deriving the asymptotic properties of the estimators for bounds and breakdown values.

First, consider the bounds for the first-stage. The plug-in estimators for the lower and the upper bound, respectively are

$$\begin{aligned}\hat{\tau}_D^{LB}(c_1, x) &= \max\{0, \hat{\tau}_{D(1)}^{LB}(c_1, x) - \hat{\tau}_{D(0)}^{UB}(c_1, x)\} \\ \hat{\tau}_D^{UB}(c_1, x) &= \hat{\tau}_{D(1)}^{UB}(c_1, x) - \hat{\tau}_{D(0)}^{LB}(c_1, x)\end{aligned}$$

The unconditional bounds are estimated by integrating the estimates of conditional bounds over the empirical distribution of covariates:

$$\hat{\tau}_D^{LB}(c_1) = \frac{1}{N} \sum_{i=1}^N \hat{\tau}_D^{LB}(c_1, X_i)$$

and

$$\hat{\tau}_D^{UB}(c_1) = \frac{1}{N} \sum_{i=1}^N \hat{\tau}_D^{UB}(c_1, X_i)$$

Now consider the bounds for the reduced form. The plug-in estimators are

$$\begin{aligned}\hat{\tau}_Y^{LB}(c_2, x) &= \hat{\tau}_{Y(D(1))}^{LB}(c_2, x) - \hat{\tau}_{Y(D(0))}^{UB}(c_2, x) \\ \hat{\tau}_Y^{UB}(c_2, x) &= \hat{\tau}_{Y(D(1))}^{UB}(c_2, x) - \hat{\tau}_{Y(D(0))}^{LB}(c_2, x)\end{aligned}$$

The unconditional bounds are obtained by integrating the estimates over the empirical distribution of covariates:

$$\widehat{\tau}_Y^{LB}(c_2) = \frac{1}{N} \sum_{i=1}^N \widehat{\tau}_Y^{LB}(c_2, X_i)$$

and

$$\widehat{\tau}_Y^{UB}(c_2) = \frac{1}{N} \sum_{i=1}^N \widehat{\tau}_Y^{UB}(c_2, X_i)$$

The asymptotic distribution of the estimator for these bounds is formalized in the proposition below:

Proposition 1. *Assume Assumptions 1-3, 5A, 5B and 6-8 hold. Then,*

$$\sqrt{N} \begin{pmatrix} \widehat{\tau}_D^{LB}(c_1) - \tau_D^{LB}(c_1) \\ \widehat{\tau}_D^{UB}(c_1) - \tau_D^{UB}(c_1) \end{pmatrix} \xrightarrow{d} \mathbf{Z}_{FS}(d, z, x, c_1)$$

and

$$\sqrt{N} \begin{pmatrix} \widehat{\tau}_Y^{LB}(c_2) - \tau_Y^{LB}(c_2) \\ \widehat{\tau}_Y^{UB}(c_2) - \tau_Y^{UB}(c_2) \end{pmatrix} \xrightarrow{d} \mathbf{Z}_{RF}(y, z, x, c_2)$$

where $\mathbf{Z}_{FS}(d, z, x, c_1)$ and $\mathbf{Z}_{RF}(y, z, x, c_2)$ are Gaussian Elements defined in the Section 1 of Appendix A.

Now consider the estimation for the breakdown point for the claim that $\tau_D \geq \mu$. We focus on that case where $\tau_D^{LB}(0) > \mu$, which implies that $c_1^* > 0$. Define the estimator for the breakdown point c_1^* as

$$\widehat{c}_1^* = \inf \{c_1 \in [0, 1] : \widehat{\tau}_D^{LB}(c_1) \leq \mu\}$$

and it is obtained by replacing the population quantities from $bf_{FS}(\mu)$ with the sample analogues defined in this section. The estimator for the breakdown frontier for the first-stage is thus $\widehat{c}_1^* = \min \left\{ \max \left\{ \widehat{bf}_{FS}(\mu), 0 \right\}, 1 \right\}$

Similarly, when it comes to the estimation for the breakdown point for the claim that $\tau_Y \geq \mu$, define the estimator for the breakdown point c_2^* as

$$\widehat{c}_2^* = \inf \{c_2 \in [0, 1] : \widehat{\tau}_Y^{LB}(c_2) \leq \mu\}$$

which is obtained by replacing the population quantities from $bf_{RF}(\mu)$ with the sample analogues. The estimator for the breakdown frontier for the first-stage is thus $\widehat{c}_2^* = \min \left\{ \max \left\{ \widehat{bf}_{FS}(\mu), 0 \right\}, 1 \right\}$,

I now provide a formal result regarding the asymptotic distribution of \widehat{c}_1^* and \widehat{c}_2^* :

Theorem 1. *Assume Assumptions 1-3, 5A, 5B and 6-8 hold. Furthermore, assume that $c_1, c_2 \in [0, \overline{C}]$. Then,*

$$\sqrt{N} (\widehat{c}_1^* - c_1^*) \xrightarrow{d} \mathbf{Z}_{FS}^{BP}$$

and

$$\sqrt{N} (\widehat{c}_2^* - c_2^*) \xrightarrow{d} \mathbf{Z}_{RF}^{BP}$$

where \mathbf{Z}_{FS}^{BP} and \mathbf{Z}_{RF}^{BP} are Gaussian random variables defined in Section 2 of Appendix A.

Finally, consider the estimators for the bounds and breakdown frontier of the LATE. The estimators for the lower and upper bound are respectively

$$\widehat{\tau}^{LB}(c_1, c_2) = \max \left\{ \frac{\widehat{\tau}_Y^{LB}(c_2)}{\widehat{\tau}_D^{UB}(c_1)}, -1 \right\}$$

and

$$\widehat{\tau}^{UB}(c_1, c_2) = \min \left\{ \frac{\widehat{\tau}_Y^{UB}(c_2)}{\widehat{\tau}_D^{LB}(c_1)}, 1 \right\}$$

The next lemma formalizes the asymptotic distribution of the bounds:

Proposition 2. *Suppose Assumptions 1-3, 5A, 5B and 6-8 hold. Then,*

$$\sqrt{N} \begin{pmatrix} \widehat{\tau}^{LB}(c_1, c_2) - \tau^{LB}(c_1, c_2) \\ \widehat{\tau}^{UB}(c_1, c_2) - \tau^{UB}(c_1, c_2) \end{pmatrix} \xrightarrow{d} \mathbf{Z}_\tau(y, d, z, x, c_1, c_2)$$

Denote the estimated breakdown frontier for the conclusion that $\tau \geq \mu$ by

$$\widehat{BF}(c_1, \mu) = \min \left\{ \max \left\{ \widehat{bf}(c_1, \mu), 0 \right\}, 1 \right\}$$

where

$$\widehat{bf}(c_1, \mu) = \frac{\sum_{i=1}^N \left\{ \widehat{P}_{Y|1,x} \widehat{P}_{1|x} + \widehat{P}_{Y|0,x} \widehat{P}_{0|x} - \mu \widehat{\tau}_D^{UB}(c_1, X_i) (\widehat{P}_{1|x} - \widehat{P}_{0|x}) \pm \sqrt{\widehat{h}(X_i, c_1, \mu)} \right\}}{\sum_{i=1}^N 2\mu \widehat{\tau}_D^{UB}(c_1, X_i)}$$

I show that the estimated breakdown frontier converges in distribution.

Theorem 2. *Let Assumptions 1-3, 5A, 5B and 6-8 hold. Furthermore, let $\mathcal{C} \subset [0, \bar{C}]$ and $\mathcal{M} \subset [-1, 1]$ be finite grids of points. Then,*

$$\sqrt{N} \left(\widehat{BF}(c_1, \mu) - BF(c_1, \mu) \right) \xrightarrow{d} \mathbf{Z}_{BF}(c_1, \mu),$$

a tight random element of $l^\infty(\mathcal{C} \times \mathcal{M})$.

Since the limiting process is non-Gaussian, inference on the breakdown values will not be based on standard errors. The processes' distribution is characterized fully by the expressions in Appendices B1 and B2, but obtaining analytical estimates of functionals of these processes is challenging. In the next subsection I give describe a bootstrap procedure that can be used to construct confidence intervals for the breakdown points and confidence bands for the breakdown frontier.

3.6.1 Bootstrap Inference

Before describing the procedure, I introduce some notation. Let $\mathcal{F}_i = (Y_i, D_i, Z_i, X_i)$ and $\mathcal{F}_{1:N} = \{\mathcal{F}_1, \dots, \mathcal{F}_N\}$. Let $\hat{\theta}$ denote the estimator of a parameter θ_0 based on $\mathcal{F}_{1:N}$. Let $\mathbf{A}_{1:N}^* \equiv \sqrt{N} (\hat{\theta}^* - \hat{\theta})$ where $\hat{\theta}^*$ is a draw from the nonparametric bootstrap distribution of $\hat{\theta}$. Suppose \mathbf{A} is the tight limiting process of $\sqrt{N} (\hat{\theta} - \theta_0)$. Bootstrap consistency is given by weak convergence in probability conditional on $\mathcal{F}_{1:N}$, that is,

$$\sup_{h \in BL_1} |\mathbb{E}[h(\mathbf{A}_{1:N}^*) | \mathcal{F}_{1:N}] - \mathbb{E}[h(\mathbf{A})]| = o_p(1)$$

where BL_1 denotes the set of Lipschitz functions into \mathbb{R} with Lipschitz constant no greater than 1. For θ_0 and $\hat{\theta}$, I focus on the parameters introduced in Section 3 and their sample analogue estimators, which are plugged-in in the bounds estimators.

Let $\mathbf{Z}_{1:N} = \sqrt{N} (\hat{\theta}^* - \hat{\theta})$. Theorem 3.6.1 of van der Vaart and Wellner (1996) implies the bootstrap consistency of $\mathbf{Z}_{1:N}$. Since the parameters of interest are Hadamard differentiable functionals of θ_0 , it follows from (FANG; SANTOS, 2018) that the nonparametric bootstrap can be used to do inference on the identified sets and breakdown values.

For the breakdown points of the first-stage and the reduced form, the bootstrap procedure can be used to construct one-sided confidence intervals as in (KLINE; SANTOS, 2013). For the breakdown frontier of the LATE, the bootstrap can be used to construct uniform one-sided confidence bands as in (MASTEN; POIRIER, 2020).

3.7 Partial Identification under joint c -dependence

The results from Sections 3 and 4 provide the breakdown analysis framework for IV settings with binary outcomes in the case where the assumption of instrument independence is replaced by a bounded dependence assumption, that consider the probability of assignment conditional on potential treatments and potential outcomes separately.

Relaxing the independence assumption in terms of c_1 and c_2 -dependence allows the researcher to conduct breakdown analysis for the share of compliers and the ITT separately, while also allowing for the possibility of assessing tradeoffs between these assumptions in the breakdown frontier for the LATE.

Despite the several desirable features of this approach, it does not provide a sharp identified set for the LATE. In this section, I derive sharp bounds for the LATE under a joint c -dependence assumption, which I define below:

Definition: Let $x \in \mathcal{S}(X)$. Let c be a scalar between 0 and 1. We say that Z is joint c -dependent with $(Y(D(z)), D(z))$ given $X = x$ if

$$\sup_{(y,d) \in \{0,1\}^2} |\mathbb{P}(Z = 1 | Y(D(z)) = y, D(z) = d, X = x) - \mathbb{P}(Z = 1 | X = x)| \leq c$$

As in Section 3, the sensitivity parameter captures deviations from the independence assumption in terms of the distance in probability units between the probability of assignment given covariates and the probability of assignment given covariates and potential quantities. Note that $c = 0$ is the case where independence holds, and the target parameters in the setting are point identified. Throughout this section, I assume joint c -dependence:

Assumption 9: Z is joint c -dependent with $(Y(D(1)), D(1))$ given X and $(Y(D(0)), D(0))$ given X . Next, I derive the sharp identified set for potential quantities and the LATE:

Theorem 3. For a random variable $Q \in \{Y, D\}$, denote its potential value associated to assignment z as $Q(z)$. Suppose Assumptions 1-3 and 6-9 hold. The sharp identified set for potential quantities is $\tau_{Q(z)}(x) \in [\tau_{Q(z)}^{LB}(c, x), \tau_{Q(z)}^{UB}(c, x)]$, where

$$\tau_{Q(z)}^{LB}(c, x) = \max \left\{ p_{Q|z,x} p_{z|x} + \left[\frac{(p_{1-z|x} - c) p_{Q|z,x} p_{z|x}}{(p_{z|x} + c) p_{1-z|x}} \right] p_{1-z|x}, p_{Q|z,x} p_{z|x} \right\}$$

and

$$\tau_{Q(z)}^{UB}(c, x) = \min \left\{ p_{Q|z,x} p_{z|x} + \left[\frac{(p_{1-z|x} + c) p_{Q|z,x} p_{z|x}}{(p_{z|x} - c) p_{1-z|x}} \right] p_{1-z|x}, p_{Q|z,x} p_{z|x} + (1 - p_{z|x}) \right\}$$

Consequently, the sharp identified set for the LATE is $\tau(x) \in [\tau^{LB}(c, x), \tau^{UB}(c, x)]$, where

$$\tau^{LB}(c, x) = \frac{\tau_{Y(D(1))}^{LB}(c, x) - \tau_{Y(D(0))}^{UB}(c, x)}{\tau_{D(1)}^{UB}(c, x) - \tau_{D(0)}^{LB}(c, x)}$$

and

$$\tau^{UB}(c, x) = \frac{\tau_{Y(D(1))}^{UB}(c, x) - \tau_{Y(D(0))}^{LB}(c, x)}{\tau_{D(1)}^{LB}(c, x) - \tau_{D(0)}^{UB}(c, x)}$$

Theorem 3 provides the partial identification results for the LATE as a function of the sensitivity parameter c . As in Section 4, the bounds for the LATE can be used to identify the breakdown point to a particular conclusion under joint c -dependence. Estimation and inference procedures for this case are similar to the ones presented in Section 6. Once again, the bounds for the unconditional LATE are identified by integrating the conditional bounds over the distribution of covariates.

When it comes to estimation, the nonparametric estimators for conditional probabilities defined in Section 6 can be used as plug-ins to build an estimator for the bounds of the LATE under joint c -dependence. The proposition below shows that the estimators of the bounds converge to Gaussian elements:

Proposition 3. *Suppose Assumptions 1-3 and 6-9 hold. Then,*

$$\sqrt{N} \begin{pmatrix} \widehat{\tau}^{LB}(c) - \tau^{LB}(c) \\ \widehat{\tau}^{UB}(c) - \tau^{UB}(c) \end{pmatrix} \xrightarrow{d} \mathbf{Z}_{j,\tau}(y, d, z, x, c)$$

Now, turn to the breakdown point for the conclusion that the LATE is equal or greater than μ , which is denoted by c^* . Let

$$\widehat{c}^* = \inf \{c \in [0, 1] : \widehat{\tau}^{LB}(c) \leq \mu\}$$

denote the estimated breakdown point. The next result formally present a result about the asymptotic distribution of \widehat{c}^* :

Theorem 4. *Suppose Assumptions 1-3 and 6-9 hold. Furthermore, assume that $c \in [0, \overline{C}]$. Then,*

$$\sqrt{N} (\widehat{c}^* - c^*) \xrightarrow{d} \mathbf{Z}_j^{BP}$$

where \mathbf{Z}_j^{BP} is a Gaussian random variable defined in Section 7 of Appendix A.

Tabela 1 – Monte Carlo Simulations

	c_1^*	c_2^*	c^*
$\mu = 0$			
Av. Bias	0.001	-0.003	0.003
Med. Bias	-0.001	-0.005	-0.002
Cover	0.932	0.927	0.954
$\mu = 0.10$			
Av. Bias	0.001	0.009	0.004
Med. Bias	0.002	0.008	-0.002
Cover	0.929	0.916	0.942
$\mu = 0.20$			
Av. Bias	0.011	0.008	0.008
Med. Bias	0.009	0.006	0.007
Cover	0.931	0.926	0.933

Note: Simulations based on 1.000 Monte Carlo experiments with sample size $N = 1.000$. One-sided CIs for were built using the procedure from (KLINE; SANTOS, 2013).

When it comes to inference, the same Bootstrap procedures introduced in Section 6.1 can be used in the case of joint c -dependence to perform valid inference over the bounds for the LATE and the breakdown point.

3.8 Monte Carlo Simulations

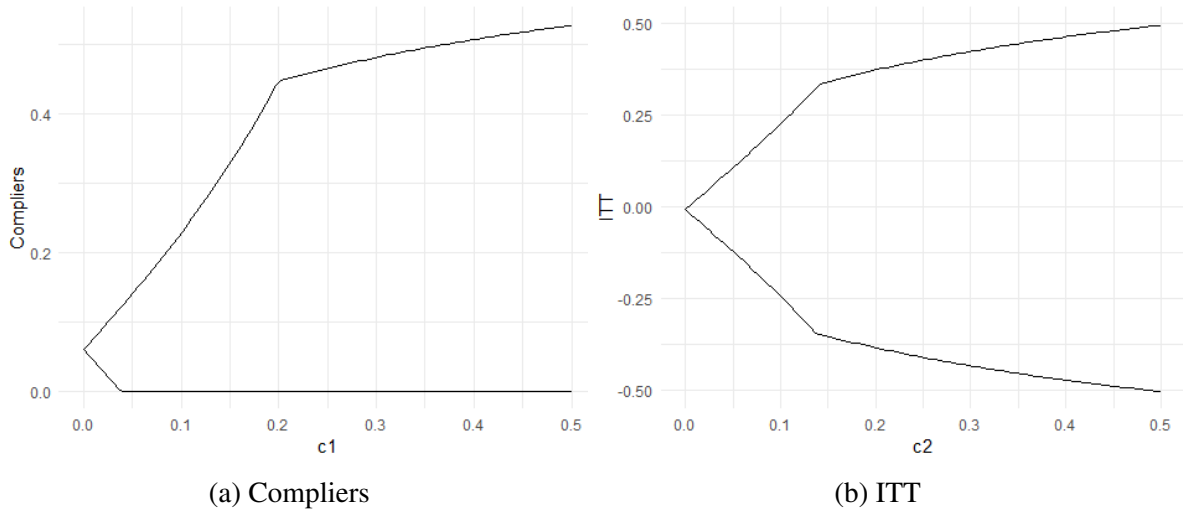
In this section I perform Monte Carlo exercises to study the properties of the estimators for the breakdown points under the DGP described in Section 5. I consider a sample size n equal to 1000 and conduct 1000 Monte Carlo simulations to study the performance of the estimators for breakdown points regarding different conclusions.

To focus the nondegenerate case, I only consider claims that yield breakdown points greater than zero.

I analyze the performance of the estimators for the breakdown points of conclusions regarding the share of compliers and the ITT under separate c_1 and c_2 -dependence, and conclusions regarding the LATE under joint c -dependence. Table 1 displays the results. The estimators are analyzed in terms of their average and median bias, and the 95 % coverage of the one-sided confidence interval from (KLINE; SANTOS, 2013).

Overall, the estimators exhibit desirable finite-sample properties, expressed in terms of close to zero finite-sample bias and empirical coverage of the confidence interval being close to the target 95 %. The performance of the estimators is stable across different values of μ .

Figure 3 - Family Size and Female Employment: Compliers and ITT



Note: Figure (a) show the identified set for the share of compliers in the Angrist & Evans (1998) setting under different values of c_1 . Figure (b) show the identified set for the ITT in the Angrist & Evans (1998) setting under different values of c_2 .

3.9 Empirical Application: Family Size and Employment

In this Section, I use the estimators from Sections 6 and 7 to perform the breakdown analysis for the results regarding family size and female employment in (ANGRIST; EVANS, 1998), using data from the US Census Public Use Microsamples married mothers aged 21–35 in 1980 with at least 2 children and oldest child less than 18.

In this setting, the dependent variable is an indicator for women who worked for pay in 1979. Treatment is an indicator for women having three or more children, and the instrument is an indicator for women whose first two children have the same sex. The authors control for age, age at the first birth, race and sex of the first two children as covariates.

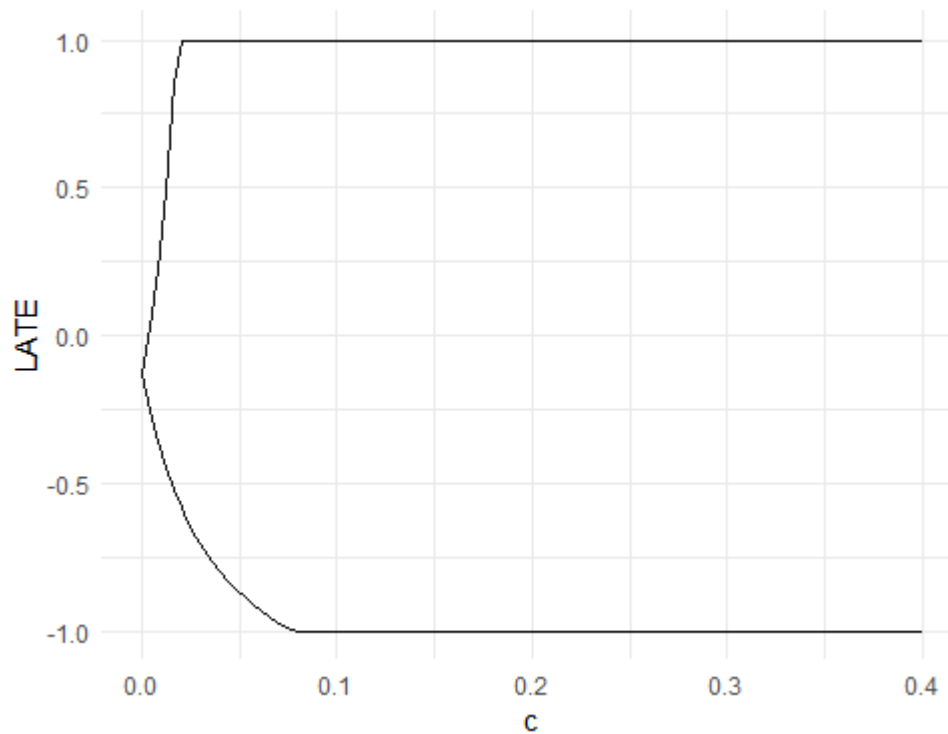
To begin the sensitivity analysis, I use selection on observables to calibrate the beliefs regarding the amount of selection on unobservables. I take the approach from (ALTONJI; ELDER; TABER, 2008) and (MASTEN; POIRIER, 2018). I partition the vector of covariates X as (X_k, X_{-k}) , where X_k is the k -th component and X_{-k} is a vector with remaining components. The measures used to calibrate the beliefs regarding deviations from independence are

$$\bar{c}_k = \sup_{x_{-k}} \sup_{x_k} |\mathbb{P}(Z = 1 | X = (x_k, x_{-k})) - \mathbb{P}(Z = 1 | X_{-k} = x_{-k})|$$

In the data, the largest value obtained from \bar{c}_k is associated to the indicator for women whose second child is a man, which was estimated to be $\bar{c}_{2nd \text{ sex}} = 0.011$.

Picture 3 shows the identified sets for the share of compliers and the ITT in the application. The share of compliers in the case where $c_1 = 0$ is approximately 0.060 and the estimated breakdown point for the conclusion that the share of compliers is greater than zero is $\hat{c}_1^* = 0.037$.

Figura 4 – Identified-set for the effects of family size on unemployment



Note: Figure 4 show the identified-set for the LATE in the Angrist & Evans (1998) setting under different values of c .

The identified set for the ITT is displayed on the left. If point identification holds, then the ITT is equal to -0.008. However, the identified set is uninformative for most of the values of c_2 . The estimated breakdown point for the conclusion that treatment effects are negative is $\hat{c}_2^* = 0.004$, and it is not statistically different from zero.

Figure 4 shows the identified set for the LATE under joint c -dependence. In the case of full independence of the instrument ($c = 0$), the LATE is equal to -0.132. The estimated breakdown point for the conclusion that the LATE is negative is $\hat{c}^* = 0.004$, and it is not statistically different from zero.

Overall, the results from the breakdown analysis suggest that the conclusions regarding the effects of family size on unemployment using same-sex siblings as the instrument are not robust to violations of independence of the instrument.

3.10 Conclusion

In this paper I discuss the partial identification of treatment effects in IV settings with binary outcomes under violations of independence. I derive identified sets for the first-stage, the reduced form and the LATE. Building on this result, I identify breakdown values for conclusions regarding these parameters. I derive the asymptotic properties for the estimators of the bounds and the breakdown values. I also derive sharp bounds for the LATE under a joint c -dependence assumption.

Monte Carlo simulations show the desirable properties of the estimators for the breakdown points and the bootstrap procedure for constructing one-sided confidence intervals. This is still a work in progress. In the empirical application I study the effects of family size on female employment and find that weak conclusions about the share of compliers to the same-sex sibling instrument and treatment effects are highly sensitive to relaxations of random assignment of the instrument.

Referências

- ABBRING, J. H.; HECKMAN, J. J. Chapter 72 econometric evaluation of social programs, part iii: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation. In: HECKMAN, J. J.; LEAMER, E. E. (Ed.). Elsevier, 2007, (Handbook of Econometrics, v. 6). p. 5145–5303. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1573441207060722>>.
- ACEMOGLU, D. et al. Democracy does cause growth. **Journal of Political Economy**, v. 127, n. 1, p. 47–100, 2019. Disponível em: <<https://doi.org/10.1086/700936>>.
- AGHION, P. et al. Innovation and Top Income Inequality. **The Review of Economic Studies**, v. 86, n. 1, p. 1–45, 06 2018. ISSN 0034-6527. Disponível em: <<https://doi.org/10.1093/restud/rdy027>>.
- ALTONJI, J. G.; ELDER, T. E.; TABER, C. R. Using selection on observed variables to assess bias from unobservables when evaluating swan-ganz catheterization. **American Economic Review**, v. 98, n. 2, p. 345–50, May 2008. Disponível em: <<https://www.aeaweb.org/articles?id=10.1257/aer.98.2.345>>.
- ANGRIST, J. D.; EVANS, W. N. Children and their parents' labor supply: Evidence from exogenous variation in family size. **The American Economic Review**, American Economic Association, v. 88, n. 3, p. 450–477, 1998. ISSN 00028282. Disponível em: <<http://www.jstor.org/stable/116844>>.
- ARELLANO, M.; BOND, S. Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. **The Review of Economic Studies**, v. 58, n. 2, p. 277–297, 04 1991. ISSN 0034-6527. Disponível em: <<https://doi.org/10.2307/2297968>>.
- ARKHANGELSKY, D.; IMBENS, G. W. Doubly robust identification for causal panel data models. **The Econometrics Journal**, v. 25, n. 3, p. 649–674, 06 2022. ISSN 1368-4221. Disponível em: <<https://doi.org/10.1093/ectj/utac019>>.
- ARKHANGELSKY, D. et al. Double-robust two-way-fixed-effects regression for panel data. **Working Paper**, 2021.
- ASSUNÇÃO, J.; GANDOUR, C.; ROCHA, R. Deter-ing deforestation in the amazon: Environmental monitoring and law enforcement. **American Economic Journal: Applied Economics**, v. 15, n. 2, p. 125–56, April 2023. Disponível em: <<https://www.aeaweb.org/articles?id=10.1257/app.20200196>>.
- BHULLER, M.; SIGSTAD, H. 2sls with multiple treatments. **Journal of Econometrics**, v. 242, n. 1, p. 105785, 2024. ISSN 0304-4076. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0304407624001313>>.
- BJÖRKLUND, A.; MOFFITT, R. The Estimation of Wage Gains and Welfare Gains in Self-Selection Models. **The Review of Economics and Statistics**, v. 69, n. 1, p. 42–49, 1987.
- BLANCHARD, E.; MATSCHKE, X. U.S. Multinationals and Preferential Market Access. **The Review of Economics and Statistics**, v. 97, n. 4, p. 839–854, 10 2015. ISSN 0034-6535. Disponível em: <https://doi.org/10.1162/REST_a_00496>.

BLANDHOL, C. et al. **When is TSLs Actually LATE?** [S.l.], 2022. (Working Paper Series, 29709). Disponível em: <<http://www.nber.org/papers/w29709>>.

BLOOM, H. S. Accounting for no-shows in experimental evaluation designs. **Evaluation Review**, v. 8, n. 2, p. 225–246, 1984. Disponível em: <<https://doi.org/10.1177/0193841X8400800205>>.

BOJINOV, I.; RAMBACHAN, A.; SHEPHARD, N. Panel experiments and dynamic causal effects: A finite population perspective. **Quantitative Economics**, v. 12, n. 4, p. 1171–1196, 2021. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.3982/QE1744>>.

BOJINOV, I.; SHEPHARD, N. Time series experiments and causal estimands: Exact randomization tests and trading. **Journal of the American Statistical Association**, Taylor Francis, v. 114, n. 528, p. 1665–1682, 2019. Disponível em: <<https://doi.org/10.1080/01621459.2018.1527225>>.

BONHOMME, S.; SAUDER, U. Recovering distributions in difference-in-differences models: A comparison of selective and comprehensive schooling. **Review of Economic and Statistics**, v. 93, n. 2, p. 479–494, 2011.

BORUSYAK, K.; HULL, P. Negative weights are no concern in design-based specifications. **Working Paper**, 2024.

BORUSYAK, K.; JARAVEL, X.; SPIESS, J. Revisiting event study designs: Robust and efficient estimation. **Working Paper**, 2021.

BRINCH, C.; MOGSTAD, M.; WISWALL, M. Beyond LATE with a discrete instrument. **Journal of Political Economy**, v. 125, n. 4, p. 985–1037, 2017.

CALLAWAY, B.; GOODMAN-BACON, A.; SANT'ANNA, P. H. C. **Difference-in-Differences with a Continuous Treatment**. arXiv, 2021. Disponível em: <<https://arxiv.org/abs/2107.02637>>.

CALLAWAY, B.; SANT'ANNA, P. Difference-in-Differences with multiple time periods. **Journal of Econometrics**, v. 225, n. 2, p. 200–230, 2021.

CARNEIRO, P.; HECKMAN, J.; VYTLACIL, E. Estimating Marginal Returns to Education. **American Economic Review**, v. 101, n. 1, p. 2754–2781, 2011.

CHAISEMARTIN, C. D.; D'HAULTEFOEUILLE, X. Fuzzy Difference-in-Differences. **Review of Economic Studies**, v. 85, n. 1, p. 995–1028, 2018.

CHAISEMARTIN, C. de; LEI, Z. More robust estimators for instrumental-variable panel designs, with an application to the effect of imports from china on us employment. **Working Paper**, 2023.

CHESHER, A.; ROSEN, A. M. What do instrumental variable models deliver with discrete dependent variables? **American Economic Review**, v. 103, n. 3, p. 557–62, May 2013. Disponível em: <<https://www.aeaweb.org/articles?id=10.1257/aer.103.3.557>>.

CINELLI, C.; HAZLETT, C. An omitted variable bias framework for sensitivity analysis of instrumental variables. **Biometrika**, p. asaf004, 01 2025. ISSN 1464-3510. Disponível em: <<https://doi.org/10.1093/biomet/asaf004>>.

CONLEY, T. G.; HANSEN, C. B.; ROSSI, P. E. Plausibly exogenous. **The Review of Economics and Statistics**, v. 94, n. 1, p. 260–272, 02 2012. ISSN 0034-6535. Disponível em: <https://doi.org/10.1162/REST_v_a_00139>.

CORNELISSEN, T. et al. From LATE to MTE: Alternative methods for the evaluation of policy interventions. **Labour Economics**, v. 41, n. 1, p. 47–60, 2016.

CORNELISSEN, T. et al. Who benefits from universal child care? estimating marginal returns to early child care attendance. **Journal of Political Economy**, v. 126, n. 6, p. 2356–2409, 2018.

DUFLO, E. Schooling and labor market consequences of school construction in indonesia: Evidence from an unusual policy experiment. **American Economic Review**, v. 91, n. 1, p. 795–811, 2001.

ESCANCIANO, J. C.; PEREZ-IZQUIERDO, T. Automatic locally robust estimation with generated regressors. **Working paper**, 2023.

FANG, Z.; SANTOS, A. Inference on directionally differentiable functions. **The Review of Economic Studies**, v. 86, n. 1, p. 377–412, 09 2018. ISSN 0034-6527. Disponível em: <<https://doi.org/10.1093/restud/rdy049>>.

FERRARA, E. L.; CHONG, A.; DURYEA, S. Soap operas and fertility: Evidence from brazil. **American Economic Journal: Applied Economics**, v. 4, n. 4, p. 1–31, 2012.

GHANEM, D.; SANT'ANNA, P. H. C.; WÜTHRICH, K. **Selection and parallel trends**. 2023.

GOODMAN-BACON, A. Difference-in-differences with variation in treatment timing. **Journal of Econometrics**, v. 225, n. 2, p. 254–277, 2021.

GREENSTONE, M.; HANNA, R. Environmental regulations, air and water pollution, and infant mortality in india. **American Economic Review**, v. 104, n. 10, p. 3038–72, October 2014. Disponível em: <<https://www.aeaweb.org/articles?id=10.1257/aer.104.10.3038>>.

HAHN, J. On the role of the propensity score in efficient semiparametric estimation of average treatment effects. **Econometrica**, [Wiley, Econometric Society], v. 66, n. 2, p. 315–331, 1998. ISSN 00129682, 14680262. Disponível em: <<http://www.jstor.org/stable/2998560>>.

HAN, S. Identification in nonparametric models for dynamic treatment effects. **Journal of Econometrics**, v. 225, n. 2, p. 132–147, 2021. ISSN 0304-4076. Themed Issue: Treatment Effect 1. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0304407620303717>>.

HAN, S. Optimal dynamic treatment regimes and partial welfare ordering. **Journal of the American Statistical Association**, Taylor Francis, v. 0, n. 0, p. 1–11, 2023.

HANSEN, M. C. et al. Humid tropical forest clearing from 2000 to 2005 quantified by using multitemporal and multiresolution remotely sensed data. **Proceedings of the National Academy of Sciences**, v. 105, n. 27, p. 9439–9444, 2008. Disponível em: <<https://www.pnas.org/doi/abs/10.1073/pnas.0804042105>>.

HECKMAN, J.; ROBB, R. Identification of Causal Effects Using Instrumental Variables. **Journal of Econometrics**, v. 30, n. 1, p. 239–267, 1985.

HECKMAN, J.; VYTLACIL, E. Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects. **Proceedings of the National Academy of Sciences of the United States of America**, v. 96, n. 8, p. 4730–4734, 1999.

- HECKMAN, J.; VYTLACIL, E. Policy-Relevant Treatment Effects. **American Economic Review**, v. 91, n. 2, p. 107–111, 2001.
- HECKMAN, J.; VYTLACIL, E. Structural Equations, Treatment Effects, and Econometric Policy Evaluation. **Econometrica**, v. 73, n. 3, p. 669–738, 2005.
- HECKMAN, J.; VYTLACIL, E. Chapter 71 econometric evaluation of social programs, part ii: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. In: HECKMAN, J. J.; LEAMER, E. E. (Ed.). [S.l.]: Elsevier, 2007, (Handbook of Econometrics, v. 6). p. 4875–5143.
- HECKMAN, J. J. Dummy endogenous variables in a simultaneous equation system. **Econometrica**, [Wiley, Econometric Society], v. 46, n. 4, p. 931–959, 1978. ISSN 00129682, 14680262. Disponível em: <<http://www.jstor.org/stable/1909757>>.
- HECKMAN, J. J.; HUMPHRIES, J. E.; VERAMENDI, G. Dynamic treatment effects. **Journal of Econometrics**, v. 191, n. 2, p. 276–292, 2016. ISSN 0304-4076. Innovations in Measurement in Economics and Econometrics. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0304407615002778>>.
- HECKMAN, J. J.; NAVARRO, S. Dynamic discrete choice and dynamic treatment effects. **Journal of Econometrics**, v. 136, n. 2, p. 341–396, 2007. ISSN 0304-4076. The interface between econometrics and economic theory. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0304407605002186>>.
- HECKMAN, J. J.; URZUA, S.; VYTLACIL, E. Understanding instrumental variables in models with essential heterogeneity. **The Review of Economics and Statistics**, The MIT Press, v. 88, n. 3, p. 389–432, 2006. ISSN 00346535, 15309142. Disponível em: <<http://www.jstor.org/stable/40043006>>.
- HERNAN, M. A.; ROBINS, J. M. Causal inference: What if. In: ROUTLEGE (Ed.). [S.l.]: CRC Press, 2023. v. 1, p. 1–312.
- HIRANO, K.; IMBENS, G. W.; RIDDER, G. Efficient estimation of average treatment effects using the estimated propensity score. **Econometrica**, v. 71, n. 4, p. 1161–1189, 2003. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-0262.00442>>.
- HOROWITZ, J. L.; MANSKI, C. F. Identification and robustness with contaminated and corrupted data. **Econometrica**, [Wiley, Econometric Society], v. 63, n. 2, p. 281–302, 1995. ISSN 00129682, 14680262. Disponível em: <<http://www.jstor.org/stable/2951627>>.
- HUDSON, S.; HULL, P.; JACK, L. Interpreting instrumented difference-in-differences. **Working Paper**, 2017.
- IMBENS, G. W.; ANGRIST, J. D. Identification and estimation of local average treatment effects. **Econometrica**, [Wiley, Econometric Society], v. 62, n. 2, p. 467–475, 1994. ISSN 00129682, 14680262. Disponível em: <<http://www.jstor.org/stable/2951620>>.
- IMBENS, G. W.; RUBIN, D. B. Regular assignment mechanisms with noncompliance: Analysis. In: _____. **Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction**. [S.l.]: Cambridge University Press, 2015. p. 511–586.

- JACKSON, C. K.; JOHNSON, R. C.; PERSICO, C. The Effects of School Spending on Educational and Economic Outcomes: Evidence from School Finance Reforms *. **The Quarterly Journal of Economics**, v. 131, n. 1, p. 157–218, 10 2015. ISSN 0033-5533. Disponível em: <<https://doi.org/10.1093/qje/qjv036>>.
- KANG, H.; PECK, L.; KEELE, L. Inference for Instrumental Variables: A Randomization Inference Approach. **Journal of the Royal Statistical Society Series A: Statistics in Society**, v. 181, n. 4, p. 1231–1254, 02 2018. ISSN 0964-1998. Disponível em: <<https://doi.org/10.1111/rssa.12353>>.
- KIRKEBOEN, L. J.; LEUVEN, E.; MOGSTAD, M. Field of Study, Earnings, and Self-Selection*. **The Quarterly Journal of Economics**, v. 131, n. 3, p. 1057–1111, 05 2016. ISSN 0033-5533. Disponível em: <<https://doi.org/10.1093/qje/qjw019>>.
- KLINE, B.; MASTEN, M. Finite population identification and design-based sensitivity analysis. **Working Paper**, 2025.
- KLINE, P.; SANTOS, A. Sensitivity to missing data assumptions: Theory and an evaluation of the u.s. wage structure. **Quantitative Economics**, v. 4, n. 2, p. 231–267, 2013. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.3982/QE176>>.
- KOLESÁR, M. Estimation in an instrumental variables model with treatment effect heterogeneity. **Working Paper**, 2013.
- MACHADO, C.; SHAIKH, A. M.; VYTLACIL, E. J. Instrumental variables and the sign of the average treatment effect. **Journal of Econometrics**, v. 212, n. 2, p. 522–555, 2019. ISSN 0304-4076. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0304407619301381>>.
- MARX, P.; TAMER, E.; TANG, X. **Parallel Trends and Dynamic Choices**. 2023.
- MASTEN, M. A.; POIRIER, A. Identification of treatment effects under conditional partial independence. **Econometrica**, v. 86, n. 1, p. 317–351, 2018. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA14481>>.
- MASTEN, M. A.; POIRIER, A. Inference on breakdown frontiers. **Quantitative Economics**, v. 11, n. 1, p. 41–111, 2020. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.3982/QE1288>>.
- MASTEN, M. A.; POIRIER, A. Salvaging falsified instrumental variable models. **Econometrica**, v. 89, n. 3, p. 1449–1469, 2021. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA17969>>.
- MOGSTAD, M.; SANTOS, A.; TORGOVITSKY, A. Using instrumental variables for inference about policy relevant treatment parameters. **Econometrica**, v. 86, n. 5, p. 1589–1619, 2018.
- MOUNTJOY, J. Community colleges and upward mobility. **American Economic Review**, v. 112, n. 8, p. 2580–2630, August 2022. Disponível em: <<https://www.aeaweb.org/articles?id=10.1257/aer.20181756>>.
- MURPHY, S. A.; LAAN, M. J. van der; ROBINS, J. M. Marginal mean models for dynamic regimes. **Journal of the American Statistical Association**, Taylor Francis, v. 96, n. 456, p. 1410–1423, 2001.

OLSEN, R. J. A least square correction for selectivity bias. **Econometrica**, v. 48, n. 7, p. 1815–1820, 1980.

PHAM, T. T.; CHEN, W. The instrumental variable method for estimating local average treatment regime effects. **Working Paper**, 2017.

RAMBACHAN, A.; ROTH, J. Design-based uncertainty for quasi-experiments. **Working Paper**, 2025.

ROBINS, J. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. **Mathematical Modelling**, v. 7, n. 9, p. 1393–1512, 1986. ISSN 0270-0255. Disponível em: <<https://www.sciencedirect.com/science/article/pii/0270025586900886>>.

ROBINS, J. A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods. **Journal of Chronic Diseases**, v. 40, p. 139S–161S, 1987. ISSN 0021-9681. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0021968187800188>>.

ROBINS, J. M.; MARK, S. D.; NEWAY, W. K. Estimating exposure effects by modelling the expectation of exposure conditional on confounders. **Biometrics**, International Biometric Society, v. 48, n. 2, p. 479–495, 1992. ISSN 0006341X, 15410420. Disponível em: <<http://www.jstor.org/stable/2532304>>.

SANT'ANNA, P.; ZHAO, J. Doubly Robust Difference-in-Differences Estimators. **Journal of Econometrics**, v. 219, n. 1, p. 101–122, 2020.

SASAKI, Y.; URA, T. Estimation and inference for policy relevant treatment effects. **Journal of Econometrics**, v. 234, n. 2, p. 394–450, 2023.

SCHULER, A.; LEE, K. J.; HUBBARD, A. Bridging binarization: Causal inference with dichotomized continuous treatments. **Working Paper**, 2024.

SHAIKH, A. M.; VYTLACIL, E. J. Partial identification in triangular systems of equations with binary dependent variables. **Econometrica**, v. 79, n. 3, p. 949–955, 2011. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA9082>>.

SHEN, S.; CHOI, J.; SEONG, D. Panel instrumental variable regression models with varying-intensity repeated treatments: Theory and the china syndrome application. **Working Paper**, 2024.

SOTRA, R. B.; SYRGKANIS, V. Dynamic local average treatment effects. **Working Paper**, 2024.

STANGO, V.; ZINMAN, J. We Are All Behavioural, More, or Less: A Taxonomy of Consumer Decision-Making. **The Review of Economic Studies**, v. 90, n. 3, p. 1470–1498, 08 2022. ISSN 0034-6527. Disponível em: <<https://doi.org/10.1093/restud/rdac055>>.

VYTLACIL, E. Independence, monotonicity, and latent index models: An equivalence result. **Econometrica**, [Wiley, Econometric Society], v. 70, n. 1, p. 331–341, 2002. ISSN 00129682, 14680262. Disponível em: <<http://www.jstor.org/stable/2692171>>.

VYTLACIL, E.; YILDIZ, N. Dummy endogenous variables in weakly separable models. **Econometrica**, v. 75, n. 3, p. 757–779, 2007. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-0262.2007.00767.x>>.

WALD, A. The fitting of straight lines if both variables are subject to error. **The Annals of Mathematical Statistics**, Institute of Mathematical Statistics, v. 11, n. 3, p. 284–300, 1940. ISSN 00034851. Disponível em: <<http://www.jstor.org/stable/2235677>>.

WANG, X. et al. Sensitivity analysis and power for instrumental variable studies. **Biometrics**, v. 74, n. 4, p. 1150–1160, 2018. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/biom.12873>>.

4 Appendix to Chapter 1

4.1 Appendix A

Proof of Theorem 1

Let's start with the first stage. Consider a shift of the value of the instrument from 0 to z . Under Assumptions 1 and 3-6,

$$\begin{aligned}
& \mathbb{E} \left[\frac{D_i T_i}{\mathbb{E}[T_i]} - \frac{D_i(1-T_i)}{(1-\mathbb{E}[T_i])} \middle| Z_i = z \right] - \mathbb{E} \left[\frac{D_i T_i}{\mathbb{E}[T_i]} - \frac{D_i(1-T_i)}{(1-\mathbb{E}[T_i])} \middle| Z_i = 0 \right] \\
&= \mathbb{P}(D_1 = 1 | Z_i = z) - \mathbb{P}(D_0 = 1 | Z_i = z) - (\mathbb{P}(D_1 = 1 | Z_i = 0) - \mathbb{P}(D_0 = 1 | Z_i = 0)) \\
&= \mathbb{P}(V_1 \leq \mu(z) | Z_i = z) - \mathbb{P}(V_0 \leq \mu(0) | Z_i = z) - (\mathbb{P}(V_1 \leq \mu(0) | Z_i = 0) - \mathbb{P}(V_0 \leq \mu(0) | Z_i = 0)) \\
&= \mathbb{P}(V_1 \leq \mu(0) | Z_i = z) + \mathbb{P}(\mu(0) < V_1 \leq \mu(z) | Z_i = z) - \mathbb{P}(V_0 \leq \mu(0) | Z_i = z) \\
&\quad - (\mathbb{P}(V_1 \leq \mu(0) | Z_i = 0) - \mathbb{P}(V_0 \leq \mu(0) | Z_i = 0)) \\
&= \mathbb{P}(V_0 \leq \mu(0) | Z_i = z) + \mathbb{P}(V_1 \leq \mu(0) < V_0 | Z_i = z) + \mathbb{P}(\mu(0) < V_1 \leq \mu(z) < V_0 | Z_i = z) \\
&\quad - \mathbb{P}(V_0 \leq \mu(0) | Z_i = z) - (\mathbb{P}(V_0 \leq \mu(0) | Z_i = 0) + \mathbb{P}(V_1 \leq \mu(0) < V_0 | Z_i = 0) - \mathbb{P}(V_0 \leq \mu(0) | Z_i = 0)) \\
&= \mathbb{P}(V_1 \leq \mu(0) < V_0) + \mathbb{P}(\mu(0) < V_1 \leq \mu(z) < V_0) - \mathbb{P}(V_1 \leq \mu(0) < V_0) \\
&= \mathbb{P}(\mu(0) < V_1 \leq \mu(z) < V_0)
\end{aligned}$$

Now consider the reduced form estimand associated to a shift of the value of the instrument from 0 to z . Under Assumptions 1-6,

$$\begin{aligned}
& \mathbb{E} \left[\frac{Y_i T_i}{\mathbb{E}[T_i]} - \frac{Y_i(1-T_i)}{(1-\mathbb{E}[T_i])} \middle| Z_i = z \right] - \mathbb{E} \left[\frac{Y_i T_i}{\mathbb{E}[T_i]} - \frac{Y_i(1-T_i)}{(1-\mathbb{E}[T_i])} \middle| Z_i = 0 \right] \\
&= \mathbb{E}[Y_1 - Y_0 | Z_i = z] - \mathbb{E}[Y_1 - Y_0 | Z_i = 0] \\
&= \mathbb{E}[Y_1(D_1(z)) - Y_0(D_0(0)) | Z_i = z] - \mathbb{E}[Y_1(D_1(0)) - Y_0(D_0(0)) | Z_i = 0] \\
&= \mathbb{E}[Y_1(1) - Y_1(0) | Z_i = z, D_1(z) > D_1(0), D_0(z) = 0] \mathbb{P}(D_1(z) > D_1(0), D_0(z) = 0 | Z_i = z) \\
&= \mathbb{E}[Y_1(1) - Y_1(0) | \mu(0) < V_1 \leq \mu(z) < V_0] \mathbb{P}(\mu(0) < V_1 \leq \mu(z) < V_0)
\end{aligned}$$

Hence, it follows that

$$\frac{\mathbb{E} \left[\frac{Y_i T_i}{\mathbb{E}[T_i]} - \frac{Y_i(1-T_i)}{(1-\mathbb{E}[T_i])} \middle| Z_i = z \right] - \mathbb{E} \left[\frac{Y_i T_i}{\mathbb{E}[T_i]} - \frac{Y_i(1-T_i)}{(1-\mathbb{E}[T_i])} \middle| Z_i = 0 \right]}{\mathbb{E} \left[\frac{D_i T_i}{\mathbb{E}[T_i]} - \frac{D_i(1-T_i)}{(1-\mathbb{E}[T_i])} \middle| Z_i = z \right] - \mathbb{E} \left[\frac{D_i T_i}{\mathbb{E}[T_i]} - \frac{D_i(1-T_i)}{(1-\mathbb{E}[T_i])} \middle| Z_i = 0 \right]} = \mathbb{E}[Y_1(1) - Y_1(0) | \mu(0) < V_1 \leq \mu(z) < V_0]$$

Taking the limit of the expression with z approaching zero as in (HECKMAN; VYTLACIL, 1999) yields

$$\frac{\partial \mathbb{E} \left[\frac{Y_i T_i}{\mathbb{E}[T_i]} - \frac{Y_i(1-T_i)}{(1-\mathbb{E}[T_i])} \middle| Z_i = z \right] / \partial z}{\partial \mathbb{E} \left[\frac{D_i T_i}{\mathbb{E}[T_i]} - \frac{D_i(1-T_i)}{(1-\mathbb{E}[T_i])} \middle| Z_i = z \right] / \partial z} = \mathbb{E} [Y_1(1) - Y_1(0) | V_1 = \mu(z) < V_0]$$

which concludes the proof.

Proof of Corollary 1

Define the cohort-specific propensity score as $\mathbb{P}(D_t = 1 | X_i = x, Z_i = z) := \pi_t(x, z)$.

Then the limit form of the first-stage identifies

$$\begin{aligned} \partial \mathbb{E} \left[\frac{D_i T_i}{\mathbb{E}[T_i]} - \frac{D_i(1-T_i)}{(1-\mathbb{E}[T_i])} \middle| Z_i = z \right] / \partial z &= \partial (\pi_1(z, x) - \pi_0(z, x)) / \partial z \\ &= \partial (\pi_1(z, x) - \pi_0(0, x)) / \partial z = \frac{\partial \pi_1(z, x)}{\partial z} \end{aligned}$$

Under the separability assumption, the conditional expectation of Y_{it} given $X_i = x$ and $Z_i = z$ can be written as

$$\begin{aligned} \mathbb{E}[Y_t | X_i = x, Z_i = z] &= x' \beta_{0t} + x' (\beta_{1t} - \beta_{0t}) \pi_t(z, x) + \mathbb{E}[U_{it} | Z_i = z] \\ &= x' \beta_{0t} + x' (\beta_{1t} - \beta_{0t}) \pi_t(z, x) + K(\pi_t(z, x)) \end{aligned}$$

where $U_{it} = U_{i1t} D_{it} + U_{i0t} (1 - D_{it})$.

Hence, it follows that under Assumptions 1-7,

$$\begin{aligned} \mathbb{E} \left[\frac{Y_i T_i}{\mathbb{E}[T_i]} - \frac{Y_i(1-T_i)}{(1-\mathbb{E}[T_i])} \middle| X_i = x, Z_i = z \right] &= x' (\beta_{01} - \beta_{00}) + x' (\beta_{11} - \beta_{01}) \pi_1(z, x) + K(\pi_1(z, x)) \\ &\quad - x' (\beta_{10} - \beta_{00}) \pi_0(z, x) - K(\pi_0(z, x)) \\ &= x' (\beta_{01} - \beta_{00}) + x' (\beta_{11} - \beta_{01}) \pi_1(z, x) + K(\pi_1(z, x)) - x' (\beta_{10} - \beta_{00}) \pi_0(0, x) - K(\pi_0(0, x)) \end{aligned}$$

and thus,

$$\partial \mathbb{E} \left[\frac{Y_i T_i}{\mathbb{E}[T_i]} - \frac{Y_i(1-T_i)}{(1-\mathbb{E}[T_i])} \middle| Z_i = z \right] / \partial z = x' (\beta_{11} - \beta_{01}) \frac{\partial \pi_1(z, x)}{\partial z} + \frac{\partial K(\pi_1(z, x))}{\partial z} \frac{\partial \pi_1(z, x)}{\partial z}$$

from which we obtain

$$\frac{\partial \mathbb{E} \left[\frac{Y_i T_i}{\mathbb{E}[T_i]} - \frac{Y_i(1-T_i)}{(1-\mathbb{E}[T_i])} \middle| Z_i = z \right] / \partial z}{\partial \mathbb{E} \left[\frac{D_i T_i}{\mathbb{E}[T_i]} - \frac{D_i(1-T_i)}{(1-\mathbb{E}[T_i])} \middle| Z_i = z \right] / \partial z} = x' (\beta_{11} - \beta_{01}) + \frac{\partial K(\pi_1(z, x))}{\partial z}$$

which concludes the proof.

Proof of Theorem 2

We begin by stating the additional assumptions required for the result to hold:

Assumption 8 (Independent and Identically Distributed Observations): The vector of observations $\{Y_i, D_i, Z_i, T_i, X_i\}_{i=1}^N$ is i.i.d as (Y, D, Z, T, X) and $\mathbb{V}(D_t | X = x, Z = z)$ is bounded on the compact support of (Z, X) , defined as $\mathcal{S}(Z, X)$.

Assumption 9 (Regularity Conditions for the First-Stage Sieve Estimator): For every K there is a nonsingular matrix of constants B_D such that for $P_D^K(z) = B_D p_D^K(z)$,

- The smallest eigenvalue of $\mathbb{E} [P_D^K(Z_i) W P_D^K(Z_i)']$ is bounded away from zero uniformly in K .
- There exists a sequence of constants $\zeta_D(K)$ that satisfies $\sup_{z \in \mathcal{S}(Z)} \|P_D^K(z)\| \leq \zeta_D(K)$ with K such that $\zeta_D(K)^2/n \rightarrow 0$ as $n \rightarrow \infty$.
- There exists $\alpha_D > 0$ such that

$$\sup_{z \in \mathcal{S}(Z)} \left| g_D(z) - P_D^K(z)' \gamma_D \right| = O(K^{-\alpha_D})$$

$$\sup_{z \in \mathcal{S}(Z)} \left| \frac{\partial g_D(z)}{\partial z} - \nabla P_D^K(z)' \gamma_D \right| = O(K^{-\alpha_D})$$

- As $n \rightarrow \infty$, $K \rightarrow \infty$ and $K/n \rightarrow 0$
- $K^{-\alpha_D} = o((K/n)^{1/2})$

Moreover, $\mathbb{E} \left[(D_t - x' \beta_{Dt} - g_{Dt}(z))^4 \middle| X = x, Z = z \right]$ is bounded and $\mathbb{V}(D_t - x' \beta_{Dt} | X = x, Z = z)$ is bounded away from zero.

Assumption 10 (Regularity Conditions for the Reduced Form Sieve Estimator): For every K there is a nonsingular matrix of constants B_Y such that for $P_Y^K(z) = B_Y p_Y^K(z)$,

- The smallest eigenvalue of $\mathbb{E} [P_Y^K(Z_i) W P_Y^K(Z_i)']$ is bounded away from zero uniformly in K .
- There exists a sequence of constants $\zeta_Y(K)$ that satisfies $\sup_{z \in \mathcal{S}(Z)} \|P_Y^K(z)\| \leq \zeta_Y(K)$ with K such that $\zeta_Y(K)^2/n \rightarrow 0$ as $n \rightarrow \infty$.

- There exists $\alpha_Y > 0$ such that

$$\sup_{z \in \mathcal{S}(Z)} \left| g_Y(z) - P_Y^K(z)' \gamma_D \right| = O(K^{-\alpha_Y})$$

$$\sup_{z \in \mathcal{S}(Z)} \left| \frac{\partial g_Y(z)}{\partial z} - \nabla P_6^K(z)' \gamma_Y \right| = O(K^{-\alpha_Y})$$

- As $n \rightarrow \infty$, $K \rightarrow \infty$ and $K/n \rightarrow 0$
- $K^{-\alpha_Y} = o((K/n)^{1/2})$

Moreover, $\mathbb{E} \left[(Y_t - x' \alpha - x' \delta (x' \beta_D + P_D^K(z)' \gamma_D) - g_{Yt}(z))^4 | X = x, Z = z \right]$ is bounded and $\mathbb{V} (Y_t - x' \alpha - x' \delta (x' \beta_D + P_D^K(z)' \gamma_D) | X = x, Z = z)$ is bounded away from zero.

We begin with the asymptotic properties of the estimator for the first-stage. Under Assumptions 1-9, it follows from Donald & Newey(1994), Newey (1997) and Li & Racine (2007) that

$$\sqrt{N} \left(\partial \widehat{\mathbb{E}} \left[\frac{D_i T_i}{\mathbb{E}[T_i]} - \frac{D_i(1 - T_i)}{(1 - \mathbb{E}[T_i])} | X_i = x, Z_i = z \right] / \partial z - \frac{\partial \pi_1(z, x)}{\partial z} \right) \xrightarrow{d} N(0, V_D)$$

where

$$V_D = N^{-1} \nabla P_D^K(z)' Q_D^{-1} \Sigma_D Q_D^{-1} \nabla P_D^K(z)$$

with

$$Q_D = \mathbb{E} \left[P_D^D(Z_i) W P_D^K(Z_i)' \right]$$

$$\Sigma_D = \mathbb{E} \left[P_D^D(Z_i) W P_D^K(Z_i)' \sigma_D^2(X_i, Z_i) \right]$$

$$\sigma_D^2(X_i, Z_i) = \mathbb{E} \left[Y_{it} - x' \beta_{Dt} - g_{Dt}(z) | X_i = x, Z_i = z \right]$$

Now, let's consider the reduced form estimator, similarly, under Assumptions 1-9, it follows from Donald & Newey(1994), Newey (1997) and Li & Racine (2007) that

$$\sqrt{N} \left(\partial \widehat{\mathbb{E}} \left[\frac{Y_i T_i}{\mathbb{E}[T_i]} - \frac{Y_i(1 - T_i)}{(1 - \mathbb{E}[T_i])} | X_i = x, Z_i = z \right] / \partial z - \partial \mathbb{E} \left[\frac{Y_i T_i}{\mathbb{E}[T_i]} - \frac{Y_i(1 - T_i)}{(1 - \mathbb{E}[T_i])} | Z_i = z \right] / \partial z \right) \xrightarrow{d} N(0, V_Y)$$

with $V_Y = V_\delta + V_{\gamma_Y}$, with V_δ and V_{γ_Y} defined analogously to V_D .

Hence, we conclude that

$$\sqrt{N} \left(\widehat{MTE}(\pi_1(z, x)) - MTE(\pi_1(z, x)) \right) \xrightarrow{d} N(0, V_{MTE(\pi_1(z, x))})$$

where $V_{MTE(\pi_1(z, x))}$ is obtained by the Delta Method.

Proof of Theorem 3

We conduct the proof using the notation previously established in Section 4

Given an estimate \hat{m}_{ij} and an estimate $\hat{P}_D(G, T, X)$ we consider the following RIPW estimator:

$$\begin{aligned}\hat{\tau}(\pi) &= \underset{\mu, \tau}{\operatorname{argmin}} \sum_{i=1}^N ((Y_i - \hat{m}_{ij}) - \mu - \tau D_i)^2 \Phi_i \\ &= \underset{\mu, \tau}{\operatorname{argmin}} \sum_{i=1}^N (\tilde{Y}_i - \mu - \tau D_i)^2 \Phi_i\end{aligned}$$

The first order conditions with respect to μ and τ are respectively

$$\sum_{i=1}^N \Phi_i (\tilde{Y}_i - \mu - \tau D_i) = 0$$

and

$$\sum_{i=1}^N \Phi_i D_i (\tilde{Y}_i - \mu - \tau D_i) = 0$$

From the first order conditions we obtain the optimal $\hat{\mu}$ as function of τ :

$$\hat{\mu}(\tau) = \Gamma_{\Phi}^{-1}(\Gamma_Y - \tau \Gamma_D)$$

Substitute $\hat{\mu}(\tau)$ into the first order condition with relation to τ and after some tedious algebraic manipulation we obtain

$$\hat{\tau}(\pi) = \frac{\Gamma_{DY} - \Gamma_{\Phi}^{-1}(\Gamma_D \Gamma_Y)}{\Gamma_{DD} - \Gamma_{\Phi}^{-1} \Gamma_D^2} \equiv \frac{\mathcal{N}}{\mathcal{D}}$$

Proof of Theorem 4

We conduct the proof using the notation introduced in Section 4. We begin with the asymptotic expansion of the RIPW estimator. First, note that $\mathcal{D}(\hat{\tau} - \tau) = \mathcal{N} - \tau \cdot \mathcal{D}$. Assumptions 2 and 12 imply that Lemma A.2 from (ARKHANGELSKY et al., 2021) holds. Hence,

$$\begin{aligned}& |(\Gamma_{DY} - \mathbb{E}[\Gamma_{DY}])(\Gamma_{\Phi} - \mathbb{E}[\Gamma_{\Phi}])| + |(\Gamma_D - \mathbb{E}[\Gamma_D])(\Gamma_Y - \mathbb{E}[\Gamma_Y])| \\ &= O_p(\operatorname{Var}(\Gamma_{DY}) + \operatorname{Var}(\Gamma_{\Phi}) + \operatorname{Var}(\Gamma_D) + \operatorname{Var}(\Gamma_Y)) = O_p(n^{-q})\end{aligned}$$

Let

$$\mathcal{V}_{i1} = \Phi_i \left\{ \mathbb{E} [\Gamma_{DY}] - \mathbb{E} [\Gamma_Y] D_i + \mathbb{E} [\Gamma_\Phi] D_i \tilde{Y}_i - \mathbb{E} [\Gamma_D] \tilde{Y}_i \right\}$$

and

$$\mathcal{V}_{i2} = \Phi_i \left\{ \mathbb{E} [\Gamma_{DD}] - \mathbb{E} [\Gamma_D] D_i + \mathbb{E} [\Gamma_\Phi] D_i^2 - \mathbb{E} [\Gamma_D] D_i \right\}$$

Then, we can write.

$$\mathcal{N} = \mathbb{E} [\Gamma_{DY}] \mathbb{E} [\Gamma_\Phi] - \mathbb{E} [\Gamma_D] \mathbb{E} [\Gamma_Y] + \frac{1}{n} \sum_{i=1}^N (\mathcal{V}_{i1} - \mathbb{E} [\mathcal{V}_{i1}]) + O_p(N^{-q})$$

and

$$\mathcal{D} = \mathbb{E} [\Gamma_{DD}] \mathbb{E} [\Gamma_\Phi] - \mathbb{E} [\Gamma_D]^2 + \frac{1}{N} \sum_{i=1}^N (\mathcal{V}_{i2} - \mathbb{E} [\mathcal{V}_{i2}]) + O_p(N^{-q})$$

Since $\mathcal{V}_i = \mathcal{V}_{i1} - \tau \cdot \mathcal{V}_{i2}$, it follows that

$$\mathcal{D}(\hat{\tau} - \tau) = \mathcal{N} - \tau \cdot \mathcal{D} = N_* + \frac{1}{N} \sum_{i=1}^N (\mathcal{V}_i - \mathbb{E} [\mathcal{V}_i]) + O_p(N^{-q})$$

where

$$N_* = \frac{1}{2n} \sum_{i=1}^N \mathbb{E} [\mathcal{V}_i] = \mathbb{E} [\Gamma_{DY}] \mathbb{E} [\Gamma_\Phi] - \mathbb{E} [\Gamma_D] \mathbb{E} [\Gamma_Y] - \tau (\mathbb{E} [\Gamma_{DD}] \mathbb{E} [\Gamma_\Phi] - \mathbb{E} [\Gamma_D]^2)$$

Furthermore, note that Assumptions 2 and 12 also imply that Lemma A.3 from (ARKHANGELSKY et al., 2021) hold, from which it follows that $1/\mathcal{D} = O_p(1)$. Define a constant $c_\pi > 0$ satisfying Assumption 2, then

$$\begin{aligned} & \frac{1}{4} \text{Var}(\mathcal{V}_{i1}) \\ \leq & \text{Var}(\Phi_i \mathbb{E} [\Gamma_{DY}]) + \text{Var}(\Phi_i \mathbb{E} [\Gamma_\Phi] D_i \tilde{Y}_i) + \text{Var}(\Phi_i \mathbb{E} [\Gamma_D] \tilde{Y}_i) + \text{Var}(\Phi_i \mathbb{E} [\Gamma_Y] D_i) \\ \leq & \mathbb{E} [\Phi_i \mathbb{E} [\Gamma_{DY}]^2] + \mathbb{E} [\Phi_i \mathbb{E} [\Gamma_\Phi]^2 D_i^2 \tilde{Y}_i^2] + \mathbb{E} [\Phi_i \mathbb{E} [\Gamma_D]^2 \tilde{Y}_i^2] + \mathbb{E} [\Phi_i \mathbb{E} [\Gamma_Y]^2 D_i^2] \\ \leq & 1/c_\pi^2 \left\{ \mathbb{E} [\Gamma_{DY}]^2 + \mathbb{E} [\Gamma_\Phi]^2 \mathbb{E} [D_i \tilde{Y}_i]^2 + \mathbb{E} [\mathbb{E} [\Gamma_{DY}] \tilde{Y}_i]^2 + \mathbb{E} [\mathbb{E} [\Gamma_Y] D_i]^2 \right\} \\ \leq & 1/c_\pi^2 \left\{ \mathbb{E} [\Gamma_{DY}]^2 + \mathbb{E} [\Gamma_\Phi]^2 \mathbb{E} [|\tilde{Y}_i|] + \left| \mathbb{E} [\Gamma_D] \mathbb{E} [\tilde{Y}_i] \right| + \left| \mathbb{E} [\Gamma_Y] \mathbb{E} [D_i] \right| \right\} \end{aligned}$$

The result follows the common support assumption and Cauchy-Schwarz inequality. By Lemma A.2 from (ARKHANGELSKY et al., 2021), we have for a generic constant C_1 that

$$Var(\mathcal{V}_{i1}) \leq C_1 \left(\mathbb{E} \left[\left\| \tilde{Y} \right\|^2 \right] + 1 \right) \leq C_1 \left(\mathbb{E} \left[\left\| \tilde{Y}(1) \right\|^2 \right] + \mathbb{E} \left[\left\| \tilde{Y}(0) \right\|^2 \right] + 1 \right)$$

Similarly, $Var(\mathcal{V}_{i2}) \leq C_2$ for a generic constant C_2 . Thus, it follows that

$$Var(\mathcal{V}_i) \leq 2Var(\mathcal{V}_{i1}) + 2\tau^2 \cdot Var(\mathcal{V}_{i2}) \leq C \left(\mathbb{E} \left[\left\| \tilde{Y}(1) \right\|^2 \right] + \mathbb{E} \left[\left\| \tilde{Y}(0) \right\|^2 \right] + 1 \right)$$

for a generic constant C which depends on c_π .

Lemma A.1 from (ARKHANGELSKY et al., 2021) implies that for any function f on the domain of \mathcal{Z}_i ,

$$Var \left[\frac{1}{N} \sum_{i=1}^N f(\mathcal{Z}_i) \right] \leq \frac{1}{2} \sum_{i=1}^N Var [f_i(\mathcal{Z}_i)] \rho_i$$

Combined with Assumption 3, the Lemma implies that

$$Var \left(\frac{1}{N} \sum_{i=1}^N \mathcal{V}_i \right) \leq \frac{1}{N^2} \sum_{i=1}^N Var(\mathcal{V}_i) \rho_i = o(1)$$

By Chebyshev's Inequality, we conclude that

$$\frac{1}{N} \sum_{i=1}^N (\mathcal{V}_i - \mathbb{E}[\mathcal{V}_i]) = o_p(1)$$

We have shown above that the asymptotic limit of $\mathcal{D}(\hat{\tau} - \tau)$ is N_* . We prove the double-robust property of $\hat{\tau}$ by showing that $\mathbb{E}[N_*] = 0$ when either of the nuisance parameters is correctly specified.

To do so, we assume for simplicity that $\tau = 0$, without loss of generality. Hence,

$$N_* = \mathbb{E}[\Gamma_{DY}] - \mathbb{E}[\Gamma_\Phi] - \mathbb{E}[\Gamma_D] \mathbb{E}[\Gamma_Y]$$

Note that $\tilde{Y} = \tilde{Y}(0) + \tau \cdot D$. Then,

$$\begin{aligned} \mathbb{E}[\Gamma_{DY}] &= \frac{1}{N} \sum_{i=1}^N \mathbb{E}[\Phi_i D_i \tilde{Y}_i] = \frac{1}{N} \sum_{i=1}^N \mathbb{E}[\Phi_i D_i (\tilde{Y}_i(0) + \tau \cdot D_i)] \\ &= \frac{1}{N} \sum_{i=1}^N \mathbb{E}[\Phi_i D_i] \mathbb{E}[\tilde{Y}(0)] + \frac{1}{N} \sum_{i=1}^N \mathbb{E}[\Phi_i D_i^2] \tau \end{aligned}$$

Analogously,

$$\mathbb{E}[\Gamma_Y] = \frac{1}{N} \sum_{i=1}^N \mathbb{E}[\Phi_i] \mathbb{E}[\tilde{Y}_i(0)] + \frac{1}{N} \sum_{i=1}^N \mathbb{E}[\Phi_i D_i] \tau$$

Combining the expressions above,

$$\begin{aligned} N_* &= \frac{1}{N} \sum_{i=1}^N \{ \mathbb{E}[\Phi_i D_i] \mathbb{E}[\Gamma_\Phi] - \mathbb{E}[\Phi_i] \mathbb{E}[\Gamma_D] \} \mathbb{E}[\tilde{Y}_i(0)] \\ &\quad + \frac{1}{N} \sum_{i=1}^N \{ \mathbb{E}[\Phi_i D_i^2] \mathbb{E}[\Gamma_\Phi] - \mathbb{E}[\Gamma_D] \mathbb{E}[\Phi_i D_i] \} \tau \end{aligned} \quad (4.1)$$

Consider the case in which the outcome model is correctly specified. In that case, by the definition of $\tilde{Y}(0)$ it follows that $\mathbb{E}[\tilde{Y}(0)] = 0$. Since $\tau = 0$, it follows naturally that $\mathbb{E}[N_*] = 0$.

Now consider the case in which the propensity score is correctly specified, $\mathbb{E}[\hat{P}] = P$. In that case,

$$\mathbb{E}[\Phi_i D_i] \equiv \mathbb{E}_{D \sim \pi}[D] = \mathbb{E}[\Gamma_D]$$

and

$$\mathbb{E}[\Phi_i] = 1 = \mathbb{E}[\Gamma_\Phi]$$

Moreover,

$$\mathbb{E}[\Phi_i D_i^2] = \mathbb{E}_{D \sim \pi}[D^2]$$

Then, we obtain

$$\mathbb{E}[\Phi_i D_i] \mathbb{E}[\Gamma_\Phi] - \mathbb{E}[\Phi_i] \mathbb{E}[\Gamma_D] = 0$$

Thus, we conclude that by Equation (9) and $\tau = 0$.

$$N_* = \frac{1}{N} \sum_{i=1}^N \mathbb{E}_{D \sim \pi} [(D - \mathbb{E}_{D \sim \pi}[D]) D] \tau$$

which has mean zero.

Inference for the RIPW Estimator

If we assume further that the bias from the nuisance parameters is $o(1/\sqrt{N})$, then Theorem 5 implies that

$$\mathcal{D}\sqrt{N}(\hat{\tau} - \tau) = \frac{1}{\sqrt{N}} \sum_{i=1}^N (\mathcal{V}_i - \mathbb{E}[\mathcal{V}_i]) + o_p(1)$$

from which it follows that

$$\frac{\mathcal{D}\sqrt{N}(\hat{\tau} - \tau)}{\sigma^*} \sim N(0, 1) \quad (4.2)$$

where $\sigma^{*2} = \frac{1}{N} \sum_{i=1}^N \text{Var}(\mathcal{V}_i)$.

We estimate the variance of $\frac{1}{\sqrt{N}} \sum_{i=1}^N (\mathcal{V}_i - \mathbb{E}[\mathcal{V}_i])$ as

$$\hat{\sigma}^{*2} = \frac{1}{N} \sum_{i=1}^N \left(\hat{\mathcal{V}}_i - \frac{1}{N} \sum_{i=1}^N \hat{\mathcal{V}}_i \right)^2 = \frac{N}{N-1} \sum_{i=1}^N \left\{ \frac{1}{N} \sum_{i=1}^N \hat{\mathcal{V}}_i^2 - \left(\frac{1}{N} \sum_{i=1}^N \hat{\mathcal{V}}_i \right)^2 \right\}$$

with $\hat{\mathcal{V}}_i$ as the plug-in estimate for \mathcal{V}_i .

This yields a Wald-type confidence interval for $\hat{\tau}$:

$$\hat{C}_{1-\alpha} = \left[\hat{\tau} - z_{1-\alpha/2} \hat{\sigma}^* / \sqrt{N} \mathcal{D}, \hat{\tau} + z_{1-\alpha/2} \hat{\sigma}^* / \sqrt{N} \mathcal{D} \right]$$

where $z_{1-\alpha}$ denotes the $(1 - \alpha)$ -th quantile in the standard normal distribution.

4.2 Appendix B

Below, we show the performance of the RIPW estimator under the misspecification of one or more nuisance parameters. Simulations follow the parameters and the DGP from Section 6.1.

Table 5 shows the results for the simulations in which only the propensity score is correct. We misspecify the control function by estimating a quadratic model for the MTE. In that case, the CF-DID estimator presents a non-negligible bias, while the RIPW estimator still shows little to no bias.

Table 6 presents the results for the case in which only the control function is correctly specified. We estimate a linear model for the MTE, using an incorrect propensity score to build the weights but the correct propensity score for the control function. The CF-DID estimators

Table 5 - Only the Propensity Score is correctly specified.

Estimator	Av. Bias	Med. Bias	RMSE	Cover	CIL
TWFE	-1.613	-1.614	1.615	0.000	0.208
DR-DID	1.611	1.612	1.615	0.327	2.632
WALD-DID	2.971	2.863	2.988	0.657	4.542
TC-WALD	-0.450	-0.439	0.449	0.768	1.822
CF-DID	-0.141	-0.143	0.222	0.793	1.072
RIPW	-0.007	-0.001	0.209	0.949	1.329

Note: Simulations based on 10,000 Monte Carlo experiments. TWFE is the two-way fixed effects estimator, DR-DID is the Doubly-Robust DiD estimator as proposed in Sant’anna and Zhao (2020), WALD-DID is the Wald-DiD estimator as used in Duflo (2001), TC-WALD is the Time-Corrected Wald Ratio proposed by De Chaisemartin and D’Haultefoeuille (2018), CF-DID and RIPW are our proposed estimator. “Av. Bias”, “Med. Bias”, “RMSE”, “Cover” and “CIL”, stand for the average simulated bias, median simulated bias, simulated root-mean-squared errors, 95% coverage probability, and 95% confidence interval length, respectively.

Table 6 - Only the Control Function is correctly specified.

Estimator	Av. Bias	Med. Bias	RMSE	Cover	CIL
TWFE	-1.613	-1.614	1.615	0.000	0.208
DR-DID	1.611	1.612	1.615	0.327	2.632
WALD-DID	2.971	2.863	2.988	0.657	4.542
TC-WALD	-0.450	-0.439	0.449	0.768	1.822
CF-DID	-0.160	-0.173	0.304	0.664	1.488
RIPW	-0.012	-0.013	0.205	0.941	1.617

Note: Simulations based on 10,000 Monte Carlo experiments. TWFE is the two-way fixed effects estimator, DR-DID is the Doubly-Robust DiD estimator as proposed in Sant’anna and Zhao (2020), WALD-DID is the Wald-DiD estimator as used in Duflo (2001), TC-WALD is the Time-Corrected Wald Ratio proposed by De Chaisemartin and D’Haultefoeuille (2018), CF-DID and RIPW are our proposed estimator. “Av. Bias”, “Med. Bias”, “RMSE”, “Cover” and “CIL”, stand for the average simulated bias, median simulated bias, simulated root-mean-squared errors, 95% coverage probability, and 95% confidence interval length, respectively.

exhibit non-negligible bias with a magnitude similar to the one displayed in Table 5. The RIPW estimator remains unbiased.

When both nuisance parameters are incorrect, all estimators have non-negligible bias, and all inference procedures are misleading. In this scenario, our estimators present smaller biases when compared to the rest, and the CF-DID seems to perform the best in this case, as shown in Table 7.

The simulations assert the desirable double-robustness property of the RIPW estimator. In terms of efficiency, however, our estimators show a similar performance. Moreover, the simulations show that we can consistently estimate the MTE curve when the nuisance parameters are correctly specified. There is no double robust procedure for estimating the MTE, as it relies on the assumption that we can model the unobservable gains from treatment. Therefore, correctly specifying the control function is paramount for the identification of marginal treatment effects. To do so, researchers must rely on economic theory to model their control functions.

Table 7 - Control Function and Propensity Score are incorrect.

Estimator	Av. Bias	Med. Bias	RMSE	Cover	CIL
TWFE	-1.613	-1.614	1.615	0.000	0.208
DR-DID	1.611	1.612	1.615	0.327	2.632
WALD-DID	2.971	2.863	2.988	0.657	4.542
TC-WALD	-0.450	-0.439	0.449	0.768	1.822
CF-DID	-0.137	-0.141	0.352	0.699	1.451
RIPW	0.174	0.266	0.334	0.799	1.652

Note: Simulations based on 10,000 Monte Carlo experiments. TWFE is the two-way fixed effects estimator, DR-DID is the Doubly-Robust DiD estimator as proposed in Sant’anna and Zhao (2020), WALD-DID is the Wald-DiD estimator as used in Duflo (2001), TC-WALD is the Time-Corrected Wald Ratio proposed by De Chaisemartin and D’Haultfoeuille (2018), CF-DID and RIPW are our proposed estimator. “Av. Bias”, “Med. Bias”, “RMSE”, “Cover” and “CIL”, stand for the average simulated bias, median simulated bias, simulated root-mean-squared errors, 95% coverage probability, and 95% confidence interval length, respectively.

4.3 Appendix C: Identification with Panel Data

In this section we present the framework that is most suited for Panel data. Suppose that for all individuals i and time-periods $t \in \{0, 1\}$ we observe an outcome variable Y_{it} , a binary treatment indicator D_{it} and an instrument Z_{it} .

The value of the instrument in $t = 0$ is assumed to be the same for all individuals, so we drop the time index from the notation and simply write $Z_i = 1$.

Let $\mathcal{S}(Z)$ denote the support of the instrument. For each $d \in \{0, 1\}$ and $z \in \mathcal{S}(Z)$, let $Y_{it}(d, z)$ denote the potential outcome for individual i in period t that he/she would have if their treatment and instrument have been set to d and z . Similarly, let $D_{it}(z)$ denote the potential treatment of individual i at period t , if the instrument were z . The observed and potential variables for individual i in period t are related through

$$Y_{it} = \sum_{d \in \{0,1\}} \sum_{z \in \mathcal{S}(Z)} 1 \{D_{it} = d, Z_{it} = z\} Y_{it}(d, z), \quad D_{it} = \sum_{z \in \mathcal{S}(Z)} 1 \{Z_{it} = z\} D_{it}(z)$$

Since $Z_{i0} = 0$ for all individuals, observed quantities for individuals in this cohort are always associated with $z = 0$. Observed outcomes are equal to $Y_{i0} = \sum_{d \in \{0,1\}} Y_{i0}(d, 0)$ and observed treatment statuses are equal to $D_{i0} = D_{i0}(0)$.

At each time-period, selection into treatment follows a single-crossing threshold model vylta:

$$D_{it} = 1 \{ \mu(Z_{it}) \geq U_{iD_t} \}$$

The marginal distribution of U_{iD_t} can be normalized to a standard uniform distribution $V_{it} \equiv F_{U_{iD_t}}(U_{iD_t})$. Hence, the choice model is represented simply by $D_{it} = 1 \{ \pi(Z_{it}) \geq V_{it} \}$.

The assumption modifies Assumption 1 from Section 2.1:

Assumption 1P (Trend in Selection) : For all $z \in \mathcal{S}(Z)$ all i , $\mathbb{P}(V_{1i} \leq \pi(z)) \geq \mathbb{P}(V_{i0} \leq \pi(z))$.

We maintain the standard exclusion restriction assumption:

Assumption 2P (Exclusion Restriction): For all $z \in \mathcal{S}(Z)$, all i , $t \in \{0, 1\}$ and $d \in \{0, 1\}$, $Y_{it}(z, d) = Y_{it}(d)$.

Assumption 3P is the modified parallel trends assumption for the reduced form and the first-stage:

Assumption 3P (Parallel Trends): For all $z \in \mathcal{S}(Z)$, $\mathbb{E}[Y_{i1}(D_{i1}(z)) - Y_{i0}(D_{i0}(0)) | Z_i = z] \perp Z_i$ and $\mathbb{P}(V_{i1} \leq \pi(z) < V_{i0} | Z_i = z) \perp Z_i$.

Finally, the technical assumptions which ensure that MTE parameters are well defined in the Panel data setting are displayed below:

Assumption 4P: The function $\mu(Z_i)$ is a nondegenerate random variable.

Assumption 5P: For $t \in \{0, 1\}$, the distribution of U_{iD_t} is absolutely continuous with respect to the Lebesgue Measure.

Assumption 6P: For $t \in \{0, 1\}$, the values of $\mathbb{E}[|Y_{it}(1)|]$ and $\mathbb{E}[|Y_{it}(0)|]$ are finite.

Identification

The MTE parameter in the Panel data case is defined as

$$MTE(\pi_1(z)) = \mathbb{E}[Y_{i1}(1) - Y_{i1}(0) | V_{i1} = \pi_1(z) < V_{i0}]$$

The MTE is the average treatment effect for individuals in period 1 who are indifferent between taking treatment or not and would not take treatment in period 0. Theorem 6 shows that the MTE can be identified with panel data under modified LIV estimand evaluated on the evolution of the outcome and treatment through time:

Theorem 5. Define $\Delta Y_i = Y_{i1} - Y_{i0}$ and $\Delta D_i = D_{i1} - D_{i0}$. Under Assumptions 1P-6P,

$$\frac{\partial \mathbb{E}[\Delta Y_i | Z_i = z] / \partial z}{\partial \mathbb{E}[\Delta D_i | Z_i = z] / \partial z} = MTE(\pi_1(z))$$

Proof of Theorem 5

We begin with the first-stage. Consider a shift of the value of the instrument from 0 to z . Under Assumptions 1P and 3P-6P,

$$\begin{aligned}
& \mathbb{E} [\Delta D_i | Z_i = z] - \mathbb{E} [\Delta D_i | Z_i = 0] \\
&= \mathbb{P} (D_{i1} = 1 | Z_i = z) - \mathbb{P} (D_{i0} = 1 | Z_i = z) - (\mathbb{P} (D_{i1} = 1 | Z_i = 0) - \mathbb{P} (D_{i0} = 1 | Z_i = 0)) \\
&= \mathbb{P} (V_{i1} \leq \mu(z) | Z_i = z) - \mathbb{P} (V_{i0} \leq \mu(0) | Z_i = z) - (\mathbb{P} (V_{i1} \leq \mu(0) | Z_i = 0) - \mathbb{P} (V_{i0} \leq \mu(0) | Z_i = 0)) \\
&= \mathbb{P} (V_{i1} \leq \mu(0) | Z_i = z) + \mathbb{P} (\mu(0) < V_{i1} \leq \mu(z) | Z_i = z) - \mathbb{P} (V_{i0} \leq \mu(0) | Z_i = z) \\
&\quad - (\mathbb{P} (V_{i1} \leq \mu(0) | Z_i = 0) - \mathbb{P} (V_{i0} \leq \mu(0) | Z_i = 0)) \\
&= \mathbb{P} (V_{i0} \leq \mu(0) | Z_i = z) + \mathbb{P} (V_{i1} \leq \mu(0) < V_{i0} | Z_i = z) + \mathbb{P} (\mu(0) < V_{i1} \leq \mu(z) < V_{i0} | Z_i = z) \\
&\quad - \mathbb{P} (V_{i0} \leq \mu(0) | Z_i = z) - (\mathbb{P} (V_{i0} \leq \mu(0) | Z_i = 0) + \mathbb{P} (V_{i1} \leq \mu(0) < V_{i0} | Z_i = 0) - \mathbb{P} (V_{i0} \leq \mu(0) | Z_i = 0)) \\
&= \mathbb{P} (V_{i1} \leq \mu(0) < V_{i0}) + \mathbb{P} (\mu(0) < V_{i1} \leq \mu(z) < V_{i0}) - \mathbb{P} (V_{i1} \leq \mu(0) < V_{i0}) \\
&= \mathbb{P} (\mu(0) < V_{i1} \leq \mu(z) < V_{i0})
\end{aligned}$$

Now consider the reduced form estimand associated to a shift of the value of the instrument from 0 to z . Under Assumptions 1P-6P,

$$\begin{aligned}
& \mathbb{E} [\Delta Y_i | Z_i = z] - \mathbb{E} [\Delta Y_i | Z_i = 0] \\
&= \mathbb{E} [Y_{i1} - Y_{i0} | Z_i = z] - \mathbb{E} [Y_{i1} - Y_{i0} | Z_i = 0] \\
&= \mathbb{E} [Y_{i1}(D_{i1}(z)) - Y_{i0}(D_{i0}(0)) | Z_i = z] - \mathbb{E} [Y_{i1}(D_{i1}(0)) - Y_{i0}(D_{i0}(0)) | Z_i = 0] \\
&= \mathbb{E} [Y_{i1}(1) - Y_{i1}(0) | Z_i = z, D_{i1}(z) > D_{i1}(0), D_{i0}(z) = 0] \mathbb{P} (D_{i1}(z) > D_{i1}(0), D_{i0}(z) = 0 | Z_i = z) \\
&= \mathbb{E} [Y_{i1}(1) - Y_{i1}(0) | \mu(0) < V_{i1} \leq \mu(z) < V_{i0}] \mathbb{P} (\mu(0) < V_{i1} \leq \mu(z) < V_{i0})
\end{aligned}$$

Hence, it follows that

$$\frac{\mathbb{E} [\Delta Y_i | Z_i = z] - \mathbb{E} [\Delta Y_i | Z_i = 0]}{\mathbb{E} [\Delta D_i | Z_i = z] - \mathbb{E} [\Delta D_i | Z_i = 0]} = \mathbb{E} [Y_{i1}(1) - Y_{i1}(0) | \mu(0) < V_{i1} \leq \mu(z) < V_{i0}]$$

Taking the limit of the expression with z approaching zero as in (HECKMAN; VYTLAČIL, 1999) yields

$$\frac{\partial \mathbb{E} [\Delta Y_i | Z_i = z] / \partial z}{\partial \mathbb{E} [\Delta D_i | Z_i = z] / \partial z} = \mathbb{E} [Y_{i1}(1) - Y_{i1}(0) | V_{i1} = \mu(z) < V_{i0}]$$

which concludes the proof.

5 Appendix to Chapter 2

5.1 Appendix A

Proof of Proposition 1

I begin with the decomposition of the first stage β_2^{FS} . We have under the (IMBENS; ANGRIST, 1994) assumptions for the first stage that

$$\begin{aligned}
 \beta_2^{FS} &= \mathbb{E}[D_{i,2}|Z_{i,2} = 1] - \mathbb{E}[D_{i,2}|Z_{i,2} = 0] \\
 &= \mathbb{P}(D_{i,2} = 1|Z_{i,2} = 1) - \mathbb{P}(D_{i,2} = 1|Z_{i,2} = 0) \\
 &= \mathbb{P}(D_{i,2}(1) = 1) - \mathbb{P}(D_{i,2}(0) = 1) \\
 &= \mathbb{P}(G_{i,2} = AT_2) + \mathbb{P}(G_{i,2} = C_2) - \mathbb{P}(G_{i,2} = AT_2) \\
 &= \mathbb{P}(G_{i,2} = C_2)
 \end{aligned}$$

For the reduced form,

$$\begin{aligned}
 \beta_2^{RF} &= \mathbb{E}[Y_{i,2}|Z_{i,2} = 1] - \mathbb{E}[Y_{i,2}|Z_{i,2} = 0] \\
 &= \mathbb{E}[Y_{i,2}|Z_{i,1:2} = (1, 1)] \mathbb{P}(Z_{i,1} = 1|Z_{i,2} = 1) + \mathbb{E}[Y_{i,2}|Z_{i,1:2} = (0, 1)] \mathbb{P}(Z_{i,1} = 0|Z_{i,2} = 1) \\
 &\quad - \mathbb{E}[Y_{i,2}|Z_{i,1:2} = (1, 0)] \mathbb{P}(Z_{i,1} = 1|Z_{i,2} = 0) - \mathbb{E}[Y_{i,2}|Z_{i,1:2} = (0, 0)] \mathbb{P}(Z_{i,1} = 0|Z_{i,2} = 0) \\
 &= \mathbb{E}[Y_{i,2}(D_{i,1:2}(1, 1))] \mathbb{P}(Z_{i,1} = 1|Z_{i,2} = 1) + \mathbb{E}[Y_{i,2}(D_{i,1:2}(0, 1))] \mathbb{P}(Z_{i,1} = 0|Z_{i,2} = 1) \\
 &\quad - \mathbb{E}[Y_{i,2}(D_{i,1:2}(1, 0))] \mathbb{P}(Z_{i,1} = 1|Z_{i,2} = 0) - \mathbb{E}[Y_{i,2}(D_{i,1:2}(0, 0))] \mathbb{P}(Z_{i,1} = 0|Z_{i,2} = 0)
 \end{aligned}$$

Add and subtract $Y_{i,2}(D_{i,1:2}(0, 0))$ to obtain

$$\begin{aligned}
 &\{\mathbb{E}[Y_{i,2}(D_{i,1:2}(1, 1)) - Y_{i,2}(D_{i,1:2}(0, 0))] + \mathbb{E}[Y_{i,2}(D_{i,1:2}(0, 0))]\} \mathbb{P}(Z_{i,1} = 1|Z_{i,2} = 1) \\
 &+ \{\mathbb{E}[Y_{i,2}(D_{i,1:2}(0, 1)) - Y_{i,2}(D_{i,1:2}(0, 0))] + \mathbb{E}[Y_{i,2}(D_{i,1:2}(0, 0))]\} \mathbb{P}(Z_{i,1} = 0|Z_{i,2} = 1) \\
 &- \{\mathbb{E}[Y_{i,2}(D_{i,1:2}(1, 0)) - Y_{i,2}(D_{i,1:2}(0, 0))] + \mathbb{E}[Y_{i,2}(D_{i,1:2}(0, 0))]\} \mathbb{P}(Z_{i,1} = 1|Z_{i,2} = 0) \\
 &\quad - \mathbb{E}[Y_{i,2}(D_{i,1:2}(0, 0))] \mathbb{P}(Z_{i,1} = 0|Z_{i,2} = 0) \\
 &= \mathbb{E}[Y_{i,2}(D_{i,1:2}(1, 1)) - Y_{i,2}(D_{i,1:2}(0, 0))] \mathbb{P}(Z_{i,1} = 1|Z_{i,2} = 1) \\
 &\quad + \mathbb{E}[Y_{i,2}(D_{i,1:2}(0, 1)) - Y_{i,2}(D_{i,1:2}(0, 0))] \mathbb{P}(Z_{i,1} = 0|Z_{i,2} = 1) \\
 &\quad - \mathbb{E}[Y_{i,2}(D_{i,1:2}(1, 0)) - Y_{i,2}(D_{i,1:2}(0, 0))] \mathbb{P}(Z_{i,1} = 1|Z_{i,2} = 0)
 \end{aligned}$$

After some algebraic manipulation, the expression can be written as

$$\begin{aligned} & \{\mathbb{E} [Y_{i,2}(D_{i,1:2}(1, 1)) - Y_{i,2}(D_{i,1:2}(1, 0))] + Y_{i,2}(D_{i,1:2}(1, 0)) - Y_{i,2}(D_{i,1:2}(0, 0))\} \mathbb{P}(Z_{i,1} = 1|Z_{i,2} = 1) \\ & + \mathbb{E} [Y_{i,2}(D_{i,1:2}(0, 1)) - Y_{i,2}(D_{i,1:2}(0, 0))] \mathbb{P}(Z_{i,1} = 0|Z_{i,2} = 1) \\ & - \mathbb{E} [Y_{i,2}(D_{i,1:2}(1, 0)) - Y_{i,2}(D_{i,1:2}(0, 0))] \mathbb{P}(Z_{i,1} = 1|Z_{i,2} = 0) \end{aligned}$$

Decomposing in terms of the groups implied by the (IMBENS; ANGRIST, 1994) assumptions for potential treatments at each period yields

$$\begin{aligned} & \sum_{\mathbf{z} \in \{0,1\}} \omega_{\mathbf{z}}(1) \mathbb{E} [Y_{i,2}(D_{i,1}(\mathbf{z}), 1) - Y_{i,2}(D_{i,1}(\mathbf{z}), 0) | G_{i,2} = C_2] \mathbb{P}(G_{i,2} = C_2) \\ & + + (\omega_1(1) - \omega_1(0)) \mathbb{E} [Y_{i,2}(1, D_{i,2}(0)) - Y_{i,2}(0, D_{i,2}(0)) | G_{i,1} = C_1] \mathbb{P}(G_{i,1} = C_1) \end{aligned}$$

Taking the ratio $\frac{\beta_2^{RF}}{\beta_2^{FS}}$ concludes the proof.

Proof of Lemma 1

Note that if $Z_{i,1}$ and $Z_{i,2}$ are independent, then $\omega_1(1) = \omega_1(0)$ and $\omega_{\mathbf{z}}(1) = \mathbb{P}(Z_{i,1} = \mathbf{z})$. The result follows trivially.

Proof of Proposition 2

First, consider the two-stage estimand for the mean potential outcome associated to full exposure. I begin with the first-stage.

Note that under the assumptions of the model,

$$\mathbb{E} \left[\frac{D_{i,1} D_{i,2} \mathbf{1}\{Z_{i,1:2} = (z_{1:2})\}}{\pi_i(z_{1:2})} \right] = \mathbb{P}(D_{i,1}(z_1) = 1, D_{i,2}(z_2) = 1)$$

Hence, the four quantities that constitute $\Delta^2 \left(\mathbb{E} \left[\frac{D_{i,1} D_{i,2} \mathbf{1}\{Z_{i,1:2} = (z_{1:2})\}}{\pi_i(z_{1:2})} \right] \right)$ can be written under the (IMBENS; ANGRIST, 1994) monotonicity assumption, respectively as

$$\begin{aligned}
\mathbb{E} \left[\frac{D_{i,1}D_{i,2}\mathbf{1}\{Z_{i,1:2} = (1, 1)\}}{\pi_i(1, 1)} \right] &= \mathbb{P}(G_{i,1:2} = AT_{1:2}) + \mathbb{P}(G_{i,1:2} = (AT_1, C_2)) \\
&+ \mathbb{P}(G_{i,1:2} = (C_1, AT_2)) + \mathbb{P}(G_{i,1:2} = C_{1:2}) \\
\mathbb{E} \left[\frac{D_{i,1}D_{i,2}\mathbf{1}\{Z_{i,1:2} = (1, 0)\}}{\pi_i(1, 0)} \right] &= \mathbb{P}(G_{i,1:2} = AT_{1:2}) + \mathbb{P}(G_{i,1:2} = (C_1, AT_2)) \\
\mathbb{E} \left[\frac{D_{i,1}D_{i,2}\mathbf{1}\{Z_{i,1:2} = (0, 1)\}}{\pi_i(0, 1)} \right] &= \mathbb{P}(G_{i,1:2} = AT_{1:2}) + \mathbb{P}(G_{i,1:2} = (AT_1, C_2)) \\
\mathbb{E} \left[\frac{D_{i,1}D_{i,2}\mathbf{1}\{Z_{i,1:2} = (0, 0)\}}{\pi_i(0, 0)} \right] &= \mathbb{P}(G_{i,1:2} = AT_{1:2})
\end{aligned}$$

Thus, the “difference-in-differences” for the modified first stage identifies $\mathbb{P}(G_{i,1:2} = C_{1:2})$.

Now let’s consider the estimand for the modified reduced form. Under the assumptions of the model,

$$\mathbb{E} \left[Y_{i,2} \frac{D_{i,1}D_{i,2}\mathbf{1}\{Z_{i,1:2} = (z_{1:2})\}}{\pi_i(z_{1:2})} \right] = \mathbb{E} [Y_{i,2}(1, 1)|D_{i,1}(z_1) = 1, D_{i,2}(z_2) = 1] \mathbb{P}(D_{i,1}(z_1) = 1, D_{i,2}(z_2) = 1)$$

Therefore, using the same reasoning as for the first stage, the estimand for the modified reduced form identifies

$$\mathbb{E} [Y_{i,2}(1, 1)|G_{i,1:2} = C_{1:2}] \mathbb{P}(G_{i,1:2} = C_{1:2})$$

and, therefore, the ratio identifies $m_2(1, 1)$. We now turn to the identification of the mean potential outcome associated to no exposure. For the first-stage, note that

$$\mathbb{E} \left[\frac{(1 - D_{i,1})(1 - D_{i,2})\mathbf{1}\{Z_{i,1:2} = (z_{1:2})\}}{\pi_i(z_{1:2})} \right] = \mathbb{P}(D_{i,1}(z_1) = 0, D_{i,2}(z_2) = 0)$$

Hence, the four quantities that constitute $\Delta^2 \left(\mathbb{E} \left[\frac{(1 - D_{i,1})(1 - D_{i,2})\mathbf{1}\{Z_{i,1:2} = (z_{1:2})\}}{\pi_i(z_{1:2})} \right] \right)$ can be written under the (IMBENS; ANGRIST, 1994) monotonicity assumption, respectively as

$$\begin{aligned}
\mathbb{E} \left[\frac{(1 - D_{i,1})(1 - D_{i,2})\mathbf{1}\{Z_{i,1:2} = (1, 1)\}}{\pi_i(1, 1)} \right] &= \mathbb{P}(G_{i,1:2} = NT_{1:2}) \\
\mathbb{E} \left[\frac{(1 - D_{i,1})(1 - D_{i,2})\mathbf{1}\{Z_{i,1:2} = (1, 0)\}}{\pi_i(1, 0)} \right] &= \mathbb{P}(G_{i,1:2} = NT_{1:2}) + \mathbb{P}(G_{i,1:2} = (NT_1, C_2)) \\
\mathbb{E} \left[\frac{(1 - D_{i,1})(1 - D_{i,2})\mathbf{1}\{Z_{i,1:2} = (0, 1)\}}{\pi_i(0, 1)} \right] &= \mathbb{P}(G_{i,1:2} = NT_{1:2}) + \mathbb{P}(G_{i,1:2} = (C_1, AT_2)) \\
\mathbb{E} \left[\frac{(1 - D_{i,1})(1 - D_{i,2})\mathbf{1}\{Z_{i,1:2} = (0, 0)\}}{\pi_i(0, 0)} \right] &= \mathbb{P}(G_{i,1:2} = AT_{1:2}) + \mathbb{P}(G_{i,1:2} = (AT_1, C_2)) \\
&+ \mathbb{P}(G_{i,1:2} = (C_1, AT_2)) + \mathbb{P}(G_{i,1:2} = C_{1:2})
\end{aligned}$$

Thus, the “difference-in-differences” for the modified first stage identifies $-\mathbb{P}(G_{i,1:2} = C_{1:2})$.

Now let’s consider the estimand for the modified reduced form. Under the assumptions of the model,

$$\mathbb{E} \left[Y_{i,2} \frac{(1 - D_{i,1})(1 - D_{i,2}) \mathbf{1}\{Z_{i,1:2} = (z_{1:2})\}}{\pi_i(z_{1:2})} \right] = \mathbb{E} [Y_{i,2}(0, 0) | D_{i,1}(z_1) = 0, D_{i,2}(z_2) = 0] \mathbb{P}(D_{i,1}(z_1) = 0, D_{i,2}(z_2) = 0)$$

Therefore, using the same reasoning as for the first stage, the estimand for the modified reduced form identifies

$$-\mathbb{E} [Y_{i,2}(0, 0) | G_{i,1:2} = C_{1:2}] \mathbb{P}(G_{i,1:2} = C_{1:2})$$

and, therefore, the ratio identifies $m_2(0, 0)$, which concludes the proof.

Proof of Lemma 2

Note that under the linear Toy model and the sustained assumptions,

$$\frac{\Delta^2 \left(\mathbb{E} \left[Y_{i,2} \frac{D_{i,1} D_{i,2} \mathbf{1}\{Z_{i,1:2} = (z_{1:2})\}}{\pi_i(z_{1:2})} \right] \right)}{\Delta^2 \left(\mathbb{E} \left[\frac{D_{i,1} D_{i,2} \mathbf{1}\{Z_{i,1:2} = (z_{1:2})\}}{\pi_i(z_{1:2})} \right] \right)} = \frac{(\beta_0 + \beta_1 + \beta_2) \alpha_1^1 \alpha_2^2}{\alpha_1^1 \alpha_2^2} = \beta_0 + \beta_1 + \beta_2$$

Next, note that

$$\frac{\Delta^2 \left(\mathbb{E} \left[Y_{i,2} \frac{(1 - D_{i,1})(1 - D_{i,2}) \mathbf{1}\{Z_{i,1:2} = (z_{1:2})\}}{\pi_i(z_{1:2})} \right] \right)}{\Delta^2 \left(\mathbb{E} \left[\frac{(1 - D_{i,1})(1 - D_{i,2}) \mathbf{1}\{Z_{i,1:2} = (z_{1:2})\}}{\pi_i(z_{1:2})} \right] \right)} = \frac{-\beta_0 \alpha_1^1 \alpha_2^2}{-\alpha_1^1 \alpha_2^2} = \beta_0$$

from which the result follows.

Proof of Proposition 3

I begin with the decomposition of the first stage β_t^{FS} . We have under Assumptions 1-4 and 6 that

$$\begin{aligned} \mathbb{E} [D_{i,t} | Z_{i,t=1}] - \mathbb{E} [D_{i,t} | Z_{i,t=0}] &= \mathbb{P}(D_{i,t} = 1 | Z_{i,t} = 1) - \mathbb{P}(D_{i,t} = 1 | Z_{i,t} = 0) \\ &= \mathbb{P}(D_{i,t}(1) = 1) - \mathbb{P}(D_{i,t}(0) = 1) = \mathbb{P}(G_{i,t} = C_t) \end{aligned}$$

For the reduced form, we have under the Law of Total Probabilities

$$\begin{aligned} & \mathbb{E}[Y_{i,t}|Z_{i,t} = 1] - \mathbb{E}[Y_{i,t}|Z_{i,t} = 0] \\ &= \sum_{\mathbf{z} \in \{0,1\}^{t-1}} \mathbb{E}[Y_{i,t}|Z_{i,1:t} = (\mathbf{z}, 1)] \mathbb{P}(Z_{i,1:t-1} = \mathbf{z}|Z_{i,t} = 1) - \sum_{\mathbf{z} \in \{0,1\}^{t-1}} \mathbb{E}[Y_{i,t}|Z_{i,1:t} = (\mathbf{z}, 0)] \mathbb{P}(Z_{i,1:t-1} = \mathbf{z}|Z_{i,t} = 0) \end{aligned}$$

Under Assumptions 1, 2 and 4, the estimand for the reduced form can be further decomposed as

$$\begin{aligned} & \sum_{\mathbf{z} \in \{0,1\}^{t-1}} \mathbb{E}[Y_{i,t}(D_{i,1:t}(\mathbf{z}, 1))] \mathbb{P}(Z_{i,1:t-1} = \mathbf{z}|Z_{i,t} = 1) \\ & - \sum_{\mathbf{z} \in \{0,1\}^{t-1}} \mathbb{E}[Y_{i,t}(D_{i,1:t}(\mathbf{z}, 0))] \mathbb{P}(Z_{i,1:t-1} = \mathbf{z}|Z_{i,t} = 0) \end{aligned}$$

Adding and subtracting $\sum_{z_t \in \{0,1\}} \sum_{\mathbf{z} \in \{0,1\}^{t-1}} \mathbb{E}[Y_{i,t}(D_{i,1:t}(\mathbf{0}))] \mathbb{P}(Z_{i,1:t-1} = \mathbf{z}|Z_{i,t} = z_t)$ yields

$$\begin{aligned} & \sum_{\mathbf{z} \in \{0,1\}^{t-1}} \mathbb{E}[Y_{i,t}(D_{i,1:t}(\mathbf{z}, 1)) + Y_{i,t}(D_{i,1:t}(\mathbf{0})) - Y_{i,t}(D_{i,1:t}(\mathbf{0}))] \mathbb{P}(Z_{i,1:t-1} = \mathbf{z}|Z_{i,t} = 1) \\ & - \sum_{\mathbf{z} \in \{0,1\}^{t-1}} \mathbb{E}[Y_{i,t}(D_{i,1:t}(\mathbf{z}, 0)) + Y_{i,t}(D_{i,1:t}(\mathbf{0})) - Y_{i,t}(D_{i,1:t}(\mathbf{0}))] \mathbb{P}(Z_{i,1:t-1} = \mathbf{z}|Z_{i,t} = 0) \\ &= \sum_{\mathbf{z} \in \{0,1\}^{t-1}} \mathbb{E}[Y_{i,t}(D_{i,1:t}(\mathbf{z}, 1)) - Y_{i,t}(D_{i,1:t}(\mathbf{0}))] \mathbb{P}(Z_{i,1:t-1} = \mathbf{z}|Z_{i,t} = 1) \\ & - \sum_{\mathbf{z} \in \{0,1\}^{t-1}} \mathbb{E}[Y_{i,t}(D_{i,1:t}(\mathbf{z}, 0)) - Y_{i,t}(D_{i,1:t}(\mathbf{0}))] \mathbb{P}(Z_{i,1:t-1} = \mathbf{z}|Z_{i,t} = 0) \end{aligned}$$

Adding and subtracting

$$\sum_{j=1}^{t-1} \sum_{\mathbf{z} \in \{0,1\}^{t-1}} \sum_{\mathbf{z} \in \{0,1\}^{t-1}} \mathbb{E}[Y_{i,t}(D_{i,1:j-1}(\mathbf{0}), D_{i,j}(1), D_{i,j+1:t}(\mathbf{0}))] \mathbb{P}(Z_{i,1:t-1} = \mathbf{z}|Z_{i,t} = z_t)$$

and using Assumption 3 yields

$$\begin{aligned} & \sum_{\mathbf{z} \in \{0,1\}^{t-1}} \{ \mathbb{E}[Y_{i,t}(D_{i,1:t-1}(\mathbf{z}), 1) - Y_{i,t}(D_{i,1:t-1}(\mathbf{z}), 0)|G_{i,t} = C_t] \mathbb{P}(G_{i,t} = C_t) \} \mathbb{P}(Z_{i,1:t-1} = \mathbf{z}|Z_{i,t} = 1) \\ & + \sum_{j=1}^{t-1} \{ \mathbb{P}(Z_{i,j} = 1|Z_{i,t} = 1) - \mathbb{P}(Z_{i,j} = 1|Z_{i,t} = 0) \} \times \\ & \mathbb{E}[Y_{i,t}(D_{i,1:j-1}(\mathbf{0}), 1, D_{i,j+1:t}(\mathbf{0})) - Y_{i,t}(D_{i,1:j-1}(\mathbf{0}), 0, D_{i,j+1:t}(\mathbf{0}))|G_{i,j} = C_j] \mathbb{P}(G_{i,j} = C_j) \end{aligned}$$

Taking the ratio concludes the proof.

Proof of Theorem 1

The proof is conducted by induction. I show that the result holds for the baseline case $p = 0$ and then show that if it holds for a general p , then it must also hold for $p + 1$.

Consider the case where $p = 0$ and $D_{i,t} = 1$. Results for the identification in the cross-sectional setting can be modified to show that under Assumptions 1-3 and 5-7, the first stage identifies

$$\begin{aligned} & \mathbb{E} \left[\frac{D_{i,t} Z_{i,t}}{\pi_{i,t}(1)} \middle| \mathcal{F}_{i,1:t-1} \right] - \mathbb{E} \left[\frac{D_{i,t}(1 - Z_{i,t})}{\pi_{i,t}(0)} \middle| \mathcal{F}_{i,1:t-1} \right] \\ &= \mathbb{P}(D_{i,t}(1) = 1) - \mathbb{P}(D_{i,t}(0) = 1) = \mathbb{P}(G_{i,t} = C_t) \end{aligned}$$

For the reduced form estimand of the modified outcome, note that under Assumptions 1-3 and 5-7,

$$\begin{aligned} & \mathbb{E} \left[\frac{Y_{i,t} D_{i,t} Z_{i,t}}{\pi_{i,t}(1)} \middle| \mathcal{F}_{i,1:t-1} \right] - \mathbb{E} \left[\frac{Y_{i,t} D_{i,t} (1 - Z_{i,t})}{\pi_{i,t}(0)} \middle| \mathcal{F}_{i,1:t-1} \right] \\ &= \mathbb{E} [Y_{i,t}(d_{1:t-1}^{obs}, 1) | D_{i,t}(1) = 1] \mathbb{P}(D_{i,t}(1) = 1) - \mathbb{E} [Y_{i,t}(d_{1:t-1}^{obs}, 1) | D_{i,t}(0) = 1] \mathbb{P}(D_{i,t}(0) = 1) \\ & \quad \mathbb{E} [Y_{i,t}(d_{1:t-1}^{obs}, 1) | G_{i,t} = C_t] \mathbb{P}(G_{i,t} = C_t) \end{aligned}$$

and the ratio identifies

$$\mathbb{E} [Y_{i,t}(d_{1:t-1}^{obs}, 1) | G_{i,t} = C_t] = m_t(1)$$

The result for the case where $D_{i,t} = 0$ can be demonstrated analogously, once we note that the first stage identifies

$$\begin{aligned} & \mathbb{E} \left[\frac{(1 - D_{i,t}) Z_{i,t}}{\pi_{i,t}(1)} \middle| \mathcal{F}_{i,1:t-1} \right] - \mathbb{E} \left[\frac{(1 - D_{i,t})(1 - Z_{i,t})}{\pi_{i,t}(0)} \middle| \mathcal{F}_{i,1:t-1} \right] \\ &= \mathbb{P}(D_{i,t}(1) = 0) - \mathbb{P}(D_{i,t}(0) = 0) = -\mathbb{P}(G_{i,t} = C_t) \end{aligned}$$

and the modified reduced form identifies

$$-\mathbb{E} [Y_{i,t}(d_{1:t-1}^{obs}, 0) | G_{i,t} = C_t] \mathbb{P}(G_{i,t} = C_t)$$

For the general case, I prove the result only for $D_{i,t-p-1} = 1$. The case for $D_{i,t-p-1} = 0$ and can be derived using the intuition from the displayed result above. I begin with the first stage. Note that under Assumption 1,

$$\begin{aligned}
& \Delta^{p+2} \left(\mathbb{E} \left[\frac{\mathbf{1}\{D_{i,t-p-1:t} = (1, \mathbf{d})\} \mathbf{1}\{Z_{i,t-p-1:t} = z_{t-p-1:t}\}}{\pi_{i,t-p-1:t}(z_{t-p-1:t})} \middle| \mathcal{F}_{i,t-p-2} \right] \right) \\
&= \Delta^{p+1} \left(\mathbb{E} \left[\frac{\mathbf{1}\{D_{i,t-p-1:t} = (1, \mathbf{d})\} \mathbf{1}\{Z_{i,t-p-1:t} = (1, z_{t-p:t})\}}{\pi_{i,t-p-1:t}(1, z_{t-p:t})} \middle| \mathcal{F}_{i,t-p-2} \right] \right) \\
&- \Delta^{p+1} \left(\mathbb{E} \left[\frac{\mathbf{1}\{D_{i,t-p-1:t} = (1, \mathbf{d})\} \mathbf{1}\{Z_{i,t-p-1:t} = (0, z_{t-p:t})\}}{\pi_{i,t-p-1:t}(0, z_{t-p:t})} \middle| \mathcal{F}_{i,t-p-2} \right] \right)
\end{aligned}$$

Assume the result holds for p . If Assumptions 2,3 and 5-7 further hold, then the modified first stage can be written as

$$\begin{aligned}
& \mathbb{P}(D_{i,t-p-1}(1) = 1, G_{i,t} = C_t) - \mathbb{P}(D_{i,t-p-1}(0) = 1, G_{i,t-p:t} = C_{t-p:t}) \\
&= \mathbb{P}(G_{i,t-p-1:t} = C_{t-p-1:t})
\end{aligned}$$

For the modified reduced form, note that under Assumption 1,

$$\begin{aligned}
& \Delta^{p+2} \left(\mathbb{E} \left[\frac{Y_{i,t} \mathbf{1}\{D_{i,t-p-1:t} = (1, \mathbf{d})\} \mathbf{1}\{Z_{i,t-p-1:t} = z_{t-p-1:t}\}}{\pi_{i,t-p-1:t}(z_{t-p-1:t})} \middle| \mathcal{F}_{i,t-p-2} \right] \right) \\
&= \Delta^{p+1} \left(\mathbb{E} \left[\frac{Y_{i,t} \mathbf{1}\{D_{i,t-p-1:t} = (1, \mathbf{d})\} \mathbf{1}\{Z_{i,t-p-1:t} = (1, z_{t-p:t})\}}{\pi_{i,t-p-1:t}(1, z_{t-p:t})} \middle| \mathcal{F}_{i,t-p-2} \right] \right) \\
&- \Delta^{p+1} \left(\mathbb{E} \left[\frac{Y_{i,t} \mathbf{1}\{D_{i,t-p-1:t} = (1, \mathbf{d})\} \mathbf{1}\{Z_{i,t-p-1:t} = (0, z_{t-p:t})\}}{\pi_{i,t-p-1:t}(0, z_{t-p:t})} \middle| \mathcal{F}_{i,t-p-2} \right] \right)
\end{aligned}$$

Assuming the result holds for p , if Assumptions 2,3 and 5-7 further hold, we can write the result displayed above as

$$\begin{aligned}
& \mathbb{E} [Y_{i,t}(d_{1:t-p-2}^{obs}, 1, \mathbf{d}) | D_{i,t-p-1}(1) = 1, G_{i,t-p:t} = C_{t-p:t}] \mathbb{P}(D_{i,t-p-1}(1) = 1, G_{i,t-p:t} = C_{t-p:t}) \\
&- \mathbb{E} [Y_{i,t}(d_{1:t-p-2}^{obs}, 1, \mathbf{d}) | D_{i,t-p-1}(0) = 1, G_{i,t-p:t} = C_{t-p:t}] \mathbb{P}(D_{i,t-p-1}(0) = 1, G_{i,t-p:t} = C_{t-p:t}) \\
&= \mathbb{E} [Y_{i,t}(d_{1:t-p-2}^{obs}, 1, \mathbf{d}) | G_{i,t-p-1:t} = C_{t-p-1:t}] \mathbb{P}(G_{i,t-p-1:t} = C_{t-p-1:t})
\end{aligned}$$

Hence, it follows that

$$\frac{\Delta^{p+2} \left(\mathbb{E} \left[\frac{Y_{i,t} \mathbf{1}\{D_{i,t-p:t} = \mathbf{d}\} \mathbf{1}\{Z_{i,t-p:t} = z_{t-p:t}\}}{\pi_{i,t-p:t}(z_{t-p:t})} \middle| \mathcal{F}_{i,t-p-1} \right] \right)}{\Delta^{p+2} \left(\mathbb{E} \left[\frac{\mathbf{1}\{D_{i,t-p:t} = \mathbf{d}\} \mathbf{1}\{Z_{i,t-p:t} = z_{t-p:t}\}}{\pi_{i,t-p:t}(z_{t-p:t})} \middle| \mathcal{F}_{i,t-p-1} \right] \right)} = m_t(\mathbf{d})$$

Which concludes the proof.

Proof of Theorem 2

If potential outcomes are bounded, then the standard Lindeberg condition holds. The first result follows from the triangular array central limit theorem, as in (BOJINOV; RAMBACHAN; SHEPHARD, 2021).

The variance $(\sigma_t(1, 0; 0))^2$ is simply

$$(\sigma_t(1, 0; 0))^2 = \frac{1}{N} \sum_{i=1}^N (\sigma_{i,t}(1, 0; 0))^2$$

where $(\sigma_{i,t}(1, 0; 0))^2$ is defined in Lemma 1 from Appendix B.

The second result follows from Theorem 3.2 in (BOJINOV; RAMBACHAN; SHEPHARD, 2021). The variance is given by

$$(\sigma(1, 0; 0))^2 = \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N (\sigma_{i,t}(1, 0; 0))^2$$

Proof of Theorem 3

The result follows the same reasoning as the one in Theorem 2. This time, however, we have

$$\left(\sigma_i(\mathbf{d}, \tilde{\mathbf{d}}; p)\right)^2 = \frac{1}{N} \sum_{i=1}^N \left(\sigma_{i,t}(\mathbf{d}, \tilde{\mathbf{d}}; p)\right)^2$$

and

$$\left(\sigma(\mathbf{d}, \tilde{\mathbf{d}}; p)\right)^2 = \frac{1}{N(T-p)} \sum_{t=p+1}^T \sum_{i=1}^N \left(\sigma_{i,t}(\mathbf{d}, \tilde{\mathbf{d}}; p)\right)^2$$

See Lemma B2 from Appendix B for the necessary results.

5.2 Appendix B

Proof of Lemma B1

Define $W_{i,t-p:t}(\mathbf{z}) = \pi_{i,t-p:t}(\mathbf{z})^{-1} \mathbf{1}\{Z_{i,t-p:t} = \mathbf{z}\}$. From Lemma A.1 in (BOJINOV; RAMBACHAN; SHEPHARD, 2021), we have

$$\begin{aligned}\mathbb{E}_R [W_{i,t-p:t}(\mathbf{z}) | \mathcal{F}_{i,t-p-1}] &= 1 \\ \mathbb{V}_R [W_{i,t-p:t}(\mathbf{z}) | \mathcal{F}_{i,t-p-1}] &= \pi_{i,t-p:t}(\mathbf{z})^{-1}(1 - \pi_{i,t-p:t}(\mathbf{z})) \\ \text{Cov}_R [W_{i,t-p:t}(\mathbf{z}), W_{i,t-p:t}(\tilde{\mathbf{z}}) | \mathcal{F}_{i,t-p-1}] &= -1\end{aligned}$$

I analyze the properties of the estimator for each stage separately before analyzing the properties of the two-stage estimator. I begin with the first stage. Define

$$u_{i,t}^{FS} = \widehat{\tau}_{i,t}^{FS} - \tau_{i,t}^{FS} = D_{i,t}(1)(W_{i,t}(1) - 1) - D_{i,t}(0)(W_{i,t}(0) - 1)$$

Lemma A.1 from (BOJINOV; RAMBACHAN; SHEPHARD, 2021) implies that $\mathbb{E}_R [u_{i,t}^{FS} | \mathcal{F}_{i,t-p-1}] = 0$. Hence, the error terms from the estimator are a martingale difference sequence and so, uncorrelated through time. Now, let's look at the variance:

$$\begin{aligned}\mathbb{V}_R [u_{i,t}^{FS} | \mathcal{F}_{i,t-p-1}] &= D_{i,t}(1)^2 \mathbb{V} [W_{i,t}(1)] + D_{i,t}(0)^2 \mathbb{V} [W_{i,t}(0)] \\ &\quad - 2D_{i,t}(1)D_{i,t}(0) \text{Cov} [W_{i,t}(1), W_{i,t}(0)] \\ &= D_{i,t}(1)^2 \pi_{i,t}(1)^{-1}(1 - \pi_{i,t}(1)) + D_{i,t}(0)^2 \pi_{i,t}(0)^{-1}(1 - \pi_{i,t}(0)) \\ &\quad + 2D_{i,t}(1)D_{i,t}(0) \\ &= \frac{D_{i,t}(1)^2}{\pi_{i,t}(1)} + \frac{D_{i,t}(0)^2}{\pi_{i,t}(0)} - (D_{i,t}(1) - D_{i,t}(0))^2 \\ &= (\gamma_{i,t}^{FS}(1, 0)(0))^2 - (D_{i,t}(1) - D_{i,t}(0))^2 = (\sigma_{i,t}^{FS}(1, 0)(0))^2\end{aligned}$$

Now, consider the reduce form. Define

$$u_{i,t}^{RF} = \widehat{\tau}_{i,t}^{RF} - \tau_{i,t}^{RF} = Y_{i,t}(d_{1:t-1}^{obs}, D_{i,t}(1))(W_{i,t}(1) - 1) - Y_{i,t}(d_{1:t-1}^{obs}, D_{i,t}(0))(W_{i,t}(0) - 1)$$

We have $\mathbb{E}_R [u_{i,t}^{RF} | \mathcal{F}_{i,t-p-1}] = 0$. Hence, the error terms from the estimator are a martingale difference sequence and so uncorrelated through time. Now, let's look at the variance:

$$\begin{aligned}\mathbb{V}_R [u_{i,t}^{RF} | \mathcal{F}_{i,t-p-1}] &= Y_{i,t}(d_{1:t-1}^{obs}, D_{i,t}(1))^2 \mathbb{V} [W_{i,t}(1)] + Y_{i,t}(d_{1:t-1}^{obs}, D_{i,t}(0))^2 \mathbb{V} [W_{i,t}(0)] \\ &\quad - 2Y_{i,t}(d_{1:t-1}^{obs}, D_{i,t}(1))Y_{i,t}(d_{1:t-1}^{obs}, D_{i,t}(0)) \text{Cov} [W_{i,t}(1), W_{i,t}(0)] \\ &= Y_{i,t}(d_{1:t-1}^{obs}, D_{i,t}(1))^2 \pi_{i,t}(1)^{-1}(1 - \pi_{i,t}(1)) + Y_{i,t}(d_{1:t-1}^{obs}, D_{i,t}(0))^2 \pi_{i,t}(0)^{-1}(1 - \pi_{i,t}(0)) \\ &\quad + 2Y_{i,t}(d_{1:t-1}^{obs}, D_{i,t}(1))Y_{i,t}(d_{1:t-1}^{obs}, D_{i,t}(0)) \\ &= \frac{Y_{i,t}(d_{1:t-1}^{obs}, D_{i,t}(1))^2}{\pi_{i,t}(1)} + \frac{Y_{i,t}(d_{1:t-1}^{obs}, D_{i,t}(0))^2}{\pi_{i,t}(0)} - (Y_{i,t}(d_{1:t-1}^{obs}, D_{i,t}(1)) - Y_{i,t}(d_{1:t-1}^{obs}, D_{i,t}(0)))^2 \\ &= (\gamma_{i,t}^{RF}(1, 0; 0))^2 - (Y_{i,t}(d_{1:t-1}^{obs}, D_{i,t}(1)) - Y_{i,t}(d_{1:t-1}^{obs}, D_{i,t}(0)))^2 = (\sigma_{i,t}^{RF}(1, 0; 0))^2\end{aligned}$$

From the results above, it follows that

$$\frac{\mathbb{E}_R [u_{i,t}^{RF} | \mathcal{F}_{i,t-p-1}]}{\mathbb{E}_R [u_{i,t}^{FS} | \mathcal{F}_{i,t-p-1}]} = 0$$

We apply the Uniform Delta Method to obtain

$$\mathbb{V}_R [u_{i,t}] = g' \left(\begin{array}{c} (\sigma_{i,t}^{RF}(1, 0; 0))^2 \\ (\sigma_{i,t}^{FS}(1, 0; 0))^2 \end{array} \right) g := (\sigma_{i,t}(1, 0; 0))^2$$

Where g is the gradient of $h(x, y) = x/y$ evaluated at $(\tau_{i,t}^{RF}, \tau_{i,t}^{FS})$, which concludes the proof.

Proof of Lemma B2

Define

$$\begin{aligned} u_{i,t}^{FS}(\mathbf{d}) &= \widehat{m}_{i,t}^{FS}(\mathbf{d}) - m_{i,t}^{FS}(\mathbf{d}) = \Delta^{p+1} (\mathbf{1} \{D_{i,t-p:t}(\mathbf{z})\} (W_{i,t-p:t}(\mathbf{z}) - 1)) \\ u_{i,t}^{RF}(\mathbf{d}) &= \widehat{m}_{i,t}^{RF}(\mathbf{d}) - m_{i,t}^{RF}(\mathbf{d}) = \Delta^{p+1} (Y_{i,t}(d_{1:t-p-1}^{obs}, \mathbf{d}) \mathbf{1} \{D_{i,t-p:t}(\mathbf{z})\} (W_{i,t-p:t}(\mathbf{z}) - 1)) \end{aligned}$$

We have $\mathbb{E}_R [u_{i,t}^{FS}(\mathbf{d}) | \mathcal{F}_{i,t-p-1}] = \mathbb{E}_R [u_{i,t}^{RF}(\mathbf{d}) | \mathcal{F}_{i,t-p-1}] = 0$ for all $\mathbf{d} \in \{0, 1\}^{p+1}$. Furthermore, define $u_{i,t}^{RF}(\mathbf{d}, \tilde{\mathbf{d}}; p) = u_{i,t}^{RF}(\mathbf{d}) + u_{i,t}^{RF}(\tilde{\mathbf{d}})$ and $u_{i,t}^{FS}(\mathbf{d}, \tilde{\mathbf{d}}; p)$ analogously. It follows that $\mathbb{E}_R [u_{i,t}^{RF}(\mathbf{d}, \tilde{\mathbf{d}}; p) | \mathcal{F}_{i,t-p-1}] = \mathbb{E}_R [u_{i,t}^{FS}(\mathbf{d}, \tilde{\mathbf{d}}; p) | \mathcal{F}_{i,t-p-1}] = 0$.

Now let's consider the variance. I prove the result by induction for the reduced form and first stage of each potential outcome separately, and build on it to derive the properties of the estimator.

First, let's consider the case when $p = 0$ and $D_{i,t} = 1$ (the case for $D_{i,t} = 0$ is analogous). For the first stage, it's been proved earlier that

$$\begin{aligned} \mathbb{V}_R [u_{i,t}^{FS}(1) | \mathcal{F}_{i,t-p-1}] &= \frac{D_{i,t}^2(1)}{\pi_{i,t}(1)} + \frac{D_{i,t}^2(0)}{\pi_{i,t}(0)} - (D_{i,t}(1) - D_{i,t}(0))^2 \\ &= (\gamma_{i,t}^{FS}(1))^2 - (D_{i,t}(1) - D_{i,t}(0))^2 = (\sigma_{i,t}^{FS}(1))^2 \end{aligned}$$

Now let's consider the reduced form. Using a reasoning similar from the one in the lag-0 case, we obtain

$$\begin{aligned}
\mathbb{V}_R [u_{i,t}^{RF}(1) | \mathcal{F}_{i,t-p-1}] &= (Y_{i,t}(d_{1:t-1}^{obs}, 1)D_{i,t}(1))^2 \mathbb{V} [W_{i,t}(1)] + (Y_{i,t}(d_{1:t-1}^{obs}, 1)D_{i,t}(0))^2 \mathbb{V} [W_{i,t}(0)] \\
&\quad - 2Y_{i,t}(d_{1:t-1}^{obs}, 1)D_{i,t}(1)Y_{i,t}(d_{1:t-1}^{obs}, 1)D_{i,t}(0) \text{Cov} [W_{i,t}(1), W_{i,t}(0)] \\
&= \frac{(Y_{i,t}(d_{1:t-1}^{obs}, 1)D_{i,t}(1))^2}{\pi_{i,t}(1)} + \frac{(Y_{i,t}(d_{1:t-1}^{obs}, 1)D_{i,t}(0))^2}{\pi_{i,t}(0)} \\
&\quad - (Y_{i,t}(d_{1:t-1}^{obs}, 1)D_{i,t}(1) - Y_{i,t}(d_{1:t-1}^{obs}, 1)D_{i,t}(0))^2 \\
&= (\gamma_{i,t}^{RF}(1))^2 - (Y_{i,t}(d_{1:t-1}^{obs}, 1)D_{i,t}(1) - Y_{i,t}(d_{1:t-1}^{obs}, 1)D_{i,t}(0))^2 = (\sigma_{i,t}^{RF}(1))^2
\end{aligned}$$

Now, I proceed by deriving the expression for a general $p + 1$, assuming that the result holds for p . For the first stage,

$$\begin{aligned}
\mathbb{V}_R [u_{i,t}^{FS}(\mathbf{d}) | \mathcal{F}_{i,t-p-1}] &= \mathbb{V}_R [\Delta^{p+2} (\mathbf{1} \{D_{i,t-p-1:t}(\mathbf{z}) = \mathbf{d}\} (W_{i,t-p-1}(\mathbf{z}) - 1))] \\
&= \mathbb{V}_R [\Delta^{p+1} (\mathbf{1} \{D_{i,t-p-1:t}(1, \mathbf{z}_-) = \mathbf{d}\} (W_{i,t-p-1}(1, \mathbf{z}_-) - 1)) \\
&\quad - \Delta^{p+1} (\mathbf{1} \{D_{i,t-p-1:t}(0, \mathbf{z}_-) = \mathbf{d}\} (W_{i,t-p-1}(0, \mathbf{z}_-) - 1))] \\
&= \mathbb{V}_R [\Delta^{p+1} (\mathbf{1} \{D_{i,t-p-1:t}(1, \mathbf{z}_-) = \mathbf{d}\} (W_{i,t-p-1}(1, \mathbf{z}_-) - 1))] \\
&\quad + \mathbb{V}_R [\Delta^{p+1} (\mathbf{1} \{D_{i,t-p-1:t}(0, \mathbf{z}_-) = \mathbf{d}\} (W_{i,t-p-1}(0, \mathbf{z}_-) - 1))] \\
&\quad - 2\Delta^{p+1} (\mathbf{1} \{D_{i,t-p-1:t}(1, \mathbf{z}_-) = \mathbf{d}\}) \Delta^{p+1} (\mathbf{1} \{D_{i,t-p-1:t}(0, \mathbf{z}_-) = \mathbf{d}\}) \text{Cov} [W_{i,t-p-1}(1, \mathbf{z}_-), W_{i,t-p-1}(0, \mathbf{z}_-)] \\
&= \sum_{\mathbf{z}_- \in \{0,1\}^{p+1}} \frac{\mathbf{1} \{D_{i,t-p-1}(1, \mathbf{z}_-) = \mathbf{d}\}}{\pi_{i,t-p-1}(1, \mathbf{z}_-)} - \Delta^{p+1} (D_{i,t-p-1}(1, \mathbf{z}_-))^2 \\
&\quad + \sum_{\mathbf{z}_- \in \{0,1\}^{p+1}} \frac{\mathbf{1} \{D_{i,t-p-1}(0, \mathbf{z}_-) = \mathbf{d}\}}{\pi_{i,t-p-1}(0, \mathbf{z}_-)} - \Delta^{p+1} (D_{i,t-p-1}(0, \mathbf{z}_-))^2 \\
&\quad + 2\Delta^{p+1} (D_{i,t-p-1}(1, \mathbf{z}_-)) \Delta^{p+1} (D_{i,t-p-1}(0, \mathbf{z}_-)) \\
&= \sum_{\mathbf{z} \in \{0,1\}^{p+2}} \frac{\mathbf{1} \{D_{i,t-p-1}(\mathbf{z}) = \mathbf{d}\}}{\pi_{i,t-p-1}(\mathbf{z})} \\
&\quad - (\Delta^{p+2} (\mathbf{1} \{D_{i,t-p-1}(1, \mathbf{z}_-) = \mathbf{d}\}) - \Delta^{p+2} (\mathbf{1} \{D_{i,t-p-1}(0, \mathbf{z}_-) = \mathbf{d}\}))^2 \\
&= (\gamma_{i,t}^{FS}(\mathbf{d}))^2 - (\Delta^{p+2} (\mathbf{1} \{D_{i,t-p-1}(\mathbf{z}) = \mathbf{d}\}))^2 \\
&= (\sigma_{i,t}^{FS}(\mathbf{d}))^2
\end{aligned}$$

Analogously, for the reduced form it can be shown that

$$\begin{aligned}
\mathbb{V}_R [u_{i,t}^{RF}(\mathbf{d}) | \mathcal{F}_{i,t-p-1}] &= \sum_{\mathbf{z} \in \{0,1\}^{p+2}} \frac{(Y_{i,t}(d_{1:t-p-2}^{obs}, \mathbf{d}) \mathbf{1} \{D_{i,t-p-1:t}(\mathbf{z}) = \mathbf{d}\})^2}{\pi_{i,t-p-1}(\mathbf{z})} \\
&\quad - (\Delta^{p+2} (Y_{i,t}(d_{1:t-p-2}^{obs}, \mathbf{d}) \mathbf{1} \{D_{i,t-p-1:t}(\mathbf{z}) = \mathbf{d}\}))^2 \\
&= (\gamma_{i,t}^{RF}(\mathbf{d}))^2 - (\Delta^{p+2} (Y_{i,t}(d_{1:t-p-2}^{obs}, \mathbf{d}) \mathbf{1} \{D_{i,t-p-1:t}(\mathbf{z}) = \mathbf{d}\}))^2 = (\sigma_{i,t}^{RF}(\mathbf{d}))^2
\end{aligned}$$

From the results above, we have

$$\mathbb{V}_R [u_{i,t}(\mathbf{d}) | \mathcal{F}_{i,t-p-1}] = g' \begin{pmatrix} (\sigma_{i,t}^{RF}(\mathbf{d}))^2 \\ (\sigma_{i,t}^{FS}(\mathbf{d}))^2 \end{pmatrix} g = (\sigma_{i,t}(\mathbf{d}))^2$$

Where g is the gradient of $h(x, y) = x/y$ evaluated at $(m_{i,t}^{RF}(\mathbf{d}), m_{i,t}^{FS}(\mathbf{d}))$

The variance for $\widehat{\tau}_{i,t}(\mathbf{d}, \tilde{\mathbf{d}}; p)$ is simply $(\sigma_{i,t}(\mathbf{d}))^2 + (\sigma_{i,t}(\tilde{\mathbf{d}}))^2$, which concludes the proof.

5.3 Appendix C

In order to derive the asymptotic properties of the estimator with an estimated propensity score, first I define the estimation error in the first stage as

$$u_{i,t}^{FS}(\mathbf{d}) = \widehat{m}_{i,t}^{FS}(\mathbf{d}) - m_{i,t}^{FS}(\mathbf{d}) = \Delta^{p+1} \left(\mathbf{1} \{D_{i,t-p:t}(\mathbf{z})\} (\widehat{W}_{i,t-p:t}(\mathbf{z}) - 1) \right)$$

which can be written as

$$\begin{aligned} u_{i,t}^{FS}(\mathbf{d}) &= \Delta^{p+1} \left(\mathbf{1} \{D_{i,t-p:t}(\mathbf{z})\} (W_{i,t-p:t}(\mathbf{z}) - 1) \right) \\ &+ \Delta^{p+1} \left(\mathbf{1} \{D_{i,t-p:t}(\mathbf{z})\} (\widehat{W}_{i,t-p:t}(\mathbf{z}) - W_{i,t-p:t}(\mathbf{z})) \right) \end{aligned}$$

If we assume that $\mathbb{E}_R [\widehat{\pi}_{i,t-p:t}(\mathbf{z}) | \mathcal{F}_{i,t-p-1}] = \pi_{i,t-p:t}(\mathbf{z})$, then it follows that $\mathbb{E}_R [\widehat{W}_{i,t-p:t}(\mathbf{z}) | \mathcal{F}_{i,t-p-1}] = W_{i,t-p:t}(\mathbf{z})$.

If we further assume that $\widehat{\pi}_{i,t-p:t}(\mathbf{z}) - \pi_{i,t-p:t}(\mathbf{z}) = O_p(N^{-1/2})$, then using (HAHN, 1998), (HIRANO; IMBENS; RIDDER, 2003) we obtain the following approximation:

$$\sqrt{N} (\widehat{m}_t^{FS}(\mathbf{d}) - m_t^{FS}(\mathbf{d})) = \frac{1}{\sqrt{N}} \sum_{i=1}^N \Delta^{p+1} \left(\mathbf{1} \{D_{i,t-p:t} = \mathbf{d}\} (W_{i,t-p:t}(\mathbf{z}) - 1) \right) + o_p(1)$$

A similar result can be shown for $\widehat{m}_t^{RF}(\mathbf{d})$. Therefore, it follows that

$$\frac{\sqrt{N} (\widehat{m}_t(\mathbf{d}) - m_t(\mathbf{d}))}{\sigma_t(\mathbf{d})} \xrightarrow{d} \mathcal{N}(0, 1)$$

where is obtained using the uniform Delta method and

$$\frac{\sqrt{N} \left(\widehat{\tau}_t(\mathbf{d}, \tilde{\mathbf{d}}; p) - \bar{\tau}_t(\mathbf{d}, \tilde{\mathbf{d}}; p) \right)}{\sigma_t(\mathbf{d}, \tilde{\mathbf{d}}; p)} \xrightarrow{d} \mathcal{N}(0, 1)$$

where $\sigma_t(\mathbf{d}, \tilde{\mathbf{d}}; p)$ is the same from the case with known propensity score.

The same can be shown for total lag- p dynamic causal effects, once we note that the following approximation holds:

$$\sqrt{N(T-p)} (\hat{m}^{FS}(\mathbf{d}) - m^{FS}(\mathbf{d})) = \frac{1}{\sqrt{N(T-p)}} \sum_{t=p+1}^T \sum_{i=1}^N \Delta^{p+1} (\mathbf{1}\{D_{i,t-p:t} = \mathbf{d}\} (W_{i,t-p:t}(\mathbf{z}) - 1)) + o_p(1)$$

5.4 Appendix D

As a robustness check, I estimate the results presented in Table 5 using different methods for the binarization of the instrument and the treatment. In Table 7 I present the results when treatment and instrument are binarized using the median of fines and cloud coverage instead of the mean.

Table 7 - Impulse Response Functions - Median

	AGR(2023)	lag-p			
		0	1	2	3
Point estimate	-0.107	-0.0504	-0.0151	0.0134	0.0196
95% CI	(-0.049, -0.001)	(-0.097, -0.002)	(-0.037, -0.001)	(-0.017, 0.041)	(-0.033, 0.058)
Baseline	0.0071	0.0071	0.0071	0.0071	0.0071
Observations	5210	5210	4689	4168	3647
Municipalities	521	521	521	521	521

Note: The dependent variable is the ratio between deforested area in a year and the municipality area. The set of control variables contains precipitation and temperature (weather), PRODES cloud coverage and other nonobservable areas (satellite visibility), and agricultural commodity prices.

As a final robustness check, I binarize the treatment using an indicator for municipalities that received a number of fines greater than zero. The results are presented in Table 8 below.

Table 8 - Impulse Response Functions - Greater than zero

	AGR(2023)	lag-p			
		0	1	2	3
Point estimate	-0.107	-0.0641	-0.0350	0.0134	0.0026
95% CI	(-0.049, -0.001)	(-0.123, -0.004)	(-0.070, -0.005)	(-0.085, 0.018)	(-0.049, 0.008)
Baseline	0.0071	0.0071	0.0071	0.0071	0.0071
Observations	5210	5210	4689	4168	3647
Municipalities	521	521	521	521	521

Note: The dependent variable is the ratio between deforested area in a year and the municipality area. The set of control variables contains precipitation and temperature (weather), PRODES cloud coverage and other nonobservable areas (satellite visibility), and agricultural commodity prices.

Overall, the estimates are fairly stable across specifications, with the exception of the impulse response function for the lag-1 dynamic causal effect, which is not statistically significant under the binarization procedure for Table 8.

6 Appendix to Chapter 3

6.1 Appendix A

Proof of Proposition 1

I begin with the first-stage. Consider the upper bound: $\widehat{\tau}_D^{UB}(c_1, x) = \widehat{\tau}_{D(1)}^{UB}(c_1, x) - \widehat{\tau}_{D(0)}^{LB}(c_1, x)$. It follows directly from Lemma 4 in Appendix B that

$$\sqrt{N} \left(\widehat{\tau}_D^{UB}(c_1, x) - \tau_D^{UB}(c_1, x) \right) \xrightarrow{d} \left(\widetilde{\mathbf{Z}}_D^{(1)}(d, 1, x, c_1) - \widetilde{\mathbf{Z}}_D^{(2)}(d, 0, x, c_1) \right)$$

Now we turn to the lower bound: $\widehat{\tau}_D^{LB}(c_1, x) = \max \left\{ 0, \widehat{\tau}_{D(1)}^{LB}(c_1, x) - \widehat{\tau}_{D(0)}^{UB}(c_1, x) \right\}$. For fixed d and c_1 , define the mapping

$$\begin{aligned} \phi_D^* : l^\infty(\{0, 1\} \times \{0, 1\} \times \mathcal{S}(X)) \times l^\infty(\{0, 1\} \times \mathcal{S}(X)) \times l^\infty(\mathcal{S}(X)) \\ \rightarrow l^\infty(\{0, 1\}, \mathcal{S}(X), \mathbb{R}) \end{aligned}$$

by

$$[\phi_D^*(\theta)](z, x) = \max \{0, \theta^{(1)}(d, z, x) - \theta^{(2)}(d, z, x)\}$$

which is comprised by the functionals $\delta_{D,1}^*(\theta) = 0$ and $\delta_{D,2}^*(\theta) = \theta^{(1)}(d, z, x) - \theta^{(2)}(d, z, x)$, with Hadamard derivatives respectively equal to 0 and $h^{(1)}(d, z, x) - h^{(2)}(d, z, x)$. Thus, the Hadamard derivative of ϕ_{FS}^* at θ_0 is

$$\phi_{D,\theta_0}^*(h) = \begin{pmatrix} \mathbf{1}(\delta_{D,1}^*(\theta_0) = \delta_{D,2}^*(\theta_0)) \max \{0, \delta_{D,2,\theta_0}^*(h)\} \\ + \mathbf{1}(\delta_{D,1}^*(\theta_0) > \delta_{D,2}^*(\theta_0)) \delta_{D,2,\theta_0}^*(h) \end{pmatrix}$$

Using the Delta Method for Hadamard differentiable functions, we obtain

$$\left[\sqrt{N} \left(\phi_D^*(\widehat{\theta}) - \phi_D^*(\theta_0) \right) \right] (z, x) \xrightarrow{d} \left[\phi_{D,\theta_0}^*(\widetilde{\mathbf{Z}}_D) \right] (z, x) \equiv \mathbf{Z}_{FS}^{*(2)}(d, z, x, c_1)$$

Combining the results yields

$$\sqrt{N} \begin{pmatrix} \widehat{\tau}_D^{UB}(c_1, x) - \tau_D^{UB}(c_1, x) \\ \widehat{\tau}_D^{LB}(c_1, x) - \tau_D^{LB}(c_1, x) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \left(\tilde{\mathbf{Z}}_D^{(1)}(d, 1, x, c_1) - \tilde{\mathbf{Z}}_D^{(2)}(d, 0, x, c_1) \right) \\ \mathbf{Z}_{FS}^{*(2)}(d, z, x, c_1) \end{pmatrix} \equiv \mathbf{Z}_{FS}^*(d, z, x, c_1)$$

It follows that the estimators for the unconditional bounds converge weakly to Gaussian elements. We have

$$\begin{aligned} \sqrt{N} \left(\widehat{\tau}_D^{LB}(c_1) - \tau_D^{LB}(c_1) \right) &\xrightarrow{d} \sum_{k=1}^K \left(\tilde{\mathbf{Z}}_D^{(1)}(d, 1, x, c_1) - \tilde{\mathbf{Z}}_D^{(2)}(d, 0, x, c_1) \right) + \sum_{k=1}^K \tau_D^{LB}(c_1, x_k) \mathbf{Z}_D^{(3)}(0, 0x_k) \\ &\equiv \mathbf{Z}_{FS}^{(1)}(d, z, x, c_1) \end{aligned}$$

and

$$\sqrt{N} \left(\widehat{\tau}_D^{UB}(c_1) - \tau_D^{UB}(c_1) \right) \xrightarrow{d} \sum_{k=1}^K \mathbf{Z}_{FS}^{*(2)}(d, z, x, c_1) + \sum_{k=1}^K \tau_D^{UB}(c_1, x_k) \mathbf{Z}_D^{(3)}(0, 0x_k) \equiv \mathbf{Z}_{FS}^{(2)}(d, z, x, c_1)$$

Now, turn to the reduced form. We begin with the upper bound: $\widehat{\tau}_Y^{UB}(c_2, x) = \widehat{\tau}_{Y(D(1))}^{UB}(c_2, x) - \widehat{\tau}_{Y(D(0))}^{LB}(c_2, x)$. It follows directly from Lemma 4 that

$$\sqrt{N} \left(\widehat{\tau}_Y^{UB}(c_2, x) - \tau_Y^{UB}(c_2, x) \right) \xrightarrow{d} \left(\tilde{\mathbf{Z}}_Y^{(1)}(y, 1, x, c_2) - \tilde{\mathbf{Z}}_Y^{(2)}(y, 0, x, c_2) \right)$$

Now we turn to the lower bound: $\widehat{\tau}_Y^{LB}(c_2, x) = \widehat{\tau}_{Y(D(1))}^{LB}(c_2, x) - \widehat{\tau}_{Y(D(0))}^{UB}(c_2, x)$.

It also follows directly from Lemma 4 that

$$\sqrt{N} \left(\widehat{\tau}_Y^{LB}(c_2, x) - \tau_Y^{LB}(c_2, x) \right) \xrightarrow{d} \left(\tilde{\mathbf{Z}}_Y^{(2)}(y, 1, x, c_2) - \tilde{\mathbf{Z}}_Y^{(1)}(y, 0, x, c_2) \right)$$

Hence, we have

$$\sqrt{N} \begin{pmatrix} \widehat{\tau}_Y^{LB}(c_2, x) - \tau_Y^{LB}(c_2, x) \\ \widehat{\tau}_Y^{UB}(c_2, x) - \tau_Y^{UB}(c_2, x) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \left(\tilde{\mathbf{Z}}_Y^{(1)}(y, 1, x, c_2) - \tilde{\mathbf{Z}}_Y^{(2)}(y, 0, x, c_2) \right) \\ \left(\tilde{\mathbf{Z}}_Y^{(2)}(y, 1, x, c_2) - \tilde{\mathbf{Z}}_Y^{(1)}(y, 0, x, c_2) \right) \end{pmatrix} \equiv \tilde{\mathbf{Z}}_{RF}^*(y, z, x, c_2)$$

It follows that the estimators for the unconditional bounds converge weakly to Gaussian elements. We have

$$\begin{aligned} \sqrt{N} (\widehat{\tau}_Y^{LB}(c_2) - \tau_Y^{LB}(c_2)) &\xrightarrow{d} \sum_{k=1}^K \left(\widetilde{\mathbf{Z}}_Y^{(1)}(d, 1, x, c_2) - \widetilde{\mathbf{Z}}_Y^{(2)}(d, 0, x, c_2) \right) + \sum_{k=1}^K \tau_Y^{LB}(c_1, x_k) \mathbf{Z}_Y^{(3)}(0, 0x_k) \\ &\equiv \mathbf{Z}_{RF}^{(1)}(d, z, x, c_2) \end{aligned}$$

and

$$\begin{aligned} \sqrt{N} (\widehat{\tau}_Y^{UB}(c_2) - \tau_Y^{UB}(c_2)) &\xrightarrow{d} \sum_{k=1}^K \left(\widetilde{\mathbf{Z}}_Y^{(2)}(d, 1, x, c_2) - \widetilde{\mathbf{Z}}_Y^{(1)}(d, 0, x, c_2) \right) + \sum_{k=1}^K \tau_Y^{UB}(c_2, x_k) \mathbf{Z}_Y^{(3)}(0, 0x_k) \\ &\equiv \mathbf{Z}_{RF}^{(2)}(d, z, x, c_2) \end{aligned}$$

which concludes the proof.

Proof of Theorem 1

The breakdown point c_1^* is defined implicitly by $\tau_D^{LB}(c_1^*) = \mu$. The function $\tau_D^{LB}(c_1)$ satisfies the Assumptions from van de Vaart (2000) due to Assumptions ????. Thus the mapping $\tau_D^{LB}(\cdot) \mapsto c_1^*$ is Hadamard differentiable tangentially to the set of Càdlàg functions on $[0, \overline{C}]$ with derivative equal to $\frac{-\partial h(c_1^*)}{\partial \tau_D^{LB}(c_1^*)}$.

Recall that $\sqrt{N} (\widehat{\tau}_D^{LB}(c_1) - \tau_D^{LB}(c_1))$ converges in distribution to a random element of $l^\infty([0, \overline{C}])$ with continuous paths.

Let $\tilde{c}_1^* = \inf \{c_1 \in [0, \overline{C}] : \widehat{\tau}_D^{LB}(c_1) \leq \mu\}$. Since $c_1 \in [0, \overline{C}]$ by monotonicity of $\tau_D^{LB}(\cdot)$, it follows that $\tau_D^{LB}(\overline{C}) \leq \mu$. By \sqrt{N} convergence of the First-stage bounds, we have

$$\mathbb{P}(\tau_D^{LB}(\overline{C}) \geq \mu) = \mathbb{P}\left(\sqrt{N}(\tau_D^{LB}(\overline{C}) - \mu) < \sqrt{N}(\widehat{\tau}_D^{LB}(\overline{C}) - \tau_D^{LB}(\overline{C}))\right) \rightarrow 0$$

Therefore, the set $\{c_1 \in [0, \overline{C}] : \widehat{\tau}_D^{LB}(c_1) \leq \mu\}$ is nonempty almost surely, which implies that $\mathbb{P}(\tilde{c}_1^* = \widehat{c}_1^*)$ approaches one as $N \rightarrow \infty$. Hence, it follows that

$$\begin{aligned} \sqrt{N}(\widehat{c}_1^* - c_1^*) &= \sqrt{N}(\widehat{c}_1^* - \tilde{c}_1^*) + \sqrt{N}(\tilde{c}_1^* - c_1^*) \\ &o_p(1) + \sqrt{N}(\tilde{c}_1^* - c_1^*) \xrightarrow{d} \mathbf{Z}_{FS}^{BP} \end{aligned}$$

Applying the same logic to the reduced form yields the result for \widehat{c}_2^* , which concludes the proof.

Proof of Proposition 2

Let $\widehat{\theta}_\tau = (\widehat{\tau}_Y^{LB}(c_2), \widehat{\tau}_Y^{UB}(c_2), \widehat{\tau}_D^{UB}(c_1), \widehat{\tau}_D^{LB}(c_1))$ and $\theta_\tau = (\tau_Y^{LB}(c_2), \tau_Y^{UB}(c_2), \tau_D^{UB}(c_1), \tau_D^{LB}(c_1))$. For fixed y, d, c_1, c_2 , define the mapping

$$\begin{aligned} \phi_\tau : l^\infty(\{0, 1\}^3 \times \mathcal{S}(X)) \times l^\infty(\{0, 1\}^2 \times \mathcal{S}(X)) \times l^\infty(\{0, 1\} \times \mathcal{S}(X)) \times l^\infty(\mathcal{S}(X)) \\ \rightarrow l^\infty(\{-1, 1\}, \mathcal{S}(X), \mathbb{R}^2) \end{aligned}$$

by

$$[\phi_\tau(\theta_\tau)](z, x) = \begin{pmatrix} \min \left\{ \frac{\theta_\tau^{(1)}(y, z, x, c_2)}{\theta_\tau^{(3)}(d, z, x, c_1)}, 1 \right\} \\ \max \left\{ \frac{\theta_\tau^{(2)}(y, z, x, c_2)}{\theta_\tau^{(4)}(d, z, x, c_1)}, -1 \right\} \end{pmatrix}$$

The Hadamard derivative for $[\delta_1(\theta)](z, x) = \frac{\theta^{(1)}(y, z, x, c_2)}{\theta^{(3)}(d, z, x, c_1)}$ is equal to

$$[\delta'_{1, \theta}(h)](z, x) = \frac{h^{(1)}(y, z, x, c_2)\theta^{(3)}(d, z, x, c_1) - \theta^{(1)}(y, z, x, c_2)h^{(3)}(d, z, x, c_1)}{(\theta^{(3)}(d, z, x, c_1))^2}$$

The Hadamard derivative for $[\delta_2(\theta)](z, x) = 1$ is equal to 0. The hadamard derivative for $[\delta_3(\theta)](z, x) = \frac{\theta^{(2)}(y, z, x, c_2)}{\theta^{(4)}(d, z, x, c_1)}$ is equal to

$$[\delta'_{3, \theta}(h)](z, x) = \frac{h^{(2)}(y, z, x, c_2)\theta^{(4)}(d, z, x, c_1) - \theta^{(2)}(y, z, x, c_2)h^{(4)}(d, z, x, c_1)}{(\theta^{(4)}(d, z, x, c_1))^2}$$

And the Hadamard derivative for $[\delta_4(\theta)](z, x) = -1$ is equal to 0.

Hence, the Hadamard directional derivative of ϕ_τ evaluated at $\theta_{\tau, 0}$ is

$$\phi'_{\tau, \theta_0}(h) = \begin{pmatrix} \mathbf{1}(\delta_{\tau, 1}(\theta_0) = 1) \max \{ \delta'_{\tau, 1, \theta_0}(h), 0 \} \\ \quad + \mathbf{1}(\delta_{\tau, 1}(\theta_0) < 1) \delta'_{\tau, 1, \theta_0}(h) \\ \mathbf{1}(\delta_{\tau, 3}(\theta_0) = -1) \min \{ \delta'_{\tau, 3, \theta_0}(h), 0 \} \\ \quad + \mathbf{1}(\delta_{\tau, 3}(\theta_0) > -1) \delta'_{\tau, 3, \theta_0}(h) \end{pmatrix}$$

By Proposition 1 and the Delta Method for Hadamard directionally differentiable functions,

$$\sqrt{N} \left(\phi_\tau(\widehat{\theta}) - \phi_\tau(\theta_0) \right) \xrightarrow{d} \left[\theta'_{\tau, \theta_0}(\tilde{\mathbf{Z}}) \right](z, x) \equiv \tilde{\mathbf{Z}}_\tau(z, x)$$

which yields the process $\mathbf{Z}_\tau(y, d, z, x, c_1, c_2)$.

Proof of Theorem 2

By Lemmas 3 and 4 and Proposition 1, the numerator converges uniformly over $\mathcal{C} \times \mathcal{M}$. By Proposition 1, the denominator also converges uniformly. Hence, applying the Delta Method we obtain

$$\sqrt{N} \left(\widehat{BF}(c_1, \mu) - BF(c_1, \mu) \right) \xrightarrow{d} \mathbf{Z}_{BF}(c_1, \mu)$$

which concludes the proof.

Proof of Theorem 3

I show the partial identification of $\tau_{Y(D(1))}(x)$ and $\tau_{Y(D(0))}(x)$. The results for $\tau_{D(1)}(x)$ and $\tau_{D(0)}(x)$ are analogous.

I begin with $\tau_{Y(D(1))}(x)$. We have

$$\tau_{Y(D(1))}(x) \equiv \mathbb{P}(Y(D(1)) = 1 | X = x) = p_{Y|1,x} p_{1|x} + \mathbb{P}(Y(D(1)) | Z = 0, X = x) p_{0|x}$$

Note that by the Law of Total Probabilities, we have

$$\begin{aligned} & \mathbb{P}(Y(D(1)) = 1 | Z = 0, X = x) \\ &= \mathbb{P}(Y(D(1)) = 1, D(1) = 1 | Z = 0, X = x) + \mathbb{P}(Y(D(1)) = 1, D(1) = 0 | Z = 0, X = x) \end{aligned}$$

First, consider $\mathbb{P}(Y(D(1)) = 1, D(1) = 1 | Z = 0, X = x)$. This joint probability can be expressed as

$$\begin{aligned} & \mathbb{P}(Y(D(1)) = 1, D(1) = 1 | Z = 0, X = x) \\ &= \frac{\mathbb{P}(Z = 0 | Y(D(1)) = 1, D(1) = 1, X = x) \mathbb{P}(Y(D(1)) = 1, D(1) = 1 | X = x)}{p_{0|x}} \\ &\leq \frac{(1 - \mathbb{P}(Z = 1 | Y(D(1)) = 1, D(1) = 1, X = x)) p_{Y,D|1,x} p_{1|x}}{(p_{1|x} - c) p_{0|x}} \\ &\leq \frac{\{1 - (p_{1|x} - c)\} p_{Y,D|1,x} p_{1|x}}{(p_{1|x} - c) p_{0|x}} \end{aligned}$$

Where the second equality follows from Proposition 5 of (MASTEN; POIRIER, 2018) and the third equality follows from joint c -dependence.

Similarly, under the same procedure it can be shown that

$$\mathbb{P}(Y(D(1)) = 1, D(1) = 0 | Z = 0, X = x) \leq \frac{\{1 - (p_{1|x} - c)\} p_{Y,1-D|1,x} p_{1|x}}{(p_{1|x} - c) p_{0|x}}$$

Combining the results yields

$$\tau_{Y(D(1))}(x) \leq p_{Y|1,x} p_{1|x} + \left[\frac{\{1 - (p_{1|x} - c)\} p_{Y|1,x} p_{1|x}}{(p_{1|x} - c) p_{0|x}} \right] p_{0|x}$$

Moreover,

$$\tau_{Y(D(1))}(x) = p_{Y|1,x} p_{1|x} + \mathbb{P}(Y(D(1)) = 1 | Z = 0, X = x) p_{0|x} \leq p_{Y|1,x} p_{1|x} + (1 - p_{1|x})$$

Combining these inequalities yields the upper bound. For the lower bound, note that under joint c -dependence and Proposition 5 from (MASTEN; POIRIER, 2018),

$$\mathbb{P}(Y(D(1)) = 1, D(1) = 1 | Z = 0, X = x) \geq \frac{\{1 - (p_{1|x} + c)\} p_{Y,D|1,x} p_{1|x}}{(p_{1|x} + c) p_{0|x}}$$

and

$$\mathbb{P}(Y(D(1)) = 1, D(1) = 0 | Z = 0, X = x) \geq \frac{\{1 - (p_{1|x} + c)\} p_{Y,1-D|1,x} p_{1|x}}{(p_{1|x} + c) p_{0|x}}$$

Combining the results yields

$$\tau_{Y(D(1))}(x) \leq p_{Y|1,x} p_{1|x} + \left[\frac{\{1 - (p_{1|x} + c)\} p_{Y|1,x} p_{1|x}}{(p_{1|x} + c) p_{0|x}} \right] p_{0|x}$$

Moreover,

$$\tau_{Y(D(1))}(x) = p_{Y|1,x} p_{1|x} + \mathbb{P}(Y(D(1)) = 1 | Z = 0, X = x) p_{0|x} \geq p_{Y|1,x} p_{1|x}$$

Combining the result yields the lower bound. When it comes to $\tau_{Y(D(0))}(x)$, under the same reasoning it can be shown that

$$\tau_{Y(D(0))}(x) \leq \left[\frac{(p_{1|x} + c) p_{Y|0,x} p_{0|x}}{(p_{0|x} - c) p_{1|x}} \right] p_{1|x} + p_{Y|0,x} p_{0,x}$$

and

$$\tau_{Y(D(0))}(x) \leq p_{1|x} + p_{Y|0,x}p_{0|x}$$

which yields the upper bound. Also, one can show that

$$\tau_{Y(D(0))}(x) \geq \left[\frac{(p_{1|x} - c)p_{Y|0,x}p_{0|x}}{(p_{0|x} + c)p_{1|x}} \right] p_{1|x} + p_{Y|0,x}p_{0|x}$$

and

$$\tau_{Y(D(0))}(x) \geq p_{Y|0,x}p_{0|x}$$

which yields the lower bound.

The proof of sharpness is the same as in Proposition 5 from (MASTEN; POIRIER, 2018), but adapted to the case of joint distribution of potential quantities.

Proof of Proposition 3

Define $\tau_D^{UB}(c, x) = \tau_{D(1)}^{UB}(c, x) - \tau_{D(0)}^{LB}(c, x)$, $\tau_D^{LB}(c, x) = \tau_{D(1)}^{LB}(c, x) - \tau_{D(0)}^{UB}(c, x)$, $\tau_Y^{UB}(c, x) = \tau_{Y(D(1))}^{UB}(c, x) - \tau_{Y(D(0))}^{LB}(c, x)$ and $\tau_Y^{LB}(c, x) = \tau_{Y(D(1))}^{LB}(c, x) - \tau_{Y(D(0))}^{UB}(c, x)$. Furthermore, define $\widehat{\tau}_D^{UB}(c, x)$, $\widehat{\tau}_D^{LB}(c, x)$, $\widehat{\tau}_Y^{UB}(c, x)$, $\widehat{\tau}_Y^{LB}(c, x)$ as the estimators for these quantities, respectively. It follows from Lemma 6 that

$$\sqrt{N} \begin{pmatrix} \widehat{\tau}_D^{UB}(c, x) - \tau_D^{UB}(c, x) \\ \widehat{\tau}_D^{LB}(c, x) - \tau_D^{LB}(c, x) \\ \widehat{\tau}_Y^{UB}(c, x) - \tau_Y^{UB}(c, x) \\ \widehat{\tau}_Y^{LB}(c, x) - \tau_Y^{LB}(c, x) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \mathbf{Z}_{j,FS}^{(1)}(d, 1, x, c) - \mathbf{Z}_{j,FS}^{(2)}(d, 0, x, c) \\ \mathbf{Z}_{j,FS}^{(2)}(d, 1, x, c) - \mathbf{Z}_{j,FS}^{(1)}(d, 0, x, c) \\ \mathbf{Z}_{j,RF}^{(1)}(d, 1, x, c) - \mathbf{Z}_{j,RF}^{(2)}(d, 0, x, c) \\ \mathbf{Z}_{j,RF}^{(2)}(d, 1, x, c) - \mathbf{Z}_{j,RF}^{(1)}(d, 0, x, c) \end{pmatrix}$$

And thus,

$$\begin{aligned}
& \sqrt{N} \begin{pmatrix} \widehat{\tau}_D^{UB}(c) - \tau_D^{UB}(c) \\ \widehat{\tau}_D^{LB}(c) - \tau_D^{LB}(c) \\ \widehat{\tau}_Y^{UB}(c) - \tau_Y^{UB}(c) \\ \widehat{\tau}_Y^{LB}(c) - \tau_Y^{LB}(c) \end{pmatrix} \\
& \xrightarrow{d} \begin{pmatrix} \sum_{k=1}^K q_{x_k} \left(\mathbf{Z}_{j,FS}^{(1)}(d, 1, x_k, c) - \mathbf{Z}_{j,FS}^{(2)}(d, 0, x_k, c) \right) + \sum_{k=1}^K \tau_D^{UB}(c, x_k) \mathbf{Z}_j^{(3)}(0, 0, 0, x_k) \\ \sum_{k=1}^K q_{x_k} \left(\mathbf{Z}_{j,FS}^{(2)}(d, 1, x_k, c) - \mathbf{Z}_{j,FS}^{(1)}(d, 0, x_k, c) \right) + \sum_{k=1}^K \tau_D^{LB}(c, x_k) \mathbf{Z}_j^{(3)}(0, 0, 0, x_k) \\ \sum_{k=1}^K q_{x_k} \left(\mathbf{Z}_{j,RF}^{(1)}(y, 1, x_k, c) - \mathbf{Z}_{j,RF}^{(2)}(y, 0, x_k, c) \right) + \sum_{k=1}^K \tau_Y^{UB}(c, x_k) \mathbf{Z}_j^{(3)}(0, 0, 0, x_k) \\ \sum_{k=1}^K q_{x_k} \left(\mathbf{Z}_{j,RF}^{(2)}(y, 1, x_k, c) - \mathbf{Z}_{j,RF}^{(1)}(y, 0, x_k, c) \right) + \sum_{k=1}^K \tau_Y^{LB}(c, x_k) \mathbf{Z}_j^{(3)}(0, 0, 0, x_k) \end{pmatrix} \\
& \equiv \mathbf{Z}_j^*(y, d, z, x, c)
\end{aligned}$$

$$\text{Let } \widehat{\Omega}_\tau = (\widehat{\tau}_Y^{UB}(c), \widehat{\tau}_Y^{LB}(c), \widehat{\tau}_D^{UB}(c), \widehat{\tau}_D^{LB}(c)) \text{ and } \Omega_\tau = (\tau_Y^{UB}(c), \tau_Y^{LB}(c), \tau_D^{UB}(c), \tau_D^{LB}(c)).$$

For fixed y, d and c , define the mapping

$$\begin{aligned}
\psi_\tau : l^\infty(\{0, 1\}^3 \times \mathcal{S}(X)) \times l^\infty(\{0, 1\}^2 \times \mathcal{S}(X)) \times l^\infty(\{0, 1\} \times \mathcal{S}(X)) \times l^\infty(\mathcal{S}(X)) \\
\rightarrow l^\infty(\{-1, 1\}, \mathcal{S}(X), \mathbb{R}^2)
\end{aligned}$$

by

$$[\psi_\tau(\Omega_\tau)](z, x) = \begin{pmatrix} \min \left\{ \frac{\Omega^{(1)}(y, z, x)}{\Omega^{(3)}(d, z, x)}, 1 \right\} \\ \max \left\{ \frac{\Omega^{(2)}(y, z, x)}{\Omega^{(4)}(d, z, x)}, -1 \right\} \end{pmatrix}$$

The mapping is comprise with four elements. We have

$$[\delta_{\tau,1}(\Omega)] = \frac{\Omega^{(2)}(y, z, x)}{\Omega^{(4)}(d, z, x)}$$

with Hadamard derivative equal to

$$\left[\delta'_{\tau,1,\Omega}(h) \right] = \frac{h^{(2)}(y, z, x)}{\Omega^{(4)}(d, z, x)} - \frac{h^{(4)}(d, z, x) \Omega^{(2)}(y, z, x)}{(\Omega^{(4)}(d, z, x))^2},$$

$$[\delta_{\tau,2}(\Omega)] = -1$$

with Hadamard derivative equal to

$$\left[\delta'_{\tau,2,\Omega}(h) \right] = 0,$$

$$[\delta_{\tau,3}(\Omega)] = \frac{\Omega^{(1)}(y, z, x)}{\Omega^{(3)}(d, z, x)}$$

with Hadamard derivative equal to

$$\left[\delta'_{\tau,3,\Omega}(h) \right] = \frac{h^{(1)}(y, z, x)}{\Omega^{(3)}(d, z, x)} - \frac{h^{(3)}(d, z, x)\Omega^{(1)}(y, z, x)}{(\Omega^{(3)}(d, z, x))^2},$$

and

$$[\delta_{\tau,4}(\Omega)] = 1$$

with Hadamard derivative equal to

$$\left[\delta'_{\tau,4,\Omega}(h) \right] = 0,$$

Therefore, the Hadamard directional derivate of ψ_τ evaluated at Ω_τ is

$$\psi'_{\tau,\Omega_\tau}(h) = \begin{pmatrix} \mathbf{1} (\delta_3(\Omega_\tau) = 1) \min \{ \delta'_{3,\tau,\Omega_\tau}(h), 0 \} \\ \quad + \mathbf{1} (\delta_3(\Omega_\tau) > 1) \delta'_{3,\tau,\Omega_\tau}(h) \\ \mathbf{1} (\delta_1(\Omega_\tau) = -1) \min \{ \delta'_{1,\tau,\Omega_\tau}(h), 0 \} \\ \quad + \mathbf{1} (\delta_1(\Omega_\tau) > -1) \delta'_{1,\tau,\Omega_\tau}(h) \end{pmatrix}$$

It follows from the Delta Method that

$$\left[\sqrt{N} \left(\psi_\tau(\widehat{\Omega}) - \psi_\tau(\Omega_\tau) \right) \right] (z, x) \xrightarrow{d} \left[\psi'_{\tau,\Omega_\tau}(\mathbf{Z}_j^*) \right] (z, x) \equiv \mathbf{Z}_{j,\tau}(z, x)$$

which yields the process $\mathbf{Z}_{j,\tau}(y, d, z, x, c)$

Proof of Theorem 4

The steps of the proof are the same as in the proof of Theorem 1.

6.2 Appendix B

In this section I derive the asymptotic properties of the estimators used as plug-ins for the estimations of the bounds. I show that these estimators converge uniformly to mean-zero Gaussian processes.

Lemma 3. *Suppose Assumptions 6 and 7 hold. Then,*

$$\sqrt{N} \begin{pmatrix} \widehat{p}_{D|z,x} - p_{D|z,x} \\ \widehat{p}_{z|x} - p_{z|x} \\ \widehat{q}_x - q_x \end{pmatrix} \xrightarrow{d} \mathbf{Z}_D(d, z, x)$$

and

$$\sqrt{N} \begin{pmatrix} \widehat{p}_{Y|z,x} - p_{Y|z,x} \\ \widehat{p}_{z|x} - p_{z|x} \\ \widehat{q}_x - q_x \end{pmatrix} \xrightarrow{d} \mathbf{Z}_Y(y, z, x)$$

which are mean-zero Gaussian Processes in $l^\infty(\{0, 1\} \times \{0, 1\} \times \mathcal{S}(X), \mathbb{R}^3)$ with covariance kernels respectively equal to $\mathbb{E}[\Sigma_D(d, z, x)\Sigma_D(d, \tilde{z}, \tilde{x})']$ and $\mathbb{E}[\Sigma_Y(y, z, x)\Sigma_Y(y, \tilde{z}, \tilde{x})']$

Proof:

By a second-order Taylor Expansion,

$$\begin{aligned} \widehat{p}_{D|z,x} - p_{D|z,x} &= \frac{\frac{1}{N} \sum_{i=1}^N \mathbf{1}(D_i = 1) \mathbf{1}(Z_i = z, X_i = x)}{\frac{1}{N} \sum_{i=1}^N \mathbf{1}(Z_i = z, X_i = x)} - \frac{\mathbb{P}(D_i = 1, Z_i = z, X_i = x)}{\mathbb{P}(Z_i = z, X_i = x)} \\ &= \frac{\frac{1}{N} \sum_{i=1}^N \mathbf{1}(D_i = 1) \mathbf{1}(Z_i = z, X_i = x) - \mathbb{P}(D_i = 1, Z_i = z, X_i = x)}{\mathbb{P}(Z_i = z, X_i = x)} \\ &\quad - \frac{p_{D|z,x}}{\mathbb{P}(Z_i = z, X_i = x)} \left(\frac{1}{N} \sum_{i=1}^N \mathbf{1}(Z_i = z, X_i = x) - \mathbb{P}(Z_i = z, X_i = x) \right) \\ &\quad + O_p \left[\left(\frac{1}{N} \sum_{i=1}^N \mathbf{1}(D_i = 1) \mathbf{1}(Z_i = z, X_i = x) - p_{D|z,x} \mathbb{P}(Z_i = z, X_i = x) \right) \right. \\ &\quad \quad \left. \cdot \left(\frac{1}{N} \sum_{i=1}^N \mathbf{1}(Z_i = z, X_i = x) - \mathbb{P}(Z_i = z, X_i = x) \right) \right] \\ &\quad + O_p \left[\left(\frac{1}{N} \sum_{i=1}^N \mathbf{1}(Z_i = z, X_i = x) - \mathbb{P}(Z_i = z, X_i = x) \right)^2 \right] \end{aligned}$$

By standard bracket entropy results (van der Vaart, 2000), the functions classes given by $\{\mathbf{1}(D_i = 1) \mathbf{1}(Z_i = z, X_i = x) : z \in \{0, 1\}, x \in \mathcal{S}(X)\}$ and $\{\mathbf{1}(Z_i = z, X_i = x) : z \in \{0, 1\}, x \in \mathcal{S}(X)\}$ are both P-Donsker. Hence, the residuals of order $O_p(N^{-1})$ uniformly over $(d, z, x) \in 1 \times \{0, 1\} \times \mathcal{S}(X)$. Using Slutkys's Theorem, we obtain the following asymptotically linear representation:

$$\widehat{p}_{D|z,x} - p_{D|z,x} = \frac{1}{N} \frac{\sum_{i=1}^N (\mathbf{1}(Z_i = z, X_i = x) (\mathbf{1}(D_i = 1) - p_{D|z,x}))}{\mathbb{P}(Z_i = z, X_i = x)} + o_p(N^{-1/2})$$

Using the same bracket entropy arguments, it follows that the linear representation is also P-Donsker. Hence, $\sqrt{N}(\widehat{p}_{D|z,x} - p_{D|z,x})$ converges to a mean-zero gaussian process with continuous paths. Using similar arguments, we obtain

$$\widehat{p}_{z|x} - p_{z|x} = \frac{1}{N} \frac{\sum_{i=1}^N \mathbf{1}(X_i = x) (\mathbf{1}(Z_i = z) - p_{z|x})}{q_x} + o_p(N^{-1/2})$$

and

$$\widehat{q}_x - q_x = \frac{1}{N} \sum_{i=1}^N \mathbf{1}(X_i = x) - q_x + o_p(N^{-1/2})$$

The covariance kernel Σ_{FS} is calculated as follows

$$\begin{aligned} & [\Sigma_{FS}(z, x, \tilde{z}, \tilde{x})]_{1,1} \\ &= \mathbb{E} \left[\frac{\mathbf{1}(Z_i = z, X_i = x) \mathbf{1}(Z_i = \tilde{z}, X_i = \tilde{x}) (\mathbf{1}(D_i = 1) - p_{D|z,x}) (\mathbf{1}(D_i = 1) - p_{D|\tilde{z},\tilde{x}})}{\mathbb{P}(Z_i = z, X_i = x) \mathbb{P}(Z_i = \tilde{z}, X_i = \tilde{x})} \right], \end{aligned}$$

$$[\Sigma_{FS}(z, x, \tilde{z}, \tilde{x})]_{1,2} = \mathbb{E} \left[\frac{\mathbf{1}(Z_i = z, X_i = \tilde{x}) (\mathbf{1}(Z_i = \tilde{z}) - p_{\tilde{z}|\tilde{x}}) (\mathbf{1}(D_i = 1) - p_{D|z,x})}{p_{z|x} q_x q_{\tilde{x}}} \right] = 0,$$

$$[\Sigma_{FS}(z, x, \tilde{z}, \tilde{x})]_{1,3} = \mathbb{E} \left[\frac{(\mathbf{1}(X_i = \tilde{x}) - q_{\tilde{x}}) \mathbf{1}(Z_i = z, X_i = x) (\mathbf{1}(D_i = 1) - p_{D|z,x})}{\mathbb{P}(Z_i = z, X_i = x)} \right] = 0,$$

$$[\Sigma_{FS}(z, x, \tilde{z}, \tilde{x})]_{2,1} = [\Sigma_{FS}(z, x, \tilde{z}, \tilde{x})]_{1,2} = 0,$$

$$[\Sigma_{FS}(z, x, \tilde{z}, \tilde{x})]_{2,2} = \mathbb{E} \left[\mathbf{1}(X_i = x) \mathbf{1}(X_i = \tilde{x}) (\mathbf{1}(Z_i = z) - p_{z|x}) (\mathbf{1}(Z_i = \tilde{z}) - p_{\tilde{z}|\tilde{x}}) \right],$$

$$[\Sigma_{FS}(z, x, \tilde{z}, \tilde{x})]_{2,3} = \mathbb{E} \left[\frac{(\mathbf{1}(X_i = \tilde{x}) - q_{\tilde{x}})(\mathbf{1}(Z_i = z) - p_{z|x})}{q_x} \right] = 0,$$

$$[\Sigma_{FS}(z, x, \tilde{z}, \tilde{x})]_{3,1} = [\Sigma_{FS}(z, x, \tilde{z}, \tilde{x})]_{1,3} = 0,$$

$$[\Sigma_{FS}(z, x, \tilde{z}, \tilde{x})]_{3,2} = [\Sigma_{FS}(z, x, \tilde{z}, \tilde{x})]_{2,3} = 0,$$

$$[\Sigma_{FS}(z, x, \tilde{z}, \tilde{x})]_{3,3} = \mathbb{E} [(\mathbf{1}(X_i = \tilde{x}) - q_{\tilde{x}})(\mathbf{1}(X_i = x) - q_x)]$$

Repeating the procedure for the estimators used in the plug-in of the reduced form yields the asymptotic distribution $\mathbf{Z}_Y(y, z, x)$, which concludes the proof.

The following lemma provides the asymptotic distributions for the bounds of potential quantities.

Lemma 4. *Suppose Assumptions 1-3 and 6-8 hold. Then,*

$$\sqrt{N} \begin{pmatrix} \widehat{\tau}_{D(z)}^{LB}(c_1, x) - \tau_{D(z)}^{LB}(c_1, x) \\ \widehat{\tau}_{D(z)}^{UB}(c_1, x) - \tau_{D(z)}^{UB}(c_1, x) \end{pmatrix} \xrightarrow{d} \tilde{\mathbf{Z}}_D(d, z, x, c_1)$$

and

$$\sqrt{N} \begin{pmatrix} \widehat{\tau}_{Y(D(z))}^{LB}(c_2, x) - \tau_{Y(D(z))}^{LB}(c_2, x) \\ \widehat{\tau}_{Y(D(z))}^{UB}(c_2, x) - \tau_{Y(D(z))}^{UB}(c_2, x) \end{pmatrix} \xrightarrow{d} \tilde{\mathbf{Z}}_Y(y, z, x, c_2)$$

both tight elements of $l^\infty(\{0, 1\} \times \{0, 1\} \times \mathcal{S}(X), \mathbb{R}^2)$.

Proof:

Let $\theta_0 = (p_{D|z,x}, p_{z|x}, q_x)$ and $\widehat{\theta} = (\widehat{p}_{D|z,x}, \widehat{p}_{z|x}, \widehat{q}_x)$. For a fixed d and fixed c_1 , define the mapping

$$\begin{aligned} \phi_D : l^\infty(\{0, 1\} \times \{0, 1\} \times \mathcal{S}(X)) \times l^\infty(\{0, 1\} \times \mathcal{S}(X)) \times l^\infty(\mathcal{S}(X)) \\ \rightarrow l^\infty(\{0, 1\}, \mathcal{S}(X), \mathbb{R}^2) \end{aligned}$$

by

$$[\phi_D(\theta)](z, x) = \left(\begin{array}{c} \min \left\{ \frac{\theta^{(1)}(d, z, x)\theta^{(2)}(z, x)}{\theta^{(2)}(z, x) - c_1}, \theta^{(1)}(d, z, x)\theta^{(2)}(z, x) + (1 - \theta^{(2)}(z, x)) \right\} \\ \max \left\{ \frac{\theta^{(1)}(d, z, x)\theta^{(2)}(z, x)}{\theta^{(2)}(z, x) + c_1}, \theta^{(1)}(d, z, x)\theta^{(2)}(z, x) \right\} \end{array} \right)$$

where $\theta^{(j)}$ is the j -th component of θ . Note that

$$[\phi_D(\theta_0)](z, x) = \left(\begin{array}{c} \tau_{D(z)}^{UB}(c_1, x) \\ \tau_{D(z)}^{LB}(c_1, x) \end{array} \right)$$

The mapping ϕ_D is comprised with max and min operators, along with four other functions. The maps $(a_1, a_2) \mapsto \min \{a_1, a_2\}$ and $(a_1, a_2) \mapsto \max \{a_1, a_2\}$ have Hadamard directional derivatives at (a_1, a_2) respectively equal to

$$h \mapsto \begin{cases} h^{(1)}, a_1 < a_2 \\ \min \{h^{(1)}, h^{(2)}\}, a_1 = a_2 \\ h^{(2)}, a_1 > a_2 \end{cases}$$

and

$$h \mapsto \begin{cases} h^{(2)}, a_1 < a_2 \\ \max \{h^{(1)}, h^{(2)}\}, a_1 = a_2 \\ h^{(1)}, a_1 > a_2 \end{cases}$$

where $h \in \mathbb{R}^2$. We begin by computing the Hadamard derivative of these functions with respect to θ and proceed by using Fang and Santos (2019) and the Chain rule for Hadamard differentiable functions to obtain the derivative of ϕ_D .

First, consider $[\delta_{D,1}(\theta)](z, x) = \frac{\theta^{(1)}(d, z, x)\theta^{(2)}(z, x)}{\theta^{(2)}(z, x) + c_1}$, which has Hadamard derivative equal to

$$[\delta'_{D,1,\theta}(h)](z, x) = \frac{\theta^{(1)}(d, z, x)h^{(2)}(z, x) + h^{(1)}(d, z, x)\theta^{(2)}(z, x)}{\theta^{(2)}(z, x) + c_1} - \frac{\theta^{(1)}(d, z, x)\theta^{(2)}(z, x)h^{(2)}(z, x)}{(\theta^{(2)}(z, x) + c_1)^2}$$

Next, $[\delta_{D,2}(\theta)](z, x) = \theta^{(1)}(d, z, x)\theta^{(2)}(z, x)$ has Hadamard derivative equal to

$$[\delta'_{D,2,\theta}(h)](z, x) = h^{(1)}(d, z, x)\theta^{(2)}(z, x) + \theta^{(1)}(d, z, x)h^{(2)}(z, x)$$

Next, we turn to the functionals inside the min operator. First, we have $[\delta_{D,3}(\theta)](z, x) = \frac{\theta^{(1)}(d, z, x)\theta^{(2)}(z, x)}{\theta^{(2)}(z, x) - c_1}$, which has Hadamard derivative equal to

$$\left[\delta'_{D,3,\theta}(h) \right](z, x) = \frac{\theta^{(1)}(d, z, x)h^{(2)}(z, x) + h^{(1)}(d, z, x)\theta^{(2)}(z, x)}{\theta^{(2)}(z, x) - c_1} - \frac{\theta^{(1)}(d, z, x)\theta^{(2)}(z, x)h^{(2)}(z, x)}{(\theta^{(2)}(z, x) - c_1)^2}$$

Finally, $[\delta_{D,4}(\theta)](z, x) = h^{(1)}(d, z, x)\theta^{(2)}(z, x) + h^{(2)}(z, x)(\theta^{(1)}(d, z, x) - 1)$

Using this notation, we write the functional ϕ_D as

$$\phi_D(\theta) = \begin{pmatrix} \min \{ \delta_{D,3}(\theta), \delta_{D,4}(\theta) \} \\ \max \{ \delta_{D,1}(\theta), \delta_{D,2}(\theta) \} \end{pmatrix}$$

Using the chain rule (Masten and Poirier, 2020), the Hadamard derivative of ϕ_D at θ_0 is

$$\phi'_{D,\theta_0}(h) = \begin{pmatrix} \mathbf{1}(\delta_{D,3}(\theta_0) < \delta_{D,4}(\theta_0)) \delta'_{D,4,\theta_0}(h) \\ +\mathbf{1}(\delta_{D,3}(\theta_0) = \delta_{D,4}(\theta_0)) \min \{ \delta'_{D,3,\theta_0}(h), \delta'_{D,4,\theta_0}(h) \} \\ +\mathbf{1}(\delta_{D,3}(\theta_0) > \delta_{D,4}(\theta_0)) \delta'_{D,3,\theta_0}(h) \\ \\ \mathbf{1}(\delta_{D,1}(\theta_0) < \delta_{D,2}(\theta_0)) \delta'_{D,1,\theta_0}(h) \\ +\mathbf{1}(\delta_{D,1}(\theta_0) = \delta_{D,2}(\theta_0)) \max \{ \delta'_{D,1,\theta_0}(h), \delta'_{D,2,\theta_0}(h) \} \\ +\mathbf{1}(\delta_{D,1}(\theta_0) > \delta_{D,2}(\theta_0)) \delta'_{D,2,\theta_0}(h) \end{pmatrix}$$

By Lemma 3, $\sqrt{N}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathbf{Z}_D(d, x)$. Using the Delta Method for Hadamard differentiable functions, we obtain

$$\left[\sqrt{N}(\phi_D(\hat{\theta}) - \phi_D(\theta_0)) \right](z, x) \xrightarrow{d} \left[\phi'_{D,\theta_0}(\mathbf{Z}_D) \right](z, x) \equiv \tilde{\mathbf{Z}}_D(z, x)$$

which concludes the proof for the bounds of potential treatments. Now, consider the bounds for potential outcomes.

Let $\theta_0 = (p_{Y|z,x}, p_{z|x}, q_x)$ and $\hat{\theta} = (\hat{p}_{Y|z,x}, \hat{p}_{z|x}, \hat{q}_x)$. For a fixed y and fixed c_2 , define the mapping

$$\begin{aligned} \phi_Y &: l^\infty(\{0, 1\} \times \{0, 1\} \times \mathcal{S}(X)) \times l^\infty(\{0, 1\} \times \mathcal{S}(X)) \times l^\infty(\mathcal{S}(X)) \\ &\rightarrow l^\infty(\{0, 1\}, \mathcal{S}(X), \mathbb{R}^2) \end{aligned}$$

by

$$[\phi_Y(\theta)](z, x) = \left(\begin{array}{c} \min \left\{ \frac{\theta^{(1)}(y, z, x)\theta^{(2)}(z, x)}{\theta^{(2)}(z, x) - c_2}, \theta^{(1)}(y, z, x)\theta^{(2)}(z, x) + (1 - \theta^{(2)}(z, x)) \right\} \\ \max \left\{ \frac{\theta^{(1)}(y, z, x)\theta^{(2)}(z, x)}{\theta^{(2)}(z, x) + c_2}, \theta^{(1)}(y, z, x)\theta^{(2)}(z, x) \right\} \end{array} \right)$$

where $\theta^{(j)}$ is the j -th component of θ . Note that

$$[\phi_Y(\theta_0)](z, x) = \left(\begin{array}{c} \tau_{Y(D(z))}^{UB}(c_2, x) \\ \tau_{Y(D(z))}^{LB}(c_2, x) \end{array} \right)$$

The mapping ϕ_Y is comprised with max and min operators, along with four other functions.

We begin by computing the Hadamard derivative of these functions with respect to θ and proceed by using Fang and Santos (2019) and the Chain rule for Hadamard differentiable functions to obtain the derivative of ϕ_Y .

First, consider $[\delta_{Y,1}(\theta)](z, x) = \frac{\theta^{(1)}(y, z, x)\theta^{(2)}(z, x)}{\theta^{(2)}(z, x) + c_2}$, which has Hadamard derivative equal to

$$[\delta'_{Y,1,\theta}(h)](z, x) = \frac{\theta^{(1)}(y, z, x)h^{(2)}(z, x) + h^{(1)}(y, z, x)\theta^{(2)}(z, x)}{\theta^{(2)}(z, x) + c_2} - \frac{\theta^{(1)}(y, z, x)\theta^{(2)}(z, x)h^{(2)}(z, x)}{(\theta^{(2)}(z, x) + c_2)^2}$$

Next, $[\delta_{Y,2}(\theta)](z, x) = \theta^{(1)}(y, z, x)\theta^{(2)}(z, x)$ has Hadamard derivative equal to

$$[\delta'_{Y,2,\theta}(h)](z, x) = h^{(1)}(y, z, x)\theta^{(2)}(z, x) + \theta^{(1)}(y, z, x) + h^{(2)}(z, x)$$

Next, we turn to the functionals inside the min operator. First, we have $[\delta_{Y,3}(\theta)](z, x) = \frac{\theta^{(1)}(y, z, x)\theta^{(2)}(z, x)}{\theta^{(2)}(z, x) - c_2}$, which has Hadamard derivative equal to

$$[\delta'_{Y,3,\theta}(h)](z, x) = \frac{\theta^{(1)}(y, z, x)h^{(2)}(z, x) + h^{(1)}(y, z, x)\theta^{(2)}(z, x)}{\theta^{(2)}(z, x) - c_2} - \frac{\theta^{(1)}(y, z, x)\theta^{(2)}(z, x)h^{(2)}(z, x)}{(\theta^{(2)}(z, x) - c_2)^2}$$

Finally, $[\delta_{Y,4}(\theta)](z, x) = h^{(1)}(y, z, x)\theta^{(2)}(z, x) + h^{(2)}(z, x)(\theta^{(1)}(y, z, x) - 1)$

Using this notation, we write the functional ϕ_{RF} as

$$\phi_Y(\theta) = \left(\begin{array}{c} \min \{ \delta_{Y,3}(\theta), \delta_{Y,4}(\theta) \} \\ \max \{ \delta_{Y,1}(\theta), \delta_{Y,2}(\theta) \} \end{array} \right)$$

Using the chain rule (Masten and Poirier, 2020), the Hadamard derivative of ϕ_Y at θ_0 is

$$\phi'_{Y,\theta_0}(h) = \begin{pmatrix} \mathbf{1}(\delta_{Y,3}(\theta_0) < \delta_{Y,4}(\theta_0)) \delta'_{Y,4,\theta_0}(h) \\ +\mathbf{1}(\delta_{Y,3}(\theta_0) = \delta_{Y,4}(\theta_0)) \min\{\delta'_{Y,3,\theta_0}(h), \delta'_{Y,4,\theta_0}(h)\} \\ +\mathbf{1}(\delta_{Y,3}(\theta_0) > \delta_{Y,4}(\theta_0)) \delta'_{Y,3,\theta_0}(h) \\ \\ \mathbf{1}(\delta_{Y,1}(\theta_0) < \delta_{Y,2}(\theta_0)) \delta'_{Y,1,\theta_0}(h) \\ +\mathbf{1}(\delta_{Y,1}(\theta_0) = \delta_{Y,2}(\theta_0)) \max\{\delta'_{Y,1,\theta_0}(h), \delta'_{Y,2,\theta_0}(h)\} \\ +\mathbf{1}(\delta_{Y,1}(\theta_0) > \delta_{Y,2}(\theta_0)) \delta'_{Y,2,\theta_0}(h) \end{pmatrix}$$

By Lemma 3, $\sqrt{N}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathbf{Z}_{RF}(d, x)$. Using the Delta Method for Hadamard differentiable functions, we obtain

$$\left[\sqrt{N}(\phi_Y(\hat{\theta}) - \phi_Y(\theta_0)) \right](z, x) \xrightarrow{d} \left[\phi'_{Y,\theta_0}(\mathbf{Z}_Y) \right](z, x) \equiv \tilde{\mathbf{Z}}_Y(z, x)$$

which concludes the proof.

The next lemma provides the asymptotic properties of the nonparametric estimators of conditional probabilities which are used as plug-ins in the estimator of the LATE bounds under joint c -dependence:

Lemma 5. *Suppose Assumptions 6 and 7 hold. Then,*

$$\sqrt{N} \begin{pmatrix} \hat{p}_{Y|z,x} - p_{Y|z,x} \\ \hat{p}_{D|z,x} - p_{D|z,x} \\ \hat{p}_{z|x} - p_{z|x} \\ \hat{q}_x - q_x \end{pmatrix} \rightarrow \mathbf{Z}_j(y, d, z, x)$$

which is a mean-zero Gaussian processes in $l^\infty(\{0, 1\}^3 \times \mathcal{S}(X), \mathbb{R}^3)$ with covariance kernel equal to $\Sigma_j = \mathbb{E}[\mathbf{Z}_j(y, d, z, x)\mathbf{Z}_j(y, d, \tilde{z}, \tilde{x})']$

Proof:

Using a Taylor expansion and bracket entropy arguments as in Lemma 3, the following asymptotically linear representations are obtained:

$$\begin{aligned}\widehat{p}_{Y|z,x} - p_{Y|z,x} &= \frac{1}{N} \frac{\sum_{i=1}^N \mathbf{1}(Z_i = z, X_i = x) (\mathbf{1}(Y_i = 1) - p_{Y|z,x})}{\mathbb{P}(Z = z, X = x)} + o_p(N^{-1/2}) \\ \widehat{p}_{D|z,x} - p_{D|z,x} &= \frac{1}{N} \frac{\sum_{i=1}^N \mathbf{1}(Z_i = z, X_i = x) (\mathbf{1}(D_i = 1) - p_{D|z,x})}{\mathbb{P}(Z = z, X = x)} + o_p(N^{-1/2}) \\ \widehat{p}_{z|x} - p_{z|x} &= \frac{1}{N} \frac{\sum_{i=1}^N \mathbf{1}(X_i = x) (\mathbf{1}(Z_i = z) - p_{z|x})}{q_x} + o_p(N^{-1/2}) \\ \widehat{q}_x - q_x &= \frac{1}{N} \sum_{i=1}^N (\mathbf{1}(X_i = x) - q_x) + o_p(N^{-1/2})\end{aligned}$$

The covariance kernel Σ_j can be calculated as follows:

$$\begin{aligned}[\Sigma_j]_{1,1} &= \mathbb{E} \left[\frac{\mathbf{1}(Z_i = z, X_i = x) \mathbf{1}(Z_i = \tilde{z}, X_i = \tilde{x}) (\mathbf{1}(Y_i = 1) - p_{Y|z,x}) (\mathbf{1}(Y_i = 1) - p_{Y|\tilde{z},\tilde{x}})}{\mathbb{P}(Z = z, X = x) \mathbb{P}(Z = \tilde{z}, X = \tilde{x})} \right], \\ [\Sigma_j]_{1,2} &= \mathbb{E} \left[\frac{\mathbf{1}(Z_i = z, X_i = x) \mathbf{1}(Z_i = \tilde{z}, X_i = \tilde{x}) (\mathbf{1}(Y_i = 1) - p_{Y|z,x}) (\mathbf{1}(D_i = 1) - p_{D|\tilde{z},\tilde{x}})}{\mathbb{P}(Z = z, X = x) \mathbb{P}(Z = \tilde{z}, X = \tilde{x})} \right] = 0, \\ [\Sigma_j]_{1,3} &= \mathbb{E} \left[\frac{\mathbf{1}(Z_i = z, X_i = \tilde{x}) (\mathbf{1}(Y_i = 1) - p_{Y|z,x}) (\mathbf{1}(Z_i = \tilde{z}) - p_{\tilde{z}|\tilde{x}})}{p_{z|x} q_x q_{\tilde{x}}} \right] = 0, \\ [\Sigma_j]_{1,4} &= \mathbb{E} \left[\frac{(\mathbf{1}(X_i = \tilde{x}) - q_{\tilde{x}}) \mathbf{1}(Z_i = z, X_i = x) (\mathbf{1}(Y_i = 1) - p_{Y|z,x})}{\mathbb{P}(Z = z, X = x)} \right] = 0, \\ [\Sigma_j]_{2,1} &= [\Sigma_j]_{1,2} = 0, \\ [\Sigma_j]_{2,2} &= \mathbb{E} \left[\frac{\mathbf{1}(Z_i = z, X_i = x) \mathbf{1}(Z_i = \tilde{z}, X_i = \tilde{x}) (\mathbf{1}(D_i = 1) - p_{D|z,x}) (\mathbf{1}(D_i = 1) - p_{D|\tilde{z},\tilde{x}})}{\mathbb{P}(Z = z, X = x) \mathbb{P}(Z = \tilde{z}, X = \tilde{x})} \right], \\ [\Sigma_j]_{2,3} &= \mathbb{E} \left[\frac{\mathbf{1}(Z_i = z, X_i = \tilde{x}) (\mathbf{1}(D_i = 1) - p_{D|z,x}) (\mathbf{1}(Z_i = \tilde{z}) - p_{\tilde{z}|\tilde{x}})}{p_{z|x} q_x q_{\tilde{x}}} \right] = 0, \\ [\Sigma_j]_{2,4} &= \mathbb{E} \left[\frac{(\mathbf{1}(X_i = \tilde{x}) - q_{\tilde{x}}) \mathbf{1}(Z_i = z, X_i = x) (\mathbf{1}(D_i = 1) - p_{D|z,x})}{\mathbb{P}(Z = z, X = x)} \right] = 0,\end{aligned}$$

$$[\Sigma_j]_{3,1} = [\Sigma_j]_{1,3} = 0,$$

$$[\Sigma_j]_{3,2} = [\Sigma_j]_{2,3} = 0,$$

$$[\Sigma_j]_{3,3} = \mathbb{E} \left[\frac{\mathbf{1}(X_i = x) \mathbf{1}(X_i = \tilde{x}) (\mathbf{1}(Z_i = z) - p_{z|x}) (\mathbf{1}(Z_i = \tilde{z}) - p_{z|\tilde{x}})}{q_x q_{\tilde{x}}} \right],$$

$$[\Sigma_j]_{3,4} = \mathbb{E} \left[\frac{(\mathbf{1}(X_i = \tilde{x}) - q_{\tilde{x}}) (\mathbf{1}(Z_i = z) - p_{z|x})}{q_x} \right] = 0,$$

$$[\Sigma_j]_{4,1} = [\Sigma_j]_{1,4} = 0,$$

$$[\Sigma_j]_{4,2} = [\Sigma_j]_{2,4} = 0,$$

$$[\Sigma_j]_{4,3} = [\Sigma_j]_{3,4} = 0,$$

$$[\Sigma_j]_{4,4} = \mathbb{E} [(\mathbf{1}(X_i = x) - q_x) (\mathbf{1}(X_i = \tilde{x}) - q_{\tilde{x}})]$$

The next lemma provides the asymptotic properties for the estimators of the bounds of potential quantities under joint c -dependence:

Lemma 6. *Suppose Assumptions 1-3 and 6-9 hold. Then,*

$$\sqrt{N} \begin{pmatrix} \widehat{\tau}_{D(z)}^{LB}(c, x) - \tau_{D(z)}^{LB}(c, x) \\ \widehat{\tau}_{D(z)}^{UB}(c, x) - \tau_{D(z)}^{UB}(c, x) \end{pmatrix} \xrightarrow{d} \tilde{\mathbf{Z}}_{j,D}(d, z, x, c)$$

and

$$\sqrt{N} \begin{pmatrix} \widehat{\tau}_{Y(D(z))}^{LB}(c, x) - \tau_{Y(D(z))}^{LB}(c, x) \\ \widehat{\tau}_{Y(D(z))}^{UB}(c, x) - \tau_{Y(D(z))}^{UB}(c, x) \end{pmatrix} \xrightarrow{d} \tilde{\mathbf{Z}}_{j,Y}(y, z, x, c)$$

both tight elements of $l^\infty(\{0, 1\} \times \{0, 1\} \times \mathcal{S}(X), \mathbb{R}^2)$.

Proof:

Let $\Omega_0 = (p_{Y|z,x}, p_{D|z,x}, p_{z|x}, q_x)$ and $\widehat{\Omega} = (\widehat{p}_{Y|z,x}, \widehat{p}_{D|z,x}, \widehat{p}_{z|x}, \widehat{q}_x)$. For fixed d and c , define the mapping

$$\begin{aligned} \psi_{FS} : l^\infty(\{0, 1\}^2 \times \mathcal{S}(X)) \times l^\infty(\{0, 1\}^2 \times \mathcal{S}(X)) \times l^\infty(\mathcal{S}(X)) \\ \rightarrow l^\infty(\{0, 1\} \times \mathcal{S}(X), \mathbb{R}^2) \end{aligned}$$

by

$$[\psi_{FS}(\Omega)](z, x) = \begin{pmatrix} \min \{ \delta_{FS,3}(\Omega), \delta_{FS,4}(\Omega) \} \\ \max \{ \delta_{FS,1}(\Omega), \delta_{FS,2}(\Omega) \} \end{pmatrix}$$

where

$$[\delta_{FS,1}(\Omega)](z, x) = \Omega^{(2)}(d, z, x)\Omega^{(3)}(z, x) + \left[\frac{(1 - \Omega^{(3)}(z, x) - c)\Omega^{(2)}(d, z, x)\Omega^{(3)}(z, x)}{(\Omega^{(3)}(z, x) + c)(1 - \Omega^{(3)}(z, x))} \right] (1 - \Omega^{(3)}(z, x))$$

which has Hadamard directional derivative equal to

$$\begin{aligned} \left[\delta'_{FS,1,\Omega}(h) \right](z, x) &= h^{(2)}(d, z, x)\Omega^{(3)}(z, x) + \Omega^{(2)}(d, z, x)h^{(3)}(z, x) \\ &+ \frac{h^{(2)}(d, z, x)(1 - \Omega^{(3)}(z, x) - c)\Omega^{(3)}(z, x) + \Omega^{(2)}(d, z, x)h^{(3)}(z, x)(1 - 2\Omega^{(3)}(z, x) - c)}{(\Omega^{(3)}(z, x) + c)} \\ &- \frac{(1 - \Omega^{(3)}(z, x) - c)\Omega^{(2)}(d, z, x)\Omega^{(3)}(d, z, x)h^{(3)}(z, x)(1 - 2\Omega^{(3)}(z, x))}{(\Omega^{(3)}(z, x) + c)^2(1 - \Omega^{(3)}(z, x))} \\ &- h^{(3)}(z, x) \left[\frac{(1 - \Omega^{(3)}(z, x) - c)\Omega^{(2)}(d, z, x)\Omega^{(3)}(z, x)}{(1 - \Omega^{(3)}(z, x))(\Omega^{(3)}(z, x) + c)} \right], \end{aligned}$$

$$[\delta_{FS,2}(\Omega)](z, x) = \Omega^{(2)}(d, z, x)\Omega^{(3)}(z, x)$$

which has Hadamard directional derivative equal to

$$\left[\delta'_{FS,2,\Omega}(h) \right](z, x) = h^{(2)}(d, z, x) + \Omega^{(3)}(z, x) + \Omega^{(2)}(d, z, x)h^{(3)}(z, x),$$

$$[\delta_{FS,3}(\Omega)](z, x) = \Omega^{(2)}(d, z, x)\Omega^{(3)}(z, x) + \left[\frac{(1 - \Omega^{(3)}(z, x) + c)\Omega^{(2)}(d, z, x)\Omega^{(3)}(z, x)}{(\Omega^{(3)}(z, x) - c)(1 - \Omega^{(3)}(z, x))} \right] (1 - \Omega^{(3)}(z, x))$$

which has Hadamard directional derivative equal to

$$\begin{aligned} & \left[\delta'_{FS,3,\Omega}(h) \right] (z, x) = h^{(2)}(d, z, x)\Omega^{(3)}(z, x) + \Omega^{(2)}(d, z, x)h^{(3)}(z, x) \\ & + \frac{h^{(2)}(d, z, x) (1 - \Omega^{(3)}(z, x) + c) \Omega^{(3)}(z, x) + \Omega^{(2)}(d, z, x) + h^{(3)}(d, z, x) (1 - 2\Omega^{(3)}(z, x) - c)}{(\Omega^{(3)}(z, x) - c)} \\ & - \frac{(1 - \Omega^{(3)}(z, x) + c) \Omega^{(2)}(d, z, x)\Omega^{(3)}(d, z, x)h^{(3)}(z, x) (1 - 2\Omega^{(3)}(z, x))}{(\Omega^{(3)}(z, x) - c)^2 (1 - \Omega^{(3)}(z, x))} \\ & - h^{(3)}(z, x) \left[\frac{(1 - \Omega^{(3)}(z, x) + c) \Omega^{(2)}(d, z, x)\Omega^{(3)}(z, x)}{(1 - \Omega^{(3)}(z, x)) (\Omega^{(3)}(z, x) - c)} \right], \end{aligned}$$

and

$$[\delta_{FS,4}(\Omega)](z, x) = \Omega^{(2)}(d, z, x)\Omega^{(3)}(z, x) + \Omega^{(3)}(z, x)$$

which has Hadamard directional derivative equal to

$$\left[\delta'_{FS,4,\Omega}(h) \right] (z, x) = h^{(2)}(d, z, x) + \Omega^{(3)}(z, x) + (\Omega^{(2)}(d, z, x) - 1)h^{(3)}(z, x)$$

Hence, the Hadamard directional derivative of the map ψ_{FS} evaluated at Ω_0 is

$$\psi'_{FS,\Omega_0}(h) = \left(\begin{array}{l} \mathbf{1} (\delta_{FS,3}(\Omega_0) < \delta_{FS,4}(\Omega_0)) \delta'_{FS,4,\Omega_0}(h) \\ + \mathbf{1} (\delta_{FS,3}(\Omega_0) = \delta_{FS,4}(\Omega_0)) \min \{ \delta'_{FS,3,\Omega_0}(h), \delta'_{FS,4,\Omega_0}(h) \} \\ + \mathbf{1} \{ \delta_{FS,3}(\Omega_0) > \delta_{FS,4}(\Omega_0) \} \delta'_{FS,3,\Omega_0}(h) \\ \\ \mathbf{1} (\delta_{FS,1}(\Omega_0) < \delta_{FS,2}(\Omega_0)) \delta'_{FS,1,\Omega_0}(h) \\ + \mathbf{1} (\delta_{FS,1}(\Omega_0) = \delta_{FS,2}(\Omega_0)) \max \{ \delta'_{FS,1,\Omega_0}(h), \delta'_{FS,2,\Omega_0}(h) \} \\ + \mathbf{1} (\delta_{FS,1}(\Omega_0) > \delta_{FS,2}(\Omega_0)) \delta'_{FS,2,\Omega_0}(h) \end{array} \right)$$

It follows from the Delta Method that

$$\left[\sqrt{N} \left(\psi_{FS}(\widehat{\Omega}) - \psi_{FS}(\Omega_0) \right) \right] \xrightarrow{d} \left[\psi'_{FS,\Omega_0}(\mathbf{Z}_j) \right] (z, x) \equiv \mathbf{Z}_{j,FS}(z, x)$$

For fixed y and c define the mapping

$$\begin{aligned} \psi_{RF} : l^\infty(\{0, 1\}^2 \times \mathcal{S}(X)) \times l^\infty(\{0, 1\}^2 \times \mathcal{S}(X)) \times l^\infty(\mathcal{S}(X)) \\ \rightarrow l^\infty(\{0, 1\} \times \mathcal{S}(X), \mathbb{R}^2) \end{aligned}$$

by

$$[\psi_{RF}(\Omega)](z, x) = \begin{pmatrix} \min \{ \delta_{RF,3}(\Omega), \delta_{RF,4}(\Omega) \} \\ \max \{ \delta_{RF,1}(\Omega), \delta_{RF,2}(\Omega) \} \end{pmatrix}$$

where

$$[\delta_{RF,1}(\Omega)](z, x) = \Omega^{(1)}(y, z, x)\Omega^{(3)}(z, x) + \left[\frac{(1 - \Omega^{(3)}(z, x) - c)\Omega^{(1)}(y, z, x)\Omega^{(3)}(z, x)}{(\Omega^{(3)}(z, x) + c)(1 - \Omega^{(3)}(z, x))} \right] (1 - \Omega^{(3)}(z, x))$$

which has Hadamard directional derivative equal to

$$\begin{aligned} \left[\delta'_{RF,1,\Omega}(h) \right](z, x) &= h^{(1)}(y, z, x)\Omega^{(3)}(z, x) + \Omega^{(1)}(y, z, x)h^{(3)}(z, x) \\ &+ \frac{h^{(1)}(y, z, x)(1 - \Omega^{(3)}(z, x) - c)\Omega^{(3)}(z, x) + \Omega^{(1)}(y, z, x) + h^{(3)}(z, x)(1 - 2\Omega^{(3)}(z, x) - c)}{(\Omega^{(3)}(z, x) + c)} \\ &- \frac{(1 - \Omega^{(3)}(z, x) - c)\Omega^{(1)}(y, z, x)\Omega^{(3)}(z, x)h^{(3)}(z, x)(1 - 2\Omega^{(3)}(z, x))}{(\Omega^{(3)}(z, x) + c)^2(1 - \Omega^{(3)}(z, x))} \\ &- h^{(3)}(z, x) \left[\frac{(1 - \Omega^{(3)}(z, x) - c)\Omega^{(1)}(y, z, x)\Omega^{(3)}(z, x)}{(1 - \Omega^{(3)}(z, x))(\Omega^{(3)}(z, x) + c)} \right], \end{aligned}$$

$$[\delta_{RF,2}(\Omega)](z, x) = \Omega^{(1)}(y, z, x)\Omega^{(3)}(z, x)$$

which has Hadamard directional derivative equal to

$$\left[\delta'_{RF,2,\Omega}(h) \right](z, x) = h^{(1)}(y, z, x) + \Omega^{(3)}(z, x) + \Omega^{(1)}(y, z, x)h^{(3)}(z, x),$$

$$[\delta_{RF,3}(\Omega)](z, x) = \Omega^{(1)}(y, z, x)\Omega^{(3)}(z, x) + \left[\frac{(1 - \Omega^{(3)}(z, x) + c)\Omega^{(1)}(y, z, x)\Omega^{(3)}(z, x)}{(\Omega^{(3)}(z, x) - c)(1 - \Omega^{(3)}(z, x))} \right] (1 - \Omega^{(3)}(z, x))$$

which has Hadamard directional derivative equal to

$$\begin{aligned} \left[\delta'_{RF,3,\Omega}(h) \right] (z, x) &= h^{(1)}(y, z, x)\Omega^{(3)}(z, x) + \Omega^{(1)}(y, z, x)h^{(3)}(z, x) \\ &+ \frac{h^{(1)}(y, z, x) (1 - \Omega^{(3)}(z, x) + c) \Omega^{(3)}(z, x) + \Omega^{(1)}(y, z, x) + h^{(3)}(z, x) (1 - 2\Omega^{(3)}(z, x) - c)}{(\Omega^{(3)}(z, x) - c)} \\ &- \frac{(1 - \Omega^{(3)}(z, x) + c) \Omega^{(1)}(y, z, x)\Omega^{(3)}(z, x)h^{(3)}(z, x) (1 - 2\Omega^{(3)}(z, x))}{(\Omega^{(3)}(z, x) - c)^2 (1 - \Omega^{(3)}(z, x))} \\ &- h^{(3)}(z, x) \left[\frac{(1 - \Omega^{(3)}(z, x) + c) \Omega^{(1)}(y, z, x)\Omega^{(3)}(z, x)}{(1 - \Omega^{(3)}(z, x)) (\Omega^{(3)}(z, x) - c)} \right], \end{aligned}$$

and

$$[\delta_{RF,4}(\Omega)](z, x) = \Omega^{(1)}(y, z, x)\Omega^{(3)}(z, x) + \Omega^{(3)}(z, x)$$

which has Hadamard directional derivative equal to

$$\left[\delta'_{RF,4,\Omega}(h) \right] (z, x) = h^{(1)}(y, z, x) + \Omega^{(3)}(z, x) + (\Omega^{(1)}(y, z, x) - 1)h^{(3)}(z, x)$$

Hence, the Hadamard directional derivative of the map ψ_{RF} evaluated at Ω_0 is

$$\psi'_{RF,\Omega_0}(h) = \left(\begin{array}{l} \mathbf{1}(\delta_{RF,3}(\Omega_0) < \delta_{RF,4}(\Omega_0)) \delta'_{RF,4,\Omega_0}(h) \\ + \mathbf{1}(\delta_{RF,3}(\Omega_0) = \delta_{RF,4}(\Omega_0)) \min \{ \delta'_{RF,3,\Omega_0}(h), \delta'_{RF,4,\Omega_0}(h) \} \\ + \mathbf{1} \{ \delta_{RF,3}(\Omega_0) > \delta_{RF,4}(\Omega_0) \} \delta'_{RF,3,\Omega_0}(h) \\ \\ \mathbf{1}(\delta_{RF,1}(\Omega_0) < \delta_{RF,2}(\Omega_0)) \delta'_{RF,1,\Omega_0}(h) \\ + \mathbf{1}(\delta_{RF,1}(\Omega_0) = \delta_{RF,2}(\Omega_0)) \max \{ \delta'_{RF,1,\Omega_0}(h), \delta'_{RF,2,\Omega_0}(h) \} \\ + \mathbf{1}(\delta_{RF,1}(\Omega_0) > \delta_{RF,2}(\Omega_0)) \delta'_{RF,2,\Omega_0}(h) \end{array} \right)$$

It follows from the Delta Method that

$$\left[\sqrt{N} \left(\psi_{RF}(\widehat{\Omega}) - \psi_{RF}(\Omega_0) \right) \right] \xrightarrow{d} \left[\psi'_{RF,\Omega_0}(\mathbf{Z}_j) \right] (z, x) \equiv \mathbf{Z}_{j,RF}(z, x)$$

which concludes the proof.