Instrumental Variable Regression with Varying-intensity Repeated Treatments

Jaerim Choi
Yonsei University
jaerimchoi@yonsei.ac.kr

Dakyung Seong
University of Sydney
dakyung.seong@sydney.edu.au

Shu Shen*
University of California, Davis
shushen@ucdavis.edu

This version: Sept., 2024

Abstract

Instrumental variable models with repeated endogenous treatments are popular in empirical research using pooled cross-sectional or short panel datasets. This paper proposes a novel semi-parametric approach that explicitly considers treatment effect dynamics by allowing for 1) path-dependency in the contemporaneous treatment effect and 2) a direct carryover effect from last period's treatment. We show that if either of these new features is present, the textbook two-stage least-squares estimator is generally invalid. We apply the proposed semi-parametric estimation and inference approach to revisit the work of Acemoglu et al. (2016). Using industry-level data, we find that the magnitude of contemporaneous impact of increased Chinese import competition on US manufacturing employment depends on an industry's past import exposure. In particular, industries with larger trade shocks in the 1990's tend to experience stronger impacts from contemporaneous trade shocks in the 2000's.

*The authors are very thankful for all the helpful comments from seminar participants at Syracuse University, the University of Toronto, the University of Queensland, the 2023 International Association of Applied Econometrics Annual Conference, the 2023 All California Econometrics Conference, the 2023 Australia New Zealand Econometric Study Group Meeting, the 18th International Symposium on Econometric Theory and Applications, the 2024 Kansas Econometrics workshop and the 2024 Econometrics in Rio Conference. An earlier version of this paper was circulated under the title "Panel Instrumental Variable Regression Models with Varying-intensity Repeated Treatments: Theory and the China Syndrome Application."

1 Introduction

Models with repeated varying-intensity endogenous treatments and external instrumental variables (IV) are popular in applied economics. For example, Autor et al. (2013) and Acemoglu et al. (2016), seminal papers on the nexus between international trade and labor economics, assess the repeated endogenous treatments of exposure to rising Chinese import competition across US local labor markets and industries. Card (2009), an influential paper in immigration studying the elasticity of substitution between immigrants and natives in local labor markets, uses a three-period panel model (1980, 1990, and 2000) to investigate the repeated endogenous treatments of local immigration shocks. Angrist (2002) studies the effect of having a high male-to-female sex ratio on second-generation immigrants' demographic and economic outcomes using data from the 1910, 1920, and 1940 Censuses. Boustan (2010) investigates white departures in northern cities in response to large Black migration from the rural South to northern cities between 1940 and 1970. Recently, Acemoglu and Restrepo (2022) use a stacked panel to investigate the impact of immigration on innovation in US counties.

Empirical analyses for such model setups are typically carried out using textbook two-stage least squares (2SLS) estimation. In this paper, we show that this popular estimator is generally invalid if the underlying data generating process (DGP) has nontrivial dynamic treatment effect features. Specifically, we find that, if the external treatment is correlated with past treatment take-ups, as observed in many empirical applications, the textbook 2SLS estimator does not converge in probability to an interpretable causal parameter.

Motivated by the invalidity result, we propose a new semi-parametric varying-coefficient model for the repeated endogenous treatment setting. Our model allows 1) the contemporaneous treatment effect to vary with last period's treatment and 2) last period's treatment to directly affect the contemporaneous outcome even when there is no contemporaneous treatment. The proposed dynamic treatment effect features are relevant to many empirical applications. In the China syndrome application in Autor et al. (2013), for example, path-dependency allows the contemporaneous effect of increased import exposure in the 2000s to vary with the import exposure in the 1990s. In the immigration and innovation application (e.g., Acemoglu and Restrepo, 2022), path-dependency allows the contemporaneous impact of immigration on innovation to depend on previous immigration flows.

We study both parametric and semi-parametric identification and estimation strategies for the proposed model. As we will illustrate, different parametric identification strategies rely on different exclusion restrictions, including contemporaneous/sequential exogeneity of the external instrument and/or sequential exogeneity of the endogenous treatment. The validity of the parametric approach is subject to functional form assumptions. For semi-parametric identification and estimation, we require a conditional contemporaneous exogeneity assumption and propose to use local general method of moments (GMM) estimation. Our benchmark estimator is an augmented local GMM estimator inspired by previous studies in the conditional GMM literature, including Cai and Li (2008), Su et al. (2013), and Bravo (2023), that uses local linear or polynomial expansion of varying coefficient functions to improve estimation efficiency. The proposed estimator is different from the literature for only partially augmenting the parameter function of interest, due to the practical consideration of first-stage rank condition. We also consider two alternative non-augmented local GMM estimators as means of robustness checks.

When applied to the China syndrome application, our new model, together with the proposed semi-parametric identification and estimation strategy, uncover rich dynamics that have not been described previously in the literature (see, e.g., Autor et al., 2013, 2014; Acemoglu et al., 2016; Autor et al., 2020a,b; Bloom et al., 2019; Feenstra et al., 2019, among many others). We find strong evidence, using the industry-level dataset in Acemoglu et al. (2016), that the contemporaneous impact of increased Chinese imports on employment in 1999-2011 depends on the import exposure in 1991-1999. The path-dependency is monotonic, with the previous import exposure magnifying the negative impact of the current trade shock. More interestingly, we find that the magnifying effect is mild for most industries, but becomes much larger when the increase in import exposure between 1991 and 1999 exceeds around 0.2 percentage points per year. Specifically, the China shock effect in the 2000s is stable and averages -0.25 percentage points when the import exposure change in the 1990s lies between 0 and 0.2, whereas the average effect increases to over -1 percentage points when the change lies between 0.2 and 0.3. The empirical finding underscores the importance of allowing for path-dependency with a flexible functional form when analyzing trade effects.

Our proposed model is related to the dynamic treatment effect literature in biostatistics, micro-econometrics, and causal inference. The biostatistics literature (e.g., Robins, 1986, 1987; Murphy et al., 2001; Murphy, 2003) has a long history of studying dynamic causal effects. The literature often assumes sequential randomization or unconfoundedness for identification. See Section III of Hernán and Robins (2023) for a recent survey on identification strategies and Bojinov et al. (2021) for some recent developments in inference. Our paper uses external instruments. In econometrics, Heckman and Navarro (2007) and Heckman et al. (2016) are seminal papers that consider identification in sequential discrete treatment models. They establish important dynamic treatment effect concepts including the direct effect, continuation values, and total longer-term effect of treatment interventions. In the panel difference-in-difference literature, Sun and Abraham (2021), Callaway and Sant'Anna (2021), and Athey and Imbens (2022), among many others, propose robust identification methods for dynamic treatment effects using parallel trend type identifying assumptions. The literature often considers repeated irreversible binary treatments. Han (2021) extends the dynamic treatment effect literature by considering general treatment settings that are binary but not necessarily irreversible. Han (2021) uses strictly exogenous excluded instruments for the purpose of nonparametric identification. Recently, Chen and Zhang (2023) study optimal dynamic treatment regimes with time-varying instruments that are randomly generated. Our paper studies varying-intensity treatments and does not require external instruments to be strictly or sequentially exogenous. On the other hand, our semi-parametric approach involves functional-form assumptions that are more restrictive than previous papers that pursue nonparametric model identification.

Using external instruments for identification, our paper is distinguished from the vast literature that uses internal instruments for short-panel IV identification, including Anderson and Hsiao (1982), Arellano and Bond (1991), Ahn and Schmidt (1995), and many others. Our model and method are also different from those proposed in the local projection instrumental variable (LP-IV) literature (see, for example, Stock and Watson, 2018 for an excellent review), which also studies dynamic treatment effects. This is because our model is particularly designed for pooled cross-sectional or short panel datasets and focuses on identifying path-dependent contemporaneous effects as well as direct carryover effects from past treatments. The LP-IV approach, on the other hand, is designed for the time series (c.f., Stock and Watson, 2018) setting and seeks to identify the long-term total effect of treatments. Within the local projection literature, our paper is more closely related to a recent work by Dube et al. (2023), who utilize cross-sectional variations in data and parallel trend type assumptions to identify dynamic heterogeneous treatment effects in the difference-in-differences setting. Again, Dube et al. (2023) focuses on repeated binary treatments and long-term total effects while our paper focuses on varying-intensity treatments, path-dependent contemporaneous effects, and long-term direct carryover effects.

The paper is organized as the following. Section 2 motivates our proposed model with direct carryover effect and path-dependent contemporaneous effect. The section also presents the parametric identification of the proposed model and shows its limitations. Section 3 studies semi-parametric identification of the proposed model and discusses various extensions of the benchmark model. Section 4 proposes relevant semi-parametric estimation and inference methods. Section 5 carries out Monte Carlo simulations. Section 6 applies the proposed method to the industry-level dataset from Acemoglu et al. (2016). The appendices include robustness checks for the empirical analysis as well as all mathematical proofs of the theoretical results.

2 Model Set-up and Parametric Identification

2.1 Model Set-up

We are interested in studying the effect of a series of repeated treatments that are of variable intensity. Treatments are potentially endogenous. The identification of treatment effects relies on the presence of external instruments. We propose to model treatment effect dynamics in the outcome equation by explicitly allowing for 1) a *direct carryover effect* from the past treatment even when there is no contemporaneous treatment and 2) a *path-dependent* contemporaneous treatment effect. Data observed for the repeated treatments could either be panel or pooled cross-sectional. We use the panel setting as the benchmark.

Let X_{it} be the scalar continuous treatment of individual i in period t and H_{it} be a $(d_{ht}-1)$-dimensional vector of additional controls. Model the outcome Y_{it} by

$$\textit{Benchmark:} \quad Y_{it} = \alpha_t(X_{i(t-1)}) + \beta_t\left(X_{i(t-1)}\right) \cdot X_{it} + H_{it}'\gamma_t + \varepsilon_{it}, \quad t = 1, \cdots, T, \tag{2.1}$$

where ε_{it} is the unobserved error term, which can include both the unobserved individual fixed effect and the idiosyncratic contemporaneous shock. The treatment series $\{X_{it}\}_{t=1}^{T}$ is potentially endogenous; T is fixed. The varying coefficient $\alpha(.)$ in model (2.1) is the direct carryover effect of last period's treatment[1], while $\beta_t(.)$ is the path-dependent contemporaneous treatment effect function. The two varying coefficients, $\alpha_t(.)$ and $\beta_t(.)$, are allowed to have unknown functional forms.

Model (2.1) nests the existing empirical strategy (e.g., Autor et al., 2013; Acemoglu et al., 2016; Boustan, 2010) where the outcome equation is given by

$$\text{Existing}: Y_{it} = \alpha_t + \beta_t X_{it} + H_{it}'\gamma_t + \varepsilon_{it}, \quad t = 1, \cdots, T. \tag{2.2}$$

In the China syndrome application, for example, Y_{it} is the change in labor market outcomes over two decades, the 1990s and the 2000s, while X_{it} is the China import shock, defined as the change in local labor market (or industry) import exposure. Before the 1990s, the US exposure to Chinese imports is negligible, implying that $X_{it} = 0$ for any $t \leq 0$. The empirical literature also uses a pooled version of equation (2.2) where the slope coefficients do not vary over time.

Allowing for a direct carryover effect, as in the proposed model (2.1), suggests the possibility of a delayed trade impact arising from long-term adjustments in US manufacturing industries, even after the import growth of Chinese products flattens out. Path-dependency in the contemporaneous treatment effect, on the other hand, allows the "China trade shock" impact in the 2000s to vary with the trade shock intensity in the 1990s. This path-dependency could arise from several factors. For example, innovation

[1]The term "direct" effect is used following Heckman et al. (2016) to emphasize that $\alpha_t(.)$ is the carryover effect of $X_{i(t-1)}$ on Y_{it} when there is no contemporaneous treatment at time t. The parameter is important for counterfactual policy analysis as discussed in previous literature, including Heckman et al. (2016) in a multi-stage sequential treatment decision setting, Cellini et al. (2010) and Hsu and Shen (2023) in a dynamic regression discontinuity design, and Gallen et al. (2023) in a repeated binary endogenous treatment setting.

activities might respond negatively (Autor et al., 2020c) to an adverse China trade shock and dampen industries' ability to cope with pressure from China in the future. Meanwhile, innovation activities may also respond positively (Bloom et al., 2016) to adverse China trade shocks, if past trade shocks have led to industry-level structural changes that help with managing future trade shocks.

In the context of Burchardi et al. (2020), allowing for a path-dependent contemporaneous effect function $\beta_t(.)$ opens the possibility that the contemporaneous impact of the "immigration shock" on innovation varies with the intensity of the "immigration shock" in the previous five-year period, possibly through positive externalities from immigrants settled down earlier. For instance, the first arrival of immigrants helps later immigrants integrate into local society (Battisti et al., 2022), thereby enabling new immigrants to focus on economic activities, including innovation. Previous immigrant inventors also positively influence the innovation production of their collaborators from the same ethnic origin, as documented in Bernstein et al. (2022).

If only pooled cross-sectional data is observed for the repeated treatments, the model in (2.1) with both direct carryover effect and path-dependent contemporaneous effect could be adapted to the following:

$$Y_{igt} = \alpha_t(X_{g(t-1)}) + \beta_t\left(X_{g(t-1)}\right) \cdot X_{gt} + H'_{igt}\gamma_t + \varepsilon_{igt}, \quad t = 1, \cdots, T. \tag{2.3}$$

The outcome, regressors, and error term are labeled by individual i, group g, and period t, while the endogenous repeated treatment decision only varies with g and t. Since the dataset is pooled cross-sectional, the outcome error ε_{igt} does not include individual fixed effects while the vector H_{igt} can potentially include group-specific fixed effects.

In the rest of the paper, without loss of generality, we focus on the benchmark panel data setting. Recall that in the China syndrome application, endogenous treatments and outcomes are defined using first-differenced variables and $T = 2$. Before moving to the next section, we link the benchmark varying-coefficient model, when defined with first-differenced variables and time-invariant parameter functions, to a corresponding model defined with level variables. Suppress the vector of additional controls for simplicity. Let the model of empirical interest be

$$Y^{\circ}_{it} = \rho\left(X^{\circ}_{i(t-1)}\right) + \beta\left(X^{\circ}_{i(t-1)}\right) \cdot X^{\circ}_{it} + \kappa_i + \epsilon^{\circ}_{it}, \quad t = 0, 1, 2, \tag{2.4}$$

where Y°_{it}, X°_{it}, and ϵ°_{it} are defined such that $Y_{it} = Y^{\circ}_{it} - Y^{\circ}_{i(t-1)}$, $X_{it} = X^{\circ}_{it} - X^{\circ}_{i(t-1)}$, and $\varepsilon_{it} = \epsilon^{\circ}_{it} - \epsilon^{\circ}_{i(t-1)}$, κ_i is the fixed effect, and $X^{\circ}_{it} = 0$ for any $t \leq 0$. The model is a varying coefficient extension of the classic fixed effect panel model $Y^{\circ}_{it} = \rho + \beta X^{\circ}_{it} + \kappa_i + \epsilon^{\circ}_{it}$.

Let $\alpha(X_{it}) = \rho(X_{it}) - \rho(X_{i(t-1)}) + \big(\beta(X_{it}) - \beta(X_{i(t-1)})\big) \cdot X_{it}$. It is clear that

$$
\begin{aligned}
Y_{i1} \equiv Y^{\circ}_{i1} - Y^{\circ}_{i0} =& \beta(0) \cdot X^{\circ}_{i1} + (\epsilon^{\circ}_{i1} - \epsilon^{\circ}_{i0}) = \alpha(X_{i0}) + \beta(X_{i0}) \cdot X_{i1} + \varepsilon_{i1}, \\
Y_{i2} \equiv Y^{\circ}_{i2} - Y^{\circ}_{i1} =& \alpha(X^{\circ}_{i1}) + \beta\left(X^{\circ}_{i1}\right) \cdot (X^{\circ}_{i2} - X^{\circ}_{i1}) + (\epsilon^{\circ}_{i2} - \epsilon^{\circ}_{i1}) \\
=& \alpha(X_{i1}) + \beta\left(X_{i1}\right) \cdot X_{i2} + \varepsilon_{i2},
\end{aligned}
$$

implying that the path-dependent contemporaneous effect function $\beta(.)$ is preserved through first-differencing.

2.2 Caveats of Ignoring Treatment Effect Dynamics

In this section, we discuss the danger of ignoring treatment effect dynamics. For illustration purposes, we temporarily suppress the role of H_{it} and assume that the external instrument Z_{it} is single-dimensional. Let $\widehat{\beta}_t$ denote the standard 2SLS estimator for β_t in the textbook linear regression model considered in (2.2), using Z_{it} to instrument X_{it}. The lemma below shows that this vanilla estimator is rarely directly interpretable when the true outcome equation features treatment effect dynamics.

Lemma 2.1 *Suppose the contemporaneous exogeneity and standard rank condition hold for the external instrument Z_{it} such that $\mathbb{E}[Z_{it}\varepsilon_{it}] = 0$ and $\mathbb{COV}(Z_{it}, X_{it}) \neq 0$ for all $t = 1, \ldots, T$. Suppose the outcome follows the benchmark model in (2.1).*

(a) If the direct carryover effect in model (2.1) is nontrivial, i.e., $\nexists$ a constant C such that $\alpha_t(.) = C$, the vanilla 2SLS estimator $\widehat{\beta}_t$ does not converge in probability to a weighted average of the path-dependent contemporaneous effect function $\beta_t(.)$.

(b) If the direct carryover effect in model (2.1) is trivial, i.e., $\exists$ a constant C such that $\alpha_t(.) = C$, or if Z_{it} is mean independent of $X_{i(t-1)}$, the vanilla 2SLS estimator $\widehat{\beta}$ is consistent for some weighted average of the contemporaneous effect function. However, the weights are not necessarily non-negative.

The first part of the lemma is clear since under model (2.1),

$$\begin{aligned}\widehat{\beta}_t \xrightarrow{p} & \frac{\mathbb{COV}\left(Z_{it}, \alpha_t(X_{i(t-1)}) + \beta_t(X_{i(t-1)})X_{it} + \varepsilon_{it}\right)}{\mathbb{COV}\left(Z_{it}, X_{it}\right)} \\ = & \mathbb{COV}\left(Z_{it}, \alpha_t(X_{i(t-1)})\right) / \mathbb{COV}\left(Z_{it}, X_{it}\right) + \mathbb{E}[w_t\left(X_{i(t-1)}\right) \cdot \beta_t\left(X_{i(t-1)}\right)].\end{aligned}$$

$w_t\left(X_{i(t-1)}\right) = \left(\mathbb{E}[Z_{it}X_{it}|X_{i(t-1)}] - \mathbb{E}[Z_{it}]\mathbb{E}[X_{it}|X_{i(t-1)}]\right) / \mathbb{COV}\left(Z_{it}, X_{it}\right)$. In general, $\widehat{\beta}_t$ does not converge in probability to a weighted average of $\beta_t(.)$ since $\mathbb{COV}(Z_{it}, \alpha_t(X_{i(t-1)}))$ is nonzero.

For intuition, we can consider the following special case of (2.1):

$$Y_{it} = \alpha_0 + \alpha_1 X_{i(t-1)} + \beta X_{it} + \varepsilon_{it}.$$

Consistency of $\widehat{\beta}_t$ in the special case requires the exclusion restriction that $\mathbb{E}[Z_{it}(\alpha_1(X_{i(t-1)} - \mathbb{E}[X_{i(t-1)}]) + \varepsilon_{it})] = \alpha_1 \mathbb{COV}(Z_{it}, X_{i(t-1)}) = 0$. When $\mathbb{COV}(Z_{it}, X_{i(t-1)}) \neq 0$, the exclusion restriction fails immediately with a nontrivial direct carryover effect (i.e., $\alpha_1 \neq 0$). In empirical studies, external instruments are often expected to be correlated with past treatment take-ups. For example, in the China syndrome literature, the external instrument is constructed using imports from China to eight other high-income countries, excluding the US. The sample correlation coefficient between Z_{i2} and X_{i1} is around 0.4 in the industry-level data. Such kind of high correlation is not surprising, since China's comparative advantage in manufacturing had mostly been in labor-intensive industries throughout both the 1990s and the 2000s.

The second part of the lemma holds because if there is no direct carryover effect or if $\mathbb{E}\left[Z_{it}|X_{i(t-1)}\right] = \mathbb{E}\left[Z_{it}\right]$, the estimator $\widehat{\beta}_t \xrightarrow{p} \mathbb{E}\left[w_t\left(X_{i(t-1)}\right) \cdot \beta_t\left(X_{i(t-1)}\right)\right]$ directly. It is easy to see that the weighting function w_t is not necessarily non-negative. Moreover, even under special DGPs where the weighting function is strictly positive, the weighted average that $\widehat{\beta}_t$ converges to may not be empirically relevant. We provide an example in the Monte Carlo simulations in Section 5.

2.3 Parametric Identification

If the functional form of varying coefficients in (2.1) is known, parametric identification could be achieved. Depending on the exact functional form, different exclusion restrictions, including contemporaneous/sequential exogeneity of the external instrument and/or sequential exogeneity of the endogenous treatment, are needed. In this section, we discuss two special cases of model (2.1) to provide intuitions. Formal assumptions for parametric identification are provided in Appendix B.

First, consider the special case where $\alpha_t(x) = \alpha_0 + \alpha_1 x$ and $\beta_t(.) = \beta$ for all t. The benchmark model in (2.2) reduces to

$$Y_{it} = \alpha_0 + \alpha_1 X_{i(t-1)} + \beta X_{it} + \varepsilon_{it}.$$

The model could be identified by a 2SLS regression of Y_{it} on $X_{i(t-1)}$ and X_{it} instrumented by $X_{i(t-1)}$ and Z_{it}, assuming contemporaneous exogeneity of the external instrument and additionally sequential exogeneity of the endogenous treatment. Alternatively, the model could be identified by a 2SLS regression of Y_{it} on $X_{i(t-1)}$ and X_{it} instrumented by $Z_{i(t-1)}$ and Z_{it}, assuming sequential exogeneity of the external instrument. The two exclusion restrictions are non-nested, but the second 2SLS strategy clearly requires a stronger rank condition.

If $\alpha_t(x) = \alpha_0 + \alpha_1 x$ and $\beta_t(x) = \beta_0 + \beta_1 x$, the benchmark model reduces to

$$Y_{it} = \alpha_0 + \alpha_1 X_{it-1} + \left(\beta_0 + \beta_1 X_{i(t-1)}\right) X_{it} + \varepsilon_{it}.$$

The model could be identified with the instrument set $(X_{i(t-1)}, Z_{it}, X_{i(t-1)}Z_{it})$ assuming sequential exogeneity of the endogenous treatment, or with $(Z_{i(t-1)}, Z_{it}, Z_{i(t-1)}Z_{it})$ assuming sequential exogeneity of the external instrument. The first strategy acknowledges two endogenous regressors while the second acknowledges three.

The use of the parametric approach faces several obstacles in empirical applications. First, the parametric approach requires functional-form knowledge of $\alpha_t(.)$ and $\beta_t(.)$. Second, the approach generates multiple endogenous regressors and can be demanding in terms of the first-stage rank condition, especially when the sample size is small. Third, the use of sequential exogeneity conditions rules out feedback effects from past random variables to contemporary outcome errors.

3 Semi-parametric Identification

Now we consider semi-parametric identification of the *benchmark* model in (2.1), assuming functional forms of the varying coefficients are unknown. Let $\omega_t(.)$ be some known function. We use $\omega_t(Z_{it})$ instead of Z_{it} as the instrument vector to allow for potential efficiency improvements. A similar approach was pursued in Cai and Li (2008) in which some practical examples of $\omega_t(.)$ are available. Let $\ddot{H}_{it} = (1\ H_{it}')'$ and $\ddot{Z}_{it} = (\omega_t(Z_{it})'\ \ddot{H}_{it}')'$. Suppose the dimension of $\omega_t(Z_{it})$ is K, then $\ddot{H}_{it}$ is $(K + d_{ht}) \times 1$. Let $\mathcal{X}_{t-1}$ denote the support of $X_{i(t-1)}$. In some applications, $\mathcal{X}_{t-1}$ may be a particular subset of empirical interest. All identification strategies discussed in this section can be extended to the pooled cross-sectional case described in model (2.3).

3.1 Identification with Exogenous Control Vector

First consider identification assuming that the control vector H_{it} is exogenous.

Assumption 3.1 (semi-parametric identification) *For all $t = 1, \ldots, T$, assume that*

(a) *(exclusion restriction I)* $\mathbb{E}[\varepsilon_{it}] = 0$ *and* $\mathbb{E}[\varepsilon_{it}|X_{i(t-1)}, Z_{it}, H_{it}] = \mathbb{E}[\varepsilon_{it}|X_{i(t-1)}]$.

(b) *(rank condition)* $\mathbb{E}[\ddot{Z}_{it}(X_{it}'\ \ddot{H}_{it}')|X_{i(t-1)} = x]$ *is full rank for all* $x \in \mathcal{X}_{t-1}$.

Assumption 3.1.(a) denotes the contemporaneous mean independence between external instruments and error terms *after* conditioning on the last treatment. For $t = 1$, the assumption reduces to $\mathbb{E}[\varepsilon_{i1}|Z_{i1}, H_{i1}] = 0$ when $X_{i0} = 0$. For $t \geq 2$, Assumption 3.1.(a) is neither weaker nor stronger than $\mathbb{E}[\varepsilon_{it}|Z_{it}, H_{it}] = 0$, which is a typical exclusion restriction used in parametric 2SLS regressions. In the China Syndrome example with industry-level data, Assumption 3.1.(a) requires that, for all US industries that experienced the same level of trade shocks in the last decade, the outcome shock is mean independent of the trade exposure shock experienced in the same industry by other high-income countries.

Assumption 3.1.(a) holds trivially if sequential exogeneity holds for the endogenous treatment, or that $\mathbb{E}[\varepsilon_{it}|X_{i(t-1)}, Z_{it}, H_{it}] = 0$. However, Assumption 3.1.(a) can also hold if $X_{i(t-1)}$, or even $Z_{i(t-1)}$, is endogenous, due to, for example, a feedback effect from last period's treatment $X_{i(t-1)}$ to the contemporaneous outcome shock ε_{it}. To see this, suppose in the China syndrome example $X_{it} = g_x(e_{it,cn}, s_{it,us})$ and $Z_{it} = g_z(e_{it,cn}, v_{it,eu})$, where trade exposures in the US and European countries are functions of the unobserved supply shock from China ($e_{it,cn}$) and the unobserved demand shock from the US/European countries ($s_{it,us}/v_{it,eu}$). Suppose the outcome equation follows model (2.1) where $\varepsilon_{it} = g_\varepsilon(X_{i(t-1)}, \epsilon_{it,us})$ allows for a feedback effect from $X_{i(t-1)}$. Nonetheless, the required conditional mean independence assumption in Assumption 3.1.(a) can still hold if trade shocks from different countries are mutually independent and $\epsilon_{it,us}$ is independent of $s_{it',us}$ for all $t' \leq t-1$.[2] Section 5 provides DGP examples that distinguishes between different exclusion restrictions.

Let $g_t(x) = \alpha_t(x) + \mathbb{E}[\varepsilon_{it}|X_{i(t-1)} = x]$. Assumption 3.1.(a) implies that for all $x \in \mathcal{X}_{t-1}$ and $t = 1, \dots, T$,

$$\begin{aligned}
&\mathbb{E}\left[\ddot{Z}_{it}\left(Y_{it} - \left(g_t(x) + \beta_t(x)X_{it} + H'_{it}\gamma_t\right)\right)|X_{i(t-1)} = x\right] \\
=&\mathbb{E}\left[\ddot{Z}_{it}\left(\mathbb{E}\left[\varepsilon_{it}|Z_{it}, H_{it}, X_{i(t-1)} = x\right] - \mathbb{E}\left[\varepsilon_{it}|X_{i(t-1)} = x\right]\right)|X_{i(t-1)} = x\right] = 0. \quad (3.1)
\end{aligned}$$

If $X_0 = 0$, the above conditional moment equality reduces to the classic unconditional moment equality for $t = 1$. For all $t \geq 2$, the above conditional moment equality implies the identification of $(g_t(x)\ \beta_t(x)\ \gamma'_t)'$, for all $x \in \mathcal{X}_{t-1}$. Whether the conditional moment equality is just-identified or over-identified depends on the dimension of $\omega_t(Z_{it})$. If one is willing to further assume sequential exogeneity of the endogenous treatment alongside Assumption 3.1, the function $g_t(.)$ reduces to $\alpha_t(.)$. Otherwise, Assumption 3.1 identifies the path-dependent contemporaneous treatment effect $\beta_t(.)$ while treating the direct carryover effect function $\alpha_t(.)$ as a nuisance parameter.

3.2 Identification with Potentially Endogenous Additional Controls

Note that Assumption 3.1.(a) requires additional controls in H_{it} to be exogenous. This restriction can be relaxed if we do not pursue separate identification of effects from these additional controls.

Assumption 3.2 (semi-parametric identification: exclusion restriction II) *Assume that* $\mathbb{E}[\varepsilon_{it}] = 0$, $\mathbb{E}[\varepsilon_{it}|X_{i(t-1)}, Z_{it}] = \mathbb{E}[\varepsilon_{it}|X_{i(t-1)}]$, *and* $\mathbb{E}[H_{it}|X_{i(t-1)}, Z_{it}] = \mathbb{E}[H_{it}|X_{i(t-1)}]$, *for all* $t = 1, \dots, T$.

[2] To see this, let $h_1(.)$ and $h_2(.)$ be any square integrable functions. Then, for all x,

$$\begin{aligned}
&\mathbb{E}\left[h_1(\varepsilon_t)h_2(Z_{it})|X_{i(t-1)} = x\right] = \mathbb{E}\left[h_1(g_\varepsilon(X_{i(t-1)}, \epsilon_{it,us}))h_2(Z_{it})|X_{i(t-1)} = x\right] \\
=&\int\int h_1(g_\varepsilon(x,\epsilon))h_2(z)f_{\epsilon_{t,us}}(\epsilon)f_{X_{i(t-1)},Z_{it}}(x,z)/f_{X_{i(t-1)}}(x)d\epsilon dz \\
=&\int h_1(g_\varepsilon(x,\epsilon))f_{\epsilon_{t,us}}(\epsilon)d\epsilon \cdot \mathbb{E}\left[h_2(Z_{it})|X_{i(t-1)} = x\right] = \mathbb{E}\left[h_1(\varepsilon_t)|X_{i(t-1)} = x\right] \cdot \mathbb{E}\left[h_2(Z_{it})|X_{i(t-1)} = x\right].
\end{aligned}$$

Assumption 3.2 allows the control vector H_{it} to be endogenous, as long as the external instrument is mean independent of both H_{it} and the error term ε_{it} after conditioning on the past treatment $X_{i(t-1)}$. Let $\widetilde{\gamma}_t(x) = \gamma_t + \mathbb{E}[H_{it}H_{it}'|X_{i(t-1)} = x]^{-1}\mathbb{E}[H_{it}\varepsilon_i|X_{i(t-1)} = x]$ for all $x \in \mathcal{X}_{t-1}$[3]. Assumption 3.2 implies that for all $x \in \mathcal{X}_{t-1}$,

$$\begin{aligned}&\mathbb{E}\left[\omega_t(Z_{it})\left(Y_{it} - \left(g_t(x) + \beta_t(x)X_{it} + H_{it}'\widetilde{\gamma}_t(x)\right)\right)|X_{i(t-1)} = x\right]\\ =&\mathbb{E}\left[\omega_t(Z_{it})\varepsilon_{it}|X_{i(t-1)} = x\right] - \mathbb{E}\left[\omega(Z_{it})\mathbb{E}[\varepsilon_{it}|X_{i(t-1)} = x]|X_{i(t-1)} = x\right] = 0. \end{aligned} \tag{3.2}$$

The first equality is explained in Appendix D. The second equality holds because under Assumption 3.2, $\mathbb{E}[\varepsilon_{it}|X_{i(t-1)}, Z_{it}] = \mathbb{E}[\varepsilon_{it}|X_{i(t-1)}]$. In addition, Assumption 3.2 implies that for all $x \in \mathcal{X}_{t-1}$

$$\begin{aligned}&\mathbb{E}\left[H_{it}\left(Y_{it} - \left(g_t(x) + \beta_t(x)X_{it} + H_{it}'\widetilde{\gamma}(x)\right)\right)|X_{i(t-1)} = x\right]\\ =&\mathbb{E}\left[H_{it}\varepsilon_{it}|X_{i(t-1)} = x\right] - \mathbb{E}\left[H_{it}H_{it}'|X_{i(t-1)} = x\right]\left(\widetilde{\gamma}(x) - \gamma(x)\right) = 0.\end{aligned}$$

Summing up, Assumption 3.2 implies that

$$\mathbb{E}\left[\ddot{Z}_{it}\left(Y_{it} - \left(g_t(x) + \beta_t(x)X_{it} + H_{it}'\widetilde{\gamma}_t(x)\right)\right)|X_{i(t-1)} = x\right] = 0. \tag{3.3}$$

Compared to the identification result of the last section, allowing H_{it} to be endogenous implies that we cannot separately identify its effect on the outcome. However, if the external instrument used to identify $\beta_t(.)$ does not move with H_{it} conditional on $X_{i(t-1)}$, the endogeneity of H_{it} does not influence the identification of $\beta_t(.)$. For identification of $\beta_t(.)$ only, there is no need to distinguish between Assumption 3.1.(a) and Assumption 3.2. In practice, researchers can choose either of them depending on whether the additional control vector in the empirical application is potentially endogenous.

4 Semi-parametric Estimation and Inference

This section studies semi-parametric estimation and inference strategies for parameter functions of interest and the average contemporaneous effect $\bar{\beta}_t(.) = \mathbb{E}[\beta_t(X_{i(t-1)}]$. Without loss of generality, we set $t = 2$.

[3]The definition implies that $H_{it}'\left(\widetilde{\gamma}_t(x) - \gamma_t(x)\right) = H_{it}'\mathbb{E}[H_{it}H_{it}'|X_{i(t-1)} = x]^{-1}\mathbb{E}[H_{it}\varepsilon_i|X_{i(t-1)} = x] \equiv \mathbb{L}[\varepsilon_i|X_{i(t-1)} = x, H_{it}]$, the population level linear projection of ε_i on H_{it} conditional on $X_{i(t-1)} = x$.

4.1 Estimation of Functional Coefficients

Let $\theta_2(x) = (\beta_2(x)\ g_2(x)\ \gamma_2'(x))'$ collect all parameters of interest. Let $\ddot{X}_{i2} = (X_{i2}'\ \ddot{H}_{i2}')'$. The identification result of equation (3.1), for the case of $t = 2$, can be summarized by the following conditional moment equality:

$$\mathbb{E}[\ddot{Z}_{i2}(Y_{i2} - \ddot{X}_{i2}'\theta_2(x))|X_{i1} = x] = 0. \tag{4.1}$$

We focus on identification results discussed in Section 3.1 since the difference between (3.1) and (3.3) lies only in the interpretation of H_{it}'s coefficients.

Local GMM estimation of the $\theta_2(.)$ function is straightforward following the above conditional moment equality. Inspired by the local linear approaches in Cai and Li (2008) and Su et al. (2013)[4], we propose to construct the benchmark estimator of $\theta_2(.)$ based on the following augmented conditional moment equality:

$$\mathbb{E}[(\ddot{Z}_{i2}'\ \ddot{H}_{i2,a}'(x)/h)'(Y_{i2} - \ddot{X}_{i2}'\theta_2(x) - \ddot{H}_{i2,a}'(x)\theta_{2,-1}^{(1)}(x))|X_{i1} = x] = 0, \tag{4.2}$$

where h is a finite positive constant, $\ddot{H}_{i2,a}(x) = \ddot{H}_{i2} \cdot (X_{i1} - x)$, and $\theta_{2,-1}^{(1)}(.)$ denotes the first derivative of $\theta_{2,-1}(.) \equiv (g_2(x)\ \gamma_2'(x))'$. Note that our augmented conditional moment equality is different from those in Cai and Li (2008) and Su et al. (2013), as equation (4.2) treats the contemporaneous effect function $\beta_2(.)$, which is the first argument of $\theta_2(.)$, differently from the rest of the coefficient vector. This is because a local linear expansion of $\beta_2(.)$ would increase the number of endogenous regressors in local GMM estimation. In empirical applications with limited sample sizes, such an expansion can be costly in terms of the rank condition and, therefore, not desired.

Let $\mathbf{\Lambda}_{\ddot{Z}_a Y}(x; h) = \mathbb{E}[(\ddot{Z}_{i2}'\ \ddot{H}_{i2,a}'(x)/h)'Y_{i2}|X_{i1} = x]$ be a $(K + 2d_{h2}) \times 1$ matrix and $\mathbf{\Lambda}_{\ddot{Z}_a \ddot{X}_a}(x; h) = \mathbb{E}[(\ddot{Z}_{i2}'\ \ddot{H}_{i2,a}'(x)/h)'(\ddot{X}_{i2}'\ \ddot{H}_{i2,a}'(x))|X_{i1} = x]$ be a $(K + 2d_{h2}) \times (1 + 2d_{h2})$ matrix. Let $\mathbf{W}(x)$ be a pre-determined $(K + 2d_{h2}) \times (K + 2d_{h2})$ weighting matrix. Let $\theta_2^a = \left(\theta_2'\ \left(\theta_{2,-1}^{(1)}\right)'\right)'$. For all $x \in \mathcal{X}_1$, define $\widehat{\theta}_2^a(x; h)$ as the solution to

$$\min_{\theta_2^a(x)} \left(\widehat{\mathbf{\Lambda}}_{\ddot{Z}_a Y}(x; h) - \widehat{\mathbf{\Lambda}}_{\ddot{Z}_a \ddot{X}_a}(x; h)\theta_2^a(x)\right)' \mathbf{W}(x) \left(\widehat{\mathbf{\Lambda}}_{\ddot{Z}_a Y}(x; h) - \widehat{\mathbf{\Lambda}}_{\ddot{Z}_a \ddot{X}_a}(x; h)\theta_2^a(x)\right),$$

[4] Recently, Bravo (2023) proposes a general local GMM estimation procedure based on a p-th order local polynomial approximation of the unknown varying-coefficient function, for $p = 0, 1, 2, \cdots$. The estimators proposed in this section could be generalized to accommodate local polynomial expansion as well.

where $\widehat{\mathbf{\Lambda}}_{\ddot{Z}_aY}(x;h)$ and $\widehat{\mathbf{\Lambda}}_{\ddot{Z}_a\ddot{X}_a}(x;h)$, respectively, are local constant estimators of conditional expectations $\mathbf{\Lambda}_{\ddot{Z}_aY}(x;h)$ and $\mathbf{\Lambda}_{\ddot{Z}_a\ddot{X}_a}(x;h)$ with kernel function $\kappa(.)$ and bandwidth h. It is clear that

$$\widehat{\theta}_2^a(x;h) = \left(\widehat{\mathbf{\Lambda}}'_{\ddot{Z}_a\ddot{X}_a}(x;h)\mathbf{W}(x)\widehat{\mathbf{\Lambda}}_{\ddot{Z}_a\ddot{X}_a}(x;h)\right)^{-1}\widehat{\mathbf{\Lambda}}'_{\ddot{Z}_a\ddot{X}_a}(x;h)\mathbf{W}(x)\widehat{\mathbf{\Lambda}}_{\ddot{Z}_aY}(x;h)$$

for all $x \in \mathcal{X}_1$. If $K = 1$, the conditional moment equality in (4.1) is just-identified, and

$$\widehat{\theta}_2^a(x;h) = \left(\widehat{\mathbf{\Lambda}}_{\ddot{Z}_a\ddot{X}_a}(x;h)\right)^{-1}\widehat{\mathbf{\Lambda}}_{\ddot{Z}_aY}(x;h).$$

Let $[.]_{[j]}$ denote the j-th element of the original vector and $[.]_{[j:j']}$ denote a subvector with j-th to j'-th elements of the original vector. We can further define

$$\widehat{\beta}_2(.;h) = \left[\widehat{\theta}_2^a(.;h)\right]_{[1]} \text{ and } \widehat{\theta}_2(.;h) = \left[\widehat{\theta}_2^a(.;h)\right]_{[1:(1+d_{h2})]}$$

as estimators of $\beta_2(.)$ and $\theta_2(.)$, respectively.

If a local linear expansion of $\theta_{2,-1}(.)$ is not sought after, the parameter function $\theta_2(.)$ could be estimated with

$$\widehat{\theta}_{2,\ell}(x;h) = \left(\widehat{\mathbf{\Lambda}}'_{\ddot{Z}\ddot{X},\ell}(x;h)\mathbf{W}(x)\widehat{\mathbf{\Lambda}}_{\ddot{Z}\ddot{X},\ell}(x;h)\right)^{-1}\widehat{\mathbf{\Lambda}}'_{\ddot{Z}\ddot{X},\ell}(x;h)\mathbf{W}(x)\widehat{\mathbf{\Lambda}}_{\ddot{Z}Y,\ell}(x;h),$$

for all $x \in \mathcal{X}_1$ and $\ell = 0, 1$, where $\widehat{\mathbf{\Lambda}}_{\ddot{Z}Y,\ell}(x;h)$ and $\widehat{\mathbf{\Lambda}}_{\ddot{Z}\ddot{X},\ell}(x;h)$ are ℓ-th order local polynomial estimators of $\mathbf{\Lambda}_{\ddot{Z}Y}(x) = \mathbb{E}[\ddot{Z}_{i2}Y_{i2}|X_{i1} = x]$ and $\mathbf{\Lambda}_{\ddot{Z}\ddot{X}}(x) = \mathbb{E}[\ddot{Z}_{i2}\ddot{X}'_{i2}|X_{i1} = x]$, respectively, with bandwidth h.

For notational simplicity, we suppress the h in $\widehat{\theta}_2(.;h)$, $\widehat{\theta}_{2,0}(.;h)$ and $\widehat{\theta}_{2,1}(.;h)$ and denote the estimators as $\widehat{\theta}_2(.)$, $\widehat{\theta}_{2,0}(.)$ and $\widehat{\theta}_{2,1}(.)$, respectively. We also set the predetermined weighting matrix $\mathbf{W}(.)$ to the identity matrix without loss of generality.

The next section studies asymptotic properties of the benchmark estimator $\widehat{\theta}_2(.)$ based on the augmented conditional moment equality (4.2) as well as the two non-augmented alternatives $\widehat{\theta}_{2,0}(.)$ and $\widehat{\theta}_{2,1}(.)$. Monte Carlo simulations of the proposed estimation and inference strategies are given in Section 4.4.

4.2 Asymptotic Properties of the Proposed Functional Estimator

Let $\mathbf{\Omega}(x) = \mathbf{\Lambda}_{\ddot{Z}\ddot{X}}(x)\left(\mathbf{\Lambda}'_{\ddot{Z}\ddot{X}}(x)\mathbf{\Lambda}_{\ddot{Z}\ddot{X}}(x)\right)^{-1}$ and $\mathbf{\Sigma}(x) = \frac{\nu_0}{f_{X_1}(x)}\mathbb{E}\left[\widetilde{\varepsilon}_{i2}^2\ddot{Z}_{i2}\ddot{Z}'_{i2}\middle|X_{i1} = x\right]$, where $\widetilde{\varepsilon}_{i2} = \varepsilon_{i2} - \mathbb{E}[\varepsilon_{i2}|X_{i1}]$. The following theorem summarizes asymptotic properties of the proposed semi-parametric estimators on $\mathcal{X}_1^*$, an interior subset of $\mathcal{X}_1$. The subset $\mathcal{X}_1^*$

would be particularly useful when empirical data are not densely distributed over the entire support of X_{i1}, as in our empirical example in Section 6. Let $\mu_k = \int u^k \kappa(u)du$ and $\mathsf{c}_f(x) = f^{(1)}_{X_1}(x)/f_{X_1}(x)$, where $f_{X_1}(x)$ is the probability density function of $X_{i(t-1)}$ evaluated at $x \in \mathcal{X}_1^*$ and $f^{(1)}_{X_1}(x)$ is its first derivative. Let $[A]_{[.,\ell]}$ denote the ℓ-th column of a matrix A and $A^{(\ell)}(.)$ denote the ℓ-th order derivative of matrix A. Let $1_{\{\mathcal{A}\}}$ be an indicator function that takes 1 if the event $\mathcal{A}$ is true and 0 if otherwise.

Theorem 4.1 *Suppose that the data* $\{Y_{i2}, X_{i1}, X_{i2}, Z_{i2}, H_{i2}\}_{i=1}^N$ *follow the benchmark model in* (2.1) *and Assumptions 3.1 (or having Assumption 3.2 replacing 3.1.(a)) and C.1 hold. Then, for all* $x \in \mathcal{X}_1^*$, *the benchmark augmented local GMM estimator* $\widehat{\theta}_2(x)$ *for* $\theta_2(x)$ *satisfies the following asymptotic property:*

$$\sqrt{Nh}\left(\widehat{\theta}_2(x) - \theta_2(x) - h^2\mu_2\mathbf{\Omega}'(x)\mathbf{B}_a(x)\right) \to_d N(0, \mathbf{\Omega}'(x)\mathbf{\Sigma}(x)\mathbf{\Omega}(x)),$$

where $\mathbf{B}_a(x) = \left([\mathbf{\Lambda}^{(1)}_{\ddot{Z}\ddot{X}}(x)]_{[.,1]} + c_f(x)[\mathbf{\Lambda}_{\ddot{Z}\ddot{X}}(x)]_{[.,1]}\right)\beta_2^{(1)}(x) + \mathbf{\Lambda}_{\ddot{Z}\ddot{X}}(x)\theta_2^{(2)}(x)/2$. *Meanwhile, the alternative non-augmented estimators* $\widehat{\theta}_{2,\ell}(x)$ *follow that*

$$\sqrt{Nh}\left(\widehat{\theta}_{2,\ell}(x) - \theta_2(x) - h^2\mu_2\mathbf{\Omega}'(x)\mathbf{B}_\ell(x)\right) \to_d N(0, \mathbf{\Omega}'(x)\mathbf{\Sigma}(x)\mathbf{\Omega}(x)),$$

where $\mathbf{B}_\ell(x) = \mathbf{\Lambda}^{(1)}_{\ddot{Z}\ddot{X}}(x)\theta_2^{(1)}(x) + \mathbf{\Lambda}_{\ddot{Z}\ddot{X}}(x)\left(\mathsf{c}_f(x)\theta_2^{(1)}(x)1(\ell=0) + \theta_2^{(2)}(x)/2\right)$ *and* $\ell = 0, 1$.

Theorem 4.1 shows that the proposed functional coefficient estimators are asymptotically normal. The asymptotic variance term of different estimators are the same, while their leading asymptotic bias terms differ without a clear ranking.[5] When $d_{h2} = 1$ and the intercept $g_2(.)$ degenerates to a constant, $\mathbf{B}_0(.) = \mathbf{B}_a(.)$. When the dimension of d_{h2} and/or the curvature of $\theta_{2,-1}(.)$ increases, however, the non-augmented estimators are more likely to have a larger asymptotic bias, as is suggested by the formula. We will demonstrate this difference in asymptotic bias using Monte Carlo simulations in Section 5.

Let $\widehat{\mathbf{\Lambda}}_{\ddot{Z}\ddot{X}}(x)$ be a consistent estimator of $\mathbf{\Lambda}_{\ddot{Z}\ddot{X}}(x)$ and $\widehat{f}_{X_1}$ be a consistent estimator of f_{X_1}. Let $\widehat{\varepsilon}_{i2}$ be the residual of individual i, where $\widehat{\varepsilon}_{i2} = Y_{i2} - \ddot{X}'_{i2}\check{\theta}_2(X_{i1})$ and $\check{\theta}_2(.)$ could be any of the three estimators for $\theta_2(.)$ discussed in Sections 4.1 and 4.2. Define

$$\widehat{\mathbf{\Sigma}}(x) = \frac{h}{N\widehat{f}^2_{X_1}(x)}\sum_{i=1}^N \widehat{\varepsilon}^2_{i2}\ddot{Z}_{i2}\ddot{Z}'_{i2}\kappa_h^2(X_{i1} - x).$$

[5]Although $\widehat{\theta}_{2,1}(.)$ always has one fewer leading asymptotic bias term than $\widehat{\theta}_{2,0}(.)$, it is not guaranteed that $|\mathbf{B}_1(.)| \leq |\mathbf{B}_0(.)|$.

The asymptotic variance stated in Theorem 4.1 could be estimated using $\widehat{\mathbf{\Lambda}}_{\ddot{Z}\ddot{X}}(x)$ and $\widehat{\mathbf{\Sigma}}(x)$. Let $\widehat{\mathbf{\Omega}}(x) = \widehat{\mathbf{\Lambda}}_{\ddot{Z}\ddot{X}}(x) \left(\widehat{\mathbf{\Lambda}}'_{\ddot{Z}\ddot{X}}(x)\widehat{\mathbf{\Lambda}}_{\ddot{Z}\ddot{X}}(x)\right)^{-1}$. The following proposition formalizes.

Proposition 4.1 *Suppose that the conditions in Theorem 4.1 hold. Then for each* $x \in \mathcal{X}_1^*$, $\widehat{\mathbf{\Omega}}'(x)\widehat{\mathbf{\Sigma}}(x)\widehat{\mathbf{\Omega}}(x) \to_p \mathbf{\Omega}'(x)\mathbf{\Sigma}(x)\mathbf{\Omega}(x)$.

4.3 Average Effects

Functional estimators discussed in the last section also suggest estimators for the average effects. Let $\vartheta_2 = \mathbb{E}[\theta_2(X_{i1})|X_{i1} \in \mathcal{X}_1^*] = p^{-1} \int_{\mathcal{X}_1^*} \theta_2(x) dF_{X_1}(x)$, where $p = \mathbb{E}[1_{\{X_{i1} \in \mathcal{X}_1^*\}}]$. Recall that we use $\check{\theta}_2(.)$ to represent any of the three $\theta_2(.)$ estimators studied in Sections 4.1 and 4.2. Let

$$\widehat{\vartheta}_2 = \frac{1}{\sum_{i=1}^N 1_{\{X_{i1} \in \mathcal{X}_1^*\}}} \sum_{i=1}^N \check{\theta}_2(X_{i1}) 1_{\{X_{i1} \in \mathcal{X}_1^*\}},$$

be the estimator for ϑ_2. Given $\widehat{\vartheta}_2$, we can define $\widehat{\bar{\beta}}_2 = \left[\widehat{\vartheta}_2\right]_{[1]}$ as an estimator for the average contemporaneous treatment effect $\bar{\beta}_2 = [\vartheta_2]_{[1]}$.

Theorem 4.2 *Suppose that conditions in Theorem 4.1 hold and the bandwidth additionally satisfies that* $Nh^4 \to 0$ *at a polynomial rate of* N. *Let* $N_s = \sum_i 1_{\{X_{i1} \in \mathcal{X}_1^*\}}$ *with* $N_s/N \to p \in (0, 1]$ *as* $N \to \infty$. *Then,* $\widehat{\vartheta}_2$ *satisfies that*

$$\sqrt{N_s}(\widehat{\vartheta}_2 - \vartheta_2) \to_d N(0, \mathbf{\Sigma}_1^* + \mathbf{\Sigma}_2^*),$$

where $\mathbf{\Sigma}_1^* = \mathsf{c}_\kappa \mathbb{E}[\mathbf{\Omega}'(X_{i1})\mathbb{E}[\ddot{\varepsilon}_{i2}^2 \ddot{Z}_{i2} \ddot{Z}'_{i2}|X_{i1}]\mathbf{\Omega}(X_{i1})|X_{i1} \in \mathcal{X}_1^*]$, $\mathbf{\Sigma}_2^* = \mathbb{V}[\theta_2(X_{i1})|X_{i1} \in \mathcal{X}_1^*]$ *and* $\mathsf{c}_\kappa = \int\int \kappa(u)\kappa(u-s)duds$. *If we further assume that* $\theta_2(x) = \theta_2$ *for all* $x \in \mathcal{X}_1^*$, *then*

$$\sqrt{N_s}(\widehat{\vartheta}_2 - \vartheta_2) \to_d N(0, \mathbf{\Sigma}_1^*).$$

The average estimator converges at a rate faster than its functional counterpart since the former uses all data with $X_{i1} \in \mathcal{X}_1^*$ while the latter only uses data in a shrinking window defined by the bandwidth h. To achieve the parametric convergence rate described in Theorem 4.2, a stronger bandwidth condition is required. Similar conditions can be found in, e.g., Su et al. (2013).

The asymptotic variance of $\widehat{\vartheta}_2$ consists of two terms: one is associated with the estimation error of $\widehat{\theta}_2(.)$ and the other with the heterogeneity of $\theta_2(.)$. If the function

$\theta_2(.)$ is not path-dependent, the second term $\mathbf{\Sigma}_2^*$ degenerates to zero. The variances $\mathbf{\Sigma}_1^*$ and $\mathbf{\Sigma}_2^*$ can be estimated, respectively, by

$$\widehat{\mathbf{\Sigma}}_1^* = \widehat{p} \cdot N^{-1} \sum_j \widehat{\varepsilon}_{j2}^2 \widehat{\zeta}(X_{j1}) \ddot{Z}_{j2} \ddot{Z}_{j2}' \widehat{\zeta}'(X_{j1}),$$

$$\widehat{\mathbf{\Sigma}}_2^* = N_s^{-1} \sum_{i:X_{i1} \in \mathcal{X}_1^*} (\check{\theta}_2(X_{i1}) - \widehat{\vartheta}_2)(\check{\theta}_2(X_{i1}) - \widehat{\vartheta}_2)',$$

where $\widehat{\zeta}(x) = N_s^{-1} \sum_{i:X_{i1} \in \mathcal{X}_1^*} \kappa_h(X_{i1} - x) \widehat{f}_{X_1}^{-1}(X_{i1}) \widehat{\mathbf{\Omega}}'(X_{i1})$ and $\widehat{p} = N_s/N$. Recall that $\widehat{\varepsilon}_{j2}$ is defined in Section 4.2.

Proposition 4.2 *Suppose that conditions in Theorem 4.2 hold. Then,* $\widehat{\mathbf{\Sigma}}_1^* \to_p \mathbf{\Sigma}_1^*$ *and* $\widehat{\mathbf{\Sigma}}_2^* \to_p \mathbf{\Sigma}_2^*$.

Remark: note that the parametric convergence rate of the average estimator also suggests a two-step estimation procedure for the partially linear benchmark model. The first step involves estimating the average effect $\widehat{\vartheta}_2$ and obtaining a modified outcome $Y_{i2}^* = Y_{i2} - \ddot{H}_{i2}' \left[\widehat{\vartheta}_2\right]_{[-1]}$, which partials out the effect of other exogenous controls. The second-step involves semi-parametric conditional GMM estimation with Y_{i2}^* being the outcome, $\ddot{X}_{i2} = (X_{i2}\ 1)'$ being the regressor set, and $\ddot{Z}_{i2} = (\omega_2(Z_{i2})'\ 1)'$ being the instrument set. The parametric convergence rate of the first-step average effect estimator implies that first-step estimation error is asymptotically negligible in this two-step estimation strategy.

4.4 Uniform Testing

Besides point-wise inference, researchers might be interested in conducting uniform tests using the proposed estimators for identified treatment effect functions. Take the path-dependent contemporaneous effect function as an example. Researchers might be interested in testing the following null hypotheses:

$$H_{0,zero}: \ \beta_t(x) = 0 \textit{ for all } x \in \mathcal{X}_{t-1};$$

$$H_{0,pos}: \ \beta_t(x) \geq 0 \textit{ for all } x \in \mathcal{X}_{t-1};$$

$$H_{0,neg}: \ \beta_t(x) \leq 0 \textit{ for all } x \in \mathcal{X}_{t-1};$$

$$H_{0,homo}: \ \exists \textit{ a constant } C \textit{ such that } \beta_t(x) = C \textit{ for all } x \in \mathcal{X}_{t-1}.$$

The null hypothesis $H_{0,zero}$ states that the contemporaneous effect is uniformly zero. The null hypotheses $H_{0,pos}$ and $H_{0,neg}$ state that the contemporaneous effect function is uniformly non-negative and uniformly non-positive, respectively. The last null hypothesis $H_{0,homo}$ states that the contemporaneous effect function reduces to a constant.

Let the alternative of each null hypothesis be that the null is incorrect. A rejection of $H_{0,zero}$ implies that the contemporaneous effect is nonzero for at least some values of the past treatment. Similarly, a rejection of $H_{0,pos}/H_{0,neg}$ implies that the contemporaneous effect is negative/positive for at least some values of the past treatment. A rejection of $H_{0,homo}$ implies that the contemporaneous effect function is path-dependent. Following the discussions in Lemma 2.1, rejecting $H_{0,homo}$ also raises questions about adopting the textbook 2SLS estimation strategy in a particular empirical application.

Uniform tests could be carried out using the intersection bounds approach described in Chernozhukov et al. (2013), since our varying-coefficient functional estimators are kernel-based. For illustration purposes, we focus on the null hypothesis $H_{0,zero}$ and set $t = 2$. We suppress the role of H_{i2} without loss of generality because the impact of H_{i2} on the outcome Y_{i2} could be partialed out using the two-step estimation strategy discussed at the end of Section 4.3.

Construct the test statistic for $H_{0,zero}$ as

$$\mathcal{T} = \sup_{x\in\mathcal{X}_1} |\frac{\widehat{\beta}_2(x)}{\widehat{\sigma}_\beta(x)}|.$$

where $\widehat{\beta}_2(x)$ is the benchmark augmented semi-parametric estimator and $\widehat{\sigma}_\beta(x)$ is its standard error obtained from the asymptotic variance estimator defined earlier. Let $\{\eta_i^*\}_{i=1}^N \sim i.i.d.\ N(0,1)$ be a sequence of pseudo-random variables independent of the sample path. The asymptotic distribution of the test statistic under the null hypothesis could be approximated by the multiplier bootstrap statistic

$$\mathcal{T}^* = \sup_{x\in\mathcal{X}_1} |\frac{\widehat{\beta}_2^*(x)}{\widehat{\sigma}_\beta(x)}|,$$

where $\widehat{\beta}_2^*(x)$ is defined the same as $\widehat{\beta}_2(.)$ except for using $\eta_i^*\widehat{\epsilon}_{i2}$ to replace the outcome variable in estimation.

Let $\widehat{c}(\alpha)$ be the $1-\alpha$ quantile of the empirical distribution of the multiplier bootstrap statistic. The decision rule of the test would be to

reject the null hypothesis $H_{0,zero}$ when $\mathcal{T} > \widehat{c}(\alpha)$.

This uniform test also controls size at α asymptotically. Similar uniform tests could also be designed for null hypotheses $H_{0,pos}$, $H_{0,neg}$, and $H_{0,homo}$.

5 Monte Carlo Simulations

In this section, we study the small sample performance of the proposed semi-parametric estimation and inference strategies using Monte Carlo simulations.

First, we compare the proposed augmented and non-augmented semi-parametric varying coefficient estimators with three parametric 2SLS estimators. For the purpose of illustration, we focus on estimating the second-period path-dependent contemporaneous effect function $\beta_2(.)$. The augmented semi-parametric estimator for $\beta_2(.)$ is $\widehat{\beta}_2(.)$, denoted as *semi-benchmark*. The non-augmented estimators are $\widehat{\beta}_{2,0}(.)$, denoted as *semi-alt-lc*, and $\widehat{\beta}_{2,1}(.)$, denoted as *semi-alt-ll*. As is discussed in Theorem 4.1, there is no clear ranking among the three semi-parametric estimators, although it may be reasonable to expect that the augmented estimator has a smaller asymptotic bias when the direct carryover effect is nontrivial and/or when the dimension of the exogenous control vector is large.

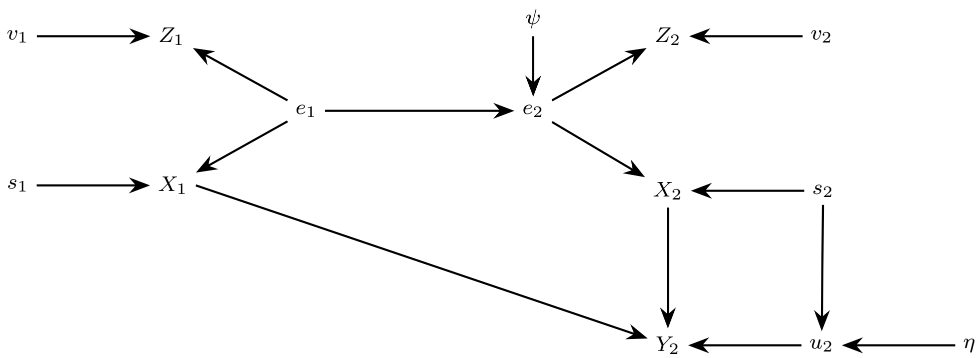
The three parametric estimators for $\beta_2(.)$ are constructed from 2SLS regressions of Y_2 on X_2 instrumented by Z_2 (*para-existing*), of Y_2 on (X_2, X_1, X_1X_2) instrumented by (Z_2, X_1, X_1Z_2) (*para-alt-1*), and of Y_2 on (X_2, X_1, X_1X_2) instrumented by (Z_2, Z_1, Z_1Z_2) (*para-alt-2*). The three parametric estimators rely on different exclusion restrictions.

For semi-parametric estimation, we follow Chernozhukov et al. (2013) and use the rule-of-thumb bandwidth:

$$h = \widehat{h}_{ROT} \times \widehat{s} \times N^{1/5-1/\varrho},$$

where $\widehat{s}$ is the standard deviation of X_1 and ϱ is a parameter. $\widehat{h}_{ROT}$ minimizes the weighted Mean Integrated Square Error of the local linear estimation of Y_2 on studentized X_1. We follow Chernozhukov et al. (2013) to use the quartic kernel function (i.e., $\kappa(s) = 15/16(1-s^2)^2 \cdot 1_{\{|s|\leq 1\}}$) and set the value of ϱ to 3.5. Simulations with $\varrho = 3.25$ and $\varrho = 3.75$ are reported as a robustness check. Note that under-smoothing and in particularly, $\varrho \leq 4$, is required for the average effect estimator constructed by $\widehat{\beta}_2(.)$, $\widehat{\beta}_{2,0}(.)$, or $\widehat{\beta}_{2,1}(.)$, to have satisfactory asymptotic properties, as discussed in Theorem 4.2.

Figure 1: The Data Generating Process, DGP A



Throughout all DGPs, first-period random variables and second-period error terms are generated following:

$$(e_1, \psi) \sim_{iid} exponential(1),\ (s_1, v_1, s_2, v_2, \eta) \sim_{iid} N(0, 0.2^2) \text{ independent of } (e_1, \psi),$$
$$X_1 = e_1 + s_1,\ Z_1 = e_1 + v_1,\ e_2 = \rho e_1 + \sqrt{1-\rho^2}\psi,\ X_2 = e_2 + s_2,\ Z_2 = e_2 + v_2.$$

For the China syndrome application, X_t/Z_t is the trade exposure in the US/European countries in decade t, $t = 1, 2$. Error terms e_t and s_t/v_t can be understood as the unobserved supply shock from China and from the US/European countries, respectively. The correlation between Z_2 and X_1 is governed by the value of ρ. Unless otherwise specified, $\rho = 1$. This simulation setup also features a right-skewed X_1 inspired by the data distribution in Section 6. In both simulation and empirical sections, we set $\mathcal{X}_1^*$ to the dense region $[0, 0.5]$.

The second-period outcome equation varies across DGPs. We generate the baseline model by fitting the empirical dataset used in Section 6 with an OLS regression. The baseline model features a quadratic functional form for the dynamic treatment effect functions. We plot the relationship between simulated random variables in Figure 1.

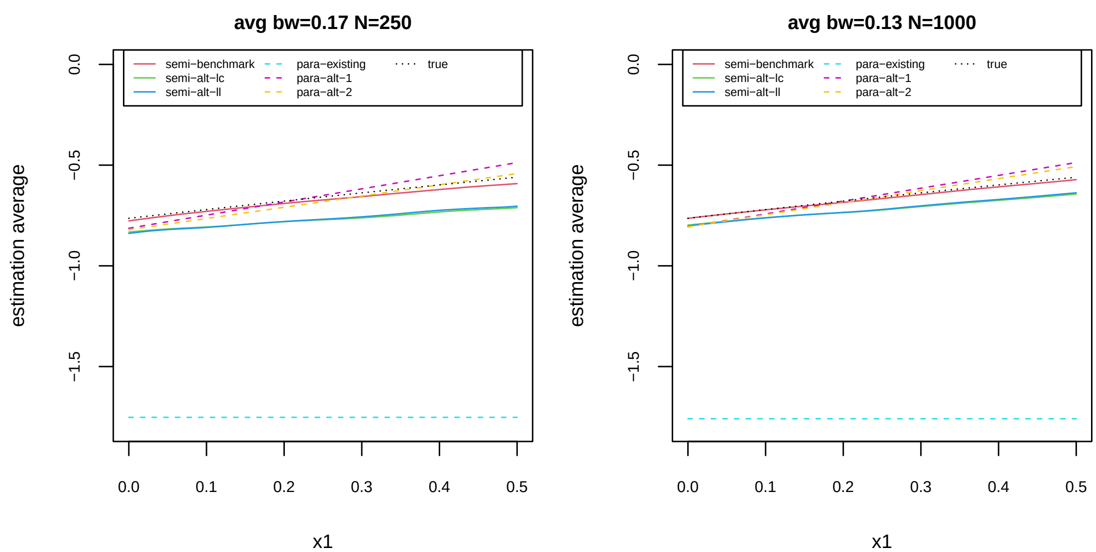
DGP A: the baseline model

$$Y_2 = -3.439 - 2.628X_1 + 0.324X_1^2 + (-0.764 + 0.440X_1 - 0.064X_1^2)X_2 + u_2,$$
$$\text{where } u_2 = 0.6\eta + 0.8s_2.$$

It is easy to see that X_2 is correlated with u_2 while all of X_1, Z_1, and Z_2 are independent of u_2. In addition, the exclusion restriction for *para-existing* is violated, since the estimator completely ignores treatment effect dynamics and, therefore, leaves the error term with a function of X_1 which is correlated with Z_2. Similarly, exclusion restrictions for both *para-alt-1* and *para-alt-2* are violated since the two models mis-specify the functional form of dynamic treatment effects. On the other hand, all semi-parametric estimators are consistent under DGP A because $(Z_2, X_1) \perp u_2$ implies the required conditional exclusion restriction.

Figure 2 compares the performance of all estimators using sample sizes $N = 250$ (left) and $N = 1,000$ (right). The plotted lines are estimation results averaged across 1,000 simulations. Figure 2 shows that the parametric estimator *para-existing* performs extremely poorly. The other two parametric estimators only suffer from mild model misspecification and, therefore, perform much better than *para-existing* under DGP A.

Figure 2: Estimates Averaged Across Simulations, DGP A



Note: Simulations are carried out 1,000 times. Semi-parametric kernel function and bandwidth are chosen following the rule-of-thumb with $\varrho = 3.5$.

Although all semi-parametric estimators are consistent under DGP A, the augmented estimator *semi-benchmark* has a better small sample performance, especially when $N = 250$. This is because the DGP has a substantial direct carryover effect, leading to extra terms in the leading asymptotic bias formula in non-augmented estimators, as described in Theorem 4.1. The augmented estimator *semi-benchmark* also outperforms all parametric estimators under DGP A.

Next, we modify the baseline model to change the form of treatment effect functions. DGP A-2 shuts down the direct carryover effect completely, while DGP A-3 exaggerates the curvature of effect functions to illustrate the drawback of parametric estimation under model misspecification.

DGP A-2: no direct carryover effect

$$Y_2 = -3.439 + (-0.764 + 0.440X_1 - 0.064X_1^2)X_2 + u_2.$$

DGP A-3: exaggerated curvature of effect functions

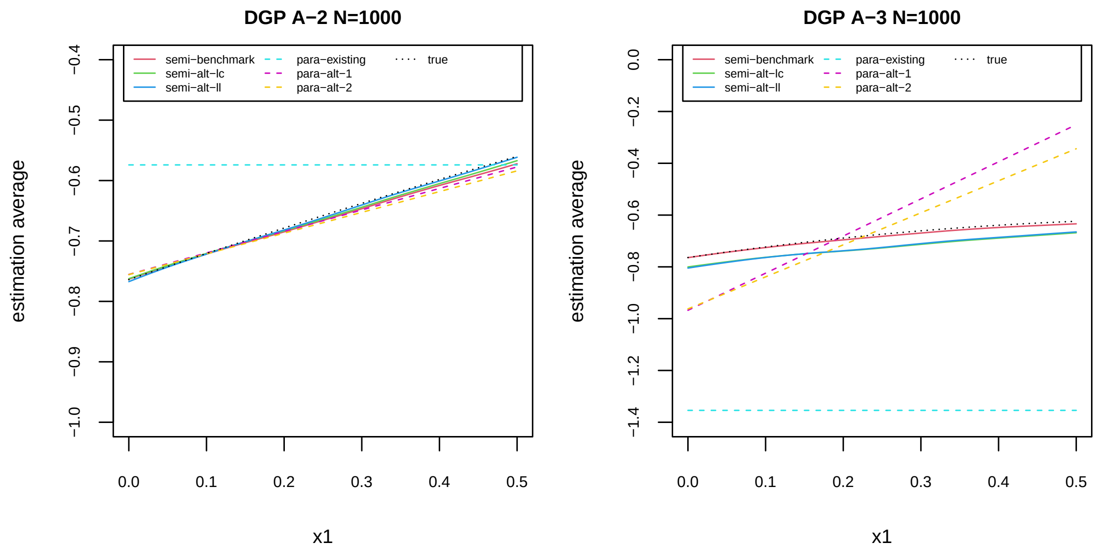
$$Y_2 = -3.439 - 2.628X_1 + 1.620X_1^2 + (-0.764 + 0.440X_1 - 0.320X_1^2)X_2 + u_2.$$

The left graph of Figure 3 reports estimation results for DGP A-2 averaged across 1,000 simulations. As is discussed in Lemma 2.1, when there is no direct carryover effect, the *para-existing* estimator is consistent to a weighted average of the true contemporaneous effect function. It is easy to derive that the weights are proportional to the values of X_1^2 under DGP A-2. This explains why the simulation average of *para-existing* reported in the left graph of Figure 3 is much higher than the simple average of $\beta_2(.)$ or an average weighted by the density of X_1. DGP A-2, therefore, points out another weakness of *para-existing*. – Even in special situations where the estimator converges to a weighted average of the true effect function, the weighted average is likely not empirically relevant.

Under DGP A-2, the three semi-parametric estimators report similar estimation results. This is expected from Theorem 4.1 because, without direct carryover effect and with only mild curvature in the contemporaneous treatment effect function, leading asymptotic bias terms of the three semi-parametric estimators (see Theorem 4.1) are very close to each other. The three semi-parametric estimators also perform better than all of their parametric competitors, even though *para-alt-1* and *para-alt-2* only suffer from mild model misspecification under DGP A-2.

The right graph of Figure 3 reports estimation results for DGP A-3. Since both the direct carryover effect function and the contemporaneous effect function have exaggerated curvatures, all parametric estimators perform poorly due to substantial model misspecification. All semi-parametric estimators continue to yield simulation averages close to the true effect function, with the augmented estimator leading the horse race.

Figure 3: Estimates Averaged Across Simulations, DGPs A-2 and A-3



Note: Simulations are carried out 1,000 times. Semi-parametric kernel function and bandwidth are chosen following the rule-of-thumb with $\varrho = 3.5$.

The next simulation experiment keeps the DGP of A-3, except for setting $\rho = 0$ when generating e_2.

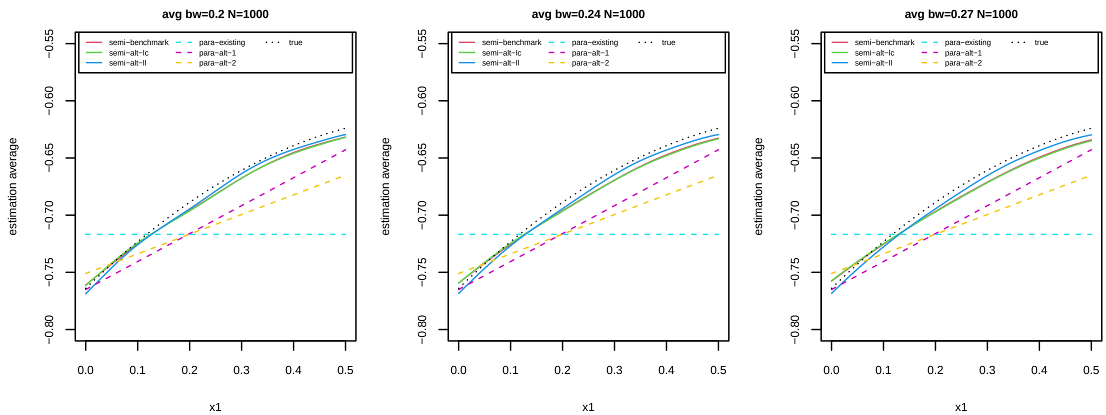
DGP B: external instrument Z_2 independent of X_1

Same as DGP A-3 except that $\rho = 0$.

DGP B breaks the link between e_1 and e_2 in Figure 1. Under the new DGP, both Z_2 and X_2 are independent of X_1. Following the derivations after Lemma 2.1, it is easy to show that the para-existing estimator converges in probability to $\mathbb{E}[\beta_2(X_1)]$ under this special DGP. In addition, under DGP B, leading asymptotic bias terms of the three semi-parametric estimators (see Theorem 4.1) are exactly the same, as $\mathbf{\Lambda}_{\ddot{\mathbf{Z}}\ddot{\mathbf{X}}}(x) = (\mathbb{E}[Z_2X_2]\ 0; 0\ 0)$ regardless of the value of x. Figure 4 reports simulation results for $N = 1,000$ using different bandwidths in semi-parametric estimation. The three semi-parametric estimates have good small sample performances. The estimators do not seem to be sensitive to the bandwidth choice.

In the last group of DGPs, we modify the baseline DGP to shut down treatment effect dynamics in the outcome equation.

Figure 4: Estimates Averaged Across Simulations, DGP B



Note: Simulations are carried out 1,000 times. Semi-parametric kernel function and bandwidth are chosen following the rule-of-thumb. The left graph uses $\varrho = 3.25$. The middle graph uses $\varrho = 3.5$. The right graph uses $\varrho = 3.75$.

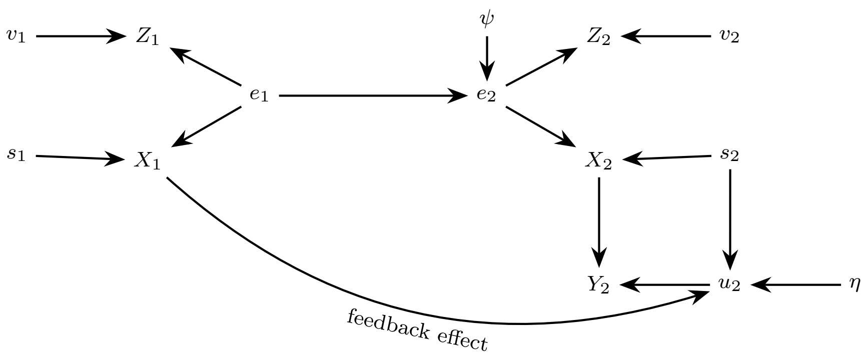
DGP C: no carryover effect or path-dependency in the contemporaneous effect

$$Y_2 = -3.439 - 0.764X_2 + u_2, \quad u_2 = 0.6\eta + 0.8s_2.$$

Under DGP C, all parametric and semi-parametric estimators are consistent. The left graph of Figure 6 reports the empirical mean squared error (MSE) of each estimator as a function of X_1, for sample size $N = 250$. Not surprisingly, the simplest parametric estimator, *para-existing*, gives the smallest MSE. The three semi-parametric estimators have efficiency loss compared to both *para-existing* and *para-alt-1* for only using data within local estimation windows. The most interesting finding is that the parametric estimator *para-alt-2* performs substantially worse than all semi-parametric estimators. This unsatisfactory small sample performance of *para-alt-2* speaks about its demanding rank condition arising from having three endogenous variables in the regression model.

DGPs C-2 and C-3, defined in the following have the same outcome equation as DGP C but with a new feedback effect from X_1 to the outcome error u_2. As is illustrated in Figure 5, the feedback effect generates a nontrivial correlation between Z_2 and u_2 as both are now correlated with X_1. As a result, all three parametric estimators are inconsistent even though there is no treatment effect dynamics in the outcome equation.

Figure 5: The Data Generating Process, DGP A



DGPs C-2 and C-3: feedback effect from X_1 to u_2

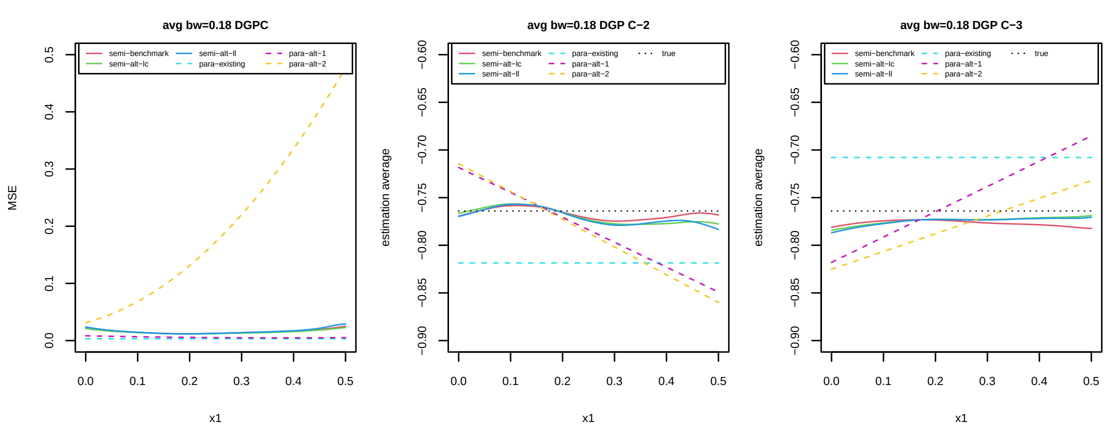
C-2: same as DGP C except that $u_2 = 0.4\eta + 0.8s_2 - 0.3\left((X_1 - \mu_X)^2 - \sigma_X^2\right)$,

C-3: same as DGP C except that $u_2 = 0.4\eta + 0.8s_2 + 0.3\left((X_1 - \mu_X)^2 - \sigma_X^2\right)$,

where μ_X and σ_X^2 are mean and variance of X_1, respectively. It is clear that $\mathbb{E}[u_2] = 0$, $\mathbb{E}[Z_2 u_2] = 0.3\mathbb{E}[Z_2((X_1-1)^2-1)] \neq 0$. The conditional exclusion restriction required for the proposed three semi-parametric estimators is still satisfied since $Z_2 \perp u_2 | X_1$.

Simulation results averaged across 1,000 simulations are summarized in the middle and right graphs of Figure 6, for $N = 250$. Similar to the case of DGP B, only the three semi-parametric estimators show good small sample performances.

Figure 6: MSE of DGP C and Estimation Average of DGPs C-2 and C-3



Note: Simulations are carried out 1,000 times. Semi-parametric kernel function and bandwidth are chosen following the rule-of-thumb with $\varrho = 3.5$.

The contemporaneous treatment effect is path-independent under all three DGPs in group C. Next, we use these DGPs to examine the size control property of the uniform test $H_{0,homo}$ discussed in Section 4.4. Table 1 reports the rejection proportion of the uniform test average across 1,000 simulations. It also reports the rejection proportion of a standard t-test constructed with the average contemporaneous effect estimator, testing if the average effect is equal to the true value. All tests are constructed using the benchmark augmented semi-parametric estimator. Table 1 shows that both the uniform test and the t-test constructed with the proposed average effect estimator have a good small sample size control, especially when $N = 1,000$.

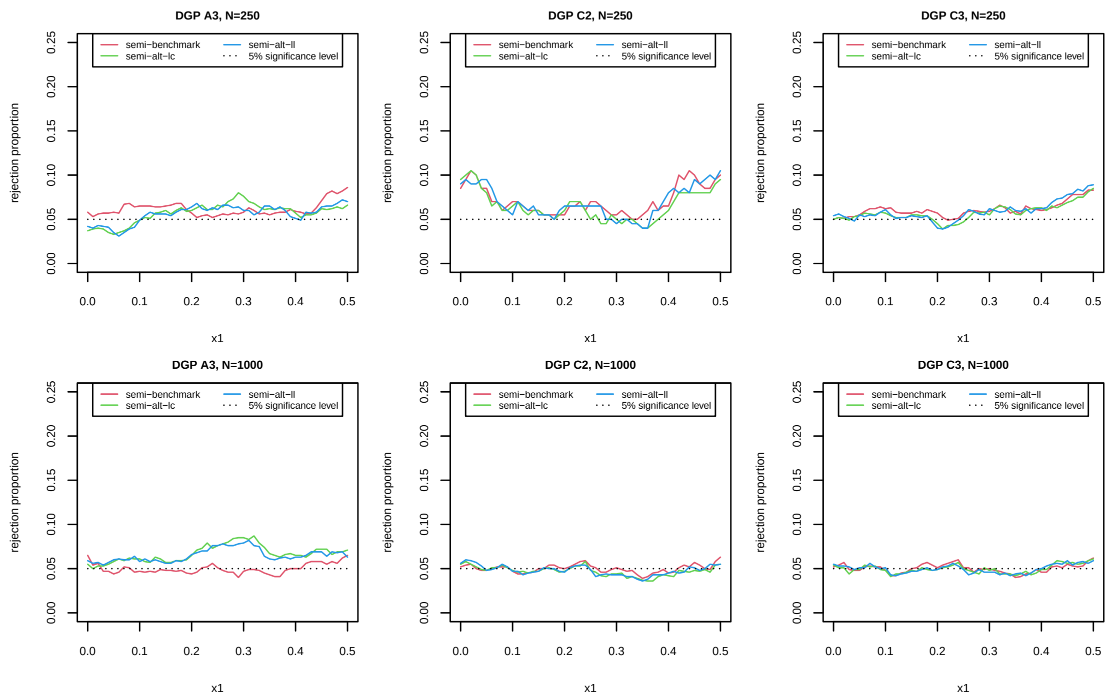
Table 1: Size Control of Proposed Tests

		$N = 250$			$N = 1,000$		
	Size	$\varrho = 3.25$	$\varrho = 3.5$	$\varrho = 3.75$	$\varrho = 3.25$	$\varrho = 3.5$	$\varrho = 3.75$
				DGP C			
Average Effect Test	0.050	0.062	0.054	0.049	0.045	0.045	0.044
Uniform Test for $H_{0,homo}$	0.050	0.054	0.050	0.058	0.026	0.025	0.031
				DGP C-2			
Average Effect Test	0.050	0.058	0.054	0.050	0.047	0.048	0.044
Uniform Test for $H_{0,homo}$	0.050	0.059	0.060	0.068	0.031	0.034	0.035
				DGP C-2			
Average Effect Test	0.050	0.059	0.053	0.052	0.048	0.047	0.045
Uniform Test for $H_{0,homo}$	0.050	0.057	0.065	0.069	0.030	0.033	0.040

Note: Simulations are carried out 1,000 times. Semi-parametric kernel function and bandwidth are chosen following the rule-of-thumb.

Last but not least, we examine the performance of the proposed asymptotic variance estimator. We choose three DGPs that are least favorable to the proposed semi-parameter estimators and carry out point-wise t-tests constructed with the estimators. Figure 7 reports the rejection proportion of each test among 1,000 simulations when the null hypothesis is true. Specifically, the null hypothesis is $H_0 : \beta_2(x) = \beta_{2,true}(x)$, for $x = 0, 0.01, \cdots, 0.49, 0.5$, where $\beta_{2,true}(.)$ is the true value. We see that the tests, especially the benchmark augmented estimator, have very good size control when $N = 1,000$. The tests have some over-rejection when $N = 250$.

Figure 7: Rejection Proportions of Point-wise t-tests Under the Null



Note: Simulations are carried out 1,000 times. $N = 250$. Semi-parametric kernel function and bandwidth are chosen following the rule-of-thumb with $\varrho = 3.5$. Significance level is 5%.

6 Empirical Application: Path-dependent China Shock Effects

Using China's spectacular, supply-driven export growth as a trade shock, combined with a first-difference IV strategy, Autor et al. (2013) investigated the impacts of import penetration on local labor market outcomes. Subsequent studies have adopted or slightly modified its empirical approach and examined the effects of the trade shock in a variety of other contexts, including worker-level outcomes (Autor et al., 2014), industry-level outcomes (Acemoglu et al., 2016), firm reorganization and relocation (Bloom et al., 2019), innovation (Autor et al., 2020b) and political outcomes (Autor et al., 2020a). In addition to import competition, Feenstra et al. (2019) revisited the analysis by adding the expansion of US exports to the framework. However, none of the aforementioned studies have yet considered treatment effect dynamics.

In this section, we revisit the China syndrome application using the proposed model with treatment effect dynamics. Our model is the first in the literature that can be used to explore path-dependency in trade effects. We use the industry-level data from Acemoglu et al. (2016).[6] The outcome of interest (Y_{it}) is the annual log employment change over 1991-1999 ($t = 1$) and over 1999-2011 ($t = 2$) in industry i. The endogenous treatment of interest (X_{it}) is the annual change in US exposure to Chinese import over the same two periods in industry i. The external instrument (Z_{it}) is the annual change in exposure to Chinese import in eight other high-income countries defined in Acemoglu et al. (2016) over the same two periods.

Investigating treatment effect dynamics is especially interesting in the China syndrome application since, the China trade shock has persisted through three distinct periods (c.f., Autor et al., 2021): the gradual beginning of China's export boom during the 1990s; the dramatic surge of China's export growth in the 2000s after its WTO accession; and China's export plateau after 2010. Across industries, the path of trade shock growth over time differs dramatically due to the natural shift in China's export composition following the growth of its economy.

The empirical literature has suggested different channels through which the contemporaneous effect of import competition could be path-dependent. For instance, if an industry was hit hard by Chinese import competition in the first decade, the industry may undergo a structural change and transform its production process from low-quality to high-quality (Bloom et al., 2016).[7] Consequently, in the second period, the industry exhibiting structural transformation may have better capabilities than other industries with no such transformation in responding to the contemporaneous China import competition. On the other hand, the industry hit hard by Chinese import competition may reduce innovation activities (Autor et al., 2020c). In such a case, the industry experiencing a slowdown in innovation may have worse competence in coping with Chinese import

[6] We use the industry-level data in Acemoglu et al. (2016) rather than the location-level data because external instruments in location-level regressions take a shift-share form, which can cause complications in inference as is explained in Borusyak et al. (2022) and Adão et al. (2019).

[7] Campbell and Mau (2021) argue that findings in Bloom et al. (2016) are driven by their empirical approach of adding one before log transformation. This approach, although popular in empirical studies, is not suitable for the trade and innovation context since many firms in sectors competing with Chinese imports had very few patents to begin with. In contrast, Campbell and Mau (2021) find no statistically significant relationship between Chinese import competition and innovation.

competition than other industries with no such innovation deceleration. Either scenario can result in path-dependent contemporaneous treatment effects in the second period. However, to our best knowledge, no such empirical framework has been proposed to estimate path-dependent trade effects.

We use three different variations of the control vector H_{it}: 1) intercept only, 2) a full set of one-digit manufacturing sector fixed effects, and 3) the sector fixed effects as well as all production controls and pre-trend controls considered in Acemoglu et al. (2016), including production workers as a share of total employment, the log average wage, the ratio of capital to value added (in 1991), computer investment as a share of total investment, high-tech equipment as a share of total investment (in 1990), and changes in the log average wage and in the industry's share of total employment over 1976–1991. The data set is a balanced panel of 392 four-digit manufacturing industries over two time periods.

An important first step in studying the potentially path-dependent trade effect is to determine the initial trade exposure treatment timing. In this application, US imports from China had been almost negligible in the 1980s and started to increase in the early 1990s, see Autor et al. (2014, Fig. 1) and Acemoglu et al. (2016, Fig. 2). Therefore, it is reasonable to consider the period from 1991 to 1999 as the first period where the treatment was given to US industries. Furthermore, around the turn of the century, China's accession to the WTO accelerated US imports from China. Given these circumstances, we focus on estimating the potentially path-dependent treatment effect in the second period (1999-2011)—that is, the impact of the China trade shock in the later period (1999-2011) depending on the magnitude of the treatment in the previous period (1991-1999).

Table 2 presents estimation results of the three parametric estimators studied in simulation studies in Section 5. Recall that the estimator *para-existing*, as is reported in columns (1), (4), and (7), is based on a 2SLS regression of Y_2 on X_2 instrumented by Z_2. The estimator is based on the simple model (2.2) which completely ignores treatment effect dynamics. The estimator *para-alt-1*, reported in columns (2), (5), and (8), is based on the 2SLS regression of Y_2 on (X_2, X_1, X_1X_2) instrumented by (Z_2, X_1, X_1Z_2). The estimator *para-alt-2*, as reported in columns (3), (6), and (9), is based on the 2SLS regression of Y_2 on (X_2, X_1, X_1X_2) instrumented by (Z_2, Z_1, Z_1Z_2). All 2SLS regressions are carried out using the same control vectors as those adopted in Acemoglu et al. (2016). Columns (1)-(3) are intercept only. Columns (4)-(6) include sector fixed effects. Columns (7)-(9) include both sector fixed effects and industry production and pretrend controls. The coefficient estimate reported in Column (1) of Table 2, -1.08, is comparable with the coefficient estimate of -1.16 reported in column (7) of Table 2 in Acemoglu et al. (2016). The two numbers are not exactly equal because Acemoglu et al. (2016) weigh observations by 1991 employment while we do not use regression weights.

Across the columns in Table 2, the contemporaneous effect estimators tend to lose statistical significance as more controls are added to the parametric 2SLS regression. In addition, parametric approaches to allowing for direct carryover effect and path-dependency in contemporaneous effect do not seem to be successful. The direct carryover coefficients and the parametric path-dependency coefficients reported in columns (5), (6), (8) and (9) do not have statistical precision.

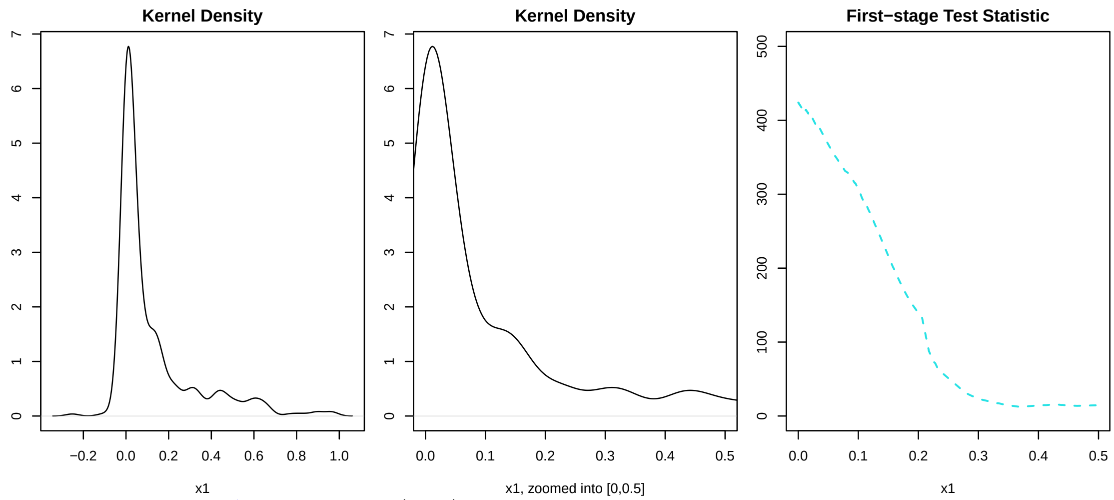
Table 2: Parametric 2SLS Regression Results

	Intercept Only			Sector FEs Only			Controls and Sector FEs		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	para-existing	*para--alt-1*	*para--alt-2*	*para--existing*	*para--alt-1*	*para--alt-2*	*para--existing*	*para--alt-1*	*para--alt-2*
X_2	-1.08***	-1.01***	-0.37	-0.47**	-0.44*	-0.11	-0.31	-0.31	-0.08
	(0.21)	(0.25)	(0.34)	(0.19)	(0.23)	(0.28)	(0.20)	(0.24)	(0.28)
X_1		-0.99***	-3.26***		-0.22	-1.46		-0.03	-0.92
		(0.38)	(1.13)		(0.33)	(1.04)		(0.33)	(1.20)
X_1X_2		0.13	0.21		0.02	0.05		0.00	0.01
		(0.09)	(0.15)		(0.08)	(0.13)		(0.07)	(0.13)
Stat	742	203	12	621	201	10	574	198	7
N	392	392	392	392	392	392	392	392	392

Note: Data are from Acemoglu et al. (2016). Parametric 2SLS regressions are carried out with *Stata*. Stat is the minimum eigenvalue statistic reported following the *ivregress 2sls* command in *Stata*. *, ** and *** indicate significance at 10%, 5%, and 1% level, respectively.

Next, we estimate the potentially path-dependent contemporaneous effect of Chinese import competition in the second decade, i.e., 1999-2011. Figure 9 focuses on the augmented *semi-benchmark* estimator. The effect function is evaluated at $[0, 0.5]$, or for industries that experienced a Chinese import exposure rise of 0-0.5 percentage points per

Figure 8: Distribution of X_1 and First-stage F Statistics



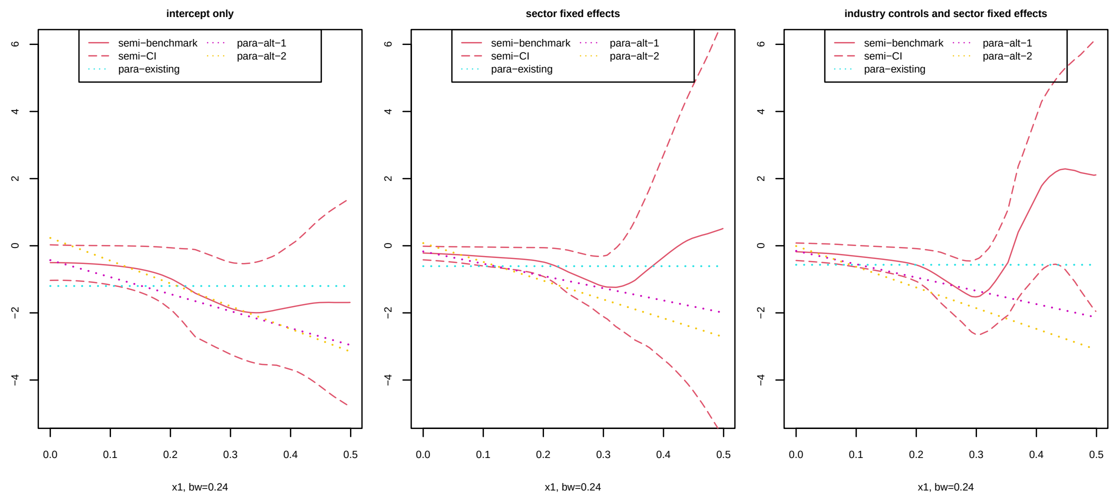
Note: Data are from Acemoglu et al. (2016). Kernel densities reported in the left and middle graphs use the *density* command in *R* and the default bandwidths. The first-stage F test statistics reported in the right graph follows from local 2SLS using the non-augmented *semi-alt-ll* approach.

year over the first decade, i.e., 1991-1999. Given the peculiar heavy right-skewedness (see Figure 8), the range reported in Figure 9 includes over 80% of the industries. Estimation results for non-augmented estimators *semi-alt-1* and *semi-alt-2* are reported in the online appendix, with qualitatively identical empirical findings.

Figure 9 shows that previous decade's Chinese import exposure magnifies the negative impact of the current decade's Chinese import exposure on employment. The magnifying tendency is fairly mild if previous decade's Chinese import exposure X_1 is small. It becomes much stronger (i.e., steeper slope) when X_1 is larger than around 0.2. Across all control vector specifications, the statistical significance of the negative semi-parametric effect estimates is preserved for $X_1 \in [0, 0.3]$. The size of the negative effect also tends to decrease as more control variables are introduced to the model. The lack of statistical precision after X_1 exceeds around 0.3 is due to sparse data and is expected from both the density graph and the first-stage F test statistics graph in Figure 8.

The proposed semi-parametric estimator is contrasted with sub-sample versions of *para-existing*, *para-alt-1*, and *para-alt-2*, using only data with $X_1 \in [0, 0.5]$. We see that the parametric approaches to allowing for dynamics in trade effects, as reported by the dot-dashed lines *para-alt-1* and *para-alt-2*, also indicate that previous decade's import

Figure 9: Second-period Contemporaneous Treatment Effect Estimates



Note: Data are from Acemoglu et al. (2016). Semi-parametric kernel function and bandwidth are chosen following the rule-of-thumb discussed in Section 5 with $\varrho = 3.5$. Robustness checks with $\varrho = 3.25$ and $\varrho = 3.75$ are reported in the appendix and see similar empirical results. The dotted lines for *para-existing*, *para-alt-1*, and *para-alt-2* in each of the three graphs correspond to the parametric approaches reported in Table 2, but carried out with the $X_1 \in [0, 0.5]$ sub-sample.

exposure magnifies the negative impact of the current decade's trade effect. However, compared to the semi-parametric results, the parametric regression fails to discover non-linearity in the path-dependent contemporaneous treatment effect function and generally reports a larger magnifying effect across all three control vector specifications.

Table 3 reports average contemporaneous treatment effect estimates constructed with the semi-parametric functional estimates reported in Figure 9 and Figure 11 in the online appendix. The first row of the table reports semi-parametric average contemporaneous effect estimates integrated over the X_1 range of $[0, 0.3]$, where the functional estimates reported in Figure 9 are quite precisely estimated. Compared to parametric estimates reported in columns (1) and (4) of Table 2, semi-parametric average effect estimates in Table 3 are around 40% smaller when no control variables are considered and 30% smaller when sector fixed effects are considered. The semi-parametric average effect estimates for the model specification with both industry production and pretrend controls and sector fixed effects are similar to the parametric result reported in column (7) of Table 2, although the latter does not have statistical significance. All semi-parametric average estimators reported in the first row of Table 3 are statistically significant at the 5% or

1% significance level. The parametric regression results (see Table 2), on the other hand, lose statistical significance as controls and fixed effects are added to the model.

Table 3: Average Contemporaneous China Shock Effect in 1999-2011

	Intercept Only			Sector FEs Only			Controls and Sector FEs		
$\mathcal{X}_1^*$	(1) *semi-benchmark*	(2) *semi-alt-lc*	(3) *semi-alt-ll*	(4) *semi-benchmark*	(5) *semi-alt-lc*	(6) *semi-alt-ll*	(7) *semi-benchmark*	(8) *semi-alt-lc*	(9) *semi-alt-ll*
$[0, 0.3]$	-0.60**	-0.67**	-0.78***	-0.30***	-0.32***	-0.38***	-0.29**	-0.31**	-0.39**
	(0.26)	(0.28)	(0.31)	(0.11)	(0.11)	(0.14)	(0.14)	(0.14)	(0.16)
$[0, 0.2]$	-0.54**	-0.61**	-0.72**	-0.26**	-0.28**	-0.34**	-0.24*	-0.25*	-0.33**
	(0.26)	(0.29)	(0.31)	(0.11)	(0.11)	(0.14)	(0.14)	(0.13)	(0.16)
$[0.2, 0.3]$	-1.42**	-1.53***	-1.70***	-0.81***	-0.87***	-0.92***	-1.01***	-1.12***	-1.30***
	(0.57)	(0.58)	(0.59)	(0.28)	(0.27)	(0.32)	(0.34)	(0.35)	(0.41)

Note: Data are from Acemoglu et al. (2016). Semi-parametric kernel function and bandwidth are chosen following the rule-of-thumb discussed in Section 5 with $\varrho = 3.5$. Robustness checks with $\varrho = 3.25$ and $\varrho = 3.75$ are reported in the appendix and see similar empirical results. *, ** and *** indicate significance at 10%, 5%, and 1% level, respectively.

The second and third rows of Table 3 report contemporaneous effect estimates averaged over the X_1 range of $[0, 0.2]$ and $[0.2, 0.3]$, respectively. Across all three model specifications, the contemporaneous China shock effect in 1999-2011 is much larger in the third row, i.e., when an industry's China shock exposure in the last decade is over 0.2 percentage points per year. In particular, when all production controls, pre-trend controls, and sector fixed effects are controlled, estimates in columns (7)-(9) suggest that a 1 percentage point increase in industry import penetration reduces domestic industry employment by about 0.25 percentage point when averaged over $X_1 \in [0, 0.2]$, and over 1 percentage point when averaged over $X_1 \in [0.2, 0.3]$. – The contemporaneous China shock effect in 1999-2011 is at least four folds as large for industries exposed to substantial China shock in the past decade, compared to industries exposed to small or moderate shock in the 1990s.

The bigger contemporaneous effect estimates of industries exposed to larger earlier shocks underscore the importance of path-dependency in the contemporaneous treatment effect in the 2000s. While our goal is not to uncover an underlying mechanism behind the magnifying effect, the result could be cautiously interpreted as evidence of decreased innovation activities caused by an earlier exposure to Chinese import penetration during the 1990s. For example, industries significantly affected by Chinese import penetration might have decreased their innovation activities in the first period (Autor et al., 2020c), which further weakens their ability to cope with Chinese import competition in the following period.

Last but not least, we carry out uniform tests to further study properties of the path-dependent contemporaneous effect function. Recall that the four uniform tests illustrated in Section 4.4 are for null hypotheses $H_{0,zero}$, $H_{0,pos}$, $H_{0,neg}$, and $H_{0,homo}$. In this application, $H_{0,zero}$ states that Chinese trade exposure in the 2000's has zero contemporaneous impact. $H_{0,pos}/H_{0,neg}$ states that the contemporaneous impact in the 2000's uniformly increases/decreases with an industry's trade exposure in the 1990's. $H_{0,homo}$ states that the contemporaneous impact from Chinese trade exposure in the 2000's does not vary with an industry's trade exposure in the 1990's. From the functional estimation results reported in Figure 9 (as well as Figures 10 and 11 in the online appendix), we expect null hypotheses $H_{0,zero}$, $H_{0,neg}$, and $H_{0,homo}$ to be rejected.

Panels A and B of Table 4 carry out the tests using the sub-sample with $X_1 \in [0, 0.5]$ and the sub-sample with $X_1 \in [0, 0.3]$, respectively. P-values reported in panel A are larger since estimation precision is lower for $X_1 \in [0.3, 0.5]$ as is shown in Figure 9. Focusing on panel B, we see that null hypotheses $H_{0,zero}$ and $H_{0,neg}$ are strongly rejected at the 5% significance level through all model specifications. The null hypothesis $H_{0,homo}$ is consistently rejected when the richest model specification is used, which allows for both controls and sector fixed effects to soak up variations in the outcome error and increase estimation precision. The null hypothesis $H_{0,neg}$ cannot be rejected, which is in line with the functional estimates reported in Figure 9.

7 Conclusion

In the paper, we propose a new panel IV model featuring treatment effect dynamics. Specifically, our new model allows for a direct carryover effect of the preceding treatment and path-dependency in the contemporaneous treatment effect. We show that in the presence of treatment effect dynamics, existing textbook 2SLS estimators become inconsistent if the external instrument is correlated with previous period's endogenous treatment. This is not uncommon in empirical settings since external instruments are often serially

Table 4: Uniform Testing Results: P-values

	Intercept Only			Sector FEs Only			Controls and Sector FEs		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	semi-benchmark	*semi-alt-lc*	*semi-alt-ll*	*semi-benchmark*	*semi-alt-lc*	*semi-alt-ll*	*semi-benchmark*	*semi-alt-lc*	*semi-alt-ll*
	Panel A: sub-sample with $X_1 \in [0, 0.5]$								
$H_{0,zero}$	0.034	0.025	0.063	0.024	0.034	0.116	0.041	0.021	0.039
$H_{0,neg}$	0.998	0.999	0.999	0.833	0.809	0.845	0.216	0.234	0.227
$H_{0,pos}$	0.014	0.013	0.029	0.012	0.017	0.063	0.022	0.010	0.018
$H_{0,homo}$	0.198	0.205	0.373	0.263	0.304	0.503	0.058	0.056	0.235
	Panel B: sub-sample with $X_1 \in [0, 0.3]$								
$H_{0,zero}$	0.020	0.016	0.051	0.013	0.017	0.101	0.000	0.004	0.023
$H_{0,neg}$	1.000	1.000	1.000	1.000	1.000	0.988	0.996	0.999	0.985
$H_{0,pos}$	0.005	0.005	0.021	0.008	0.009	0.055	0.000	0.002	0.011
$H_{0,homo}$	0.076	0.090	0.194	0.046	0.075	0.153	0.005	0.008	0.031

Note: Data are from Acemoglu et al. (2016). Semi-parametric kernel function and bandwidth are chosen following the rule-of-thumb discussed in Section 5 with $\varrho = 3.5$. Robustness checks with $\varrho = 3.25$ and $\varrho = 3.75$ are reported in the appendix and see similar empirical results. Bootstraps are carried out 1000 times for each simulation.

correlated in data. To address this issue, we propose a novel semi-parametric identification procedure and three local GMM estimators for parameters in the model, including a benchmark augmented estimator and two alternatives. We study asymptotic properties of the suggested estimators and show that proposed estimators have satisfactory small sample performance in Monte Carlo simulations. When applied to revisit the seminal study by Acemoglu et al. (2016) on the China syndrome, our proposed method reveals important empirical findings that have not been discovered previously. In particular, we find that the contemporaneous impact of increased Chinese import competition on US manufacturing employment is magnified by the import exposure in the preceding decade. The size of the magnifying effect is mild if the last decade's import exposure was small or moderate. But the interaction between the past and current trade shocks becomes much more significant when the import exposure over the last decade exceeds 0.2 percentage points per year.

Appendix A: Robustness Checks for the Empirical Analysis

In this section, we report robustness checks of the empirical results. Table 5 repeats parametric regressions reported in Table 2 but with different sub-sample definitions.

Table 5: Robustness Checks: Parametric 2SLS Regression Results

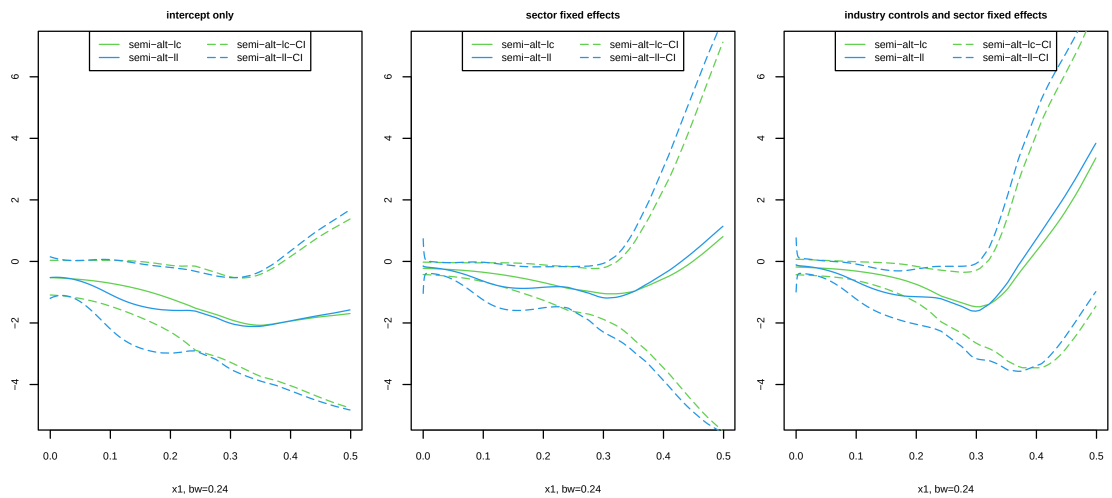
	Intercept Only			Sector FEs Only			Controls and Sector FEs		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	para-existing	*para-alt-1*	*para-alt-2*	*para-existing*	*para-alt-1*	*para-alt-2*	*para-existing*	*para-alt-1*	*para-alt-2*
	Panel A: sub-sample with $X_1 \in [0, 0.5]$								
X_2	-1.20***	-0.43*	0.23	-0.61***	-0.17***	0.08	-0.57***	-0.15	-0.01
	(0.21)	(0.26)	(0.33)	(0.18)	(0.21)	(0.27)	(0.18)	(0.21)	(0.25)
X_1		-1.53	-11.81*		0.79	0.73		1.95	6.08
		(2.44)	(6.88)		(2.15)	(6.19)		(2.11)	(6.67)
X_1X_2		-5.06***	-6.78**		-3.65**	-5.59**		-3.96***	-6.17**
		(1.64)	(3.09)		(1.41)	(2.62)		(1.38)	(2.60)
Stat	711	119	11	604	104	9	587	97	7
N	307	307	307	307	307	307	307	307	307
	Panel B: sub-sample with $X_1 \in [0, 0.3]$								
X_2	-0.87***	-0.44*	1.12	0.23	-0.19	0.08	-0.44***	-0.15	0.10
	(0.20)	(0.25)	(0.85)	(0.17)	(0.21)	(0.47)	(0.17)	(0.20)	(0.39)
X_1		-5.56	-47.85***		-1.05	-20.10		1.62	7.57
		(3.50)	(18.11)		(3.01)	(13.31)		(2.92)	(14.59)
X_1X_2		-4.01**	-12.71		-3.47**	-4.60		-3.96***	-6.29
		(1.88)	(12.14)		(1.49)	(7.16)		(1.45)	(6.63)
F Stat	737	169	3	642	169	3	628	154	3
N	277	277	277	277	277	277	277	277	277

Note: Data are from Acemoglu et al. (2016). Parametric 2SLS regressions are carried out with *Stata*. F Stat is the minimum eigenvalue statistic reported following the *ivregress 2sls* command in *Stata*. *, ** and *** indicate significance at 10%, 5%, and 1% level, respectively.

Figure 10 reports semi-parametric functional estimation results using alternative non-augmented local GMM estimation methods. Compared to Figure 9, which reports results using the augmented approach, estimates in Figure 10 lead to similar empirical findings of magnifying contemporaneous impact of Chinese import exposure.

Last but not least, we use Figure 10 and Tables 6–7 to demonstrate that semi-parametric estimation results reported in Section 6 are not sensitive to alternative bandwidth choices.

Figure 10: Robustness Checks: Second-period Contemporaneous Treatment Effect Estimates



Note: Data are from Acemoglu et al. (2016). Semi-parametric kernel function and bandwidth are chosen following the rule-of-thumb discussed in Section 5 with $\varrho = 3.5$. The blue dotted line of *para-existing* and purple dashed line of *para-alt-2* in each graph correspond to the parametric approaches reported in Table 2 but for a sub-sample with $X_1 \in [0, 0.5]$.

Appendix B: Parametric Identification

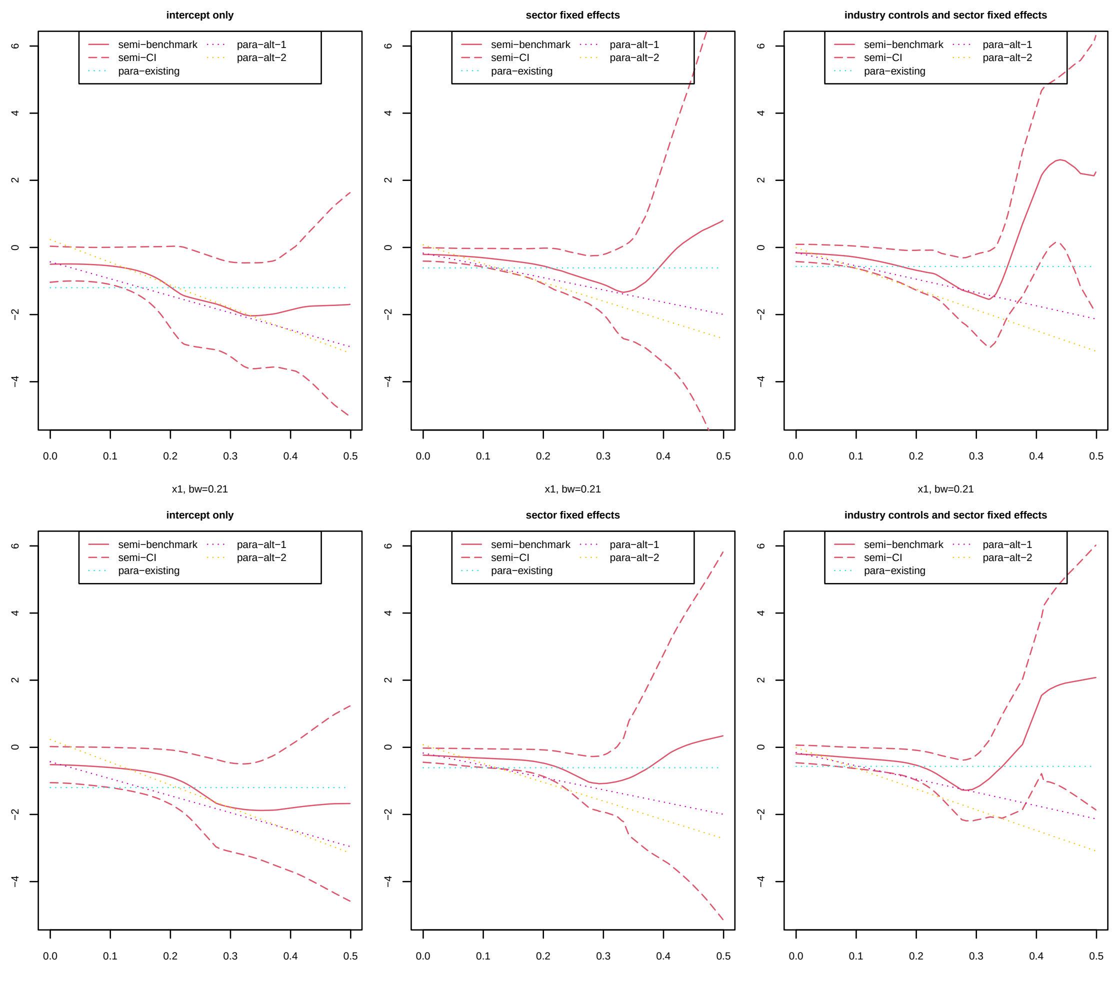
Without loss of generality, we discuss the case $T = 2$ for parametrically identifying the *benchmark* model (2.1).

Let $\boldsymbol{\nu}_{i,s:t} = (\nu_{is}' \ \ldots \ \nu_{it}')'$ denote the random vector that stacks $\nu_{i\ell}$ from period s to period t. Let $\vee$ denote the larger and $\wedge$ the smaller of two numbers. Let $f_{it} = \omega_t(\mathbf{Z}_{i,[(t-s)\vee 1]:t})$ be the d_{ft}-dimensional vector generated by external instruments from period $[(t-s) \vee 1]$ to period t with known function $\omega_t(.)$. Then unknown parameters in model (2.1) are identified through classic parametric 2SLS or GMM estimation strategies under the following assumptions. For $t = 1$, the enodgeneous regressor is X_{i1} and the external instrument set is f_{i1}. For $t = 2$, the enodgeneous regressor is $\psi_{i,d_\beta} X_{i2}$ and the external instrument set is f_{i2}.

Assumption B.1 (parametric identification) *Assume that*

1. *(known functional form)* $\beta_2(X_{i1}) = \psi_{i,d_\beta}' \eta_{d_\beta}$, *where* $\psi_{i,d_\beta} = (1 \ \psi_2(X_{i1}) \ \ldots \ \psi_{d_\beta}(X_{i1}))'$ *and* η_{d_β} *is a* d_β*-dimensional parameter vector;*

Figure 11: Robustness Checks: Second-period Contemporaneous Treatment Effect Estimates



Note: Data are from Acemoglu et al. (2016). Semi-parametric kernel function and bandwidth are chosen following the rule-of-thumb discussed in Section 5 with $\varrho = 3.25$ in the top panel and $\varrho = 3.75$ in the bottom panel. The blue dotted line of *para-existing* and purple dashed line of *para-alt-2* in each graph correspond to the parametric approaches reported in Table 2 but for a sub-sample with $X_1 \in [0, 0.5]$.

2. *(exclusion restriction)* $\mathbb{E}[\varepsilon_{it}|\mathbf{Z}_{i,[(t-s)\vee 1]:t}, H_{it}] = 0$, *for* $s = 0, 1$ *and* $t = 1, 2$*;*

3. *(rank condition)* $\mathbb{E}[(f'_{i1}\ \ H'_{i1})'(X'_{i1}\ \ H'_{i1})]$ *and* $\mathbb{E}[(f'_{i2}\ \ H'_{i2})'(X'_{i1}\ \ (\psi_{i,d_\beta} X_{i2})'\ \ H'_{i2})]$ *are both of full rank.*

Assumption B.1.1 is a standard parametric functional form assumption. Assumptions B.1.2 and B.1.3 are standard exclusion restriction and rank condition for parametric IV regressions. If $s = 1$, the assumption reduces to $E[\varepsilon_{i1}|Z_{i1}, H_{i1}] = 0$ and

Table 6: Average Contemporaneous China Shock Effect in 1999-2011: Robustness Checks with Alternative Bandwidths

$\mathcal{X}_1^*$	Intercept Only			Sector FEs Only			Controls and Sector FEs		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	semi-benchmark	*semi-alt-lc*	*semi-alt-ll*	*semi-benchmark*	*semi-alt-lc*	*semi-alt-ll*	*semi-benchmark*	*semi-alt-lc*	*semi-alt-ll*
	Panel B: $\varrho = 3.25$								
$[0, 0.3]$	-0.60**	-0.66**	-0.78**	-0.29***	-0.32***	-0.39***	-0.28**	-0.31**	-0.41**
	(0.25)	(0.27)	(0.31)	(0.11)	(0.11)	(0.14)	(0.14)	(0.13)	(0.16)
$[0, 0.2]$	-0.54**	-0.59**	-0.72**	-0.25**	-0.28**	-0.35**	-0.23*	-0.24*	-0.35**
	(0.26)	(0.27)	(0.32)	(0.11)	(0.11)	(0.14)	(0.14)	(0.14)	(0.16)
$[0.2, 0.3]$	-1.54**	-1.62**	-1.68***	-0.80***	-0.89***	-0.83***	-0.97***	-1.22***	-1.20***
	(0.63)	(0.63)	(0.61)	(0.19)	(0.19)	(0.23)	(0.24)	(0.25)	(0.30)
	Panel B: $\varrho = 3.75$								
$[0, 0.3]$	-0.61**	-0.69**	-0.77***	-0.30***	-0.33***	-0.41***	-0.29**	-0.31**	-0.39
	(0.27)	(0.30)	(0.29)	(0.11)	(0.12)	(0.14)	(0.14)	(0.13)	(0.53)
$[0, 0.2]$	-0.56**	-0.64**	-0.70**	-0.27**	-0.30***	-0.37***	-0.25*	-0.26*	-0.32
	(0.27)	(0.31)	(0.30)	(0.11)	(0.12)	(0.14)	(0.14)	(0.13)	(0.55)
$[0.2, 0.3]$	-1.30***	-1.46***	-1.75***	-0.77***	-0.82***	-0.97***	-0.90***	-0.97***	-1.29***
	(0.48)	(0.50)	(0.51)	(0.24)	(0.24)	(0.28)	(0.29)	(0.30)	(0.38)

Note: Data are from Acemoglu et al. (2016). Semi-parametric kernel function and bandwidth are chosen following the rule-of-thumb discussed in Section 5 with $\varrho = 3.25$ in panel A and $\varrho = 3.75$ in panel B. *, ** and *** indicate significance at 10%, 5%, and 1% level, respectively.

$E[\varepsilon_{i2}|Z_{i1}, Z_{i2}, H_{i2}] = 0$. If pre-intervention periods of the external instrument (e.g., Z_{i0}, $Z_{i(-1)}$) are observed, Assumption B.1.2 could also be modified to utilize such information.

Appendix C: Additional Assumptions and Some Useful Lemmas

The following assumption is required for the asymptotic results stated in Section 4.

Assumption C.1 *(a) The observations* $\{Y_{i2}, X_{i1}, X_{i2}, Z_{i2}, H_{i2}\}_{i=1}^N$ *are i.i.d.*

(b) The density function $f_{X_1}(.)$ *of* X_{i1} *is twice continuously differentiable with bounded derivatives and bounded away from zero on* $\mathcal{X}_1$*, the compact support of* X_{i1}*.*

(c) The function $\theta_2(.)$ *is three times continuously differentiable on* $\mathcal{X}_1$*.*

(d) The kernel function $\kappa(.)$ *is a symmetric density function with compact support.*

Table 7: Uniform Testing Results: Robustness Checks with Alternative Bandwidths

	Intercept Only			Sector FEs Only			Controls and Sector FEs		
	(1)	(2)	(3)	(4)	(5)	(6)			
	semi-benchmark	*semi-alt-lc*	*semi-alt-ll*	*semi-benchmark*	*semi-alt-lc*	*semi-alt-ll*	*semi-benchmark*	*semi-alt-lc*	*semi-alt-ll*
	Panel A-1: $\varrho = 3.25$, sub-sample with $X_1 \in [0, 0.5]$								
$H_{0,zero}$	0.067	0.041	0.098	0.044	0.044	0.143	0.128	0.029	0.043
$H_{0,neg}$	0.997	0.998	0.999	0.854	0.838	0.845	0.126	0.137	0.166
$H_{0,pos}$	0.036	0.017	0.057	0.024	0.023	0.079	0.061	0.013	0.020
$H_{0,homo}$	0.261	0.260	0.406	0.347	0.414	0.536	0.150	0.072	0.241
	Panel A-2: $\varrho = 3.25$, sub-sample with $X_1 \in [0, 0.3]$								
$H_{0,zero}$	0.043	0.030	0.078	0.022	0.030	0.134	0.007	0.006	0.026
$H_{0,neg}$	0.999	1.000	1.000	0.999	1.000	1.000	0.996	1.000	1.000
$H_{0,pos}$	0.025	0.010	0.047	0.014	0.013	0.075	0.003	0.002	0.012
$H_{0,homo}$	0.103	0.133	0.263	0.066	0.124	0.252	0.010	0.015	0.055
	Panel B-1: $\varrho = 3.75$, sub-sample with $X_1 \in [0, 0.5]$								
	sub-sample with $X_1 \in [0, 0.5]$								
$H_{0,zero}$	0.034	0.030	0.047	0.047	0.020	0.047	0.069	0.017	0.636
$H_{0,neg}$	0.999	0.997	0.999	0.812	0.805	0.846	0.300	0.295	0.515
$H_{0,pos}$	0.013	0.015	0.019	0.026	0.009	0.023	0.039	0.008	0.319
$H_{0,homo}$	0.164	0.196	0.274	0.231	0.237	0.351	0.093	0.066	0.502
	Panel B-2: $\varrho = 3.75$, sub-sample with $X_1 \in [0, 0.3]$								
$H_{0,zero}$	0.016	0.016	0.036	0.024	0.014	0.042	0.003	0.006	0.629
$H_{0,neg}$	0.999	1.000	1.000	1.000	1.000	0.922	1.000	0.999	0.869
$H_{0,pos}$	0.006	0.006	0.015	0.012	0.007	0.019	0.002	0.004	0.315
$H_{0,homo}$	0.076	0.080	0.126	0.061	0.051	0.024	0.009	0.006	0.326

Note: Data are from Acemoglu et al. (2016). Semi-parametric kernel function and bandwidth are chosen following the rule-of-thumb discussed in Section 5 with $\varrho = 3.5$. Robustness checks with $\varrho = 3.25$ and $\varrho = 3.75$ are reported in the appendix and see similar empirical results. Bootstraps are carried out 1000 times for each simulation.

(e) *The matrix* $\Lambda_{\ddot{Z}\ddot{X}}(.)$ *is twice continuously differentiable on* $\mathcal{X}_1$, *and* $\mathbb{E}[\tilde{\varepsilon}_{i2}^2 \ddot{Z}_{i2}\ddot{Z}_{i2}' | X_{i1} = .]$ *is Lipschitz continuous on* $\mathcal{X}_1$.

(f) *There exists an* $s > 2$ *such that* $\sup_{x \in \mathcal{X}_1} \mathbb{E}[\|\ddot{Z}_{i2}\|^{2s} | X_{i1} = x] < \infty$, $\sup_{x \in \mathcal{X}_1} \mathbb{E}[|X_{i2}|^{2s} | X_{i1} = x] < \infty$ *and* $\sup_{x \in \mathcal{X}_1} \mathbb{E}[|Y_{i2}|^{2s} | X_{i1} = x] < \infty$ *and* $N^{2\delta - 1} h \to \infty$ *for some* $\delta < 1 - s^{-1}$.

(g) $h \to \infty$, $Nh^3 \to \infty$, $Nh^7 \to 0$ *as* $N \to \infty$.

Next, we define some useful notations. Let ι_j be a $(\ell + 1)$-dimensional vector whose

j-th element is one and all the rest elements are non-zero. Let $\widetilde{X}_{i1}'(x) = (1\ (X_{i1} - x)\ \ldots (X_{i1} - x)^\ell)$. For all values of x, the ℓ-th order local polynomial estimator of $\mathbf{\Lambda}_{\ddot{Z}Y}(x)$ is defined as

$$\widehat{\mathbf{\Lambda}}_{\ddot{Z}Y}(x) = N^{-1} \sum_i \kappa_h(X_{i1} - x) \ddot{Z}_{i2} Y_{i2} \widetilde{X}_{i1}'(x) \widehat{\mathbf{M}}(x)^{-1} \iota_1,$$

with $\widehat{\mathbf{M}}(x) = N^{-1} \sum_i \kappa_h(X_{i1} - x) \widetilde{X}_{i1}(x) \widetilde{X}_{i1}'(x)$. Similarly, the l-th order local polynomial estimator of $\left[\widehat{\mathbf{\Lambda}}_{\ddot{Z}\ddot{X}}(x)\right]_{\cdot,j}$ is defined as

$$\left[\widehat{\mathbf{\Lambda}}_{\ddot{Z}\ddot{X}}(x)\right]_{\cdot,j} = N^{-1} \sum_i \kappa_h(X_{i1} - x) \ddot{Z}_{i2} \ddot{X}_{i2,j} \widetilde{X}_{i1}'(X) \widehat{\mathbf{M}}(x)^{-1} \iota_1,$$

where $\ddot{X}_{i2,j}$ is the j-th element of $\ddot{X}_{i2}$. We suppress ℓ in estimator definitions in this appendix.

Define $\widehat{\mathbf{\Lambda}}_{\ddot{Z}_a Y}(x)$ and $\left[\widehat{\mathbf{\Lambda}}_{\ddot{Z}_a \ddot{X}_a}(x)\right]_{\cdot,j}$ similar to $\widehat{\mathbf{\Lambda}}_{\ddot{Z}Y}(x)$ and $\left[\widehat{\mathbf{\Lambda}}_{\ddot{Z}\ddot{X}}(x)\right]_{\cdot,j}$ with $\ell = 0$ and $(\ddot{Z}_{i2}'\ \ddot{H}_{i2,a}'/h)'$ and $(\ddot{X}_{i2}'\ \ddot{H}_{i2,a}'/h)'$ replacing $\ddot{Z}_{i2}$ and $\ddot{X}_{i2}$, respectively.

To derive asymptotic properties of the above estimators, we first introduce several matrix notations. We use $\mathbf{0}_{d_1,d_2}$ to denote the $d_1 \times d_2$ matrix of zeros. Let $\widetilde{\mathbf{X}}_1(x)$ be a $N \times (\ell + 1)$ matrix whose i-th row is given by $\widetilde{X}_{i1}'(x)$ and $\mathbf{K}(x)$ be a $N \times N$ diagonal matrix whose diagonal entries are given by $\{\kappa_h(X_{i1} - x)\}_{i=1}^N$. For notational convenience, hereafter let d_A denote the dimension of the vector A. Moreover, let $\widetilde{Y}_{i2} = \ddot{Z}_{i2} Y_{i2}$ and $\widetilde{X}_{i2} = \text{vec}(\ddot{Z}_{i2} \ddot{X}_{i2}')$ and $\widetilde{\mathbf{Y}}_2$ and $\widetilde{\mathbf{X}}_2$ be $N \times d_{\ddot{Z}}$ and $N \times d_{\ddot{Z}} d_{\ddot{X}}$ matrices collecting the vectors $\widetilde{Y}_{i2}$ and $\widetilde{X}_{i2}$. Then,

$$\begin{aligned}
\widehat{\mathbf{M}}(x) &= N^{-1} \widetilde{\mathbf{X}}_1'(x) \mathbf{K}(x) \widetilde{\mathbf{X}}_1(x), \\
\widehat{\mathbf{\Lambda}}_{\ddot{Z}Y}(x) &= N^{-1} \widetilde{\mathbf{Y}}_2' \mathbf{K}(x) \widetilde{\mathbf{X}}_1(x) \widehat{\mathbf{M}}(x)^{-1} \iota_1, \\
\text{vec}(\widehat{\mathbf{\Lambda}}_{\ddot{Z}\ddot{X}})(x) &= N^{-1} \widetilde{\mathbf{X}}_2' \mathbf{K}(x) \widetilde{\mathbf{X}}_1(x) \widehat{\mathbf{M}}(x)^{-1} \iota_1.
\end{aligned}$$

For each integer j, we let $\mathbf{M}_j = (\mu_{i+k+j-2})_{1 \le i,k \le \ell+1}$ with $\mu_k = \int u^k \kappa(u) du$ for an integer k, and let $\mathbf{M} \equiv \mathbf{M}_0$. It is easy to see that and $\mathbf{M}_j \iota_s = \mathbf{M}_{j-1} \iota_{s+1}$ for $s = 1, \ldots, \ell$. The matrix $\mathbf{D}$ is defined by a $(\ell+1) \times (\ell+1)$ diagonal matrix whose diagonal entries are given by $\{1\ h\ \cdots h^\ell\}$. Moreover, we let $\mathbf{L}_{\ddot{Z}\ddot{X}}(x)$ be a $d_{\ddot{X}} d_{\ddot{Z}} \times (\ell + 1)$ matrix whose ith column is given by $\text{vec}(\mathbf{\Lambda}_{\ddot{Z}\ddot{X}}^{(i-1)})/(i-1)!$. Following classic kernel derivations in, for example, Fan and Gijbels (1996), we know that, for all $x \in \mathcal{X}_1$, under the conditions of Theorem 4.1,

$$\mathbf{D}^{-1}\widehat{\mathbf{M}}(x)\mathbf{D}^{-1} = \mathbf{M} f_{X_1}(x) + h\mathbf{M}_1 f_{X_1}^{(1)}(x) + O_p(a_h), \tag{C.1}$$

$$\text{vec}(\widehat{\mathbf{\Lambda}}_{\ddot{Z}\ddot{X}}(x)) = \text{vec}(\mathbf{\Lambda}_{\ddot{Z}\ddot{X}}(x)) + O_p((Nh)^{-1/2}) + O_p(h_\ell), \tag{C.2}$$

where $a_h = (Nh)^{-1/2} + h^2 = o(h)$ since $Nh^3 \to \infty$, and h_ℓ is h^2 for both local constant local linear estimation.

In addition, let $\mathbf{D}_a = \text{diag}(\mathbf{I}_{d_{\ddot{X}}}, h\mathbf{I}_{d_{\ddot{H}}})$. Applying Lemma A.1 of Su et al. (2013), we have that, for each $x \in \mathcal{X}_1$,

$$\mathbf{D}_a^{-1}\widehat{\Lambda}_{\ddot{Z}_a\ddot{X}_a}(x) = f_{X_1}(x)\text{diag}\left(\mathbf{\Lambda}_{\ddot{Z}\ddot{X}}(x), \mu_2\mathbf{\Lambda}_{\ddot{H}\ddot{H}}(x)\right) + o_p(1), \tag{C.3}$$

where $\mathbf{\Lambda}_{\ddot{H}\ddot{H}}(x) = \mathbb{E}[\ddot{H}_{i2}\ddot{H}_{i2}'|X_{i1} = x]$.

The following two lemmas state useful results for proving the Theorems and Propositions of the main paper.

Lemma C.1 *Let* $\widehat{\mathbf{B}}_{a,1}(x) = N^{-1}\sum_i \kappa_h(X_{i1} - x)\ddot{Z}_{i2}X_{i2}(\beta_2(X_{i1}) - \beta_2(x))$ *and* $\widehat{\mathbf{B}}_{a,2}(x) = N^{-1}\sum_i \kappa_h(X_{i1} - x)\ddot{Z}_{i2}\ddot{H}_{i2}'(\theta_{2,-1}(X_{i1}) - \theta_{2,-1}(x) - \theta_{2,-1}^{(1)}(x)(X_{i1} - x))$*. Suppose conditions in Theorem 4.1 hold. Then,*

$$\widehat{\mathbf{B}}_{a,1}(x) + \widehat{\mathbf{B}}_{a,2}(x) = f_{X_1}(x)h^2\mu_2\mathbf{B}_a(x) + O_p(ha_h). \tag{C.4}$$

Lemma C.2 *Let* $\widehat{\mathbf{B}}(x) = N^{-1}\sum_i \ddot{Z}_{i2}\ddot{X}_{i2}'\theta_2(X_{i1})\widetilde{X}_{i1}'(x)\kappa_h(X_{i1} - x)\widehat{\mathbf{M}}^{-1}(x)\iota_1 - \widehat{\mathbf{\Lambda}}_{\ddot{Z}\ddot{X}}(x)\theta_2(x)$*. Suppose that conditions in Theorem 4.1 hold. Then,*

$$\widehat{\mathbf{B}}(x) = h^2\mu_2\mathbf{B}_0 + O_p(h^3) \text{ if } \ell = 0 \text{ and } \widehat{\mathbf{B}}(x) = h^2\mu_2\mathbf{B}_1 + O_p(h^3) \text{ if } \ell = 1,$$

with $\mathbf{B}_0$ *and* $\mathbf{B}_1$ *defined in Theorem 4.1.*

Lemma C.3 *Suppose that the conditions in Theorem 4.1 hold. Then, the following holds* *uniformly for* $x \in \mathcal{X}_1$.

$$\sup_{x\in\mathcal{X}_1} \|(\mathbf{I}_{d_{\ddot{X}}}\ \ \mathbf{0}_{d_{\ddot{X}},d_{\ddot{H}}})\widehat{\mathbf{\Omega}}'_a(x) - (f_{X_1}^{-1}(x)\mathbf{\Omega}'(x)\ \ \ \mathbf{0}_{d_{\ddot{X}},d_{\ddot{H}}})\| = O_p(c_h + h^2) \tag{C.5}$$

$$\sup_{x\in\mathcal{X}_1} \left\|\mathbf{D}^{-1}\widehat{\mathbf{M}}(x)\mathbf{D}^{-1} - \left(\mathbf{M}f_{X_1}(x) + h\mathbf{M}_1 f_{X_1}^{(1)}(x)\right)\right\| = O_p(c_h + h^2) \tag{C.6}$$

$$\sup_{x\in\mathcal{X}_1} \|\mathbf{\Psi}_a(x)\| = O_p(c_h), \quad \sup_{x\in\mathcal{X}_1} \|\mathbf{\Psi}(x)\| = O_p(c_h), \tag{C.7}$$

$$\sup_{x\in\mathcal{X}_1} \|\widehat{\mathbf{\Psi}}(x) - \mathbf{\Psi}(x)\| = O_p(c_h(c_h + h)), \tag{C.8}$$

$$\sup_{x\in\mathcal{X}_1} \left\|vec(\widehat{\mathbf{\Lambda}}_{\ddot{Z}\ddot{X}}(x) - \mathbf{\Lambda}_{\ddot{Z}\ddot{X}}(x))\right\| = O_p(c_h + h^2), \tag{C.9}$$

$$\sup_{x\in\mathcal{X}_1} \|(\widehat{\mathbf{\Omega}}'(x) - \mathbf{\Omega}'(x))\mathbf{\Psi}(x)\| = O_p(c_h(c_h + h^2)), \tag{C.10}$$

$$\sup_{x\in\mathcal{X}_1} \|\widehat{\mathbf{B}}(x)\| = O_p(hc_h + h^2), \tag{C.11}$$

$$\sup_{x\in\mathcal{X}_1} \|(\mathbf{I}_{d_{\ddot{X}}}\ \ \ \mathbf{0}_{d_{\ddot{X}},d_{\ddot{H}}})\widehat{\mathbf{B}}_a(x)\| = O_p(hc_h + h^2), \tag{C.12}$$

$$\sup_{x\in\mathcal{X}_1} \|(\mathbf{0}_{d_{\ddot{H}},d_{\ddot{X}}}\ \ \ \mathbf{I}_{d_{\ddot{H}}})\widehat{\mathbf{B}}_a(x)\| = O_p(hc_h + h), \tag{C.13}$$

where $c_h = (\log(1/h)/(Nh))^{1/2}$, $\widehat{\mathbf{\Omega}}_a(x) = \left(\widehat{\mathbf{\Lambda}}'_{\ddot{Z}_a\ddot{X}_a}(x)\widehat{\mathbf{\Lambda}}_{\ddot{Z}_a\ddot{X}_a}(x)\right)^{-1}\widehat{\mathbf{\Lambda}}'_{\ddot{Z}_a\ddot{X}_a}(x)$, $\mathbf{\Psi}_a(x) = N^{-1}\sum_i \kappa_h(X_{i1}-x)(\ddot{Z}'_{i2}\ \ddot{H}'_{i2,a}/h)'\widetilde{\varepsilon}_{i2}$, $\widehat{\mathbf{\Psi}}(x) = N^{-1}\sum_i \kappa_h(X_{i1}-x)\ddot{Z}_{i2}\widetilde{\varepsilon}_{i2}\widetilde{X}'_{i1}(x)\widehat{\mathbf{M}}^{-1}(x)\iota_1$, $\mathbf{\Psi}(x) \equiv \frac{1}{N}\sum_i \kappa_h(X_{i1}-x)\widetilde{\varepsilon}_{i2}\ddot{Z}_{i2}\widetilde{X}'_{i1}(x)\mathbf{D}^{-1}\mathbf{M}^{-1}\iota_1 f_{X_1}^{-1}(x)$, *and* $\widehat{\mathbf{B}}_a(x) = \widehat{\mathbf{\Lambda}}_{\ddot{Z}_aY}(x) - \widehat{\mathbf{\Lambda}}_{\ddot{Z}_a\ddot{X}_a}(x)\theta_2^a(x) - \mathbf{\Psi}_a(x)$.

Proof of Lemma C.1

Let $\widetilde{\mathbf{B}}_{a,1}(x) = N^{-1}\sum_i \kappa_h(X_{i1}-x)(X_{i1}-x)\ddot{Z}_{i2}X_{i2}\beta_2^{(1)}(x)$ and $\widetilde{\mathbf{B}}_{a,2}(x) = (2N)^{-1}\sum_i \kappa_h(X_{i1}-x)(X_{i1}-x)^2\ddot{Z}_{i2}X_{i2}\beta_2^{(2)}(x)$. We have

$$\widehat{\mathbf{B}}_{a,1}(x) = \widetilde{\mathbf{B}}_{a,1}(x) + \widetilde{\mathbf{B}}_{a,2}(x) + \mathbf{R}_{\mathbf{B}_{a,1}}(x). \tag{C.14}$$

Under Assumption C.1.(c), by standard point-wise convergence arguments of kernel estimators,

$$\begin{aligned}\widetilde{\mathbf{B}}_{a,1}(x) &= \mathbb{E}\left[\kappa_h(X_{i1}-x)\left[\mathbf{\Lambda}_{\ddot{Z}\ddot{X}}(X_{i1})\right]_{\cdot,1}(X_{i1}-x)\right]\beta_2^{(1)}(x) + O_p(h(Nh)^{-1/2}) \\ &= \mu_2 h^2 f_{X_1}(x)\left([\mathbf{\Lambda}^{(1)}_{\ddot{Z}\ddot{X}}(x)]_{\cdot,1} + c_f(x)[\mathbf{\Lambda}_{\ddot{Z}\ddot{X}}(x)]_{\cdot,1}\right)\beta_2^{(1)}(x) + O_p(ha_h), \quad \text{(C.15)} \\ \widetilde{\mathbf{B}}_{a,2}(x) &= 2^{-1}\mathbb{E}\left[\kappa_h(X_{i1}-x)[\mathbf{\Lambda}_{\ddot{Z}\ddot{X}}(X_{i1})]_{\cdot,1}(X_{i1}-x)^2\right]\beta_2^{(2)}(x) + O_p(h^2(Nh)^{-1/2}) \\ &= 2^{-1}h^2\mu_2 f_{X_1}(x)[\mathbf{\Lambda}_{\ddot{Z}\ddot{X}}(x)]_{\cdot,1}\beta_2^{(2)}(x) + O_p(h^2a_h). \quad \text{(C.16)}\end{aligned}$$

and $\mathbf{R}_{\mathbf{B}_{a,1}}(x) = O_p(h^3)$.

Similarly,

$$\begin{aligned}
\widehat{\mathbf{B}}_{a,2}(x) =&(2N)^{-1}\sum_i \kappa_h(X_{i1}-x)\ddot{Z}_{i2}\ddot{H}_{i2}'\theta_{2,-1}^{(2)}(x)(X_{i1}-x)^2\\
&+(6N)^{-1}\sum_i \kappa_h(X_{i1}-x)\ddot{Z}_{i2}\ddot{H}_{i2}'\theta_{2,-1}^{(3)}(\xi_i)(X_{i1}-x)^3\\
=&2^{-1}\mu_2 h^2 f_{X_1}(x)\left[\mathbf{\Lambda}_{\ddot{Z}\ddot{X}}(x)\right]_{\cdot,-1}\theta_{2,-1}^{(2)}(x)+O_p(h^2a_h)+O_p(h^3).
\end{aligned} \tag{C.17}$$

The desired result is obtained from (C.15), (C.16) and (C.17) and that $[\mathbf{\Lambda}_{\ddot{Z}\ddot{X}}(x)]_{\cdot,1}\beta_2^{(2)}(x)+[\mathbf{\Lambda}_{\ddot{Z}\ddot{X}}(x)]_{\cdot,-1}\theta_{2,-1}^{(2)}(x) = \mathbf{\Lambda}_{\ddot{Z}\ddot{X}}(x)\theta_2^{(2)}(x)$.

Proof of Lemma C.2:

Given that $\widehat{\mathbf{\Lambda}}_{\ddot{Z}\ddot{X}}(x)\theta_2(x) = (\theta_2(x)\otimes\mathbf{I}_{d_{\ddot{Z}}})'\text{vec}(\widehat{\mathbf{\Lambda}}_{\ddot{Z}\ddot{X}}(x))$, we find that

$$\begin{aligned}
\widehat{\mathbf{B}}(x) &= N^{-1}\sum_i\Big((\theta_2(X_{i1})-\theta_2(x))\otimes\mathbf{I}_{d_{\ddot{Z}}}\Big)'\widetilde{X}_{i2}\iota_1'\widehat{\mathbf{M}}^{-1}(x)\widetilde{X}_{i1}(x)\kappa_h(X_{i1}-x)\\
&= \widetilde{\mathbf{B}}_1(x)+\widetilde{\mathbf{B}}_2(x)+\mathbf{R}_{\mathbf{B}}(x),
\end{aligned} \tag{C.18}$$

where

$$\begin{aligned}
\widetilde{\mathbf{B}}_1(x) &= (\theta_2^{(1)}(x)\otimes\mathbf{I}_{d_{\ddot{Z}}})'N^{-1}\sum_i h\left(\frac{X_{i1}-x}{h}\right)\widetilde{X}_{i2}\iota_1'\widehat{\mathbf{M}}^{-1}(x)\widetilde{X}_{i1}(x)\kappa_h(X_{i1}-x)\\
&\equiv (\theta_2^{(1)}(x)\otimes\mathbf{I}_{d_{\ddot{Z}}})'N^{-1}\sum_i\widehat{b}_{i1}(x),\\
\widetilde{\mathbf{B}}_2(x) &= (\theta_2^{(2)}(x)/2\otimes\mathbf{I}_{d_{\ddot{Z}}})'N^{-1}\sum_i h^2\left(\frac{X_{i1}-x}{h}\right)^2\widetilde{X}_{i2}\iota_1'\widehat{\mathbf{M}}^{-1}(x)\widetilde{X}_{i1}(x)\kappa_h(X_{i1}-x)\\
&\equiv (\theta_2^{(2)}(x)/2\otimes\mathbf{I}_{d_{\ddot{Z}}})'N^{-1}\sum_i\widehat{b}_{i2}(x),\\
\mathbf{R}_{\mathbf{B}}(x) &= N^{-1}\sum_i(\theta_2^{(3)}(\xi_i)/6\otimes\mathbf{I}_{d_{\ddot{Z}}})'h^3\left(\frac{X_{i1}-x}{h}\right)^3\widetilde{X}_{i2}\iota_1'\widehat{\mathbf{M}}^{-1}(x)\widetilde{X}_{i1}(x)\kappa_h(X_{i1}-x).
\end{aligned}$$

where $\mathbf{R}_{\mathbf{B}}(x)$ is the Taylor expansion remainder term with ξ_i lying between x and X_{i1} for all i.

For the term $\widetilde{\mathbf{B}}_1(x)$, we know that

$$\begin{aligned}
&\mathbb{E}[N^{-1}\sum_i\widehat{b}_{i1}(x)|X_{i1}] = N^{-1}\sum_i\text{vec}(\mathbf{\Lambda}_{\ddot{Z}\ddot{X}}(X_{i1}))\,(X_{i1}-x)\,\widetilde{X}_{i1}'(x)\widehat{\mathbf{M}}^{-1}(x)\iota_1\kappa_h(X_{i1}-x)\\
&= N^{-1}\sum_i\left(\mathbf{L}_{\ddot{Z}\ddot{X}}(x)\widetilde{X}_{i1}(x)\,(X_{i1}-x)+h^{\ell+2}\frac{1}{(\ell+1)!}\text{vec}(\mathbf{\Lambda}^{(\ell+1)}(x))\left(\frac{X_{i1}-x}{h}\right)^{\ell+2}\right)\\
&\qquad\times\widetilde{X}_{i1}'(x)\widehat{\mathbf{M}}^{-1}(x)\iota_1\kappa_h(X_{i1}-x)+O_p(h^{\ell+3}).
\end{aligned}$$

Let $\mathbf{H}(x)$ be a diagonal matrix whose ith entry is given by $X_{i1} - x$. First, note that

$$\begin{aligned}
& N^{-1}\sum_i \mathbf{L}_{\ddot{Z}\ddot{X}}(x)\widetilde{X}_{i1}(x)\left(X_{i1}-x\right)\widetilde{X}_{i1}'(x)\widehat{\mathbf{M}}^{-1}(x)\iota_1\kappa_h(X_{i1}-x) \qquad \text{(C.19)}\\
&= N^{-1}\mathbf{L}_{\ddot{Z}\ddot{X}}(x)\widetilde{\mathbf{X}}_1'(x)\mathbf{H}(x)\mathbf{K}(x)\widetilde{\mathbf{X}}_1(x)\widehat{\mathbf{M}}^{-1}(x)\iota_1\\
&= N^{-1}\sum_{s=1}^{\ell+1}[\mathbf{L}_{\ddot{Z}\ddot{X}}(x)]_s\iota_s'\widetilde{\mathbf{X}}_1'(x)\mathbf{H}(x)\mathbf{K}(x)\widetilde{\mathbf{X}}_1(x)\widehat{\mathbf{M}}^{-1}(x)\iota_1\\
&= N^{-1}\sum_{s=1}^{\ell}[\mathbf{L}_{\ddot{Z}\ddot{X}}(x)]_s\iota_{s+1}'\widetilde{\mathbf{X}}_1'(x)\mathbf{K}(x)\widetilde{\mathbf{X}}_1(x)\widehat{\mathbf{M}}^{-1}(x)\iota_1\\
&\quad + N^{-1}[\mathbf{L}_{\ddot{Z}\ddot{X}}(x)]_{\ell+1}\iota_{\ell+1}'\widetilde{\mathbf{X}}_1'(x)\mathbf{H}(x)\mathbf{K}(x)\widetilde{\mathbf{X}}_1(x)\widehat{\mathbf{M}}^{-1}(x)\iota_1\\
&= N^{-1}[\mathbf{L}_{\ddot{Z}\ddot{X}}(x)]_{\ell+1}\iota_{\ell+1}'\mathbf{D}\mathbf{D}^{-1}\widetilde{\mathbf{X}}_1'(x)\mathbf{H}(x)\mathbf{K}(x)\widetilde{\mathbf{X}}_1(x)\mathbf{D}^{-1}\left(\mathbf{D}^{-1}\widehat{\mathbf{M}}(x)\mathbf{D}^{-1}\right)^{-1}\iota_1\\
&= h^{\ell+1}[\mathbf{L}_{\ddot{Z}\ddot{X}}(x)]_{\ell+1}\iota_{\ell+1}'\left(\mathbf{M}_1 f_{X_1}(x) + h\mathbf{M}_2 f_{X_1}^{(1)}(x) + o_p(h)\right)\\
&\qquad\qquad \times\left(f_{X_1}^{-1}(x)\mathbf{M}^{-1} + h\mathsf{c}_f(x)f_{X_1}^{-1}(x)\mathbf{M}^{-1}\mathbf{M}_1\mathbf{M}^{-1} + o_p(h)\right)\iota_1.
\end{aligned}$$

The third equality holds because $\mathbf{H}(x)\widetilde{\mathbf{X}}_1(x)\iota_s = \widetilde{\mathbf{X}}(x)\iota_{s+1}$ for $s = 1, \dots, \ell$. The fourth equality holds because $N^{-1}\iota_{s+1}'\widetilde{\mathbf{X}}_1'(x)\mathbf{K}(x)\widetilde{\mathbf{X}}_1(x)\widehat{\mathbf{M}}^{-1}(x)\iota_1 = \iota_{s+1}'\widehat{\mathbf{M}}(x)^{-1}\widehat{\mathbf{M}}(x)\iota_1 = \iota_{s+1}'\iota_1 = 0$ for all $s = 1, ..., \ell$. The last equality holds from standard kernel derivations.

Furthermore, given symmetry of the kernel function $\kappa(.)$, the (i,k)-th element of $\mathbf{M}$ is zero if $i + k$ is odd. This implies that the adjoint matrix of the $(1,k)$-th element of $\mathbf{M}$ is singular if k is even and, therefore, $\mathbf{M}^{-1}\iota_1$ is a $(\ell + 1) \times 1$ vector with all even elements equal to zero. On the other hand, $\iota_{\ell+1}'\mathbf{M}_j = (\int u^{\ell+j}\kappa(u)du \ ... \ \int u^{2\ell+j}\kappa(u)du)$ have zero odd elements when $\ell + j$ is odd and zero even elements if $\ell + j$ is even. Therefore, $\iota_{\ell+1}'\mathbf{M}_1\mathbf{M}^{-1}\iota_1 = 0$ when ℓ is even. Using similar arguments, we know that $\iota_{\ell+1}'\mathbf{M}_1\mathbf{M}^{-1}\mathbf{M}_1\mathbf{M}^{-1}\iota_1 = 0$ when ℓ is even and $\iota_{\ell+1}'\mathbf{M}_2\mathbf{M}^{-1}\iota_1 = 0$ when ℓ is odd.

Therefore, we have

$$(\text{C.19}) = h^{\ell+1}[\mathbf{L}_{\ddot{Z}\ddot{X}}(x)]_{\ell+1}\iota_{\ell+1}'\mathbf{M}_1\mathbf{M}^{-1}\iota_1 + O_p(h^{\ell+2}), \qquad \text{(C.20)}$$

if ℓ is odd and

$$(\text{C.19}) = h^{\ell+2}[\mathbf{L}_{\ddot{Z}\ddot{X}}(x)]_{\ell+1}\mathsf{c}_f(x)\iota_{\ell+1}'\mathbf{M}_2\mathbf{M}^{-1}\iota_1 + O_p(h^{\ell+3}), \qquad \text{(C.21)}$$

if ℓ is even. Similarly, the other component of $\mathbb{E}[N^{-1}\sum_i \widehat{b}_{i1}(x)|X_{i1}]$ follows

$$
\begin{aligned}
&N^{-1}\sum_i h^{\ell+2}\frac{1}{(\ell+1)!}\mathrm{vec}(\mathbf{\Lambda}^{(\ell+1)}(x))\left(\frac{X_{i1}-x}{h}\right)^{\ell+2}\widetilde{X}_{i1}'\widehat{\mathbf{M}}^{-1}(x)\iota_1\kappa_h(X_{i1}-x)\\
&=h^{\ell+2}\frac{1}{(\ell+1)!}\mathrm{vec}(\mathbf{\Lambda}^{(\ell+1)}(x))\iota_{\ell+1}'\mathbf{M}_2\mathbf{M}^{-1}\iota_1+O_p(h^{\ell+3}),
\end{aligned}
$$

if ℓ is even and $O_p(h^{\ell+3})$ if ℓ is odd.

For the term $\widetilde{\mathbf{B}}_2(x)$, we know that

$$
\begin{aligned}
&\mathbb{E}[N^{-1}\sum_i \widehat{b}_{i2}(x)|X_{i1}]\\
=&N^{-1}\sum_{s=1}^{\ell+1}[\mathbf{L}_{\ddot{Z}\ddot{X}}(x)]_s\iota_s'\widetilde{\mathbf{X}}_1'(x)\mathbf{H}^2(x)\mathbf{K}(x)\widetilde{\mathbf{X}}_1(x)\mathbf{D}^{-1}\left(\mathbf{D}^{-1}\widehat{\mathbf{M}}(x)\mathbf{D}^{-1}\right)^{-1}\iota_1+O_p(h^{\ell+3})\\
=&N^{-1}[\mathbf{L}_{\ddot{Z}\ddot{X}}(x)]_\ell\iota_{\ell+1}'\mathbf{D}\mathbf{D}^{-1}\widetilde{\mathbf{X}}_1'(x)\mathbf{H}(x)\mathbf{K}(x)\widetilde{\mathbf{X}}_1(x)\mathbf{D}^{-1}\left(\mathbf{D}^{-1}\widehat{\mathbf{M}}(x)\mathbf{D}^{-1}\right)^{-1}\iota_1\\
&+N^{-1}[\mathbf{L}_{\ddot{Z}\ddot{X}}(x)]_{\ell+1}\iota_{\ell+1}'\mathbf{D}\mathbf{D}^{-1}\widetilde{\mathbf{X}}_1'(x)\mathbf{H}_2(x)\mathbf{K}(x)\widetilde{\mathbf{X}}_1(x)\mathbf{D}^{-1}\left(\mathbf{D}^{-1}\widehat{\mathbf{M}}(x)\mathbf{D}^{-1}\right)^{-1}\iota_1\\
&+O_p(h^{\ell+3}).
\end{aligned}
$$

The first term in the RHS is not relevant for $\ell=0$. When relevant, it reduces to $h^{\ell+1}[\mathbf{L}_{\ddot{Z}\ddot{X}}(x)]_\ell\iota_{\ell+1}'\mathbf{M}_1\mathbf{M}^{-1}\iota_1+O_p(h^{\ell+2})$ when ℓ is odd and $h^{\ell+2}[\mathbf{L}_{\ddot{Z}\ddot{X}}(x)]_\ell\mathsf{c}_f(x)\iota_{\ell+1}'\mathbf{M}_2\mathbf{M}^{-1}\iota_1+O_p(h^{\ell+3})$ when ℓ is even. The second term reduces to $h^{\ell+2}[\mathbf{L}_{\ddot{Z}\ddot{X}}(x)]_{\ell+1}\iota_{\ell+1}'\mathbf{M}_2\mathbf{M}^{-1}\iota_1+O_p(h^{\ell+3})$ when ℓ is even and $O_p(h^{\ell+3})$ when ℓ is odd.

Using similar derivations, one can show that $\mathbb{V}[N^{-1}\sum_i(\widehat{b}_{i1}(x)+\widehat{b}_{i2}(x))|X_{i1}]=O_p(h/N)=O_p(h^4)$ if $Nh^3\to\infty$ and that $\mathbf{R}_{\mathbf{B}}(x)=O_p(h^3)$ under the uniform boundedness conditions.

Summing up, we conclude that

$$
\begin{aligned}
\widehat{\mathbf{B}}(x)=&(\theta_2^{(1)}(x)\otimes\mathbf{I}_{d_{\ddot{Z}}})'\cdot h^2\mathrm{vec}(\mathbf{\Lambda}_{\ddot{Z}\ddot{X}}^{(1)}(x))\iota_2'\mathbf{M}_1\mathbf{M}^{-1}\iota_1\\
&+(\theta_2^{(2)}(x)/2\otimes\mathbf{I}_{d_{\ddot{Z}}})'\cdot h^2\mathrm{vec}(\mathbf{\Lambda}_{\ddot{Z}\ddot{X}}(x))\iota_2'\mathbf{M}_1\mathbf{M}^{-1}\iota_1+o_p(h^2)\\
&=h^2\mu_2\left(\mathbf{\Lambda}_{\ddot{Z}\ddot{X}}^{(1)}(x)\theta_2^{(1)}(x)+\mathbf{\Lambda}_{\ddot{Z}\ddot{X}}(x)\theta_2^{(2)}(x)/2\right)+O_p(h^3),
\end{aligned}
$$

when $\ell=1$ and

$$
\widehat{\mathbf{B}}(x)=h^2\mu_2\left[\left(\mathsf{c}_f(x)\mathbf{\Lambda}_{\ddot{Z}\ddot{X}}(x)+\mathbf{\Lambda}_{\ddot{Z}\ddot{X}}^{(1)}(x)\right)\theta_2^{(1)}(x)+\mathbf{\Lambda}_{\ddot{Z}\ddot{X}}(x)\theta_2^{(2)}(x)/2\right]+O_p(h^3)
$$

when $\ell=0$, which corresponds to the statement in the lemma.

Proof of Lemma C.3

To show the first result (C.6) of the lemma, we notice that for all $x \in \mathcal{X}_1^*$,

$$\mathbf{D}_a\widehat{\boldsymbol{\Omega}}'_a(x) = \left(\mathbf{D}_a^{-1}\widehat{\boldsymbol{\Lambda}}'_{\ddot{Z}_a\ddot{X}_a}(x)\widehat{\boldsymbol{\Lambda}}_{\ddot{Z}_a\ddot{X}_a}(x)\mathbf{D}_a^{-1}\right)^{-1}\mathbf{D}_a^{-1}\widehat{\boldsymbol{\Lambda}}'_{\ddot{Z}_a\ddot{X}_a}(x),$$

while $\widehat{\boldsymbol{\Lambda}}_{\ddot{Z}_a\ddot{X}_a}(.)$ has a uniform convergence result that

$$\widehat{\boldsymbol{\Lambda}}_{\ddot{Z}_a\ddot{X}_a}(x)\mathbf{D}_a^{-1} = \text{diag}(\boldsymbol{\Lambda}_{\ddot{Z}\ddot{X}}(x), \mu_2\boldsymbol{\Lambda}_{\ddot{H}\ddot{H}}(x)) + O_p(c_h + h^2),$$

following from (A.1) of Fan and Huang (2005). The uniform convergence result leads to (C.6) since $\boldsymbol{\Lambda}_{\ddot{Z}\ddot{X}}(.)$ and $\mu_2\boldsymbol{\Lambda}_{\ddot{H}\ddot{H}}(.)$ have eigenvalues uniformly bounded and bounded away from zero and bounded from above. Then, if we let $\widehat{\mathbf{A}}_x$ (resp. $\mathbf{A}_x$) denote the denominator of $\mathbf{D}_a\widehat{\boldsymbol{\Omega}}_a(x)$ (resp. that of $\boldsymbol{\Omega}_a(x)$), we have $\mathbf{A}_x^{-1}$ is bounded uniformly in $x \in \mathcal{X}_1$ by the invertibility of Schur complement. In addition,

$$\sup_{x\in\mathcal{X}_1}\|\widehat{\mathbf{A}}_x^{-1} - \mathbf{A}_x^{-1}\| = \sup_{x\in\mathcal{X}_1}\|\widehat{\mathbf{A}}_x^{-1}(\mathbf{A}_x - \widehat{\mathbf{A}}_x)\mathbf{A}_x^{-1}\| \le O_p(1)\sup_{x\in\mathcal{X}_1}\|\mathbf{A}_x - \widehat{\mathbf{A}}_x\| = O_p(c_h + h^2).$$

Equation (C.6) then follows.

The result in (C.6) follows from the uniform convergences that $\sup_x \left\|\mathbf{D}^{-1}\widehat{\mathbf{M}}(x)\mathbf{D}^{-1} - \mathbb{E}[\mathbf{D}^{-1}\widehat{\mathbf{M}}(x)\mathbf{D}^{-1}]\right\| = O_p(c_h)$ and $\sup_x \left\|\mathbb{E}[\mathbf{D}^{-1}\widehat{\mathbf{M}}(x)\mathbf{D}^{-1}] - \mathbf{M}f_{X_1}(x) - h\mathbf{M}_1 f_{X_1}^{(1)}(x)\right\| = O_p(h^2)$. The former is a trivial extension of the uniform convergence result stated in Mack and Silverman (1982) for local constant estimation to local polynomial estimation. The latter follows from standard kernel bias derivation and uniform boundedness conditions stated in the assumptions.

The first part of (C.7) is a trivial consequence of Lemma A.1 in Fan and Huang (2005). The second part is similarly obtained by using $\sup_{x\in\mathcal{X}_1}|f_{X_1}^{-1}(x)| < \infty$ (Assumption C.1.(b)) and the fact that

$$\sup_{x\in\mathcal{X}_1}\|N^{-1}\sum_i \widetilde{\varepsilon}_{i2}\ddot{Z}_{i2}\widetilde{X}'_{i1}(x)\mathbf{D}^{-1}\kappa_h(X_{i1} - x)\| = O_p(c_h). \tag{C.22}$$

The result in (C.8) is obtained as a consequence of (C.6) and (C.22); specifically, we have

$$\begin{aligned}
&\|\widehat{\boldsymbol{\Psi}}(x) - \boldsymbol{\Psi}(\mathbf{x})\| \\
&\le \|N^{-1}\sum_i \widetilde{\varepsilon}_{i2}\ddot{Z}_{i2}\widetilde{X}'_{i1}(x)\mathbf{D}^{-1}\kappa_h(X_{i1} - x)\|\|(\mathbf{D}^{-1}\widehat{\mathbf{M}}(x)\mathbf{D}^{-1})^{-1} - \mathbf{M}^{-1}f_{X_1}^{-1}(x)\| \\
&\le O_p(c_h)\|(\mathbf{D}^{-1}\widehat{\mathbf{M}}(x)\mathbf{D}^{-1})^{-1}\|\|\mathbf{D}^{-1}\widehat{\mathbf{M}}(x)\mathbf{D}^{-1} - \mathbf{M}f_{X_1}(x)\|\|\mathbf{M}^{-1}f_{X_1}^{-1}(x)\| \\
&= O_p(c_h)O_p(c_h + h),
\end{aligned} \tag{C.23}$$

uniformly in $x \in \mathcal{X}_1$. The second inequality follows from $A^{-1} - B^{-1} = B^{-1}(A - B)A^{-1}$ for matrices A and B.

To prove (C.9), we define $\upsilon_i = \text{vec}(\ddot{Z}_{i2}\ddot{X}_{i2}' - \Lambda_{\ddot{Z}\ddot{X}}(X_{i1}))$ and then, given that $\mathbf{\Lambda}_{\ddot{Z}\ddot{X}}(.)$ is twice continuously differentiable, we have

$$\begin{aligned}
&\text{vec}(\widehat{\mathbf{\Lambda}}_{\ddot{Z}\ddot{X}}(x) - \mathbf{\Lambda}_{\ddot{Z}\ddot{X}}(x)) \\
=& N^{-1}\widetilde{\mathbf{X}}_2'\mathbf{K}(x)\widetilde{\mathbf{X}}_1(x)\widehat{\mathbf{M}}(x)^{-1}\iota_1 - \mathbf{L}_{\ddot{Z}\ddot{X}}(x)N^{-1}\widetilde{\mathbf{X}}_1'(x)\mathbf{K}(x)\widetilde{\mathbf{X}}_1(x)\widehat{\mathbf{M}}^{-1}(x)\iota_1 \\
=& N^{-1}(\widetilde{\mathbf{X}}_2 - \widetilde{\mathbf{X}}_1(x)\mathbf{L}_{\ddot{Z}\ddot{X}}'(x))'\mathbf{K}(x)\widetilde{\mathbf{X}}_1(x)\mathbf{D}^{-1}\left(\mathbf{D}^{-1}\widehat{\mathbf{M}}(x)\mathbf{D}^{-1}\right)^{-1}\mathbf{D}^{-1}\iota_1.
\end{aligned}$$

By (C.6) and Assumption C.1.(b) on the density function, we know that $\left(\mathbf{D}^{-1}\widehat{\mathbf{M}}(x)\mathbf{D}^{-1}\right)^{-1} = O_p(1)$ uniformly in $x \in \mathcal{X}_1$. In addition, we have

$$N^{-1}(\widetilde{\mathbf{X}}_2 - \widetilde{\mathbf{X}}_1(x)\mathbf{L}_{\ddot{Z}\ddot{X}}'(x))'\mathbf{K}(x)\widetilde{\mathbf{X}}_1(x)\mathbf{D}^{-1} = \mathtt{r}_{11}(x) + \mathtt{r}_{12}(x), \tag{C.24}$$

$$\begin{aligned}
\text{where } \mathtt{r}_{11}(x) =& N^{-1}\sum_i \upsilon_i\widetilde{X}_{i1}'(x)\mathbf{D}^{-1}\kappa_h(X_{i1} - x), \\
\mathtt{r}_{12}(x) =& 1_{\{\ell=0\}}N^{-1}h\sum_i \text{vec}(\mathbf{\Lambda}_{\ddot{Z}\ddot{X}}^{(1)}(x))\left(\frac{X_{i1} - x}{h}\right)\kappa_h(X_{i1} - x) \\
&+ N^{-1}h^2\sum_i \text{vec}(\mathbf{\Lambda}_{\ddot{Z}\ddot{X}}^{(2)}(\xi_{i1}))\left(\frac{X_{i1} - x}{h}\right)^2\widetilde{X}_{i1}'(x)\mathbf{D}^{-1}\kappa_h(X_{i1} - x),
\end{aligned}$$

by the Taylor expansion for some ξ_{i1} between X_{i1} and x. Because $\mathbb{E}[\upsilon_i|X_{i1}] = 0$ and $\sup_{x\in\mathcal{X}_1}\mathbb{E}[|\nu_i\|^s|X_{i1} = x] < \infty$, we follow the uniform convergence results in Mack and Silverman (1982) and Fan and Huang (2005, Lemma A.1) and obtain that $\mathtt{r}_{11}(.) = O_p(c_h)$ uniformly on $\mathcal{X}_1$. Similarly,

$$\begin{aligned}
\mathtt{r}_{12}(x) =& h1_{\{\ell=0\}}\text{vec}(\mathbf{\Lambda}_{\ddot{Z}\ddot{X}}^{(1)}(x))\mathbb{E}\left[\left(\frac{X_{i1} - x}{h}\right)\kappa_h(X_{i1} - x)\right] \\
&+ h^2\mathbb{E}\left[\left(\frac{X_{i1} - x}{h}\right)^2\text{vec}(\mathbf{\Lambda}_{\ddot{Z}\ddot{X}}^{(2)}(\xi_{i1}))\widetilde{X}_{i1}'(x)\mathbf{D}^{-1}\kappa_h(X_{i1} - x)\right] + O_p(c_h) \\
=& h^2(\text{vec}(\mathbf{\Lambda}_{\ddot{Z}\ddot{X}}^{(2)}(x)) + o_p(1))(f_{X_1}(x)\iota_1'\mathbf{M}_2 + O_p(h)) + O_p(c_h + h^2),
\end{aligned}$$

uniformly in $x \in \mathcal{X}_1$. The last equality follows from uniform boundedness conditions in Assumptions C.1.(b) and C.1.(e). Then, the desired result is followed.

$$N^{-1}(\widetilde{\mathbf{X}}_2 - \widetilde{\mathbf{X}}_1(x)\mathbf{L}_{\ddot{Z}\ddot{X}}'(x))'\mathbf{K}(x)\widetilde{\mathbf{X}}_1(x)\mathbf{D}^{-1}\left(\mathbf{D}^{-1}\widehat{\mathbf{M}}(x)\mathbf{D}^{-1}\right)^{-1}\mathbf{D}^{-1}\iota_1 = O_p(c_h + h^2).$$

We then focus on (C.10). Because of (C.9) and Assumption 3.1, we have $\|\widehat{\mathbf{\Omega}}(x) - \mathbf{\Omega}(x)\| = O_p(c_h + h^2)$ uniformly in $x \in \mathcal{X}_1$ since the inverse of $\mathbf{\Lambda}'_{\ddot{Z}\ddot{X}}(x)\mathbf{\Lambda}_{\ddot{Z}\ddot{X}}(x)$ is uniformly bounded in $x \in \mathcal{X}_1$. Then, the desired result is obtained by combining this with (C.7).

To study the last part, we recall that $\widehat{\mathbf{B}}(x)$ can be decomposed into $\widetilde{\mathbf{B}}_1(x)$, $\widetilde{\mathbf{B}}_2(x)$ and $\mathbf{R}_{\mathbf{B}}(x)$, each of which is defined in the proof of Lemma C.2. We then let

$$N^{-1}\sum_i b_{i1}(x) = N^{-1}\sum_i h\left(\frac{X_{i1}-x}{h}\right)\widetilde{X}_{i2}\iota_1'\mathbf{M}^{-1}\mathbf{D}^{-1}\widetilde{X}_{i1}(x)\kappa_h(X_{i1}-x)f_{X_1}^{-1}(x),$$

$$N^{-1}\sum_i b_{i2}(x) = N^{-1}\sum_i h^2\left(\frac{X_{i1}-x}{h}\right)^2\widetilde{X}_{i2}\iota_1'\mathbf{M}^{-1}\mathbf{D}^{-1}\widetilde{X}_{i1}(x)\kappa_h(X_{i1}-x)f_{X_1}^{-1}(x).$$

Because of the uniform boundedness conditions in Assumptions C.1.(b), C.1.(c) and C.1.(f) and the uniform convergence result in Lemma A.1 in Fan and Huang (2005), we find that

$$\begin{aligned}
&N^{-1}\sum_i \left(\frac{X_{i1}-x}{h}\right)\widetilde{X}_{i2}\widetilde{X}_{i1}'\mathbf{D}^{-1}\kappa_h(X_{i1}-x)\\
&= \mathrm{vec}\left(\mathbf{\Lambda}_{\ddot{Z}\ddot{X}}(x)\right)\iota_1'\mathbf{M}_1 f_{X_1}(x)\\
&\quad + h f_{X_1}(x)\left(\mathsf{c}_f(x)\mathrm{vec}\left(\mathbf{\Lambda}_{\ddot{Z}\ddot{X}}(x)\right)\iota_1'\mathbf{M}_1 + \mathrm{vec}\left(\mathbf{\Lambda}^{(1)}_{\ddot{Z}\ddot{X}}(x)\right)\iota_1'\mathbf{M}_2\right) + O_p(c_h+h^2),
\end{aligned}$$

uniformly in $x \in \mathcal{X}_1$. By using similar arguments in proving (C.20) and (C.21), we have $\iota_1'\mathbf{M}_1\mathbf{M}^{-1}\iota_1 = 0$ and thus $N^{-1}\sum_i b_{i1}(x)$ satisfies the following:

$$N^{-1}\sum_i b_{i1}(x) = h^2\mathrm{vec}\left(\mathbf{\Lambda}^{(1)}_{\ddot{Z}\ddot{X}}(x)\right)\iota_1'\mathbf{M}_2\mathbf{M}^{-1}\iota_1 + O_p(h(c_h+h^2)) = O_p(hc_h+h^2). \quad \text{(C.25)}$$

Similarly, the following holds uniformly in $x \in \mathcal{X}_1$:

$$N^{-1}\sum_i \left(\frac{X_{i1}-x}{h}\right)^2\widetilde{X}_{i2}\widetilde{X}_{i1}'\mathbf{D}^{-1}\kappa_h(X_{i1}-x) = \mathrm{vec}\left(\mathbf{\Lambda}_{\ddot{Z}\ddot{X}}(x)\right)\iota_1'\mathbf{M}_2 f_{X_1}(x) + O_p(c_h+h),$$

from which it is deduced that

$$N^{-1}\sum_i b_{i2}(x) = h^2\mathrm{vec}\left(\mathbf{\Lambda}_{\ddot{Z}\ddot{X}}(x)\right)\iota_1'\mathbf{M}_2\mathbf{M}^{-1}\iota_1 + O_p(h^2(c_h+h)) = O_p(h^2c_h+h^2). \quad \text{(C.26)}$$

Then, the followings are deduced from Assumption C.1.(c), (C.6), (C.25), and (C.26):

$$\widetilde{\mathbf{B}}_1(x) = (\theta_1^{(1)}(x)\otimes\mathbf{I}_{d_{\ddot{Z}}})'N^{-1}\sum_i b_{i1}(x) + o_p(hc_h+h^2) = O_p(hc_h+h^2), \quad \text{(C.27)}$$

$$\widetilde{\mathbf{B}}_2(x) = (\theta_1^{(2)}(x)/2\otimes\mathbf{I}_{d_{\ddot{Z}}})'N^{-1}\sum_i b_{i2}(x) + o_p(h^2c_h+h^2) = O_p(h^2c_h+h^2), \quad \text{(C.28)}$$

uniformly in $\mathcal{X}_1$. Using similar derivations, one can show that $\mathbf{R}_{\mathbf{B}}(.) = o_p(hc_h + h^2)$. By combining these results, the last part of the lemma is obtained.

Since $\widehat{\mathbf{B}}_a(x) = \widehat{\mathbf{B}}_{a,1}(x) + \widehat{\mathbf{B}}_{a,2}(x)$, results in (C.12) and (C.13) can be proven by using Lemma A.1 in Fan and Huang (2005). Specifically, we have

$$
\begin{aligned}
[\widehat{\mathbf{B}}_a(x)]_{1:d_{\ddot{X}}} &= N^{-1}\sum_i \kappa_h(X_{i1}-x)\ddot{Z}_{i2}\ddot{X}_{i2}'(\theta_2(X_{i1}) - \theta_2(x) - \theta_2^{(1)}(x)(X_{i1}-x)) \\
&\quad + hN^{-1}\sum_i \kappa_h(X_{i1}-x)\ddot{Z}_{i2}X_{i2}\beta_2^{(1)}(x)\frac{X_{i1}-x}{h} \\
&= h^2\mathbb{E}\left[\kappa_h(X_{i1}-x)\ddot{Z}_{i2}\ddot{X}_{i2}'\theta_2^{(2)}(\xi_i)\left(\frac{X_{i1}-x}{h}\right)^2\right] + O_p(hc_h),
\end{aligned}
$$

where $\xi_i \in [X_{i1}, x]$ for each i. Then the desired result follows from the uniform boundedness of $\theta_2^{(2)}(.)$ and $\mathbf{\Lambda}_{\ddot{Z}\ddot{X}}(.)$. Similarly, the last $d_{\ddot{H}}$ rows of $\widehat{\mathbf{B}}_a(x)$ satisfy that

$$
\begin{aligned}
&[\widehat{\mathbf{B}}_a(x)]_{(d_{\ddot{X}}+1):(d_{\ddot{X}}+d_{\ddot{H}})} \\
&= h^2\mathbb{E}\left[\kappa_h(X_{i1}-x)\ddot{H}_{i2}\ddot{X}_{i2}'\theta_2^{(2)}(\xi_i)\left(\frac{X_{i1}-x}{h}\right)^3\right] \\
&\quad + h\mathbb{E}\left[\kappa_h(X_{i1}-x)\ddot{H}_{i2}X_{i2}\beta_2^{(1)}(x)\left(\frac{X_{i1}-x}{h}\right)^2\right] + O_p(hc_h) \\
&= O_p(hc_h + h), \qquad \text{(C.29)}
\end{aligned}
$$

where the last last follows from that $\mathbb{E}\left[\kappa_h(X_{i1}-x)\ddot{H}_{i2}X_{i2}(X_{i1}-x)^2/h^2\right] = \mu_2\mathbb{E}[\ddot{H}_{i2}X_{i2}|X_{i1} = x] + O(h)$. This concludes the proof.

Appendix D: Proofs of Theorems and Propositions

Proof of equality (3.2)

The first equality in equation (3.2) holds because

$$
\begin{aligned}
&\mathbb{E}\left[\omega_t(Z_{it})\left(Y_{it} - \left(\beta_t(x)X_{it} + H_{it}'\widetilde{\gamma}_t(x)\right)\right)|X_{i(t-1)} = x\right] \\
=&\mathbb{E}\left[\omega_t(Z_{it})\left(\varepsilon_{it} - H_{it}'(\widetilde{\gamma}_t(x) - \gamma_t(x))\right)|X_{i(t-1)} = x\right] \\
=&\mathbb{E}\left[\omega_t(Z_{it})\varepsilon_{it}|X_{i(t-1)} = x\right] \\
&- \mathbb{E}\left[\omega(Z_{it})\mathbb{E}[\mathbb{L}[\varepsilon_{it}|X_{i(t-1)} = x, \widetilde{H}_{it}]|Z_{it}, X_{i(t-1)} = x]|X_{i(t-1)} = x\right] \\
=&\mathbb{E}\left[\omega_t(Z_{it})\varepsilon_{it}|X_{i(t-1)} = x\right] \\
&- \mathbb{E}\left[\omega_t(Z_{it})\mathbb{E}[\mathbb{L}[\varepsilon_{it}|X_{i(t-1)} = x, \widetilde{H}_{it}]|X_{i(t-1)} = x]|X_{i(t-1)} = x\right],
\end{aligned}
$$

where the first two equalities hold respectively from the outcome equation in (2.1) and by the law of iterated expectations. The third equality holds because $\mathbb{E}[\widetilde{H}_{it}|X_{i(t-1)}, Z_{it}] = \mathbb{E}[\widetilde{H}_{it}|X_{i(t-1)}]$ by Assumption 3.2 while the fourth holds because $\mathbb{E}[\mathbb{L}[\varepsilon_{it}|X_{i(t-1)} = x, \widetilde{H}_i]|X_{i(t-1)} = x] = \mathbb{E}[\varepsilon_{it}|X_{i(t-1)} = x]$ by the definition of linear projection. To see this, one just needs to show by block matrix inversion that $\mathbb{E}[\mathbb{L}[Y|X]] = \mathbb{E}[Y]$, where $\mathbb{L}[Y|X] = (1\ X')(\mathbb{E}[(1\ X')'(1\ X')])^{-1}(\mathbb{E}[(1\ X')'Y])$ for any scalar random variable Y and random vector X.

The second equality in equation (3.2) is explained in the main text.

Proof of Theorem 4.1

We prove the theorem without loss of generality with an identity GMM weighting matrix.

First of all, the augmented local GMM estimator $\widehat{\theta}_2^a$ satisfies that

$$\begin{aligned}\widehat{\theta}_2^a(x) - \theta_2^a(x) &= (\mathbf{I}_{d_{\ddot{X}}}\ \mathbf{0}_{d_{\ddot{X}},d_{\ddot{H}}})\left(\widehat{\theta}_2^a(x) - \theta_2^a(x)\right) \\ &= (\mathbf{I}_{d_{\ddot{X}}}\ \mathbf{0}_{d_{\ddot{X}},d_{\ddot{H}}})\widehat{\mathbf{\Omega}}_a'(x)\left(\widehat{\mathbf{\Lambda}}_{\ddot{Z}_aY}(x) - \widehat{\mathbf{\Lambda}}_{\ddot{Z}_a\ddot{X}_a}(x)\theta_2^a(x)\right),\end{aligned} \tag{D.1}$$

By (C.3), we know that

$$(\mathbf{I}_{d_{\ddot{X}}}\ \mathbf{0}_{d_{\ddot{X}},d_{\ddot{H}}})\widehat{\mathbf{\Omega}}_a'(x) = \left(f_{X_1}^{-1}(x)\mathbf{\Omega}'(x)\ \ \mathbf{0}_{d_{\ddot{X}},d_{\ddot{H}}}\right) + o_p(1). \tag{D.2}$$

Meanwhile, the first $d_{\ddot{X}}$ elements of $\widehat{\mathbf{\Lambda}}_{\ddot{Z}_aY}(x) - \widehat{\mathbf{\Lambda}}_{\ddot{Z}_a\ddot{X}_a}(x)\theta_2^a(x)$ satisfies that

$$\sqrt{Nh}[\widehat{\mathbf{\Lambda}}_{\ddot{Z}_aY}(x) - \widehat{\mathbf{\Lambda}}_{\ddot{Z}_a\ddot{X}_a}(x)\theta_2^a(x)]_{[1:d_{\ddot{X}}]} = \sqrt{Nh}h^2\mu_2\mathbf{B}_a(x) + \sqrt{Nh}[\mathbf{\Psi}_a(x)]_{[1:d_{\ddot{X}}]} + o_p(1).$$

By Lyapunov's central limit theorem, for each $x \in \mathcal{X}_1$,

$$\sqrt{Nh}[\mathbf{\Psi}_a(x)]_{1:d_{\ddot{X}}} \to_d \mathcal{N}(0, f_{X_1}^2(x)\mathbf{\Sigma}(x)). \tag{D.3}$$

Since $\sqrt{Nh}[\mathbf{\Lambda}_{\ddot{Z}_aY} - \mathbf{\Lambda}_{\ddot{Z}_a\ddot{X}_a}\theta_2^a(x)]_{[(d_{\ddot{X}}+1):(d_{\ddot{X}}+d_{\ddot{H}})]} = O_p(1)$, we have that

$$\begin{aligned}\sqrt{Nh}(\widehat{\theta}_2^a(x) - \theta_2(x) - f_{X_1}^{-1}h^2\mu_2\mathbf{\Omega}'(x)\mathbf{B}_a(x)) &= f_{X_1}^{-1}(x)\mathbf{\Omega}'(x)[\mathbf{\Psi}_a(x)]_{[1:d_{\ddot{X}}]} + o_p(1) \\ &\to_d \mathcal{N}(0, \mathbf{\Omega}'^{-1}(x)\mathbf{\Sigma}(x)\mathbf{\Omega}(x)).\end{aligned}$$

Next, we focus on the alternative non-augmented estimators $\widehat{\theta}_{2,0}(x)$ and $\widehat{\theta}_{2,1}(x)$. For both $\ell = 0, 1$, $\theta_2(x) = \widehat{\mathbf{\Omega}}'(x)\widehat{\Lambda}_{\ddot{Z}\ddot{X}}(x)\theta_2(x)$. Since $\widehat{\mathbf{\Lambda}}_{\ddot{Z}Y}(x) = N^{-1}\sum_i \kappa_h(X_{i1} -$

$x)\ddot{Z}_{i2}\ddot{X}'_{i2}\theta_2(X_{i1})\widetilde{X}'_{i1}(x)\widehat{\mathbf{M}}^{-1}(x)\iota_1 + N^{-1}\sum_i \kappa_h(X_{i1}-x)\ddot{Z}_{i2}\widetilde{\varepsilon}_{i2}\widetilde{X}'_{i1}(x)\widehat{\mathbf{M}}^{-1}(x)\iota_1$. The difference between $\widehat{\theta}_2(x)$ and $\theta_2(x)$ can be written as:

$$\widehat{\theta}_{2,\ell}(x) - \theta_2(x) = \widehat{\mathbf{\Omega}}'(x)\widehat{\mathbf{B}}(x) + \widehat{\mathbf{\Omega}}'(x)\widehat{\mathbf{\Psi}}(x). \tag{D.4}$$

From (C.2), we have

$$\widehat{\mathbf{\Omega}}(x) - \mathbf{\Omega}(x) = O_p((Nh)^{-1/2}) + o_p(h_\ell), \tag{D.5}$$

where $\mathbf{\Omega}(x) = \mathbf{\Lambda}_{\ddot{Z}\ddot{X}}(x)\left(\mathbf{\Lambda}'_{\ddot{Z}\ddot{X}}(x)\mathbf{\Lambda}_{\ddot{Z}\ddot{X}}(x)\right)^{-1}$ as is defined in Theorem 4.1.

Next, rewrite $\widehat{\mathbf{\Psi}}(x)$ as follows.

$$\begin{aligned}\widehat{\mathbf{\Psi}}(x) &= \left(\iota'_1(\mathbf{D}^{-1}\widehat{\mathbf{M}}\mathbf{D}^{-1})^{-1}\otimes \mathbf{I}_{d_{\ddot{Z}}}\right) N^{-1}\sum_i \widetilde{\varepsilon}_{i2}\kappa_h(X_{i1}-x)(\mathbf{D}^{-1}\widetilde{X}_{i1}(x)\otimes \mathbf{I}_{d_{\ddot{Z}}})\ddot{Z}_{i2}\\ &\equiv \left(\iota'_1(\mathbf{D}^{-1}\widehat{\mathbf{M}}\mathbf{D}^{-1})^{-1}\otimes \mathbf{I}_{d_{\ddot{Z}}}\right) N^{-1}\sum_i \varphi_i(x),\end{aligned}$$

where $\varphi_i(x)$ is a $d_{\ddot{Z}}\times d_{\ddot{X}}$ dimensional vector. Note that

$$\begin{aligned}&(\iota'_1(\mathbf{M}f_{X_1}(x))^{-1}\otimes \mathbf{I}_{\ddot{Z}})\mathbb{V}(\sqrt{h}\varphi_i(x))((\mathbf{M}f_{X_1}(x))^{-1}\iota_1\otimes \mathbf{I}_{\ddot{Z}})\\ &= f^{-2}_{X_1}(x)\mathbb{E}\left[h\kappa^2_h(X_{i1}-x)\iota'_1\mathbf{M}^{-1}\mathbf{D}^{-1}\widetilde{X}_{i1}(x)\widetilde{X}'_{i1}(x)\mathbf{D}^{-1}\mathbf{M}^{-1}\iota_1\mathbb{E}[\widetilde{\varepsilon}^2_{i2}\ddot{Z}_{i2}\ddot{Z}'_{i2}|X_{i1}]\right]\\ &= f^{-1}_{X_1}(x)\mathbb{E}[\widetilde{\varepsilon}^2_{i2}\ddot{Z}_{i2}\ddot{Z}'_{i2}|X_{i1}=x]\iota'_1\mathbf{M}^{-1}\mathbf{M}_\kappa\mathbf{M}^{-1}\iota_1 + o(1),\end{aligned} \tag{D.6}$$

where $\mathbf{M}_\kappa$ is a $(\ell+1)\times(\ell+1)$ matrix whose (i,j)-th component is ν_{i+j-2}. Note that $\iota'_1\mathbf{M}^{-1}\mathbf{M}_\kappa\mathbf{M}^{-1}\iota_1 = \nu_0$ for both $\ell = 0, 1$.

Then, because of (D.6), the convergence of $\mathbf{D}^{-1}\widehat{\mathbf{M}}\mathbf{D}^{-1}$ in (C.1), the Cramér-Wold device, the Lyapunov's central limit theorem, and Slutsky's theorem, we have the weak convergence result that

$$\sqrt{Nh}\widehat{\mathbf{\Psi}}(x) \to_d \mathcal{N}\left(0, \Sigma(x)\right).$$

Together with (D.4) and (D.5), the weak convergence result of the theorem is proven.

Proof of Proposition 4.1

The point-wise consistency of $\widehat{\mathbf{\Omega}}(x)$ has been established in (D.5). In this proof, we focus on the consistency of $\widehat{\mathbf{\Sigma}}(x)$. Let $\widehat{f}_{X_1}(x)$ be a consistent estimator of $f_{X_1}(x)$. Then,

$$\begin{aligned}
\widehat{\mathbf{\Sigma}}(x) &= \widehat{f}_{X_1}^{-2}(x)\frac{h}{N}\sum_i \widetilde{\varepsilon}_{i2}^2 \ddot{Z}_{i2}\ddot{Z}_{i2}'\kappa_h^2(X_{i1}-x) + \mathbf{R}_\Sigma(x) \\
&= f_{X_1}^{-2}(x)\mathbb{E}[\widetilde{\varepsilon}_{i2}^2 \ddot{Z}_{i2}\ddot{Z}_{i2}' h\kappa_h^2(X_{i1}-x) + \mathbf{R}_\Sigma(x) + o_p(1) \\
&= f_{X_1}^{-2}(x)\int \mathbb{V}(\widetilde{\varepsilon}_{i2}\ddot{Z}_{i2}|X_{i1}=x+uh) f_{X_1}(x+uh)\kappa^2(u)du + \mathbf{R}_\Sigma(x) + o_p(1) \\
&= f_{X_1}^{-1}(x)\mathbb{V}(\widetilde{\varepsilon}_{i2}\ddot{Z}_{i2}|X_{i1}=x)\int \kappa^2(u)du + \mathbf{R}_\Sigma(x) + o_p(1) \\
&= \mathbf{\Sigma}(x) + \mathbf{R}_\Sigma(x) + o_p(1),
\end{aligned} \tag{D.7}$$

where $\mathbf{R}_\Sigma(x)$ is the sum of two terms, $\mathbf{R}_{\Sigma,1}(x)$ and $\mathbf{R}_{\Sigma,2}(x)$, defined by

$$\begin{aligned}
\mathbf{R}_{\Sigma,1}(x) &:= \frac{h}{N}\sum_i \left(\ddot{X}_{i2}'(\widehat{\theta}_2(X_{i1})-\theta_2(X_{i1}))\right)^2 \ddot{Z}_{i2}\ddot{Z}_{i2}'\kappa_h^2(X_{i1}-x), \\
\mathbf{R}_{\Sigma,2}(x) &:= \frac{2h}{N}\sum_i \left(\widetilde{\varepsilon}_{i2}\ddot{X}_{i2}'(\widehat{\theta}_2(X_{i1})-\theta_2(X_{i1}))\right) \ddot{Z}_{i2}\ddot{Z}_{i2}'\kappa_h^2(X_{i1}-x).
\end{aligned}$$

The second and third equalities of (D.7) are obtained from the standard arguments on the point-wise consistency of the kernel estimator and kernel estimation derivations. The fourth equality is obtained from Assumptions C.1.(b) and C.1.(f). The rest of the proof then focuses on showing that $\mathbf{R}_\Sigma(x)$ is $o_p(1)$.

The first remainder $\mathbf{R}_{\Sigma,1}(x)$ is bounded above as follows.

$$\begin{aligned}
\|\mathbf{R}_{\Sigma,1}(x)\| &\leq 2N^{-1}\sum_i \|\ddot{X}_{i2}\|^2\|\ddot{Z}_{i2}\|^2\|\widehat{\theta}_2(X_{i1})-\theta_2(X_{i1})\|^2 h\kappa_h^2(X_{i1}-x) \\
&\leq \sup_{x\in\mathcal{X}_1}\|\widehat{\theta}_2(x)-\theta_2(x)\|^2 N^{-1}\sum_i \|\ddot{X}_{i2}\|^2\|\ddot{Z}_{i2}\|^2 h\kappa_h^2(X_{i1}-x).
\end{aligned}$$

Then, because of the results in (D.4), (D.5), and (C.8), and Lemma C.3, we find that

$$\sup_{x\in\mathcal{X}_1}\|\widehat{\theta}_2(x)-\theta_2(x)\| + o_p(1) \leq \sup_{x\in\mathcal{X}_1}\|\widehat{\mathbf{\Omega}}(x)\| \sup_{x\in\mathcal{X}_1}\left(\|\widehat{\mathbf{B}}(x)\| + \|\widehat{\mathbf{\Psi}}(x)\|\right) = o_p(1). \tag{D.8}$$

Moreover, because of the Markov's inequality and Assumptions C.1.(a), C.1.(d), and C.1.(f), we have that $N^{-1}\sum_i \|\ddot{X}_{i2}\|^2\|\ddot{Z}_{i2}\|^2 h\kappa_h^2(X_{i1}-x) = O_p(1)$. Hence, by combining these, we have uniformly in $x \in \mathcal{X}_1$

$$\|\mathbf{R}_{\Sigma,1}(x)\| = o_p(1). \tag{D.9}$$

By using similar arguments, we find that uniformly in $x \in \mathcal{X}_1$,

$$\|\mathbf{R}_{\Sigma,2}(x)\| \leq o_p(1) N_s^{-1} \sum_{i:X_{i1}\in\mathcal{X}_1} \|\widetilde{\varepsilon}_{i2}\| \|\ddot{X}_{i2}\| \|\ddot{Z}_{i2}\|^2 h \kappa_h^2(X_{i1}-x) = o_p(1). \tag{D.10}$$

Thus, the desired result is given from (D.5), (D.7), (D.9) and (D.10).

Proof of Theorem 4.2

Unless otherwise specified, all summations in the proof over i are with respect to $X_{i1} \in \mathcal{X}_1^*$, while all summations over j are with respect to the full sample, or $j = 1, ..., N$. Let $\widetilde{\vartheta}_2 = N_s{}^{-1} \sum_{i:X_{i1}\in\mathcal{X}_1^*} \theta_2(X_{i1})$ and $\widehat{p} = N_s/N$. The estimator $\widehat{\vartheta}_2$ satisfies

$$\sqrt{N_s}(\widehat{\vartheta}_2 - \vartheta_2) = \sqrt{N_s}(\widehat{\vartheta}_2 - \widetilde{\vartheta}_2) + \sqrt{N_s}(\widetilde{\vartheta}_2 - \vartheta_2). \tag{D.11}$$

Because $\widetilde{\vartheta}_2 - \vartheta_2 = N_s^{-1}\sum_{i:X_{i1}\in\mathcal{X}_1}(\theta_2(X_{i1}) - \vartheta_2) = N_s^{-1}\sum_i(\theta_2(X_{i1}) - \mathbb{E}[\theta_2(X_{i1})|X_{i1} \in \mathcal{X}_1])$, and by the CLT, we have that the second summand satisfies

$$\sqrt{N_s}(\widetilde{\vartheta}_2 - \vartheta_2) \to_d N(0, \mathbb{V}[\theta_2(X_{i1})|X_{i1} \in \mathcal{X}_1^*]) \overset{d}{=} N(0, \mathbf{\Sigma}_2^*). \tag{D.12}$$

Then, we focus on the first summand in (D.11). Because of (C.5) and (C.7) in Lemma C.3, we find that

$$\begin{aligned}
\sqrt{N_s}(\widehat{\vartheta}_2 - \widetilde{\vartheta}_2) &= N_s^{-1/2} \sum_i (\mathbf{I}_{d_{\ddot{X}}} \;\; \mathbf{0}_{d_{\ddot{H}}})(\widehat{\theta}_2^a(X_{i1};h) - \theta_2^a(X_{i1})) \\
&= N_s^{-1/2} \sum_i f_{X_1}^{-1}(X_{i1}) \left(\mathbf{\Omega}'(X_{i1}) \;\; \mathbf{0}_{d_{\ddot{H}}}\right) \mathbf{\Psi}_a(X_{i1};h) \\
&\quad + N_s^{-1/2} \sum_i \left(\widehat{\mathbf{\Omega}}_a(X_{i1};h) - f_{X_1}^{-1}(X_{i1})(\mathbf{\Omega}'(X_{i1}) \;\; \mathbf{0}_{d_{\ddot{H}}})\right) \mathbf{\Psi}_a(X_{i1};h) \\
&\quad + N_s^{-1/2} \sum_i \left(\widehat{\mathbf{\Omega}}_a(X_{i1};h) - f_{X_1}^{-1}(X_{i1})(\mathbf{\Omega}'(X_{i1}) \;\; \mathbf{0}_{d_{\ddot{H}}})\right) \widehat{\mathbf{B}}_a(X_{i1};h) \\
&\quad + N_s^{-1/2} \sum_i f_{X_1}^{-1}(X_{i1}) \mathbf{\Omega}'(X_{i1}) [\widehat{\mathbf{B}}_a(X_{i1};h)]_{[1:d_{\ddot{X}},.]} \\
&= N_s^{-1/2} \sum_i f_{X_1}^{-1}(X_{i1}) \mathbf{\Omega}'(X_{i1}) \left[\mathbf{\Psi}_a(X_{i1};h)\right]_{[1:d_{\ddot{X}}]} + o_p(1), \\
&= N^{-1} N_s^{-1/2} \sum_j \sum_{i\neq j} f_{X_1}^{-1}(X_{i1}) \kappa_h(X_{i1} - X_{j1}) \mathbf{\Omega}'(X_{i1}) \ddot{Z}_{j2} \widetilde{\varepsilon}_{j2} + o_p(1). \\
&= (N_s/N)^{1/2} \cdot N^{-1/2} \sum_{j=1}^N \zeta(X_{j1}) \widetilde{\varepsilon}_{j2} \ddot{Z}_{j2} + o_p(1),
\end{aligned} \tag{D.13}$$

where $\zeta(X_{j1}) = N_s^{-1}\sum_{i\neq j} f_{X_1}^{-1}(X_{i1})\kappa_h(X_{i1}-X_{j1})\mathbf{\Omega}'(X_{i1})$. The third and fourth equalities follow from Lemma C.3 and that $N^{-1/2}\sum_i f_{X_1}^{-1}(X_{i1})\mathbf{\Omega}'(X_{i1})\ddot{Z}_{i2}\widetilde{\varepsilon}_{i2} = O_p(1)$, because of Markov's inequality and the independence of $\varepsilon_{i2}|(X_{i1},\ddot{Z}_{i2})$ across i.

Let $\overline{\zeta}(x) = \mathbb{E}[\zeta(x)|X_{i1}\in\mathcal{X}_1^*] = \mathbb{E}[\mathbf{\Omega}'(X_{i1})\kappa_h(X_{i1}-x)f_{X_1}^{-1}(X_{i1})|X_{i1}\in\mathcal{X}_1^*]$. It is easy to show that $\overline{\zeta}(x) = p^{-1}\int_{\mathcal{X}_1^*}\mathbf{\Omega}'(t)\kappa_h(t-x)dt$ and therefore $\overline{\zeta}(X_{j1}) = p^{-1}\int_{\mathcal{X}_1^*}\kappa_h(t-X_{j1})\mathbf{\Omega}'(t)dt$. Since $\mathbb{E}[\zeta(X_{j1})\widetilde{\varepsilon}_{j2}\ddot{Z}_{j2}] = 0$ by the exclusion restriction, we know that $\mathbb{E}[\overline{\zeta}(X_{j1})\widetilde{\varepsilon}_{j2}\ddot{Z}_{j2}] = 0$ and $\mathbb{E}[(\zeta(X_{j1})-\overline{\zeta}(X_{j1}))\widetilde{\varepsilon}_{j2}\ddot{Z}_{j2}] = 0$ as well. In addition,

$$N^{-1}\mathbb{E}[\|\sum_j(\zeta(X_{j1})-\overline{\zeta}(X_{j1}))\widetilde{\varepsilon}_{j2}\ddot{Z}_{j2}\|^2] \leq \mathbb{E}\left[\|\zeta(X_{j1})-\overline{\zeta}(X_{j1})\|^2\mathbb{E}[\widetilde{\varepsilon}_{j2}^2\|\ddot{Z}_{j2}\|^2|X_{j1}]\right]$$
$$\leq O(1)\mathbb{E}[\|\zeta(X_{j1})-\overline{\zeta}(X_{j1})\|^2] = O((Nh)^{-1}). \tag{D.14}$$

Then, by the Markov's inequality, we have

$$N^{-1/2}\sum_{j=1}^N\zeta(X_{j1})\widetilde{\varepsilon}_{j2}\ddot{Z}_{j2} - N^{-1/2}\sum_{j=1}^N\overline{\zeta}(X_{j1})\widetilde{\varepsilon}_{j2}\ddot{Z}_{j2} = o_p(1). \tag{D.15}$$

Combining this with (D.13), we have

$$\sqrt{N_s}(\widehat{\vartheta}_2-\widetilde{\vartheta}_2) = (N_s/N)^{1/2}\cdot N^{-1/2}\sum_{j=1}^N\overline{\zeta}(X_{j1})\widetilde{\varepsilon}_{j2}\ddot{Z}_{j2} + o_p(1). \tag{D.16}$$

with $\mathbb{E}[\overline{\zeta}(X_{j1})\widetilde{\varepsilon}_{j2}\ddot{Z}_{j2}] = 0$ and

$$\begin{aligned}
&\mathbb{E}[\overline{\zeta}(X_{j1})\widetilde{\varepsilon}_{j2}^2\ddot{Z}_{j2}\ddot{Z}_{j2}'\overline{\zeta}'(X_{j1})]\\
&= p^{-2}\int\int_{\mathcal{X}_1^*}\int_{\mathcal{X}_1^*}\kappa_h(t-X_{j1})\kappa_h(\widetilde{t}-X_{j1})\mathbf{\Omega}'(t)\mathbb{V}(\widetilde{\varepsilon}_{j2}\ddot{Z}_{j2}|X_{j1})\mathbf{\Omega}(\widetilde{t})dtd\widetilde{t}dF_{X_1}(X_{j1})\\
&= p^{-2}\int\int\kappa(u)\kappa(u-s)u(u-s)'duds\\
&\quad\times\int_{\mathcal{X}_1^*}\mathbf{\Omega}'(w)\mathbb{V}(\widetilde{\varepsilon}_{j2}\ddot{Z}_{j2}|X_{j1}=w)\mathbf{\Omega}(w)dF_{X_1}(w)+o(1)\\
&= p^{-1}\mathbf{\Sigma}_1^*+o(1).
\end{aligned}$$

Moreover, we note that, because $\overline{\zeta}(.)$ is uniformly bounded,

$$\mathbb{E}[\|\overline{\zeta}(X_{j1})\|^{2+\delta}\|\varepsilon_{j2}\ddot{Z}_{j2}\|^{2+\delta}] \leq O(1)\mathbb{E}[\|\overline{\zeta}(X_{j1})\|^{2+\delta}] = O(1),$$

and by combining it with the boundedness of the variance of $\overline{\zeta}(X_{j1})\widetilde{\varepsilon}_{j2}\ddot{Z}_{j2}$, one can show that the Lyapunov's condition holds. Therefore, the following is obtained by applying the Lyapunov CLT to (D.16) and $N_s/N\to p$;

$$(N_s/N)^{1/2}\cdot N^{-1/2}\sum_j\overline{\zeta}(X_{j1})\widetilde{\varepsilon}_{j2}\ddot{Z}_{j2}\to_d N(0,\mathbf{\Sigma}_1^*) \tag{D.17}$$

Then, the desired result follows from (D.12) and (D.17) and the fact that $\widetilde{\vartheta}_2 - \vartheta_2$ is a function of X_{i1} only and thus $\mathbb{E}[(\widetilde{\vartheta}_2 - \vartheta_2)'\zeta(X_{i1})\ddot{Z}_{i2}\widetilde{\varepsilon}_{i2}] = 0$.

We then focus on $\widehat{\theta}_{2,\ell}$. We notice that

$$\sqrt{N_s}(\widehat{\vartheta}_{2,\ell} - \vartheta_2) = \sqrt{N_s}(\widehat{\vartheta}_{2,\ell} - \widetilde{\vartheta}_2) + \sqrt{N_s}(\widetilde{\vartheta}_2 - \vartheta_2). \tag{D.18}$$

For notational convenience, we for the moment let the scalar-valued random variable $\varsigma_\ell(X_{i1}, X_{j1})$ be defined as follows.

$$\varsigma_\ell(X_{i1}, X_{j1}) = \kappa_h(X_{j1} - X_{i1})f_{X_1}^{-1}(X_{i1}) \cdot \widetilde{X}'_{j1}(X_{i1})\mathbf{D}^{-1}\mathbf{M}^{-1}\iota_1.$$

In our case of $\ell = 0, 1$, the random variable $\varsigma_\ell(X_{i1}, X_{j1})$ reduces to $\kappa_h(X_{i1} - X_{j1})f_{X_1}^{-1}(X_{i1})$ for any i and j, because $\iota_1'\mathbf{M}^{-1}\mathbf{D}^{-1}\widetilde{X}_{j1}(.) = 1$ for any j.

Then, the following holds because of results in Lemma C.3:

$$\begin{aligned}
\sqrt{N_s}(\widehat{\vartheta}_{2,\ell} - \widetilde{\vartheta}_2) =& N_s^{-1/2}\sum_i \left(\widehat{\theta}_{2,\ell}(X_{i1}) - \theta_2(X_{i1})\right) \\
=& N_s^{-1/2}\sum_i \left(\widehat{\mathbf{\Omega}}'(X_{i1})\widehat{\mathbf{B}}'(X_{i1}) + \widehat{\mathbf{\Omega}}'(X_{i1})\widehat{\mathbf{\Psi}}'(X_{i1})\right) \\
=& N_s^{-1/2}\sum_i \left(\widehat{\mathbf{\Omega}}'(X_{i1})\widehat{\mathbf{\Psi}}'(X_{i1})\right) + O_p(\sqrt{N}(hc_h + h^2)) \\
=& N_s^{-1/2}\sum_i \mathbf{\Omega}'(X_{i1})\mathbf{\Psi}(X_{i1}) \\
&+ N_s^{-1/2}\sum_i (\widehat{\mathbf{\Omega}}'(x) - \mathbf{\Omega}'(x))\mathbf{\Psi}(x) \\
&+ N_s^{-1/2}\sum_i \widehat{\mathbf{\Omega}}'(X_{i1})(\widehat{\mathbf{\Psi}}(x) - \mathbf{\Psi}(x)) + O_p(\sqrt{N}(hc_h + h^2)) \\
=& N_s^{-1/2}\sum_i \mathbf{\Omega}'(X_{i1})\mathbf{\Psi}(X_{i1}) + O_p(\sqrt{N}(c_h + h)^2) \\
=& N^{-1}N_s^{-1/2}\sum_j\sum_i \widetilde{\varepsilon}_{j2}\mathbf{\Omega}(X_{i1})'\ddot{Z}_{j2}\varsigma_\ell(X_{i1}, X_{j1}) + o_p(1) \\
=& N^{-1}N_s^{-1/2}\sum_j\sum_{i\neq j} \widetilde{\varepsilon}_{j2}\mathbf{\Omega}(X_{i1})'\ddot{Z}_{j2}\varsigma_\ell(X_{i1}, X_{j1}) + o_p(1).
\end{aligned}$$

The second last equality holds from $\sqrt{N}(c_h + h)^2 = o(1)$ under the rate condition in Assumption C.1 and the additional condition $Nh^4 = o(1)$ stated in the theorem. The last equality holds because

$$N^{-1}N_s^{-1/2}\sum_{j=1}^{N}\mathbf{\Omega}'(X_{j1})\ddot{Z}_{j2}\widetilde{\varepsilon}_{j2}\varsigma_\ell(X_{j1},X_{j1})$$
$$\leq \sup_{x\in\mathcal{X}_1}\left\|f_{X_1}^{-1}(x)\right\|\sup_{x\in\mathcal{X}_1}\left\|\mathbf{\Omega}'(X_{j1})\right\|\iota_1'\mathbf{M}^{-1}\iota_1\kappa_h(0)N^{-1}N_s^{-1/2}\sum_{j=1}^{N}\ddot{Z}_{j2}\widetilde{\varepsilon}_{i2}=o_p(1).$$

Let $\zeta_\ell(X_{j1})=N_s^{-1}\sum_{i:i\neq j}\mathbf{\Omega}'(X_{i1})\varsigma_\ell(X_{i1},X_{j1})$. For $\ell=0,1$, we have $\mathbb{E}[\zeta_\ell(x)|X_{i1}\in\mathcal{X}_1^*]=\mathbb{E}[\mathbf{\Omega}'(X_{i1})\kappa_h(X_{i1}-x)f_{X_1}^{-1}(X_{i1})\widetilde{X}_{j1}(X_{i1})'\mathbf{D}^{-1}\mathbf{M}^{-1}\iota_1|X_{i1}\in\mathcal{X}_1^*]=\mathbb{E}[\mathbf{\Omega}'(X_{i1})\kappa_h(X_{i1}-x)f_{X_1}^{-1}(X_{i1})|X_{i1}\in\mathcal{X}_1^*]=\overline{\zeta}(x)$. Then, by using similar arguments in (D.14) and (D.15), we have

$$\sqrt{N_s}(\widehat{\vartheta}_{2,\ell}-\widetilde{\vartheta}_2)=(N_s/N)^{1/2}\cdot N^{-1/2}\sum_{j=1}^{N}\overline{\zeta}(X_{j1})\widetilde{\varepsilon}_{j2}\ddot{Z}_{j2}+o_p(1). \tag{D.19}$$

Therefore, the desired result for the alternative estimator is given as a consequence of (D.12) and (D.17).

Proof of Proposition 4.2

The suggested estimator of $\mathbf{\Sigma}_1^*$ is as follows,

$$\widehat{\mathbf{\Sigma}}_1^*=\widehat{p}\cdot N^{-1}\sum_j\widehat{\varepsilon}_{j2}^2\widehat{\zeta}(X_{j1})\ddot{Z}_{j2}\ddot{Z}_{j2}'\widehat{\zeta}'(X_{j1}), \tag{D.20}$$

where $\widehat{\zeta}(x)=N_s^{-1}\sum_{i:X_{i1}\in\mathcal{X}_1}\widehat{f}_{X_1}^{-1}(X_{i1})\kappa_h(X_{i1}-x)\widehat{\mathbf{\Omega}}'(X_{i1})$ and $\widehat{\varepsilon}_{i2}=\widetilde{\varepsilon}_{i2}-\ddot{X}_{i2}'(\widehat{\theta}_2(X_{i1})-\theta_2(X_{i1}))$. Then, because of (C.6) in Lemma C.3, (D.5), and the fact that $\iota_1'\mathbf{M}^{-1}\widetilde{X}_{j1}(.)$ reduces to 1 for all j the following holds uniformly in $x\in\mathcal{X}_1$:

$$\|\widehat{\zeta}(x)-\zeta(x)\|\leq O_p(1)\sup_{\widetilde{x}\in\mathcal{X}_1^*}\left(|\widehat{f}_{X_1}^{-1}(\widetilde{x})-f_{X_1}^{-1}(\widetilde{x})|+\|\widehat{\mathbf{\Omega}}(\widetilde{x})-\mathbf{\Omega}(\widetilde{x})\|\right)=o_p(1). \tag{D.21}$$

In addition, because of Lemma A.1 in Fan and Huang (2005), $\zeta(.)$ satisfies that

$$\sup_{x\in\mathcal{X}_1}\|\zeta(x)-\overline{\zeta}(x)\|=O_p(c_h), \tag{D.22}$$

The uniform convergence results in (D.21) and (D.22) imply that

$$\sup_{x\in\mathcal{X}_1}\|\widehat{\zeta}(x)-\overline{\zeta}(x)\|=o_p(1). \tag{D.23}$$

We for the moment define $\widetilde{\mathbf{\Sigma}}_1^*$ as follows.

$$\widetilde{\mathbf{\Sigma}}_1^*=\widehat{p}\cdot N^{-1}\sum_j\overline{\zeta}(X_{j1})\widehat{\varepsilon}_{j2}^2\ddot{Z}_{j2}\ddot{Z}_{j2}'\overline{\zeta}'(X_{j1}).$$

Then, by the LLN for the i.i.d. data and the probability limit of $\widehat{p}$, we have

$$\widetilde{\boldsymbol{\Sigma}}_1^* \to_p \boldsymbol{\Sigma}_1^*. \tag{D.24}$$

Moreover, because of (D.8), (D.23), (D.24), Assumption C.1.(c), the uniform boundedness of $\overline{\zeta}(.)$, and $N^{-1}\sum_j \|\ddot{X}_{j2}\|^2\|\ddot{Z}_{j2}\|^2 = O_p(1)$ (Assumption C.1.(f)), we find that

$$\|\widehat{\boldsymbol{\Sigma}}_1^* - \widetilde{\boldsymbol{\Sigma}}_1^*\| \leq O_p(1) \sup_{x\in\mathcal{X}_1} \left(\|\widehat{\zeta}(x) - \overline{\zeta}(x)\| + \|\widehat{\theta}_2(x) - \theta_2(x)\|\right) = o_p(1). \tag{D.25}$$

Therefore, the consistency of $\widehat{\boldsymbol{\Sigma}}_1^*$ can be obtained from (D.24) and (D.25).

The suggested estimator of $\boldsymbol{\Sigma}_2^*$ is given by

$$\widehat{\boldsymbol{\Sigma}}_2^* = N_s^{-1} \sum_{i:X_{i1}\in\mathcal{X}_1} (\widehat{\theta}_2(X_{i1}) - \widehat{\vartheta}_2)(\widehat{\theta}_2(X_{i1}) - \widehat{\vartheta}_2)'. \tag{D.26}$$

Let $\widetilde{\boldsymbol{\Sigma}}_2^*$ be defined by

$$\widetilde{\boldsymbol{\Sigma}}_2^* = N_s^{-1} \sum_{i:X_{i1}\in\mathcal{X}_1} (\theta_2(X_{i1}) - \vartheta_2)(\theta_2(X_{i1}) - \vartheta_2)', \tag{D.27}$$

and then because of the LLN, it satisfies that

$$\widetilde{\boldsymbol{\Sigma}}_2^* - \boldsymbol{\Sigma}_2^* = o_p(1). \tag{D.28}$$

Note that $\widehat{\theta}_2(.) - \widehat{\vartheta}_2 = \widehat{\theta}_2(.) - \theta_2(.) + \theta_2(.) - \vartheta_2 + \vartheta_2 - \widehat{\vartheta}_2$ and because of the uniform boundedness of $\theta_2(.)$, (D.8) and Theorem 4.2, we find that

$$\|\widehat{\boldsymbol{\Sigma}}_2^* - \widetilde{\boldsymbol{\Sigma}}_2^*\| \leq O_p(1) \left(\|\widehat{\vartheta}_2 - \vartheta_2\| + \sup_{x\in\mathcal{X}_1^*} \|\widehat{\theta}_2(x) - \theta_2(x)\|\right) = o_p(1). \tag{D.29}$$

Then the consistency of $\widehat{\boldsymbol{\Sigma}}_2^*$ can be established by using (D.28) and (D.29).

References

Acemoglu, D., D. Autor, D. Dorn, G. H. Hanson, and B. Price (2016): "Import competition and the great US employment sag of the 2000s," Journal of Labor Economics, 34, S141–S198.

Acemoglu, D. and P. Restrepo (2022): "Demographics and automation," The Review of Economic Studies, 89, 1–44.

Adão, R., M. Kolesár, and E. Morales (2019): "Shift-Share Designs: Theory and Inference," The Quarterly Journal of Economics, 134, 1949–2010.

Ahn, S. C. and P. Schmidt (1995): "Efficient estimation of models for dynamic panel data," Journal of Econometrics, 68, 5–27.

Anderson, T. and C. Hsiao (1982): "Formulation and estimation of dynamic models using panel data," Journal of Econometrics, 18, 47–82.

Angrist, J. (2002): "How do sex ratios affect marriage and labor markets? Evidence from America's second generation," Quarterly Journal of Economics, 117, 997–1038.

Arellano, M. and S. Bond (1991): "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations," The Review of Economic Studies, 58, 277–297.

Athey, S. and G. W. Imbens (2022): "Design-based Analysis in Difference-in-Differences Settings with Staggered Adoption," Journal of Econometrics.

Autor, D., D. Dorn, and G. Hanson (2013): "The China syndrome: Local labor market effects of import competition in the United States," American Economic Review, 103, 2121–2168.

Autor, D., D. Dorn, G. Hanson, and K. Majlesi (2020a): "Importing political polarization? The electoral consequences of rising trade exposure," American Economic Review, 110, 3139–3183.

Autor, D., D. Dorn, and G. H. Hanson (2021): "On the persistence of the China shock," Working Paper 29401, National Bureau of Economic Research.

Autor, D., D. Dorn, G. H. Hanson, G. Pisano, and P. Shu (2020b): "Foreign Competition and Domestic Innovation: Evidence from US Patents," American Economic Review: Insights, 2, 357–374.

——— (2020c): "Foreign competition and domestic innovation: Evidence from US patents," American Economic Review: Insights, 2, 357–374.

Autor, D. H., D. Dorn, G. H. Hanson, and J. Song (2014): "Trade adjustment: Worker-level evidence," Quarterly Journal of Economics, 129, 1799–1860.

Battisti, M., G. Peri, and A. Romiti (2022): "Dynamic effects of co-ethnic networks on immigrants' economic success," The Economic Journal, 132, 58–88.

Bernstein, S., R. Diamond, A. Jiranaphawiboon, T. McQuade, and B. Pousada (2022): "The contribution of high-skilled immigrants to innovation in the United States," Working Paper 30797, National Bureau of Economic Research.

Bloom, N., M. Draca, and J. Van Reenen (2016): "Trade induced technical change? The impact of Chinese imports on innovation, IT and productivity," The Review of Economic Studies, 83, 87–117.

Bloom, N., K. Handley, A. Kurmann, and P. Luck (2019): "The impact of chinese trade on us employment: The good, the bad, and the apocryphal," .

Bojinov, I., A. Rambachan, and N. Shephard (2021): "Panel experiments and dynamic causal effects: A finite population perspective," Quantitative Economics, 12(4), 1171–1196.

Borusyak, K., P. Hull, and X. Jaravel (2022): "Quasi-Experimental Shift-Share Research Designs," The Review of Economic Studies, 89, 181–213.

Boustan, L. P. (2010): "Was postwar suburbanization "white flight"? Evidence from the black migration," The Quarterly Journal of Economics, 125, 417–443.

Bravo, F. (2023): "Local polynomial estimation of nonparametric general estimating equations," Statistics & Probability Letters, 197, 109805.

Burchardi, K. B., T. Chaney, T. A. Hassan, L. Tarquinio, and S. J. Terry (2020): "Immigration, innovation, and growth," Tech. rep., National Bureau of Economic Research.

Cai, Z. and Q. Li (2008): "Nonparametric estimation of varying coefficient dynamic panel data models," Econometric Theory, 24, 1321–1342.

Callaway, B. and P. H. Sant'Anna (2021): "Difference-in-differences with multiple time periods," Journal of Econometrics.

Campbell, D. L. and K. Mau (2021): "On "Trade induced technical change: The impact of Chinese imports on innovation, IT, and productivity"," The Review of Economic Studies, 88, 2555–2559.

Card, D. (2009): "Immigration and Inequality," American Economic Review, 99, 1–21.

Cellini, S. R., F. Ferreira, and J. Rothstein (2010): "The value of school facility investments: Evidence from a dynamic regression discontinuity design," The Quarterly Journal of Economics, 125, 215–261.

Chen, S. and B. Zhang (2023): "Estimating and improving dynamic treatment regimes with a time-varying instrumental variable," Journal of the Royal Statistical Society Series B: Statistical Methodology, 85, 427–453.

Chernozhukov, V., S. Lee, and A. M. Rosen (2013): "Intersection bounds: Estimation and inference," Econometrica, 81, 667–737.

Dube, A., D. Girardi, O. Jordá, and A. M. Taylor (2023): "A Local Projections Approach to Difference-in-Differences," Working Paper 31184, National Bureau of Economic Research.

Fan, J. and I. Gijbels (1996): Local polynomial modelling and its applications, Monographs on statistics and applied probability 66, London: Chapman and Hall/CR, first edition. ed.

Fan, J. and T. Huang (2005): "Profile likelihood inferences on semiparametric varying-coefficient partially linear models," Bernoulli, 11, 1031 – 1057.

Feenstra, R. C., H. Ma, and Y. Xu (2019): "US exports and employment," Journal of International Economics, 120, 46–58.

Gallen, Y., J. S. Joensen, E. R. Johansen, and G. F. Veramendi (2023): "The Labor Market Returns to Delaying Pregnancy," SSRN working paper.

Han, S. (2021): "Identification in nonparametric models for dynamic treatment effects," Journal of Econometrics, 225, 132–147.

Heckman, J. J., J. E. Humphries, and G. Veramendi (2016): "Dynamic treatment effects," Journal of Econometrics, 191, 276–292, innovations in Measurement in Economics and Econometrics.

Heckman, J. J. and S. Navarro (2007): "Dynamic discrete choice and dynamic treatment effects," Journal of Econometrics, 136, 341–396.

Hernán, M. A. and J. M. Robins (2023): Causal inference: What If, CRC Press.

Hsu, Y.-C. and S. Shen (2023): "Dynamic Regression Discontinuity Under Treatment Effect Heterogeneity," University of California, Davis, Working Paper.

Mack, Y. P. and B. W. Silverman (1982): "Weak and strong uniform consistency of kernel regression estimates," Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, 61, 405–415.

Murphy, S. (2003): "Optimal dynamic treatment regimes," Journal of the Royal Statistical Society Series B: Statistical Methodology, 65(2), 331–355.

Murphy, S., van der Laan, J. M.J., Robins, and C. P. P. R. Group (2001): "Marginal mean models for dynamic regimes," Journal of the American Statistical Association, 96(456), 1410–1423.

Robins, J. (1986): "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, 7(9-12), 1393–1512.

——— (1987): "A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods," Journal of chronic diseases, 40, 139S–161S.

Stock, J. H. and M. W. Watson (2018): "Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments," The Economic Journal, 128, 917–948.

Su, L., I. Murtazashvili, and A. Ullah (2013): "Local linear GMM estimation of functional coefficient IV models with an application to estimating the rate of return to schooling," Journal of Business & Economic Statistics, 31, 184–207.

Sun, L. and S. Abraham (2021): "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," Journal of Econometrics, 225, 175–199.