Regression for Spatial Models with Interference

Christopher Adjaho *

February 19, 2025
[Click for latest version]

Abstract

We consider the problem of inference in a parametric equilibrium model of trade. In our setting with multiple locations and multiple sectors, common trade shocks generate differential exposure among locations, spillovers between locations, and heterogeneous responses in the outcome. When the data generating process (DGP) admits an infeasible representation of these outcomes in terms of a set of sufficient statistics for the spillovers, we can instead incorporate interference into regressions using own exposure and a function of neighbor exposure and an interactions matrix, although this may lead to misspecification of the regression model for the true underlying DGP. We show that the estimands of the linear model can be expressed as linear combinations of sector-location causal effects, or location causal effects. Depending on the extent of misspecification, the OLS estimators possibly fail to converge and we provide conditions for (i) the estimators to be consistent for the estimands, and (ii) valid inference using a variance estimator that agrees with a class of structural models of gravity trade and produces confidence intervals with asymptotically correct coverage. In contrast to cluster inference methods used in the empirical literature, which assume independence between groups of locations, we show that our conditions allow for a single cluster of locations exhibiting network dependence, but require the total misspecification error to grow at a rate smaller than the sample size.

*Department of Economics, New York University. ca2384@nyu.edu. I thank my advisor Konrad Menzel for continuous support and helpful comments throughout the project. I also thank Tim Christensen, Elena Manresa, Yassine Sbai Sassi, Sharon Traiberman, Quang Vuong, and participants of the NYU Econometrics seminar for helpful comments and feedback.

1 Introduction

This paper addresses the issue of making inferences in a class of equilibrium models using empirical specifications that only approximately capture the true interdependencies between units. In our setting, common observable shocks generate differential responses in unit outcomes but the equilibrium model, which describes how units interact with one another following the shocks, does not admit an estimable representation of outcomes in terms of variables capturing spillovers – or *interference* – between units. This paper proposes using a linear regression as an empirical specification for relating outcomes with the set of common shocks in a way that accounts for interference between units, while recognizing possible misspecification of the regression for the true, underlying structure of interference.

Representation of a data generating process (DGP) in terms of a set of variables that serve as sufficient statistics for interference is common within the quantitative spatial literature (Redding and Rossi-Hansberg 2017). When these variables are observable, one can control for them and estimate the effect of shocks on outcomes that already accounts for spillover and feedback effects. The problem of making inferences in settings where common shocks generate differential exposure to shocks, creating heterogeneous outcomes, is known in the econometrics literature as shift-share inference. We focus attention on these settings to speak to problems that deal with units being subject to observable macro shocks, such as when locations experience exogenous changes to trade conditions (D. H. Autor, Dorn, and Hanson 2013). When the researcher would like to account for potential interference from other units but the DGP does not admit estimable sufficient statistics, we can use linear regressions to approximate the total effect of the shocks on the outcomes.

This paper is closest to Adao, Kolesár, and Morales 2019 which proposes a method for inference in these regression designs, but we make corrections that account for the presence of equilibrium effects. We ask the following questions: What restrictions on the data, and on a model for the data, can we make if we want to do valid inference on the regression when there are spillovers and possible misspecification. In standard regression problems, we commonly make restrictions on the data only; for example, the regressors and errors should have finite means or variances. We explicitly require restrictions on both the data and a DGP because the latter will define what a spillover is and, therefore, what it means for the linear regression model to be misspecified. In our setting, the misspecification arises from two sources: (i) misspecification of treatment effect heterogeneity, and (ii) misspecification of bilateral interactions. The former is in fact present in the problem of Adao, Kolesár, and Morales 2019, and we show that just like their paper, consistency of the OLS estimator can be obtained with fairly weak assumptions about errors in treatment effect heterogeneity, and no additional restriction on this misspecification is needed for inference. The interesting case concerns our assumptions about (ii). This concern arises from functional form misspecification of interference of unit j in unit i's outcome. Allowing for arbitrary interactions on the true structure possibly threatens consistency of the OLS estimator and we provide sufficient conditions as well as a distribution theory that allows construction of confidence intervals that account for the true structure of interactions among units.

This paper is organized as follows: Section 2 describes the equilibrium model thought to generate the data and govern interactions between locations in the sample. We show that it does not admit a feasible sufficient statistics representation for the estimation of the effect of shocks on endogenous outcomes accounting for interference from other units; this 'first-best world' isn't attainable [1]. We therefore derive a reduced-form representation of the outcome which we call the potential outcome under the experiment of assigning common shocks to each unit. We show that it can equivalently be thought of as a problem of interference, whereby the potential outcome for a given unit depends on other units in the sample.

Section 3 describes the econometric model and we can think of our regression as using an exogenous mapping as a way of incorporating interference into the

[1] I use single quotes because, as we will see shortly, even if the model were to admit an estimable sufficient statistic, it would require constant causal effects.

outcome equation similar to how one would obtain an exogenous instrument for interference when the model admits an observable endogenous sufficient statistic. Thus, to the extent that we view the regression as a first attempt of approximating the outcome, where the data is governed by a model of gravity trade with unknown parameters, we attempt to uncover the restrictions on general equilibrium responses required to do inference on these linear estimands. We show that the projection coefficients are a linear combination of heterogenous treatment effects. Section 4 analyzes the statistical properties of the OLS estimators. It shows that the OLS estimator possibly fails to converge under arbitrary interaction structures. We provide conditions to achieve consistency under restrictions on the rate of growth of misspecification errors for general equilibrium effects on the outcomes. We also provide conditions for a distribution theory and propose a valid variance estimator for the construction of confidence intervals. Validity of these inference methods require stronger assumptions on misspecification errors, but are valid both conditionally and unconditionally on sample characteristics, and allow for a single cluster of locations exhibiting network dependence. Section 5 concludes.

1.1 Related literature

The statistical literature on interference, i.e. violations to the Stable Unit Treatment Value Assumption (SUTVA) of causal inference, focused on restrictive settings. For example, Hudgens and Halloran 2008 use the concept of partial and stratified interference in which only treatment assignment of units in a unit's own group and only the proportion of treated units affect outcome. Other papers have focused on units that fall within a distance $K > 0$ affecting outcomes, for example Alzubaidi and Higgins 2023. Such assumptions would be incompatible with knowledge of general equilibrium spillovers. A recent wave of papers focusing on more general settings include, for example, Leung 2022 or Sävje 2021 that treat the prior literature as misspecified. In the former paper, a unit's outcome is allowed to depend on distant alters, but with decreasing effect on the one's outcome, termed Approximate Neighborhood Interference (ANI). We avoid use of ANI and use of distance-based measures, but consider misspecified exposure mappings.

In the trade literature, Donaldson and Hornbeck 2016, in a seminal paper, represent equilibrium outcomes as functions of a correct and observable sufficient statistics. While they are able to estimate a causal effect, it is constant and their representation requires a single sector. We use multiple sectors, focus on sets of heterogeneous effects, and settings where these sufficient statistics are unobservable.

Shift-share regressions have been used in the trade literature to study the impact of trade shocks on local labor market outcomes accounting for spillovers. Adao, Arkolakis, and Esposito 2019, in an extension of D. H. Autor, Dorn, and Hanson 2013, tackle the question of how China trade shocks affect U.S. local employment and wages. They argue that estimates accounting for spillovers provide empirical evidence that the aggregate impact of China trade shocks on labor market outcomes are larger than what the literature had previously estimated due to being amplified by spatial linkages and larger still than predictions of existing quantitative spatial models. D. Autor, Dorn, and Hanson 2021, using the same regression, find weak evidence of gravity-based spillovers of trade shocks across locations on employment measures. Acemoglu et al. 2016 also use a differential-exposures design similar to the previous papers where the unit of analysis is an industry and the outcome is a measure for employment, but without the common shock. In their paper, industries are linked in input-output space and exposure to China trade shocks propagates upstream to affect output and hence employment. By relaxing the assumption of independence between locations, as is done in D. H. Autor, Dorn, and Hanson 2013, and instead allowing for spatial linkages, it is no longer necessary to assume that locations are small open economies: changes in economic conditions in a given location following a trade shock can be transmitted to other regions.

In another paper that combines equilibrium with shift-share designs, Adao, Arkolakis, and Esposito 2019 attempt to estimate the aggregate and differential impact of exposure variables on outcomes. They do so by using a general equilibrium gravity trade model to derive a reduced-form equation for the change in wages

$$\hat{w}_i = \beta_{ii}(\theta)X_i + \sum_{j\neq i}^{n} \beta_{ij}(\theta)X_j, \quad (1)$$

and estimate the deep parameters θ and thus the matrix of elasticities $\{\beta_{ij}(\theta)\}_{i,j}$. This paper differs from Adao, Arkolakis, and Esposito 2019 in the estimand of interest and the perspective we adopt about the disconnect between estimates obtained in the empirical literature, and those obtained by existing quantitative spatial models with regards to the effects of common trade shocks. Their paper argues for reconciliation between large empirical estimates of China import shocks and those obtained using general equilibrium spatial models, which they argue often fail to replicate their empirical counterparts[2]. However, this paper takes the maintained hypothesis that the equilibrium model is the true DGP and that it is the empirical papers that must be reconciled with it due to misspecification. Indeed, this is particularly true because while Adao, Arkolakis, and Esposito 2019 argue that regression estimates are large, one could question the validity of the inference method used: Forming independent clusters based on, for example, geographic proximity could fail to account for more complex spatial dependencies. In spatial settings, units are certain to be interconnected and clustering could invalidate conclusions about the statistical significance of the empirical magnitudes. Furthermore, the reduced form representation (1) of the DGP reveals aggregate causal effects, but it is unclear what regression estimands reveal and, therefore, what quantitative models are supposed to be matching in terms of the estimated responses in the outcome, even with a random assignment assumption for the shocks. We attempt to be clear on our sampling assumptions and population of interest. For example, we use a finite-population model in which the sole source of randomness comes from the observed shocks.

[2]The introduction to Adao, Arkolakis, and Esposito 2019 states "the credibility of the model's aggregate impact is severely undermined if it yields reduced-form elasticities that are inconsistent with their empirical counterparts."

2 Model

2.1 Overview of problem

Let n denote the number of locations in our sample, and S the number of sectors. Suppose the outcome of interest in location i, Y_i, is related to a set of common, observable, sector shocks $\mathcal{X} = \{\mathcal{X}_s\}_{s=1}^S$, and we are interested in their effect on the outcome that accounts for spillovers – or *interference* – between locations:

$$Y_i = f_{i,n}(\mathcal{X}_1, ..., \mathcal{X}_S), \tag{2}$$

where $f_{i,n}$ is an unknown function depending on the other $n-1$ locations in the sample. In some cases, it is possible to represent outcomes in terms of a small set of observable sufficient statistics for interference, $T_i = \{T_{i,r}\}_{r=1}^K$, that are functions of the shocks[3]. Then

$$Y_i = h_i(T_{i1}, .., T_{iK}) \tag{3}$$

depends on other locations only through T_i. Empirical estimation of estimands defined by (3) will account for spillovers between units. When these sufficient statistics are not observable, we can combine Equation 2 with an empirical specification whose variables are observable and account for dependence on other locations, $Y_i = \tilde{h}_{i,n}(\tilde{T}_{i1}, ..., \tilde{T}_{ik})$, although this empirical specification is possibly misspecified for the true outcome model (2) and (3): conditional on $\tilde{T}_i$, Y_i still depends on n (misspecification of 3), and $\tilde{h}_{i,n} \neq f_{i,n}$ (misspecification of 2). This paper asks the following question: What restrictions on the data, and on a model for the data (2), can we make if we want to do valid inference on $\tilde{h} = \{\tilde{h}_{i,n}\}_{i=1}^n$ when there are spillovers following common shocks and our empirical specification $\tilde{h}$ is misspecified?

We focus attention on the case when $\tilde{h}$ is a linear regression, set $k = 2$, and define our observable variables to be $\tilde{T}_{i1} = X_i = \sum_{s=1}^S w_{is}\mathcal{X}_s$ with sector weights $(w_{is})_{s=1}^S$ such that $\sum_{s=1}^S w_{is} \leq 1$ for all i, and $\tilde{T}_{i2} = \sum_{j=1}^n G_{ij}X_j$ where $G = \{G_{ij}\}$ is a bi-

[3]'Small' in the sense of $K < n$ and $K/n \to 0$ as the sample size grows.

lateral interactions matrix such that $\sum_{j=1}^{n} G_{ij} = 1$, $G_{ii} = 0$ for all i. Our empirical specification is then

$$Y_i = \beta_1 X_i + \beta_2 \sum_{j \neq i}^{n} G_{ij} X_j + \varepsilon_i. \tag{4}$$

Equation 4 is the empirical specification that has been used by some of the papers cited in the previous section.

2.2 Parametric model

Consider a parametric model $\mathcal{M} = \mathcal{M}(\theta)$, where θ are structural parameters. Let $N = n + 1$ consist of n locations of interest and a foreign region F. Let $(\omega_1, ..., \omega_N)$ be the vector of endogenous wages with $w_F = 1$ so that statements about wage changes are relative to F. The model $\mathcal{M}$ is defined by a system

$$\mathbb{D}_i(\omega_1, ..., \omega_n | \{\tau_{ijs}\}, A, \tilde{A}) = 0, \tag{5}$$

for all $i = 1, ..., n$, $A = \{A_s, A_{Fs}\}_{s=1}^{S}$ is a $2S \times 1$ vector of exogenous common (to locations $i = 1, ..., n$) labor productivity, and $\tilde{A} = \{\tilde{A}_{is}\}_{i,s}^{n,S}$ is a $nS \times 1$ vector of idiosyncratic exogenous productivity. For location i we have $A_{is} = A_s \tilde{A}_{is}$. The vector of bilateral "iceberg" trade costs $\{\tau_{ijs}\}_{i,j,s}^{N,N,S}$ is dimension $N^2 S \times 1$ and will be fixed in our analysis: $\tau_{ii,s} = 1$ $\forall i$ and otherwise $\tau_{ij,s} > 1$. Due to Walras' law, stacking (5) we have a system of $N - 1$ equations: $\mathbb{D}(\omega | \tau, A, \tilde{A}) = \mathbf{0}$.

Let $x_{ij,s}$ be the share of j's total imports shipped from i and $\gamma_s > 0$ the share of spending each region allocates to sector s, and W_i the total income of region i. Our excess demand function is equal to excess revenue defined as total revenue in i minus total income in i:

$$\mathbb{D}_i(\omega|\tau, A, \tilde{A}; \theta) = \sum_{j=1}^{N}\sum_{s=1}^{S} x_{ij,s}\gamma_s W_j - W_i$$
$$= \sum_{j=1}^{N}\sum_{s=1}^{S} \frac{\tau_{ij,s}^{1-\sigma}\omega_i^{1-\sigma}A_s^{\sigma-1}\tilde{A}_{is}^{\sigma-1}}{\sum_{o=1}^{n}\tau_{oj,s}^{1-\sigma}\omega_o^{1-\sigma}A_s^{\sigma-1}\tilde{A}_{os}^{\sigma-1} + \tau_{Fj,s}^{1-\sigma}A_{Fs}^{\sigma-1}}\gamma_s \upsilon_j \omega_j^{1+\phi} - \upsilon_i \omega_i^{1+\phi}. \tag{6}$$

The formulation (6) can be motivated from a gravity trade model with a representative firm producing a good in each sector using only labor and individuals with Cobb-Douglas preferences over sectoral composite goods that are produced using a CES aggregator over goods produced in different locations. Consequently, $x_{ij,s} \in (0,1)$ for all (i,j) pairs so long as $\tau_{ij,s} < \infty$. While we can include labor supply shocks by allowing the elements $\{\upsilon_i\}_i$ to be random, we hold them fixed in our analysis. We also abstract from productivity externalities. The parameter vector $\theta = (\sigma, \phi)'$ contains the trade elasticity $\sigma - 1$ where $\sigma > 1$, and $\phi > 1$ is the elasticity of labor supply.

2.3 Sufficient Statistics Approach

It is often the case in spatial models that we seek a representation of the endogenous outcomes of the model in terms of a set of sufficient statistics that capture the general equilibrium effects of changes in fundamentals beyond i. Following this literature, define $CMA_{j,s} \equiv \sum_{o=1}^{N}\tau_{oj,s}^{1-\sigma}w_o^{1-\sigma}A_{os}^{\sigma-1}$ as the consumer market access in location j and sector s which represents its access to cheap products. In equilibrium, we have

$$\upsilon_i \omega_i^{\phi+\phi} = \omega_i^{1-\sigma}\sum_{s}^{S}\gamma_s A_{is}\sum_{j=1}^{N}\tau_{ij,s}^{1-\sigma}CMA_{j,s}^{-1}W_j$$
$$= \omega_i^{1-\sigma}\sum_{s=1}^{S}\gamma_s A_{is} FMA_{i,s},$$

where $FMA_{i,s}$ is firm market access in sector s for goods from i and is increasing in cheap access to large markets with few trade partners. We can then rewrite this equation in terms of the endogenous wage outcome in i as

$$\ln(\omega_i) = \kappa \ln\left(\sum_{s=1}^{S} \gamma_s A_{is} FMA_{i,s}\right) + \epsilon_i, \tag{7}$$

with $\kappa = 1/(\phi + \sigma)$ and $\epsilon_i = -\ln v_i/(\phi + \sigma)$. Alternatively, we are interested in changes in these variables. Let $\hat{\omega}_i \equiv \ln \omega'_i - \ln \omega_i$ denote the change in wage between the initial equilibrium, ω, and the new one, ω' so that we may write

$$\hat{\omega}_i = \kappa \ln\left(\sum_{s}\left(\frac{\gamma_s A'_{is} FMA_{i,s}}{\sum_s \gamma_s A_{is} FMA_{is}}\right)\exp\left(\widehat{FMA}_{is}\right)\right) + \hat{\epsilon}_i,$$

The market access variables are sufficient statistics for interference: conditional on values for $FMA_{i,s}$, the wage for i depend only on fundamentals in i. In addition, as is typically the case in these class of models, they are endogenous: the market access for i depends on ω_i which is a function of the labor supply shock v_i. Researchers may wish to find observable approximations to these variables and find an estimator for κ. However, as seen in (7), the market access variable enters multiplicatively with the unobservable productivity shocks. Unless we assume a single sector, this model will not give rise to an estimable sufficient-statistics representation of outcomes in terms of variables capturing exposure to neighbors. Relatedly, the market access variables themselves depend on the unobservable network $\{\tau_{oj,s}\}$. Furthermore, κ might not represent the estimand of interest: We obtain constant causal effects from changes in the composite market access variable as opposed to an estimand informative for a set of heterogeneous effects. Such a set could be defined by using the sector market access terms $FMA_{i,s}$, although one would need a method that can handle potentially a large set of endogenous variables and nonlinearity of the outcome in those variables.

In the next section, we give a reduced form representation for the wage in i in terms of the observable shocks of interest.

2.4 Reduced-form representation

Let $R^{t_0} = \text{diag}(R_1^{t_0}, ..., R_n^{t_0})$ be the matrix of initial revenues for locations. A first-order log-linearization of (5) around the initial equilibrium implies:

$$
\begin{aligned}
\mathbf{0} = &\underbrace{(R^{t_0})^{-1}}_{n\times n}\underbrace{\nabla_{\log\omega}\mathbb{D}(\omega^{t_0}|\tau, A^{t_0}, \tilde{A}^{t_0};\theta)}_{n\times n}\underbrace{\hat{\omega}}_{n\times 1} \\
&+ (R^{t_0})^{-1}\underbrace{\nabla_{\log\tilde{A}}\mathbb{D}(\omega^{t_0}|\tau, A^{t_0}, \tilde{A}^{t_0};\theta)}_{n\times nS}\underbrace{\hat{\tilde{A}}}_{nS\times 1} \\
&+ (R^{t_0})^{-1}\underbrace{\nabla_{\log A}\mathbb{D}(\omega^{t_0}|\tau, A^{t_0}, \tilde{A}^{t_0};\theta)}_{n\times S}\underbrace{\hat{A}}_{S\times 1} \\
&+ (R^{t_0})^{-1}\underbrace{\nabla_{\log\chi}\mathbb{D}(\omega^{t_0}|\tau, A^{t_0}, \tilde{A}^{t_0};\theta)}_{n\times S}\underbrace{\hat{\chi}}_{S\times 1},
\end{aligned}
$$

where $\hat{\chi} = \{\hat{A}_{Fs}\}_s$ denotes our shocks of interest. In the trade literature, these productivity shocks often correspond to improvements in China's productivity owing to reform-driven transition to an open market economy. For example, Galle, Rodríguez-Clare, and Yi 2023 specify the China trade shock as innovations to China productivity that is consistent with observed changes in U.S. manufacturing imports from China. For now, we ignore the linearization error. Assuming $(R^{t_0})^{-1}\nabla_{\log\omega}\mathbb{D}(\omega^{t_0}|\tau, A^{t_0}, \tilde{A}^{t_0};\theta)$ is nonsingular we have

$$
\hat{\omega}(\hat{\chi}) = \hat{\omega}(0) + g(\omega^{t_0}, \tau, A^{t_0}, \tilde{A}^{t_0};\theta)\hat{\chi}.
$$

This will give a linear-in-shocks representation for equilibrium outcomes. We first verify that the representation is well-defined.

Proposition 2.1. *For $\sigma + \phi > 1$, $\kappa = 1/(\sigma + \phi)$, and $P_{ij} = \rho_{ij}^{t_0} \propto \partial\mathbb{D}_i/\partial \log w_j$, the matrix $I_n - \kappa P$ is nonsingular, hence $B \equiv -(R^{t_0})^{-1}\nabla_{\log\omega}\mathbb{D}(\omega^{t_0}|\tau, A^{t_0}, \tilde{A}^{t_0}, \chi^{t_0};\theta)$ has an inverse, which we denote $\beta \equiv B^{-1}$. Furthermore, β has a series representation: $\beta_{ij} = \frac{1}{\sigma+\phi}\left(\mathbb{1}_{[j=i]} + \sum_{m=1}^{\infty}\kappa^m P_{ij}^{(m)}\right)$ where $P_{ij}^{(m)}$ is the (i,j) element of the matrix P^m. Let $\{a_1, a_2, ..., a_{m-1}\}$ for $m \geq 2$ be indices. Then $\kappa^m P_{ij}^{(m)} = \kappa^m \sum_{a_{m-1}=1}^{n}\sum_{a_{m-2}=1}^{n}\cdots$*

$\cdot \sum_{a_1=1}^{n} \rho_{i,a_{m-1}}^{t_0} \cdots \rho_{a_2,a_1}^{t_0} \rho_{a_1,j}^{t_0}$ *where* $\rho_{i,a_{m-1}}^{t_0} \cdots \rho_{a_2,a_1}^{t_0} \rho_{a_1,j}^{t_0} \propto (\partial \mathbb{D}_i / \partial \log w_{a_{m-1}}) \cdots (\partial \mathbb{D}_{a_2} / \partial \log w_{a_1}) (\partial \mathbb{D}_{a_1} / \partial \log w_j)$ *gives the m-th order effect of changes in wage in j on i.*

Appendix A.1 shows that the term we call the first-order effect, $P_{ij} \equiv \rho_{ij}^{t_0}$, is a sum of a direct interaction between i and j, and a term that represents how i interacts with a region k and how j interacts with k. That is, it includes all direct interactions, but is not limited to only the direct interaction between i and j. The proof of Proposition 2.1 also reveals that $\partial \mathbb{D}_i(\omega^{t_0})/\partial \log \omega_i < 0$ so that $\partial \hat{\omega}_i / \partial \hat{\omega}_j > 0$ for all j. We can write the i-th change in wage as $\hat{\omega}_i = \sum_{j \neq i}^{n} B_{ii}^{-1} B_{ij} \hat{\omega}_j + g_i(A, \tilde{A}, \chi)$, with the magnitude of the sum of the coefficients $\sum_{j \neq i}^{n} |B_{ii}|^{-1} |B_{ij}| < 1$ by Diagonal Dominance. This gives a linear-in-means representations to the structural model. The reduced-form gives the potential outcome representation:

Proposition 2.2. *Under the experiment of assigning S shocks $\{\hat{\chi}_s\}_{s=1}^{S}$ to all units, we can write potential change in wage from the initial equilibrium for location i as*

$$\hat{\omega}_i(\hat{\chi}_1, ..., \hat{\chi}_S) = \hat{\omega}_i(0) + \sum_{s=1}^{S} l_{i,s}^{t_0} \psi_{is} \hat{\chi}_s + \sum_{j \neq i}^{n} \sum_{s=1}^{S} l_{j,s}^{t_0} \psi_{ijs} \hat{\chi}_s \tag{8}$$

where $l_{i,s}^{t_0}$ is the initial labor share that i allocates to sector s, $r_{jk,s}^{t_0}$ is the share of j's sector s revenue that comes from k, $\psi_{ijs} = -\beta_{ij}(\sigma - 1) \sum_{k=1}^{N} r_{jk,s}^{t_0} x_{Fk,s}^{t_0}$ for $j = 1, ..., n$ and β_{ij} is given in Proposition 2.1.

Equation (8) shows us that the causal effect on $\hat{\omega}_i$ of a unit change in χ_s (or equivalently A_{Fs}) from the initial equilibrium is $l_{i,s}^{t_0} \psi_{is} + \sum_{j \neq i}^{n} l_{j,s}^{t_0} \psi_{ijs}$ which, importantly, depends on general equilibrium interactions ψ_{ijs}, which is a function of β_{ij} and the network τ as the two combine to determine the strength of spillovers. The term $\zeta_{js} \equiv -(\sigma - 1) \sum_{k=1}^{N} r_{jk,s}^{t_0} x_{Fk,s}^{t_0} \hat{\chi}_s$ measures how a negative import shock to j affects i: it is the total shock to j from increased sector s imports from the foreign region F by its trading partners, with spillovers to i occurring through β_{ij}. If we assume that this import shock is independent of location (i.e. same for all j) then we get a potential outcomes model that is linear in exposures as in Equation 1: $\sum_{j \neq i} \sum_s l_{j,s}^{t_0} \psi_{ijs} \hat{\chi}_s = \sum_{j \neq i} \beta_{ij} \sum_s l_{js}^{t_0} \zeta_s$. This is a similar assumption made in Adao, Arkolakis, and Esposito 2019. Nonetheless, our problem does not satisfy the SUTVA assumption in standard causal inference settings. If we further assume that β_{ii} is constant across i, then $\psi_i = \psi$ then the second term of (8) is ψX_i. Our potential outcome model is also a natural extension of the one found in Adao, Kolesár, and Morales 2019 to the case with interference. Unlike in that paper, we do not remove bilateral costs $\tau = \{\tau_{ijs}\}$ because of the importance of geography in explaining the spatial distribution of economic activity as studied in, for example, Allen and Arkolakis 2014. Even in the absence of τ, $\zeta_{js} = -(\sigma - 1)\beta_{ij} x^{t_0}_{F,s} \hat{\chi}_s$ where $x^{t_0}_{F,s}$ is the common share of imports from F. We obtain the same expression for interference, $\sum_{j \neq i} \beta_{ij} \sum_s l^{t_0}_{js} \zeta_s$ so that the potential outcomes model of Adao, Kolesár, and Morales 2019 would be misspecified when we use productivity shocks.

3 Linear Regression

We now return to our empirical model that specifies a linear relationship for interference. We first restate the true outcome model for location i, corresponding to (2), given the random vector $(\mathcal{X}_1, ..., \mathcal{X}_S)$:

$$Y_i = Y_i(\mathcal{X}_1, ..., \mathcal{X}_S) = Y_i(0) + \sum_{s=1}^{S} w_{is}\psi_{is}\mathcal{X}_s + \sum_{j \neq i}^{n} \sum_{s=1}^{S} w_{js}\psi_{ijs}\mathcal{X}_s. \tag{9}$$

Remark 3.1. *In a standard setting with a binary treatment $D_i \in \{0, 1\}$, the outcome variable is a linear function of the treatment because $Y_i = Y_i(D_i) = Y_i(0) + D_i(Y_i(1) - Y_i(0)) = Y_i(0) + D_i\tau_i$, where τ_i captures heterogenous treatment effects. With interference and, moreover, with continuous treatment, it need not be true that the potential outcome for i is linear in the treatment variables. We avoid ad-hoc linear potential outcomes through the motivating structural model of Section 2.*

Let $Z_{n,i} \equiv \sum_{j \neq i}^{n} G_{ij} X_j$. The regression model is $Y_i = \beta_1 X_i + \beta_2 Z_{n,i} + \varepsilon_i$. The shocks satisfy a random-assignment assumption $E[\mathcal{X}_s | \mathcal{F}] = 0, \; s = 1, ..., S$ for some conditioning set $\mathcal{F}$ that describes the characteristics of a population containing the n locations in our sample. Given the form of this specification and the structural model of Section 2, we take the population of interest to be the n units in the sample and define our conditioning set

$$\mathcal{F} \equiv (\{Y_i(0)\}_{i=1}^{n}, \{w_{is}\}_{i=1,s=1}^{n,S}, \{\psi_{ijs}\}_{i=1,j=1,s=1}^{n,n,S}),$$

where we keep the dependence of this set on n implicit. The gravity trade model of Section 2 describes interactions between n units and our outcome variable $Y_i(X_1, ..., X_n)$ is realized after units are drawn into the sample. Thus, this problem is inherently a finite-population model in that we remove sampling uncertainty of locations. A similar point is made in Manski 1993. One way to introduce sampling uncertainty of locations, where the sample is drawn from a larger superpopulation, is to introduce dependence of Y_i on population objects such as $E[X_i|R_i]$ where R_i denotes the reference group for i. Such a model might be motivated from a structural model with a continuum of locations. Instead, the model of Section 2 describes interactions between n units in S sectors with an $n \times n$ matrix characterizing spatial links. This also means that our estimands are indexed by (n, S), so $\beta_1 = \beta_1^{(n,S)}$ and $\delta_1 = \delta_1^{(n,S)}$. However, we may still wish to introduce uncertainty through resampling of other shocks aggregated in $Y_i(0)$ or through resampling of locations, and hence the weight matrix $\mathcal{W}$. Each would be isomorphic to sampling from a superpopulation with an outcome model following (4) because the set $\mathcal{F}$ gives a description of locations in our sample, and introducing randomness to its elements changes the description of units in the sample. Nonetheless, our estimands will continue to be indexed by (n, S). Introducing uncertainty over locations (given n) gives rise to uncertainty over general equilibrium effects $\{\psi_{ijs}\}$. Conditioning on general equilibrium effects through ψ_{ijs} allows us to speak to papers such as Acemoglu et al. 2016 that choose pre-China shock input-output spatial links that are unlikely to be endogenous to the subsequent shock.

Assumption 3.1. *The shocks satisfy (i)* $E[\mathcal{X}_s|\mathcal{F}] = 0$ *(ii)* $Cov(\mathcal{X}_s, \mathcal{X}_t|\mathcal{F}) = 0$ *for all pairs* (s, t) *(iii)* $\exists (i, s)$ *such that* $\forall (c_1, c_2) \in \mathbb{R}^2$ *with* $(c_1, c_2) \neq 0$, $c_1 w_{is} - c_2 \sum_{j\neq i}^{n} G_{ij} w_{js} \neq$

0.

Assumption 3.1(i) allows us to avoid the use of an intercept, which is a normalization, and avoid controls. It says that the sector shocks are randomly assigned in the sense of being mean-independent from other unobservable shocks in $Y(0)$ as well as the shares $\mathscr{W}$ and general equilibrium interactions. It is a weaker requiremet than the familiar $Y_i(x_1, ..., x_S) \perp \mathcal{X}_s$ that also rules out selection. Let $\ddot{X}_{ni} \equiv X_i - \delta Z_{ni}$ the residual from the population projection of X on the linear space $\{\tilde{\delta} Z_n : \tilde{\delta} \in \mathbb{R}\}$ and similarly $\ddot{Z}_{ni} \equiv Z_{ni} - \gamma X_i$. The third assumption ensures the estimands are well-defined along any sequence of finite populations; we need non-zero variation in the variables $\ddot{X}_n$ and $\ddot{Z}_n$, i.e. linear independence in the variables $\ddot{X}_n$ and $\ddot{Z}_n$, or no multicolinearity.
Our projection coefficients are defined to be the best linear predictors, in the sense of solving the following problem

$$(\beta_1, \beta_2)' := \arg\min_b \mathbb{E}\left[n^{-1} \sum_{i=1}^{n} (Y_i - b_1 X_i - b_2 Z_{ni})^2 \middle| \mathcal{F} \right]. \tag{10}$$

The minimization (10) arises from a projection of the n vector Y onto the subspace spanned by X and Z_n and is consistent with the projection estimands (β_1, β_2) being constant across i. The following proposition shows that these estimands are linear combinations of location-sector partial effects, or of location partial effects.

Proposition 3.1. *Let $Y_i(x)$ be the potential outcome of unit i at value $x = (x_1, ..., x_S)$ from (9) with $\partial Y_i(x)/\partial x_s \equiv w_{is}\psi_{is} + \sum_{j \neq i}^{n} w_{js}\psi_{ijs}$ the partial effect of the sector s shock, $\mathcal{X}_s$. Under Assumption 3.1, the projection coefficient β_1 in the linear regression model (4) has the form*

$$\beta_1 = \sum_{i=1}^{n} \sum_{s=1}^{S} \lambda_{is}^{(n)} \frac{\partial Y_i(x)}{\partial x_s}, \tag{11}$$

where $\lambda_{is}^{(n)} = E[\ddot{X}_{ni}\mathcal{X}_s|\mathcal{F}] / \sum_{i=1}^{n} E[\ddot{X}_{ni}^2|\mathcal{F}]$.

Alternatively, the projection coefficient can be expressed in the form

$$\beta_1 = \mathbb{E}\left[\sum_{i=1}^{n} \lambda_i^{(n)} \left(Y_i(\mathcal{X}_1, ..., \mathcal{X}_S) - Y_i(0)\right) \middle| \mathcal{F}\right], \tag{12}$$

where $\lambda_i^{(n)} = \ddot{X}_{ni} / \sum_{i=1}^n E[\ddot{X}_{ni}^2|\mathcal{F}]$ *is mean zero. Therefore,* β_1 *is the expected value of all linear combinations of treatment effects* $Y_i(\mathcal{X}_1, ..., \mathcal{X}_S) - Y_i(0)$ *that occur with positive probability, where the weights of the linear combination are mean zero.*

The implication for β_1 in Proposition 3.1 is a generalization of Proposition 1 in Adao, Kolesár, and Morales 2019 who find that, in general, β_1 will not be a convex combination of heterogeneous treatment effects. The focus on an estimand arising from unit shifts in $\{\mathcal{X}_s\}_s$ seems a natural choice of estimand because it implies a change in exposure of $\sum_s w_{is} \leq 1$. When the latter holds with equality for all i, we can treat such an estimand as the effect of a unit shift in the value of the exposure vector X in the outcome equation (4). The proof of the first part of Proposition 3.1 in Appendix A.2 shows that if the weights sum to 1, then we must necessarily have both negative and positive weights. It is also possible, however, that the weights are all negative or all positive, or a combination of both. In the absence of constant treatment effects, it is unclear how β can be informative for the set $\{\partial Y_i(x)/\partial x_s\}_{i,s}$. While we are unable to obtain a convex combination of heterogenous treatment effects with misspecified empirical specifications, it should be noted that models that admit observable sufficient statistics assume constant effects; the problem of obtaining a convex combination of heterogeneous effects persists.
A more appropriate term for β would be, what the interference literature calls, *exposure effect*[4]. However, our mapping $\tilde{T}_i = \tilde{T}(i, G, \{\mathcal{X}_s\}_{s=1}^S) = \sum_{j\neq i} G_{ij} \sum_s w_{js}\mathcal{X}_s$ is a function of the vector of common shocks instead of an n-vector of treatment assignments, and doesn't contrast conditional averages of the outcome. In that literature, it has been shown[5] that, with misspecified exposure mappings, the exposure effect is not necessarily a causal estimand.

[4]Which unfortunately conflicts with our terminology in this paper.
[5]For example, in the paper Leung 2024

4 Asymptotic Properties

We now discuss large sample properties of the OLS estimator of $\beta = (\beta_1, \beta_2)$. The first observation is that we have spatial dependence in the observations in the sense of dependence between the elements of the vector $(Y_1, ..., Y_n)$. One might suspect that this dependence in the outcome is due to locations receiving the same common shock, but it remains the case that if we analyze the behavior of $Y_i|(X = x)$ where $X = (X_1, ..., X_n)$, then locations will exhibit spatial dependence. Indeed, if we condition on the set $\{(\mathcal{X}_1, ..., \mathcal{X}_S) : X = x\}$, the conditional projection error in (4) is defined as

$$\begin{aligned}\varepsilon_i|(\mathcal{F}, X = x) &= Y_i - \beta_1 x_i - \beta_2 \sum_{j \neq i}^{n} G_{ij} x_j \\ &= Y_i(0) + \sum_{s=1}^{S} w_{is} \underbrace{(\psi_{is} - \beta_1)}_{\text{T.E. het. misspec.}} \mathcal{X}_s^{(x)} + \sum_{j \neq i}^{n} \sum_{s=1}^{S} w_{js} \underbrace{(\psi_{ijs} - \beta_2 G_{ij})}_{\text{G.E. effect misspec.}} \mathcal{X}_s^{(x)}\end{aligned} \tag{13}$$

Similar to Adao, Kolesár, and Morales 2019 the statistical properties of the regression residual ε_i depend on potential outcomes $Y_i(0)$, the shifters $\{\mathcal{X}_s\}_s$, and the shares $\{w_{is}\}_s$ and parameters $\{\psi_{is}\}_s$. However, when allowing for general equilibrium effects, we now have dependence on parameters $\{\psi_{ijs}\}_{js}$ and all neighbor shares $\{w_{js}\}_{j,s}^{n,S}$. Whereas in Adao, Kolesár, and Morales 2019 two locations with similar exposure could have correlated residuals, in our setting we have a richer correlation structure due to the third term. Even if $G_{ij} = 0$ for a given j, there could still be correlation between ε_i and ε_j due to $\{\psi_{ijs}\}_s$ which reflects general equilibrium interactions[6]. We can decompose the error into three terms: (i) the contribution of other shocks through $Y_i(0)$ (ii) the contribution due to misspecification of treatment effect heterogeneity, $\psi_{is} - \beta_1$ (iii) the contribution due to mis-

[6] In a social interactions model with only exogenous effects, (Y_i, X_i) can be treated as independent from (Y_j, X_j) when $G_{ij} = 0$. This isn't the case in our setting because of general equilibrium interactions.

specification of general equilibrium effects, $\psi_{ijs} - \beta_2 G_{ij}$. The first two were shown in Adao, Kolesár, and Morales 2019 – although the misspecification of treatment effect heterogeneity is more serious here due to higher order equilibrium effects. Suppose we assumed unobservable treatment effect heterogeneity, but with a direct treatment effect that is constant conditional on the observable weights. Then $\varepsilon_i|(X = x) = Y_i(0) + \alpha x_i + \sum_{j\neq i} w_{js}(\psi_{ijs} - \beta_2 G_{ij})\mathcal{X}_s^{(x)}$ where $\alpha = \psi - \beta_1$. It also follows that in general, $E[\varepsilon_i|\mathcal{F}, X] \neq 0$, or even unconditionally $E[\varepsilon_i|X] \neq 0$. Recall that β_{ij} is the sum of m-th order effects $\kappa^m P_{ij}^{(m)}$, which in turn are the sums of the products $\rho^{t_0}_{i,a_{m-1}} \cdots \rho^{t_0}_{a_2,a_1}\rho^{t_0}_{a_1,j} \propto \left(\partial\mathbb{D}_i/\partial \log w_{a_{m-1}}\right)\cdots\left(\partial\mathbb{D}_{a_2}/\partial \log w_{a_1}\right)\left(\partial\mathbb{D}_{a_1}/\partial \log w_j\right)$ for indices $\{a_1, a_2, ..., a_{m-1}\}$ and $m \geq 2$. Limiting misspecification in the error ε_i amounts to limiting the growth rate of this sum across m. Furthermore, locations with similar shares and similar GE effects will form clusters, although clustering on this unobserved heterogeneity is infeasible.

We strengthen our assumption on the dependence structure of the common shocks and consider an asymptotic framework with $S \to \infty$. Using diverging sector sizes as the index for our asymptotic framework has been used in the shift-share literature by Borusyak, Hull, and Jaravel 2022 and Adao, Kolesár, and Morales 2019 to deal with non-iid variables that are dependent across locations. In our setting, the extent of dependence is greater due to equilibrium effects creating essentially one cluster. Using large-S asymptotics with independence across sectors will allow us to keep network dependencies between locations, while allowing for a tractable framework due to forming independent groups using sector labels. First, let $n_s \equiv \sum_{i=1}^n w_{is}$ be the size of sector s in the population, and let $\mathcal{D}$ denote the data.

Assumption 4.1. *(Restrictions on $\mathcal{D}$)*
(i) Conditional on the set $\mathcal{F}$, the common shocks $\{\mathcal{X}_{Ss}\}_{s=1}^S$ form an independent triangular array (ii) For some $\eta \in (0,1]$, $E[|\mathcal{X}_{Ss}^2|^{1+\eta}|\mathcal{F}]$ is uniformly bounded across S (iii) $\limsup_{S\to\infty}\max_{\leq s\leq S} n_s < \infty$ *(iv)* $\liminf_{S\to\infty}\min_{s\leq S} n_s > 0$ *(v)* $n^{-1}\mathbb{E}[\sum_i^n Z_{ni}^2|\mathcal{F}] \to Q_Z > 0$ *(vi)* $n^{-1}\mathbb{E}[\sum_{i=1}^n X_i^2|\mathcal{F}] \to Q_X > 0$
(vii) $\sum_{s=1}^S \left(\sum_{i=1}^n \sum_{j\neq i}^n G_{ij} w_{js}\right)^{1+\eta} = o\left(n^{1+\eta}\right)$ *for* $\eta \in (0,1]$.

Assumption 4.1(ii) will be satisfied in settings where we think of the shocks as productivity shocks in locations with a regular geography. Assumption 4.1(iii) has the implication that $n \to \infty$. Indeed, we know $\sum_{s=1}^{S} n_s \leq n$ which implies $\max_s n_s / \sum_{t=1}^{S} n_t \geq \max_s n_s/n$. Suppose n is fixed as $S \to \infty$. Under the assumption, n_s is arbitrarily small for all $s \leq S$, which is a degenerate problem because the regressors X_i, Z_{ni} will be arbitrarily small for all i, which threatens identification of the estimands (Assumption 3.1(iii)) in the limit. So we in fact have $n = n(S)$. Assumptions 4.1(iii) and (iv) together imply that n and S are of the the same order when $\sum_s n_s$ and n are of the same order, which we assume because $\sum_s n_s = n$ whenever $\sum_s w_{is} = 1$. Assumption 4.1(iii) implies that the maximum sector is eventually bounded: there exists $\bar{S}$ such that for all $S \geq \bar{S}$, $\max_{s \leq S} n_s < \bar{C}$ for some finite $\bar{C}$. Together with (iv), there are finite constants $\underline{C}, \bar{C}$ such that all sectors are eventually bounded by these constants. Part (vii) imposes a restriction on the regressor Z_n, and in particular, the moments of $Z_n' Z_n$ in terms of its rate of growth. For example, when $\eta = 0$ so that we only require two moments for the shocks, this assumption states that $\sum_{s=1}^{S} \sum_{i=1}^{n} \sum_{j \neq i}^{n} G_{ij} w_{js} = o(n)$ which implies $\mathbb{E}[n^{-1} Z_n' Z_n | \mathcal{F}] \to 0$ and $n^{-1} Z_n' Z_n \to^p 0$. With $\eta = 1$, which assumes four moments for the regressor Z_n, we allow for the expected value to be order n, but have $\mathbb{E}[(n^{-1} Z_n' Z_n)^2 | \mathcal{F}] \to 0$. The case with $\eta > 0$ falls in between, so is weaker than the case with $\eta = 0$ but stronger than $\eta = 1$. Nonetheless, we may think of bounds absent our assumption: using the c_r-inequality, $\left(n^{-1} \sum_{i=1}^{n} \sum_{j \neq i}^{n} G_{ij} w_{js}\right)^{1+\eta} \leq n^{\eta} \sum_{i=1}^{n} \left(n^{-1} \sum_{j \neq i}^{n} G_{ij} w_{js}\right)^{1+\eta} = n^{\eta} \sum_{i=1}^{n} \left((n_s/n) \sum_{j=1}^{n} G_{ij} w_{js}/n_s\right)^{1+\eta}$. Using Jensen's inequality for convex functions, the latter term is less than or equal to $n^{\eta} \sum_{i=1}^{n} \sum_{j=1}^{n} (w_{js}/n_s) \left(G_{ij} n_s/n\right)^{1+\eta} = (n_s^{\eta}/n) \sum_{i=1}^{n} \sum_{j=1}^{n} w_{js} G_{ij}^{1+\eta}$. Using $\sum_s w_{is} \leq 1$ for each i we have $\sum_s \left(n^{-1} \sum_{i=1}^{n} \sum_{j \neq i}^{n} G_{ij} w_{js}\right)^{1+\eta} \leq (\max_s n_s) n^{-1} \sum_{i=1}^{n} \sum_{j=i}^{n} G_{ij}^{1+\eta} \leq \max_s n_s$. We conclude that $\sum_s \left(n^{-1} \sum_{i=1}^{n} \sum_{j \neq i}^{n} G_{ij} w_{js}\right)^{1+\eta}$ is bounded above by a constant, and possibly diverges depending on the assumption we make on $\max_s n_s$; Part (iv) bounds the maximum sector size.

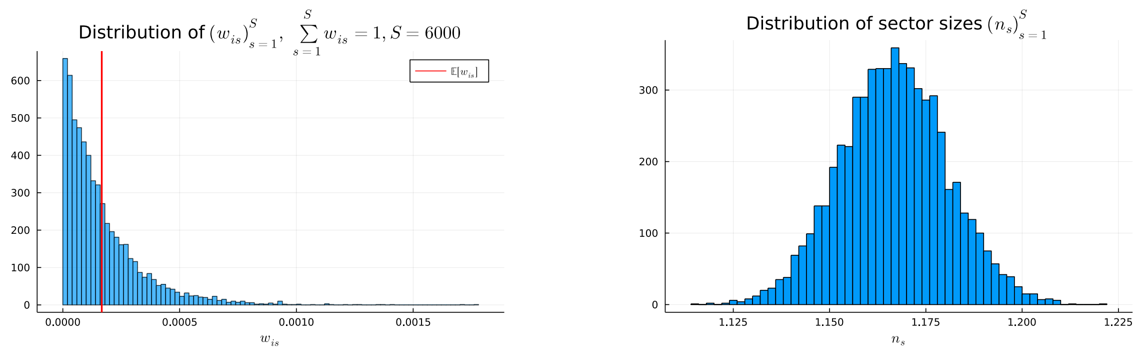
Example 4.1. *Restrictions on Z_n are restrictions on the interactions between network dependence and sector weights. Suppose we allow for any weight matrix $\mathbb{W}$. Then one example in which Part (viii) is satisfied is when all interactions happen with neighbors to the left: $G_{i,j} = \mathbb{1}[j = i-1]$ and $G_{1,n} = 1$, which is a sparse network. Then $Y_i = \beta_1 X_i + \beta_2 X_{i-1} + \varepsilon_i$, and*
$\sum_s \left(n^{-1}\sum_{i=1}^n \sum_{j\neq i}^n G_{ij} w_{js}\right)^{1+\eta} = \sum_s (n^{-1} n_s)^{1+\eta} \leq (n^{-1}\sum_s n_s)(\max_s n_s/n)^\eta \to 0.$

Example 4.2. *Let $S < n$ and suppose that for $i = 1, ..., S$ $w_{is} = \mathbb{1}[s = i]$, i.e. the first S locations specialize in a single sector, which corresponds to a sparse weight matrix $\mathbb{W}$. We impose no restriction on the remaining $n - S$ locations. Then*

$$\begin{aligned}\sum_{s=1}^S \left(n^{-1}\sum_{i=1}^n \sum_{j\neq i}^n G_{ij} w_{js}\right)^{1+\eta} &= \sum_{s=1}^S \left(n^{-1}\sum_{i\neq s}^n G_{is} + n^{-1}\sum_{i=1}^n \sum_{j=S+1}^n G_{ij} w_{js}\right)^{1+\eta} \\ &\leq 2\sum_{s=1}^S \left(n^{-1}\sum_{i\neq s}^n G_{is}\right)^{1+\eta} + 2\sum_{s=1}^S \left(n^{-1}\sum_{i=1}^n \sum_{j=S+1}^n G_{ij} w_{js}\right)^{1+\eta} \\ &\leq 2n^{-1}\sum_{i=1}^n \sum_{s=1}^S G_{is}(\max_{i\leq n} G_{is})^\eta + 2n^{-1}\sum_{i=1}^n \sum_{s=1}^S \left(\sum_{j=S+1}^n G_{ij} w_{js}\right)^{1+\eta}.\end{aligned}$$

Suppose that $G_{ij} = 1/(n-1)\mathbb{1}[j \neq i]$ for all $j \neq i$ so that there is uniformity in the network and location i's outcome Y_i depends on average exposure: $Y_i = \beta_1 X_i + \beta_2 \frac{1}{n-1}\sum_{j\neq i}^n X_j + \varepsilon_i$. Then both terms converge to zero and Assumption 4.1(viii) would be satisfied.

Example 4.3. *Suppose $(w_{is})_{s=1}^S \sim Dir(1/S, ..., 1/S)$ and $Y_i = \beta_1 X_i + \beta_2 \frac{1}{n-1}\sum_{j\neq i}^n X_j + \varepsilon_i$. We draw the vector of sector weights i.i.d across locations. Figure 1 shows the distribution of these weights for a given i, as well as a realization of sector sizes from its distribution. In this example, weights are not degenerate at 0 or 1, but contain sufficient sparsity to satisfy our assumption given a uniform interactions matrix G.*



(a) Distribution of sector weights

(b) Distribution of sector sizes

Figure 1: Example satisfying sparsity of GW

4.1 Consistency

Having defined our estimands in the previous section, we turn to the question of what restrictions on the model $\mathcal{M}$ we can impose in order for our estimators to perform well, in the sense of being 'correct' in large samples. Recall that unlike in standard linear models where the estimands β_1, β_2 are fixed constants, our problem is one in which they are indexed by (n, S) (or more precisely, by S). While this changes how we define consistency – it is now defined as $\hat{\beta} - \beta^{(S)} \to^p 0$ instead of $\hat{\beta} \to^p \beta$ – the intuition remains the same: when (n, S) is large, our estimand $\beta^{(n,S)}$ characterizes the parameter of interest that is relevant to the population of n locations and S sectors, and our estimator $\hat{\beta}$ will be close to this value in a probabilistic sense.

Let $W_i = (X_i, Z_{ni})$ so that $\hat{\beta} - \beta = (W'W)^{-1}W'\varepsilon$.

Assumption 4.2. *(Restrictions on $\mathcal{M}$)*
(i) $|Y_i(0)| \leq C_y$ for some $C_y > 0$ and for all i (ii) For all $s \leq S$, and all i, $|\psi_{is} - \beta_1^{(S)}| \leq C_{\psi,1}$ for some $C_{\psi,1} > 0$ (iii) $|\psi_{ijs} - \beta_2^{(S)} G_{ij}| < C_{\psi,2}$ for some $C_{\psi,2} > 0$ (iv) $\sum_{i=1}^{n} \sum_{j \neq i}^{n} \max_{s \leq S} |\psi_{ijs} - \beta_2^{(S)} G_{ij}| = o(n)$ (v) $|\beta_1^{(S)}|, |\beta_2^{(S)}| \leq C_\beta$ for some $C_\beta > 0$.

In contrast to Assumption 4.1 which imposes restrictions on the data, Assumption 4.2 are assumptions on the primitives of the structural model $\mathcal{M}$. These are strong assumptions that are needed for consistency when the only sources of randomness are the common shocks $\{\mathcal{X}_s\}_s$. Part (i) imposes bounded potential outcome $Y_i(0)$ and we could relax this to uniformly bounded first and second moments if we allow for a setting with additional shocks. Part (ii) bounds treatment effect misspecification; The same assumption is made in Adao, Kolesár, and Morales 2019. It allows for non-negligible total misspecification in the limit in the sense of $\liminf_{S\to\infty}\sum_{s=1}^{S}\sum_{i=1}^{n}|\psi_{is}-\beta_1^{(S)}|>0$. Part (iii) and (iv) restricts G.E effects. Part (iii) assumes bounded support while (iv) restricts its rate of growth. In particular, it implies that average misspecification error should be zero, and that in each location, the minimum pairwise G.E. effect is captured by first-order observables $\{G_{ij}\}_{j\neq i}$. To see this, we can show that given (i,s), and for any $\epsilon>0$, $a_{Ss}^{(i)}=\min_{j\neq i}|\psi_{ijs}-\beta_2^{(S)}G_{ij}|$ is such that $0<a_{Ss}^{(i)}<\epsilon$ for all $i\leq n(S)$ for S sufficiently large. Therefore, $\lim_{S\to\infty}a_{Ss}^{(i)}=0$ exists for each (i,s) pair; that is, the minimum pairwise G.E. effect is captured by the observable interaction matrix. Nonetheless, we still allow for non-neglible G.E. effect misspecification in the limit with the caveat that they do not grow too fast in the sense of the total misspecification error in each sector increasing at a slower rate than n. While the misspecification error must diverge slower than the sample size, in terms of restrictions on the primitives of the model $\mathcal{M}(\theta)$, we can allow for total G.E. effects to increase at the rate of the sample: $\sum_{i=1}^{n}\sum_{j\neq i}^{n}|\psi_{ijs}|=O(n)$. Papers that have used regression (4) acknowledge the interference channel, but cluster locations into groups which has the benefit of allowing for arbitrary equilibrium interactions within clusters, but at the cost of needing to assume zero effects outside the cluster. Our assumptions instead allow for more general interactions among locations, but requires restricting their magnitudes and rate of growth.

Proposition 4.1. *Under Assumption 4.1, Assumption 4.2 and Proposition A.1, (i)* $\hat{\beta}-\beta^{(S)}=o_p(1)$*, with* $\beta_1^{(S)}=\sum_{i=1}^{n}\sum_{s=1}^{S}\lambda_{is}\frac{\partial Y_i(x)}{\partial x_s}$ *and* $\beta_2^{(S)}=\sum_{i=1}^{n}\sum_{s=1}^{S}\tilde{\lambda}_{is}\frac{\partial Y_i(x)}{\partial x_s}$ *and weights given in Proposition 3.1.*

Recall from Section 2 that $\psi_{ijs}=-\beta_{ij}(\sigma-1)\sum_{k=1}^{N}r_{jk,s}^{t_0}x_{Fk,s}^{t_0}$ for $j=1,...,n$ where β_{ij} captures the G.E. effects and is the sum of all higher-order effects between i and j in the sense of how changes in wages in j result in changes in excess demand in i. For $|\psi_{ijs} - \beta_2 G_{ij}| \to 0$, we would need $|\beta_{ij}|$ to be such that the sum of all effects $\sum_{m=2}^{\infty} \kappa^m P_{ij}^{(m)} \to 0$, with $P_{ij}^{(m)}$ the transmission using $m-1$ neighbors. Since $|\psi_{ijs}| \leq (\sigma - 1)|\beta_{ij}|$, we can state a sufficient condition as: $n^{-1} \sum_{i=1}^{n} \sum_{j\neq}^{n} |\beta_{ij}| \leq C^*$. The terms β_{ij} is a function of $\sum_{m=1}^{\infty} \kappa^m P_{ij}^{(m)}$ where $0 < \kappa < 1$ and $\kappa^m P_{ij}^{(m)} = \kappa^m \sum_{a_{m-1}=1}^{n} \sum_{a_{m-2}=1}^{n} \cdots \sum_{a_1=1}^{n} \rho_{i,a_{m-1}}^{t_0} \cdots \rho_{a_2,a_1}^{t_0} \rho_{a_1,j}^{t_0}$ and $\sum_{j=1}^{n} \rho_{ij}^{t_0} < \kappa^{-1}$ for all n, so it must be that as $n \to \infty$, $\rho_{ij}^{t_0} \to 0$. Suppose $\rho_{ij}^{t_0} \leq C_\rho$ uniformly in (i,j,n). Then $|\beta_{ij}| \leq \kappa C_\rho + \kappa \lim_{R\to\infty} \sum_{m=2}^{R} C_\rho^{m-1}$. If $C_\rho \in (0,1)$, the latter geometric series converges, say to $\tilde{C}_\rho$, and $|\beta_{ij}| \leq 2\kappa \tilde{C}_\rho$ uniformly in (i,j,n). To think about when this could occur, suppose location i was specialized in sector 1, $w_{i1} = 1$ and suppose that each other location was equally important in its initial revenue in sector 1. Let $N = n+1$. We can show that $\rho_{ij}^{t_0} = \frac{(1+\phi)}{N} + \frac{(\sigma-1)}{N} \sum_{k=1}^{N} x_{jk,1}^{t_0}$ where $\sum_{k=1}^{N} x_{jk,1}^{t_0}$ measures j's importance in sector 1, with $x_{jk,1}^{t_0} \in (0,1)$. The first term will converge to zero for large N, and the second term will be less than 1 when $\sum_{k=1}^{N} x_{jk,1}^{t_0} < \frac{1}{\sigma-1} N$.

4.2 Inference

We now turn to the problem of developing a distribution theory for the OLS estimators of β_1, β_2 in order to do hypothesis testing and construct confidence intervals with asymptotically correct coverage that accounts for equilibrium interactions between locations. Our inference methods will be valid conditionally, and therefore, also valid unconditionally in the sense of being valid for any joint distribution of sample characteristics $\{Y_i(0), (w_{is})_{s=1}^{S}, (\psi_{ijs})_{j=1,s=1}^{n,S}\}_{i=1}^{n}$. However, the price paid will be stronger conditions than is likely necessary, and one might for example be able to relax the restriction on the rate of growth of the misspecification errors by adopting an unconditional approach, although such an approach will need to be motivated by structural arguments that explain what the underlying causes are that give rise to randomness in equilibrium interactions. We begin with the following assumptions:

Assumption 4.3. *(Restrictions on $\mathcal{D}$) (i) For some $\eta > 0$ $\mathbb{E}[|\mathcal{X}_{Ss}|^{4+\eta}|\mathcal{F}]$ is uniformly* *bounded across* (S, s) *(ii)* $\sum_{s=1}^{S}\left(\sum_{i=1}^{n}\sum_{j\neq i}^{n} G_{ij}w_{js}\right)^2 = O(n)$ *(iii) For some* $\eta > 0$, $\sum_{s=1}^{S}\left(\sum_{i=1}^{n}\sum_{j\neq i}^{n} G_{ij}w_{js}\right)^{2+\eta} = o(n^{(2+\eta)/2})$.

Assumption 4.4. *(Restrictions on* $\mathcal{M}$*) (i)* $\sum_{i=1}^{n}\sum_{j\neq i}^{n} \max_{s\leq S}|\psi_{ijs} - \beta_2^{(S)}G_{ij}| = o(\sqrt{n})$

Assumptions 4.3 and 4.4 and strengthen requirements on the data and on the model $\mathcal{M}$ from those previously. Part (4.3i) assumes moments greater than four for the common sector shocks, which is similar to what one would find in linear regression models with fixed regressors. Part (4.4i) requires a slower rate of divergence for the total misspecification error, although still allowing for valid inference with non-negligible misspecification. The rate of growth allowed for the primitive equilibrium effects remains the same, $\sum_{i=1}^{n}\sum_{j\neq i}^{n}|\psi_{ijs}| = O(n)$ for each s, but now these effects are better approximated by the observable interactions matrix. Interestingly, we do not require any additional restrictions on treatment effect heterogeneity misspecification, and it can remain fixed as we move along the sequence of finite population models. Part (4.3ii) and (4.3iii) restrict features of the regressor Z_n. Indeed, (iii) strengthens Assumption 4.1(viii) for the case $\eta = 1$, which previously only required this term be $o(n^2)$. We require order n in order to scale $\hat{\beta}-\beta$ by a common factor $r_S = \sqrt{n(S)}$ and obtain a variance-stabilizing transformation. Notice that by definition $n^{-1}\sum_{i=1}^{n}\sum_{j\neq i}^{n} G_{ij}w_{js} \leq 1$ and we previously assumed this term converges to zero, we now have $n^{-1/2}\sum_{i=1}^{n}\sum_{j\neq i}^{n} G_{ij}w_{js} \to 0$. Such a condition may arise from a sparser network structure. Part (iv) is analogous to Assumption 4.1(viii) but for powers beyond 2, and now with a stricter requirement that the terms diverge slower than $n^{(2+\eta)/2}$ instead of $n^{2+\eta}$. Together with Part (ii), it is analogous to those of the previous section in which $\eta = 0$ was the knife-edge case and of order n; here $\eta = 0$ corresponds to the second power.

Theorem 4.1. *Let* S *denote the number of sectors and* $n = n(S)$ *the number of locations in our sample. Let* $W_{n,i} = (X_i, \sum_{j\neq i}^{n} G_{ij}X_j)'$. *Under Assumptions 4.1 and Assumptions 4.3, 4.4, and Proposition A.2,* $\sqrt{n}\mathcal{V}_n^{-1/2}(W_n'W_n/n)(\hat{\beta} - \beta^{(S)}) = N(0, I) + o_p(1)$, *where* $\mathcal{V}_n := var\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_{n,i}\varepsilon_i | \mathcal{F}\right)$.

Remark 4.1. *(Normalizing sequence) Recall that in our asymptotic framework,* $n = n(S)$*. Under our assumptions, we could scale our sequence using the total number of sectors,* $\sqrt{S}$*, the sum of squared sector sizes,* $\sqrt{\sum_{s=1}^S n_s^2}$*, or* $\sqrt{\sum_{s=1}^S \left(\sum_{i=1}^n \sum_{j\neq i}^n G_{ij} w_{js}\right)^2}$*. The performance of these different scaling sequences in terms of finite sample coverage depend on the specific case, but each will be asymptotically valid.*

Theorem 4.1 establishes joint convergence of the least squares estimator, so that $(W'W/n)^{-1}\hat{\mathcal{V}}_n(W'W/n)^{-1}$ is an estimator of the asymptotic variance of $\sqrt{n}(\hat{\beta} - \beta)$ where $\hat{\mathcal{V}}_n$ satisfies $\hat{\mathcal{V}}_n - \mathcal{V}_n \geq o_p(1)$, which allows for potentially conservative inference. Since ϵ is a function of primitives, we use OLS residuals, $\hat{\epsilon}$. Observe that $\mathcal{V}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n W_{n,i}\varepsilon_i = \frac{1}{\sqrt{n}} \sum_{s=1}^S R_s \mathcal{X}_s$ where $R_s := \sum_{i=1}^n W_{ni}\varepsilon_i$ is a 2×1 random vector. Since $\varepsilon_i = \varepsilon_i(\mathcal{X}_1, ..., \mathcal{X}_S)$, the random vectors $\{R_s\mathcal{X}_s\}_{s=1}^S$ are dependent. So we need an estimator for

$$\mathcal{V}_n = n^{-1}\sum_{s=1}^S \mathbb{E}[R_s R_s' \mathcal{X}_s^2 | \mathcal{F}] - n^{-1}\sum_{s=1}^S \mathbb{E}[R_s\mathcal{X}_s|\mathcal{F}]\mathbb{E}[(R_s\mathcal{X}_s)'|\mathcal{F}] + n^{-1}\sum_{s\neq t}\mathrm{Cov}(R_s\mathcal{X}_s, R_t\mathcal{X}_t|\mathcal{F}),$$

where an estimator for the first term in the scalar case with no cross-unit interactions, $n^{-1}\sum_{s=1}^S R_s^2\mathcal{X}_s^2$ gives a conservative estimator as in Adao, Kolesár, and Morales 2019.

5 Conclusion

With heterogeneity between locations in response to common trade shocks and an infeasible representation of the data generating process in terms of a set of observable sufficient statistics, we can instead incorporate interference into regressions. In this paper, we present a statistical theory for the estimators of the linear models to achieve consistency when misspecification errors of the regression for the DGP are non-negligible, which also hints to directions of failure of the OLS estimator in these equilibrium settings. Our theory provides conditions for inference on these pseudo-true estimands; importantly, we allow for one cluster of locations exhibiting network dependence, but require that on average, misspecification errors are zero. An interesting avenue for future work involves providing necessary conditions for consistency of the OLS estimators when equilibrium effects are present.

References

Acemoglu, Daron, David Autor, David Dorn, Gordon H Hanson, and Brendan Price (2016). “Import competition and the great US employment sag of the 2000s”. In: *Journal of Labor Economics* 34.S1, S141–S198.

Adao, Rodrigo, Costas Arkolakis, and Federico Esposito (2019). *General equilibrium effects in space: Theory and measurement*. Tech. rep. National Bureau of Economic Research.

Adao, Rodrigo, Michal Kolesár, and Eduardo Morales (2019). “Shift-share designs: Theory and inference”. In: *The Quarterly Journal of Economics* 134.4, pp. 1949–2010.

Alj, Abdelkamel, Rajae Azrak, and Guy Mélard (2014). “On conditions in central limit theorems for martingale difference arrays”. In: *Economics letters* 123.3, pp. 305–307.

Allen, Treb and Costas Arkolakis (2014). “Trade and the Topography of the Spatial Economy”. In: *The Quarterly Journal of Economics* 129.3, pp. 1085–1140.

Alzubaidi, Samirah and Michael J Higgins (2023). “Estimation of Causal Effects Under K-Nearest Neighbors Interference”. In: *arXiv preprint arXiv:2307.15204*.

Autor, David, David Dorn, and Gordon H Hanson (2021). *On the persistence of the China shock*. Tech. rep. National Bureau of Economic Research.

Autor, David H, David Dorn, and Gordon H Hanson (2013). “The China syndrome: Local labor market effects of import competition in the United States”. In: *American economic review* 103.6, pp. 2121–2168.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel (2022). “Quasi-experimental shift-share research designs”. In: *The Review of economic studies* 89.1, pp. 181–213.

Donaldson, Dave and Richard Hornbeck (2016). “Railroads and American economic growth: A “market access” approach”. In: *The Quarterly Journal of Economics* 131.2, pp. 799–858.

Galle, Simon, Andrés Rodríguez-Clare, and Moises Yi (2023). “Slicing the pie: Quantifying the aggregate and distributional effects of trade”. In: *The Review of Economic Studies* 90.1, pp. 331–375.

Hudgens, Michael G and M Elizabeth Halloran (2008). “Toward causal inference with interference”. In: *Journal of the American Statistical Association* 103.482, pp. 832–842.

Leung, Michael P (2022). “Causal inference under approximate neighborhood interference”. In: *Econometrica* 90.1, pp. 267–293.

— (2024). “Causal Interpretation of Estimands Defined by Exposure Mappings”. In: *arXiv preprint arXiv:2403.08183*.

Manski, Charles F (1993). “Identification of endogenous social effects: The reflection problem”. In: *The review of economic studies* 60.3, pp. 531–542.

Redding, Stephen J and Esteban Rossi-Hansberg (2017). “Quantitative spatial economics”. In: *Annual Review of Economics* 9.1, pp. 21–58.

Sävje, Fredrik (2021). “Causal inference with misspecified exposure mappings”. In: *arXiv preprint arXiv:2103.06471* 116.

A Proofs

A.1 Proofs for Section 2

1. *Derivation of reduced-form of $\mathcal{M}(\theta)$*: Let $R^{t_0} = \text{diag}(R_1^{t_0}, ..., R_n^{t_0})$ be the matrix of initial revenues for locations and let $X_{ij,s} = x_{ij,s}\gamma_s W_j$ define bilateral sales from i to j in sector s. We have

$$\frac{\partial \mathbb{D}_i(\omega)}{\partial \log \chi_s} = -(\sigma - 1)\sum_{j=1}^{N} X_{ij,s}x_{Fj,s} \tag{14}$$

$$\frac{\partial \mathbb{D}_i(\omega)}{\partial \log A_s} = (\sigma - 1)\sum_{j=1}^{N} X_{ij,s}(1 - x_{Fj,s}) \tag{15}$$

$$\frac{\partial \mathbb{D}_i(\omega)}{\partial \log \tilde{A}_{js}} = (\sigma - 1)X_{ij,s}(\mathbb{1}_{[j=i]} - x_{jj,s}) \tag{16}$$

$$\frac{\partial \mathbb{D}_i(\omega)}{\partial \log \omega_j} = (1 + \phi)\sum_{s=1}^{S} X_{ij,s} + (\sigma - 1)\sum_{k=1}^{N}\sum_{s=1}^{S} X_{ik,s}x_{jk,s} \tag{17}$$

$$\frac{\partial \mathbb{D}_i(\omega)}{\partial \log \omega_i} = (1 - \sigma)\sum_{j=1}^{N}\sum_{s=1}^{S} X_{ij,s} + (1 + \phi)\sum_{s=1}^{S} X_{ii,s} + (\sigma - 1)\sum_{k=1}^{N}\sum_{s=1}^{S} X_{ik,s}x_{ik,s} - (1 + \phi)v_i\omega_i^{1+\phi}. \tag{18}$$

If we normalize by $R_i^{t_0}$, define $r_{ij,s}^{t_0} \equiv X_{ij,s}^{t_0} / \sum_d X_{id,s}^{t_0}$ as j's share in i's revenue in sector s, and let $l_{i,s}^{t_0} \equiv \sum_d X_{id,s}^{t_0} / R_i^{t_0}$ denote the share of sector s in total revenue, which is also equal to the share of labor in sector s, we get

$$\frac{1}{R_i^{t_0}}\frac{\partial \mathbb{D}_i(\omega^{t_0})}{\partial \log \chi_s} = -(\sigma-1) l_{i,s}^{t_0} \sum_{j=1}^{N} r_{ij,s}^{t_0} x_{Fj,s}^{t_0} \tag{19}$$

$$\frac{1}{R_i^{t_0}}\frac{\partial \mathbb{D}_i(\omega^{t_0})}{\partial \log A_s} = (\sigma-1) l_{i,s}^{t_0} \sum_{j=1}^{N} r_{ij,s}^{t_0} (1 - x_{Fj,s}^{t_0}) \tag{20}$$

$$\frac{1}{R_i^{t_0}}\frac{\partial \mathbb{D}_i(\omega^{t_0})}{\partial \log \tilde{A}_{js}} = (\sigma-1) l_{i,s}^{t_0} r_{ij,s}^{t_0} (\mathbb{1}_{[j=i]} - x_{jj,s}) \tag{21}$$

$$\frac{1}{R_i^{t_0}}\frac{\partial \mathbb{D}_i(\omega^{t_0})}{\partial \log \omega_j} = (1+\phi) \sum_{s=1}^{S} l_{i,s}^{t_0} r_{ij,s}^{t_0} + (\sigma-1) \sum_{s=1}^{S} l_{i,s}^{t_0} \sum_{k=1}^{N} r_{ik,s} x_{jk,s} \tag{22}$$

$$\frac{1}{R_i^{t_0}}\frac{\partial \mathbb{D}_i(\omega^{t_0})}{\partial \log \omega_i} = (1-\sigma) + (1+\phi) \sum_{s=1}^{S} l_{i,s}^{t_0} r_{ii,s}^{t_0} + (\sigma-1) \sum_{s=1}^{S} l_{i,s}^{t_0} \sum_{k=1}^{N} r_{ik,s}^{t_0} x_{ik,s}^{t_0} - (1+\phi). \tag{23}$$

We now discuss invertibility of $B \equiv -(R^{t_0})^{-1} \nabla_{\log \omega} \mathbb{D}(\omega^{t_0} | \tau, A^{t_0}, \tilde{A}^{t_0}, \chi^{t_0}; \theta)$. We first write $B_{ij} \equiv -(1/R_i^{t_0}) \cdot \partial \mathbb{D}_i(\omega^{t_0}) / \partial \log \omega_j = (\sigma + \phi) \mathbb{1}_{[j=i]} - \rho_{ij}^{t_0}$ where $\rho_{ij}^{t_0} > 0$ is given by $\rho_{ij}^{t_0} = (1+\phi) \sum_{s=1}^{S} l_{i,s}^{t_0} r_{ij,s}^{t_0} + (\sigma-1) \sum_{s=1}^{S} l_{i,s}^{t_0} \sum_{k=1}^{N} r_{ik,s}^{t_0} x_{jk,s}^{t_0}$. For a pair of locations (i,j), we can write $\rho_{ij}^{t_0}$ as a sum of a direct interaction between i and j, and another term that represents how i interacts with a region k and how j interacts with k. A sufficient condition for nonsingularity is diagonal dominance: $|B_{ii}| > \sum_{j \neq i}^{n} |B_{ij}|$ which has interpretation that the effect of a wage change in own-market in partial equilibrium is greater than the sum of the cross-effects from other markets. Rewrite $B = K(I_n - \kappa P)$ where $K = (\sigma + \phi)$, $\kappa = 1/K \in (0, 0.5)$, and $P_{ij} \equiv \rho_{ij}^{t_0}$.

Proof of Proposition 2.1. We first show $I_n - \kappa P$ is diagonally dominant. We show that

$$|1 - \rho_{ii}/K| - \sum_{j \neq i}^{n} |\rho_{ij}/K| = 1 - \rho_{ii}/K - \sum_{j \neq i}^{n} \rho_{ij}/K.$$

The terms in the summation follow from $\rho_{ij} > 0$ for all i, j. Consider

$$\begin{aligned}
1 - \rho_{ii}/K &= 1 - \frac{1}{\sigma + \phi}\left((1+\phi)\sum_{s=1}^{S} l_{i,s}^{t_0} r_{ii,s}^{t_0} + (\sigma - 1)\sum_{s=1}^{S} l_{i,s}^{t_0} \sum_{k=1}^{N} r_{ik,s}^{t_0} x_{ik,s}^{t_0}\right) \\
&\geq 1 - \frac{1}{\sigma + \phi}\left((1+\phi)\cdot 1 + (\sigma - 1)\cdot 1\right) \\
&= 0,
\end{aligned}$$

since we can bound the terms in the summation above by 1, and $\sigma + \phi > 0$. Therefore, we can drop the absolute value notation. We can rewrite the difference now as

$$1 - \frac{1+\phi}{\sigma+\phi}\sum_{s=1}^{S} l_{i,s}^{t_0}\left(r_{ii,s}^{t_0} + \sum_{j\neq i}^{n} r_{ij,s}^{t_0}\right) - \frac{\sigma - 1}{\sigma + \phi}\sum_{s=1}^{S} l_{i,s}^{t_0}\sum_{k=1}^{N} r_{ik,s}^{t_0}\left(x_{ik,s}^{t_0} + \sum_{j\neq i}^{n} x_{jk,s}^{t_0}\right).$$

Notice that the terms in the first parenthesis is equal to $1 - r_{iF,s}^{t_0}$ and similarly the terms in the second parenthesis equal $1 - x_{Fk,s}^{t_0}$. Using the fact that $\sum_{k=1}^{N} r_{ik,s}^{t_0} = 1$ and $\sum_{s=1}^{S} l_{i,s}^{t_0} \leq 1$ we have that

$$\begin{aligned}
|1 + \rho_{ii}/K| - \sum_{j\neq i}^{n} |\rho_{ij}/K| \geq 1 - \frac{1+\phi}{\sigma+\phi}\cdot 1 - \frac{\sigma-1}{\sigma+\phi}\cdot 1 &+ \left(\frac{1+\phi}{\sigma+\phi}\right)\sum_{s=1}^{S} l_{i,s}^{t_0} r_{iF,s}^{t_0} \\
&+ \left(\frac{\sigma-1}{\sigma+\phi}\right)\sum_{s=1}^{S} l_{i,s}^{t_0}\sum_{k=1}^{N} r_{ik,s}^{t_0} x_{Fk,s}^{t_0} \\
&> 0.
\end{aligned}$$

We conclude that $I_n - \kappa P$ is diagonally dominant.
Furthermore, since $|\kappa| = \kappa < 1$ it follows that $(I_n - \kappa P)^{-1} = \sum_{m=1}^{\infty} \kappa^m P^m$ if we can prove that the maximum absolute eigenvalue of κP is bounded above by 1. We have that $\forall i \sum_{j=1}^{n} \rho_{ij} = (1+\phi)\sum_{s=1}^{S} l_{i,s}^{t_0} \sum_{j=1}^{n} r_{ij,s}^{t_0} + (\sigma - 1)\sum_{s=1}^{S} l_{i,s}^{t_0} \sum_{k=1}^{N} r_{ik,s}^{t_0} \sum_{j=1}^{n} x_{jk,s}^{t_0}$. Recall that $\sum_{j=1}^{n} r_{ij,s}^{t_0} = 1 - r_{iF,s}^{t_0} < 1$ and $\sum_{j=1}^{n} x_{jk,s}^{t_0} = 1 - x_{Fk,s}^{t_0} < 1$ and both imply that $\sum_{j=1}^{n} \rho_{ij}^{t_0} < K$ and so $\sum_{j=1}^{n} \kappa\rho_{ij}^{t_0} < 1$. Since κP is a positive matrix, the Perron-Frobenius eigenvalue satisfies $0 < \min_i \sum_j \kappa\rho_{ij}^{t_0} \leq \lambda \leq \max_j \sum_j \kappa\rho_{ij}^{t_0} < 1$.

Therefore, $|\lambda_i| < 1 \, \forall i = 1, ..., N_c$ so the growth of $(\kappa P)^m$ is controlled in the sense of $\sum_{m=0}^{\infty} \kappa^m P^m = (I_n - \kappa P)^{-1}$ is well-defined and $\lim_{m\to\infty} \kappa^m P^m = 0$. □

Proof of Proposition 2.2.

$$\hat{\omega}_i(\hat{\chi}_1, ..., \hat{\chi}_S) = \sum_{j=1}^{n} \beta_{ij} \left(\sum_{k,s} \frac{1}{R_j^{t_0}} \frac{\partial \mathbb{D}_j(\omega^{t_0})}{\partial \log \tilde{A}_{ks}} \right) \hat{\tilde{A}}_{ks} + \sum_{j=1}^{n} \beta_{ij} \left(\sum_{s} \frac{1}{R_j^{t_0}} \frac{\partial \mathbb{D}_j(\omega^{t_0})}{\partial \log A_s} \right) \hat{A}_s + \sum_{j=1}^{n} \beta_{ij} \left(\sum_{s} \frac{1}{R_j^{t_0}} \frac{\partial \mathbb{D}_j(\omega^{t_0})}{\partial \log \chi_s} \right) \hat{\chi}_s \tag{24}$$

is well-defined, so we can now substitute in the expressions for the derivatives given in (19)-(23). □

A.2 Proofs for Section 3

Proof of Proposition 3.1. The expression for the projection coefficient in a model with multiple regressors is $\beta_1 = \sum_{i=1}^{n} E[\ddot{X}_{ni} Y_i | \mathcal{F}] / \sum_{i=1}^{n} E[\ddot{X}_{ni}^2 | \mathcal{F}]$ and similarly for β_2. Using Equation 8 we get

$$\beta_1 = \frac{\sum_{i=1}^{n} E[\ddot{X}_{ni} Y_i(0) | \mathcal{F}]}{\sum_{i=1}^{n} E[\ddot{X}_{ni}^2 | \mathcal{F}]} + \frac{\sum_{i=1}^{n} E[\ddot{X}_{ni} \sum_{s=1}^{S} w_{is} \psi_{is} \mathcal{X}_s | \mathcal{F}]}{\sum_{i=1}^{n} E[\ddot{X}_{ni}^2 | \mathcal{F}]} + \frac{\sum_{i=1}^{n} E[\ddot{X}_{ni} \sum_{j\neq i}^{n} \sum_{s=1}^{S} w_{js} \psi_{ijs} \mathcal{X}_s | \mathcal{F}]}{\sum_{i=1}^{n} E[\ddot{X}_{ni}^2 | \mathcal{F}]}. \tag{25}$$

Observe that

$$\sum_{s=1}^{S} \sum_{i=1}^{n} \lambda_{is}^{(n)} = \sum_{s=1}^{S} \sum_{i=1}^{n} \frac{\text{var}(\mathcal{X}_s | \mathcal{F})}{\sum_{i=1}^{n} E[\ddot{X}_{ni}^2 | \mathcal{F}]} \left(w_{is} - \delta \sum_{j\neq i}^{n} G_{ij} w_{js} \right), \tag{26}$$

where $\delta > 0$. The case with $\delta = 0$ corresponds to the weights obtained by Adao, Kolesár, and Morales 2019. Unlike them, the weights above may be negative depending on the signs and magnitudes of $\{w_{is} - \delta \sum_{j\neq i}^{n} G_{ij} w_{js}\}_{i=1}^{n}$, each of which belongs to the interval $(-\delta, 1)$. Furthermore, the weights can only sum to 1 if $\sum_i \sum_s E\left[\ddot{X}_{ni}\mathcal{X}_s\right] - \sum_i E\left[\ddot{X}_{ni}^2\right] = 0$, i.e. $\sum_i \sum_s (w_{is} - \delta \sum_{j \neq i} G_{ij} w_{js})\sigma_s^2 = \sum_i \sum_s (w_{is} - \delta \sum_{j \neq i} G_{ij} w_{js})^2 \sigma_s^2$. The RHS is positive. Suppose that $w_{is} - \delta \sum_{j \neq i}^n G_{ij} w_{js} > 0$ for all i, s. Then we have $\sum_{i,s} (w_{is} - \delta \sum_{j \neq i} G_{ij} w_{js})\sigma_s^2 [1 - (w_{is} - \delta \sum_{j \neq i} G_{ij} w_{js})] = 0$. However, each term in parentheses is strictly positive, so this equality cannot hold. Suppose that $w_{is} - \delta \sum_{j \neq i}^n G_{ij} w_{js} < 0$ for all i, s. We again immediately get a contradiction. Thus, it must be that for the weights to sum to 1, there exists i, s such that $w_{is} - \delta \sum_{j \neq i}^n G_{ij} w_{js}$ is positive, and others for which it is negative. However, this implies that that some of the weights will be negative, and so β_1 cannot be a convex combination of the treatment effects.

Suppose that for some $c > 0$ we have $\sum_{i,s} \lambda_{is} = -c$. We can rewrite this as requiring $\sum_{i,s} (w_{is} - \delta \sum_{j \neq i} G_{ij} w_{js})\sigma_s^2 [1 + c(w_{is} - \delta \sum_{j \neq i} G_{ij} w_{js})] = 0$. Clearly, we cannot have $w_{is} - \delta \sum_{j \neq i} G_{ij} w_{js} > 0$ for all (i, s), which is just to say that we cannot have $\lambda_{is} > 0$ for all (i, s), which is vacuous given the summation equals $-c$. Suppose $w_{is} - \delta \sum_{j \neq i} G_{ij} w_{js} < 0$ for all (i, s), i.e. $\lambda_{is} < 0$ for all (i, s). The first case is when there exists (i, s) such that $1 + c(w_{is} - \delta \sum_{j \neq i} G_{ij} w_{js}) = 0$ which implies $w_{is} - \delta \sum_{j \neq i} G_{ij} w_{js} = -1/c$ for any $c > 1/\delta$. If this previous condition does not hold for each i and each s, we need that there exists (i, s) such that $0 < 1 + c(w_{is} - \delta \sum_{j \neq i} G_{ij} w_{js}) < 1$ and (i', s') such that $1 - \delta c < 1 + c(w_{i's'} - \delta \sum_{j \neq i'} G_{i'j} w_{js'}) < 0$. Combining these, we get $-\delta < w_{i's'} - \delta \sum_{j \neq i'} G_{i'j} w_{js'} < -1/c < w_{is} - \delta \sum_{j \neq i} G_{ij} w_{js} < 0$ with $c > 1/\delta$. When $\sum_{i,s} \lambda_{is} = c$, $c \neq 1$, we can show that we can have $\lambda_{is} > 0$ for all (i, s) but there must exist (i, s) with $0 < w_{is} - \delta \sum_{j \neq i} G_{ij} w_{js} < 1/c$ and (i', s') with $1/c < w_{i's'} - \delta \sum_{j \neq i'} G_{i'j} w_{js'} < 1$ with $c > 1$. Both arguments with positive and negative sums, as well as sums equal to one, imply that for values of $c \in [-1/\delta, 1]$ we must necessarily have both positive and negative weights. It is also possible to show that we have a mixture of both positive and negative weights when, for example, $c > 1$ so that the existence $c > 1$ does not allow us to infer that all weights are positive.

Finally, we may also represent β_1 as:

$$\begin{aligned}\beta_1 &= \sum_{i=1}^{n} \mathbb{E}\left[\frac{\ddot{X}_{ni}}{\mathbb{E}[\ddot{X}_n' \ddot{X}_n | \mathcal{F}]} \left(\sum_{s=1}^{S} w_{is} \psi_{is} \mathcal{X}_s + \sum_{j \neq i}^{n} w_{js} \psi_{ijs} \mathcal{X}_s \right) \middle| \mathcal{F} \right] \\ &= \sum_{i=1}^{n} \mathbb{E}[\lambda_i^{(n)} (Y_i(\mathcal{X}_1, ..., \mathcal{X}_S) - Y_i(0)) | \mathcal{F}],\end{aligned}$$

with location weights $\lambda_i^{(n)} = \ddot{X}_{ni} / \mathbb{E}[\ddot{X}_n' \ddot{X}_n | \mathcal{F}]$ satsifying $\mathbb{E}[\lambda_i^{(n)} | \mathcal{F}] = 0$. □

A.3 Proofs for Section 4

We first show that $W'W$ converges to a positive definite matrix.

Proposition A.1. *Under Assumption 4.1, $W'W > 0$.*

1. *claim*: $Z_n' Z_n = O_p(n)$. Let $\theta_{st} = \sum_{i=1}^n \sum_{j \neq i}^n \sum_{k \neq i}^n G_{ij} G_{ik} w_{js} w_{kt}$ so that $Z_n' Z_n = \sum_{s=1}^S \theta_{ss} \mathcal{X}_s^2 + \sum_{s<t} (\theta_{st} + \theta_{ts}) \mathcal{X}_s \mathcal{X}_t$. Consider the expectation of the absolute value of the first term:

$$\begin{aligned}\mathbb{E}\left[\left| \sum_{s=1}^{S} \theta_{ss} \mathcal{X}_s^2 \right| \middle| \mathcal{F} \right] &= \mathbb{E}\left[\sum_{s=1}^{S} \theta_{ss} \mathcal{X}_s^2 \middle| \mathcal{F} \right] \\ &= C \sum_s \theta_{ss}, \\ &\leq \sum_s \sum_{i=1}^n \sum_{j \neq i}^n \sum_{k \neq i}^n G_{ij} G_{ik} w_{ks}, \\ &\leq \sum_{i=1}^n \sum_{k \neq i}^n G_{ik}, \\ &= n,\end{aligned}$$

   where the second line uses uniformly bounded second moment implied from Assumption 4.1(ii), the third line uses $w_{js} \leq 1$, and the fourth line uses $\sum_s w_{ks} \leq 1$ and $\sum_{j \neq i}^n G_{ij} = 1$. For a sequence of random variables $\{A_S\}_{S=1}^\infty$, we know that $\mathbb{E}[|A_S|^\alpha] = O(n_S)$ implies $A_S = O_p(n_S^{1/\alpha})$, so it follows that

$\sum_s \theta_{ss}\mathcal{X}_s^2 = O_p(n)$.

Consider the variance of the second term which satisfies

$$\text{var}\left(\sum_{s<t}(\theta_{st}+\theta_{ts})\mathcal{X}_s\mathcal{X}_t \middle| \mathcal{F}\right) \le C^2 \sum_{s,t=1}^{S}\sum_{i,i'=1}^{n}\sum_{j,k\neq i}\sum_{j;,k'\neq i'} G_{ij}G_{ik}G_{i'j'}G_{i'k'}(w_{js}w_{kt}w_{j's}w_{k't} + w_{js}w_{kt}w_{j't}w_{k's} + w_{jt}w_{ks}w_{j's}w_{k't} + w_{jt}w_{ks}w_{j't}w_{k's}).$$

For each of the four terms above, we will remove the t index as well as using that weights $w_{it} \le 1$ and $\sum_{j\neq i} G_{ij} = 1$. The four terms will then be less than

$$\sum_s \sum_{i,i'}^{n} \sum_{j,j'\neq i,i'} G_{ij}G_{i'j'}w_{js}w_{j's} + \sum_s \sum_{i,i'}^{n} \sum_{j,k'\neq i,i'} G_{ij}G_{i'k'}w_{js}w_{k's} + \sum_s \sum_{i,i'}^{n} \sum_{k,j'\neq i,i'} G_{ik}G_{i'j'}w_{ks}w_{j's} + \sum_s \sum_{i,i'}^{n} \sum_{k,k'\neq i,i'} G_{ik}G_{i'k'}w_{ks}w_{k's},$$

which in turn can be bounded above by

$$\sum_{i,i'=1}^{n}\sum_{j,j'\neq i,i'} G_{ij}G_{i'j'} + \sum_{i,i'=1}^{n}\sum_{j,k'\neq i,i'} G_{ij}G_{i'k'} + \sum_{i,i'=1}^{n}\sum_{k,j'\neq i,i'} G_{ik}G_{i'j'} + \sum_{i,i'=1}^{n}\sum_{k,k'\neq i,i'} G_{ik}G_{i'k'},$$

each of which is less than n^2. It follows that $|\sum_{s<t}(\theta_{st}+\theta_{ts})\mathcal{X}_s\mathcal{X}_t| = O_p(n)$ and therefore $Z_n'Z_n = O_p(n)$. Note that we can go further and appeal to Assumption 4.1(viii) to conclude that this variance term is of order smaller than n^2.

2. *claim*: $n^{-1}Z_n'X - \mathbb{E}[n^{-1}Z_n'X|\mathcal{F}] = o_p(1)$. The first term of this difference is $\sum_s \left(\sum_i \sum_{j\neq i} G_{ij}w_{js}w_{is}\right)(\mathcal{X}_s^2 - \text{var}(\mathcal{X}_s|\mathcal{F}))$. By the Bahr-Esseen inequality

we have for some $\eta \in (0,1)$

$$\begin{aligned}
&\mathbb{E}\left[\left|n^{-1}\sum_{s=1}^{S}\left(\sum_{i=1}^{n}\sum_{j\neq i}G_{ij}w_{js}w_{is}\right)(\mathcal{X}_s^2 - \text{var}(\mathcal{X}_s|\mathcal{F}))\right|^{1+\eta}\middle|\mathcal{F}\right] \\
&\preceq n^{-(1+\eta)}\sum_{s=1}^{S}\left|\sum_{i=1}^{n}\sum_{j\neq i}^{n}G_{ij}w_{js}w_{is}\right|^{1+\eta} \\
&\leq n^{-(1+\eta)}\sum_{s=1}^{S}(n_s^2)^{1+\eta} \\
&\leq \left(\frac{(\max_s n_s)^2}{n}\right)^{\eta}\frac{\sum_s n_s}{n} \\
&= o(1),
\end{aligned}$$

where the last line is due to Assumption 4.1(iv). Consider the variance of second term of of this difference:

$$\begin{aligned}
\text{var}\left(\sum_{s<t}\left(\sum_{i=1}^{n}\sum_{j\neq i}^{n}G_{ij}w_{js}w_{it}\right)\mathcal{X}_s\mathcal{X}_t\middle|\mathcal{F}\right) &\leq C^2\sum_{s,t=1}^{S}\sum_{i,i'=1}^{n}\sum_{j,j'\neq i,i'}^{n}G_{ij}G_{i'j'}w_{js}w_{j's}w_{it}w_{i't}, \\
&\leq C^2\sum_{t=1}^{S}\left(\sum_{i=1}^{n}\sum_{i'=1}^{n}w_{it}w_{i't}\right) \\
&= O\left(\sum_{t}n_t^2\right).
\end{aligned}$$

The arguments for the third term are identical. Therefore, the variance of both terms are $o(n^2)$ implying that the terms are $o_p(n)$ by Assumption 4.1(iii).

Proof of Proposition A.1. Consider the first term, $Z_n' Z_n$:

$$\sum_{i=1}^{n} Z_{ni}^2 = \sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{k\neq i}^{n}\sum_{s=1}^{S}\sum_{t=1}^{S} G_{ij}G_{ik}w_{js}w_{kt}\mathcal{X}_s\mathcal{X}_t$$
$$= \sum_{s=1}^{S}\left(\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{k\neq i}^{n} G_{ij}G_{ik}w_{js}w_{ks}\right)\mathcal{X}_s^2 + \sum_{s<t}\left(\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{k\neq i}^{n} G_{ij}G_{ik}(w_{js}w_{kt}+w_{jt}w_{ks})\right)\mathcal{X}_s\mathcal{X}_t$$
$$= \sum_{s=1}^{S}\theta_{ss}(\mathcal{X}_s^2 - \text{var}(\mathcal{X}_s|\mathcal{F})) + \sum_{s<t}(\theta_{st}+\theta_{ts})\mathcal{X}_s\mathcal{X}_t + \sum_{s=1}^{S}\theta_{ss}\text{var}(\mathcal{X}_s|\mathcal{F}) \tag{27}$$

where the second set of summations is a sum of $(S-1)S/2$ terms and has expectation zero conditional on $\mathcal{F}$. We show in the appendix that this term is of order n.

Consider the first term of (27). By Assumption 3.1 (vi) and the fact that $|\sum_s \theta_{ss}\mathcal{X}_s^2| = \sum_s \theta_{ss}\mathcal{X}_s^2$ it is $O_p(n)$. Using the von Bahr-Esseen inequality, we can show that $\mathbb{E}\left[|n^{-1}\sum_s \theta_{ss}(\mathcal{X}_s^2 - \text{var}(\mathcal{X}_s|\mathcal{F}))|^{1+\eta}|\mathcal{F}\right] \preceq \sum_s \left(n^{-1}\sum_{i=1}^n\sum_{j\neq i}^n G_{ij}w_{js}\right)^{1+\eta}$ which converges to zero by Assumption 4.1(viii). Next, we can show that $\text{var}(n^{-1}\sum_{s<t}\theta_{st}\mathcal{X}_s\mathcal{X}_t|\mathcal{F}) \preceq \sum_s \left(n^{-1}\sum_{i=1}^n\sum_{j\neq i}^n G_{ij}w_{js}\right)^2$ with each term in the parentheses between 0 and 1. By Assumption 4.1(viii), the term on the right-hand side converges to zero, which implies $n^{-1}\sum_{s<t}(\theta_{st}+\theta_{ts})\mathcal{X}_s\mathcal{X}_t = o_p(1)$. Therefore,

$$n^{-1}\sum_{i=1}^{n} Z_{ni}^2 - n^{-1}E\left[\sum_{i=1}^{n} Z_{ni}^2 \middle| \mathcal{F}\right] \to^p 0.$$

Next, consider the term

$$\sum_{i=1}^{n} Z_{ni} X_i = \sum_{i=1}^{n} \left(\sum_{j \neq i}^{n} \sum_{s=1}^{S} G_{ij} w_{js} \mathcal{X}_s \right) \left(\sum_{t=1}^{S} w_{it} \mathcal{X}_t \right)$$
$$= \sum_{s=1}^{S} \left(\sum_{i=1}^{n} \sum_{j \neq i}^{n} G_{ij} w_{js} w_{is} \right) \mathcal{X}_s^2 + \sum_{s<t} \left(\sum_{i=1}^{n} \sum_{j \neq i}^{n} G_{ij} (w_{js} w_{it} + w_{jt} w_{is}) \right) \mathcal{X}_s \mathcal{X}_t. \tag{28}$$

The expectation of the first term is $O\left(\sum_s n_s\right)$. We show in the Appendix that the variances of the second and third terms are of order $O\left(\sum_s n_s^2\right)$ so by Assumption 4.1(iii) these terms are $o_p(n)$ and conclude that $n^{-1} \sum_{i=1}^{n} Z_{ni} X_i - \mathbb{E}[n^{-1} \sum_{i=1}^{n} Z_{ni} X_i | \mathcal{F}] = o_p(1)$. □

Proof of Proposition 4.1. Convergence requires proving $W'\varepsilon \to^p 0$. Firstly, we have $\mathbb{E}[W'\varepsilon|\mathcal{F}] = 0$. To see this, observe that $\mathbb{E}[X'\varepsilon|\mathcal{F}] = \mathbb{E}[(\gamma Z_n' + \ddot{X}_n')\varepsilon|\mathcal{F}]$. Furthermore, $\mathbb{E}[\ddot{X}'\varepsilon|\mathcal{F}] = E[\ddot{X}'(Y - X\beta_1 - Z_n\beta_2)|\mathcal{F}] = -E[\ddot{X}'Z_n|\mathcal{F}]\delta - E[\ddot{X}'Z_n|\mathcal{F}]\beta_2 = 0$ since $E[\ddot{X}'Z_n|\mathcal{F}] = 0$. Therefore, $\mathbb{E}[X'\varepsilon|\mathcal{F}] = \gamma \mathbb{E}[Z_n'\varepsilon|\mathcal{F}]$. Likewise, $\mathbb{E}[Z_n'\varepsilon|\mathcal{F}] = \delta \mathbb{E}[X'\varepsilon|\mathcal{F}]$. Therefore, if $1 - \delta\gamma \neq 0$, then $\mathbb{E}[X'\varepsilon|\mathcal{F}] = \mathbb{E}[Z_n'\varepsilon|\mathcal{F}] = 0$. This restriction is satisfied when the Cauchy-Schwarz inequality is strict: $|\mathbb{E}[X'Z_n|\mathcal{F}]| < \sqrt{\mathbb{E}[Z_n'Z_n|\mathcal{F}]\mathbb{E}[X'X|\mathcal{F}]}$, which will be true under Assumption 3.1(iii). We have

$$X'\varepsilon = \sum_{s} \left(\sum_{i=1}^{n} w_{is} Y_i(0) \right) \mathcal{X}_s + \sum_{s,t} \left(\sum_{i=1}^{n} w_{is} w_{it} (\psi_{it} - \beta_1) \right) \mathcal{X}_s \mathcal{X}_t$$
$$+ \sum_{s,t} \left(\sum_{i=1}^{n} w_{is} \sum_{j \neq i}^{n} w_{jt} (\psi_{ijt} - \beta_2 G_{ij}) \right) \mathcal{X}_s \mathcal{X}_t. \tag{29}$$

The first term of (29) is conditionally mean zero with variance $\sum_s \left(\sum_{i,i'} w_{is} w_{i's} Y_i(0) Y_{i'}(0) \right) \text{var}(\mathcal{X}_s|\mathcal{F}$ which is of the same order as $\sum_s n_s^2 = o(n^2)$ by Assumption 4.1(iii). Therefore, this first term converges to zero after normalizing by n.

Consider the second term of (29):

$$\sum_{s=1}^{S}\left(\sum_{i=1}^{n} w_{is}^2(\psi_{is}-\beta_1)\right)\mathcal{X}_s^2+\sum_{s<t}\left(\sum_{i=1}^{n} w_{is}w_{it}(\psi_{it}-\beta_1)\right)\mathcal{X}_s\mathcal{X}_t+\sum_{s<t}\left(\sum_{i=1}^{n} w_{it}w_{is}(\psi_{is}-\beta_1)\right)\mathcal{X}_s\mathcal{X}_t.$$

Let $A_S \preceq B_S$ denote $A_S = O(B_S)$. If we focus on the first part, but demeaned, and use the von Bahr-Esseen inequality with $\eta < 1$ we get

$$\begin{aligned}
\mathbb{E}\left[\left|\sum_{s=1}^{S}\left(\sum_{i=1}^{n} w_{is}^2(\psi_{is}-\beta_1)\right)(\mathcal{X}_s^2-\text{var}(\mathcal{X}_s|\mathcal{F}))\right|^{1+\eta}\middle|\mathcal{F}\right] &\preceq \sum_{s=1}^{S}\left|\sum_{i=1}^{n} w_{is}^2(\psi_{is}-\beta_1)\right|^{1+\eta}\\
&\leq \sum_{s=1}^{S}\left(\sum_{i=1}^{n} w_{is}^2|\psi_{is}-\beta_1|\right)^{1+\eta}\\
&\leq C_\psi^{1+\eta}\sum_{i=1}^{S}\left(\sum_{i=1}^{n} w_{is}^2\right)^{1+\eta}\\
&\leq C_\psi^{1+\eta}(\max_s n_s)^\eta\sum_{s=1}^{S} n_s\\
&= o(n^{1+\eta}),
\end{aligned}$$

which implies that the demeaned term converges in probability to zero after normalizing by n. The variance of the second part is

$$\begin{aligned}
\text{var}\left(\sum_{s<t}\left(\sum_{i=1}^{n} w_{is}w_{it}(\psi_{it}-\beta_1)\right)\mathcal{X}_s\mathcal{X}_t\middle|\mathcal{F}\right) &\preceq \sum_{s,t=1}^{S}\left(\sum_{i=1}^{n} w_{is}w_{it}(\psi_{it}-\beta_1)\right)^2\\
&= \sum_{s,t=1}^{S}\sum_{i,i'=1}^{n} w_{is}w_{it}w_{i's}w_{i't}(\psi_{it}-\beta_1)(\psi_{i't}-\beta_1)\\
&\leq C_\psi^2\sum_{s,t=1}^{S}\sum_{i,i'=1}^{n} w_{is}w_{it}w_{i's}w_{i't}\\
&\preceq \sum_s n_s^2\\
&= o(n^2).
\end{aligned}$$

The same arguments apply to the third part of the second term. Therefore, the normalized second term converges to the limit of $(\sum_s n_s)^{-1}\mathbb{E}\left[\sum_{s=1}^{S}\left(\sum_{i=1}^{n} w_{is}^2(\psi_{is}-\beta_1)\right)\mathcal{X}_s^2|\mathcal{F}\right]$. Finally, consider the third term of (29):

$$\sum_{s=1}^{S}\left(\sum_{i=1}^{n} w_{is}\sum_{j\neq i}^{n} w_{js}(\psi_{ijs}-\beta_2 G_{ij})\right)\mathcal{X}_s^2+\sum_{s<t}\left(\sum_{i=1}^{n} w_{is}\sum_{j\neq i}^{n} w_{jt}(\psi_{ijt}-\beta_2 G_{ij})\right)\mathcal{X}_s\mathcal{X}_t$$
$$+\sum_{s<t}\left(\sum_{i=1}^{n} w_{it}\sum_{j\neq i}^{n} w_{js}(\psi_{ijs}-\beta_2 G_{ij})\right)\mathcal{X}_s\mathcal{X}_t.$$

If we follow the same arguments as before and use the von Bahr-Esseen inequality for demeaned first part, we get for $\eta\in(0,1)$

$$\begin{aligned}
&\mathbb{E}\left[\left|\sum_{s=1}^{S}\left(\sum_{i=1}^{n} w_{is}\sum_{j\neq i}^{n} w_{js}(\psi_{ijs}-\beta_2 G_{ij})\right)(\mathcal{X}_s^2-\text{var}(\mathcal{X}_s|\mathcal{F}))\right|^{1+\eta}\middle|\mathcal{F}\right]\\
&\preceq\sum_s\left(\sum_{i=1}^{n} w_{is}\sum_{j\neq i}^{n} w_{js}|\psi_{ijs}-\beta_2 G_{ij}|\right)^{1+\eta}\\
&\leq\sum_s\left(\sum_{i=1}^{n} w_{is}C_{i,\psi}+|\beta_2^{(S)}|n_s^2\right)^{1+\eta}\\
&\preceq\sum_s n_s^{1+\eta}\left[\left(\sum_{i=1}^{n}\tilde{w}_{is}C_{i,\psi}\right)^{1+\eta}+\left(|\beta_2^{(S)}|n_s\right)^{1+\eta}\right]\\
&\leq(\max_s n_s)^{\eta}\sum_{i=1}^{n}C_{i,\psi}^{1+\eta}+2^{\eta}C_{\beta}^{1+\eta}\sum_s(n_s^2)^{1+\eta}\\
&=o\left(n^{1+\eta}\right)
\end{aligned}$$

where the first line uses Assumption 4.1(ii), the second line uses the triangle inequality, the third line uses Loeve's c_r inequality, the fourth line uses Jensen's inequality for the weights $\tilde{w}_{is}=w_{is}/n_s$ and Assumption 4.2(iv) and (v), and the last line follows from Assumption 4.1(iv) and Assumption 4.2(iv), together with the Norm Monotonicity Inequality. To see this last point, we know that for

$0 < 1+r < 1+\eta$ and $C_{i,\psi} > 0$, $\left(n^{-(1+\eta)} \sum_{i=1}^{n} C_{i,\psi}^{1+\eta}\right)^{1/(1+\eta)} \leq n^{-1} \left(\sum_{i=1}^{n} C_{i,\psi}^{1+r}\right)^{1/(1+r)} = \left(n^{-(1+r)} \sum_{i=1}^{n} C_{i,\psi}^{1+r}\right)^{1/(1+r)}$, the latter of which converges to zero for $r = 0$ under Assumption 4.2(iv) while Assumption 4.1(iv) was important for bounding $\limsup \max n_s$. Observe too that we need $\eta > 0$ in this proof, i.e $2 + \eta$ moments for the shocks. Consider the variance of the second part (with an identical argument for third part):

$$\text{var}\left(\sum_{s<t}\left(\sum_{i=1}^{n} w_{is} \sum_{j\neq i}^{n} w_{jt}(\psi_{ijt} - \beta_2 G_{ij})\right) \mathcal{X}_s \mathcal{X}_t \middle| \mathcal{F}\right) \preceq \sum_{s,t=1}^{S} \sum_{i,i'=1}^{n} \sum_{j,j'\neq i,i'}^{n} w_{is} w_{i's} w_{jt} w_{j't} (\psi_{ijt} - \beta_2 G_{ij})(\psi_{i'j'}$$

Unlike before, bounding the difference $|\psi_{ijs} - \beta_2 G_{ij}|$ by a constant would not, on its own, deliver us the result we need. Observe that $\beta_2^2 \sum_{s,t=1}^{S} \sum_{i,i'=1}^{n} \sum_{j,j'\neq i,i'}^{n} w_{is} w_{i's} w_{jt} w_{j't} G_{ij} G_{i'j'} \leq \beta_2^2 \sum_s \sum_{i,i'=1}^{n} w_{is} w_{i's} = \beta_2^2 \sum_s n_s^2$. By Assumption 4.2(vi) the latter is $o\left((\sum_s n_s^2)^2\right)$. Observe too that

$$\sum_{s,t=1}^{S} \sum_{i,i'=1}^{n} \sum_{j,j'\neq i,i'}^{n} w_{is} w_{i's} w_{jt} w_{j't} |\psi_{ijt}||\psi_{i'j't}| = \sum_{s=1}^{S} \sum_{i,i'=1}^{n} w_{is} w_{i's} \left(\sum_{t=1}^{S} \left[\sum_{j\neq i}^{n} w_{jt}|\psi_{ijt}| \cdot \sum_{j'\neq i'}^{n} w_{j't}|\psi_{i'j't}|\right]\right).$$

By the Cauchy-Schwarz inequality, we have

$$\begin{aligned}\sum_{t=1}^{S} \left[\sum_{j\neq i}^{n} w_{jt}|\psi_{ijt}| \cdot \sum_{j'\neq i'}^{n} w_{j't}|\psi_{i'j't}|\right] &\leq \sqrt{\sum_{t=1}^{S} \left(\sum_{j\neq i} w_{jt}|\psi_{ijt}|\right)^2} \sqrt{\sum_{t=1}^{S} \left(\sum_{j'\neq i'} w_{j't}|\psi_{i'j't}|\right)^2} \\ &= C_{i,\psi} C_{i',\psi},\end{aligned}$$

where we have written each of the terms in square roots as

$$\begin{aligned}\sum_{t=1}^{S} \sum_{j\neq i}^{n} \sum_{k\neq i}^{n} w_{jt} w_{kt} |\psi_{ijt}||\psi_{ikt}| &\leq \left(\sum_{t=1}^{S} \sum_{j\neq i}^{n} w_{jt}|\psi_{ijt}|\right) \sum_{k\neq i}^{n} \max_t |\psi_{ikt}| \\ &= \left(\sum_{j\neq i}^{n} \max_t |\psi_{ijt}|\right)^2 \equiv C_{i,\psi}^2.\end{aligned}$$

If we combine the above lines we can write

$$\sum_{s,t=1}^{S}\sum_{i,i'=1}^{n}\sum_{j,j'\neq i,i'}^{n} w_{is}w_{i's}w_{jt}w_{j't}|\psi_{ijt}||\psi_{i'j't}| \leq \sum_{s=1}^{S}\sum_{i,i'=1}^{n} w_{is}w_{i's}C_{i,\psi}C_{i',\psi} \leq \sum_{i,i'=1}^{n} C_{i,\psi}C_{i',\psi}$$
$$= \left(\sum_{i=1}^{n} C_{i,\psi}\right)^2$$
$$= o(n^2).$$

Turning to the cross-terms we get

$$\sum_{s,t=1}^{S}\sum_{i,i'=1}^{n}\sum_{j,j'\neq i,i'}^{n} w_{is}w_{i's}w_{jt}w_{j't}|\psi_{ijt}\beta_2 G_{i'j'}| \leq \sum_{s,t=1}^{S}\left(\sum_{i,i'=1}^{n}\sum_{j\neq i}^{n} w_{is}w_{i's}w_{jt}|\psi_{ijt}\beta_2|\right)$$
$$= |\beta_2^{(S)}|\sum_{s=1}^{S}\sum_{i,i'=1}^{n} w_{is}w_{i's}\left(\sum_{t=1}^{S}\sum_{j\neq i}^{n} w_{jt}|\psi_{ijt}|\right)$$
$$\leq |\beta_2^{(S)}|\sum_{s=1}^{S}\sum_{i=1}^{n} C_{i,\psi}w_{is}n_s$$
$$\leq C_\beta \max_s n_s \sum_{i=1}^{n} C_{i,\psi}$$
$$= o(n^2).$$

We conclude that $\operatorname{var}\left(\sum_{s<t}\left(\sum_{i=1}^n w_{is}\sum_{j\neq i}^n w_{jt}(\psi_{ijt}-\beta_2 G_{ij})\right)\mathcal{X}_s\mathcal{X}_t \,\middle|\, \mathcal{F}\right) = o(n^2)$ which in turn concludes the proof that $n^{-1}X'\varepsilon - \mathbb{E}[n^{-1}X'\varepsilon|\mathcal{F}] = n^{-1}X'\varepsilon \to^p 0.$ $\square$

Consider the term $Z_n'\varepsilon$:

$$Z_n'\varepsilon = \sum_{s=1}^{S}\left(\sum_{i=1}^{n}\sum_{j\neq i}^{n} G_{ij}w_{js}Y_i(0)\right)\mathcal{X}_s + \sum_{s,t}\left(\sum_{i=1}^{n}\sum_{j\neq i}^{n} G_{ij}w_{js}w_{it}(\psi_{it}-\beta_1)\right)\mathcal{X}_s\mathcal{X}_t$$
$$+ \sum_{s,t}\left(\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{k\neq i}^{n} G_{ij}w_{js}w_{kt}(\psi_{ikt}-\beta_2 G_{ik})\right)\mathcal{X}_s\mathcal{X}_t. \tag{30}$$

The variance of the first term of (30) is

$$\begin{aligned}
\operatorname{var}\left(\sum_{s=1}^{S}\left(\sum_{i=1}^{n}\sum_{j\neq i}^{n} G_{ij}w_{js}Y_i(0)\right)\mathcal{X}_s \,\middle|\, \mathcal{F}\right) &\preceq \sum_{s}\left(\sum_{i,i'=1}^{n}\sum_{j,j'\neq i,i'}^{n} G_{ij}G_{i'j'}w_{js}w_{j's}Y_i(0)Y_{i'}(0)\right)\\
&\leq C_y^2\sum_{s}\left(\sum_{i=1}^{n}\sum_{j\neq i}^{n} G_{ij}w_{js}\right)^2\\
&= o(n^2),
\end{aligned}$$

so that the normalized term converges to zero, its mean. We can write the next term of (30) as a sum of three terms: (i) $\sum_s\left(\sum_{i=1}^n\sum_{j\neq i}^n G_{ij}w_{js}w_{is}(\psi_{is}-\beta_1)\right)\mathcal{X}_s^2$, (ii) $\sum_{s<t}\left(\sum_{i=1}^n\sum_{j\neq i}^n G_{ij}w_{js}w_{it}(\psi_{it}-\beta_1)\right)\mathcal{X}_s\mathcal{X}_t$, (iii)$\sum_{s<t}\left(\sum_{i=1}^n\sum_{j\neq i}^n G_{ij}w_{jt}w_{is}(\psi_{is}-\beta_1)\right)\mathcal{X}_s\mathcal{X}_t$ with the second and third term being identical in their proofs. We can show using arguments similar to those above and using Assumption 4.2(ii) that, when normalized by n, the demeaned first term, as well as each of the second and third converges in probability to zero.

We can write the third term of (30) as the sum of (i) $\sum_s\left(\sum_{i=1}^n\sum_{j\neq i}^n\sum_{k\neq i}^n G_{ij}w_{js}w_{ks}(\psi_{iks}-\beta_2 G_{ik})\right)\mathcal{X}_s$ (ii) $\sum_{s<t}\left(\sum_{i=1}^n\sum_{j\neq i}^n\sum_{k\neq i}^n G_{ij}w_{js}w_{kt}(\psi_{ikt}-\beta_2 G_{ik})\right)\mathcal{X}_s\mathcal{X}_t$, (iii) $\sum_{s<t}\Big(\sum_{i=1}^n\sum_{j\neq i}^n\sum_{k\neq i}^n G_{ij}w_{jt}w_{ks}(\psi_{iks}$ with the second and third having identical arguments. Using von Bahr-Esseen, we can show that (i) can be bounded above by $C_{\psi,2}^{1+\eta}(\max_s n_s)^{1+\eta}\sum_s(\sum_{i=1}^n\sum_{j\neq i}G_{ij}w_{js})^{1+\eta} = o(n^{1+\eta})$ when $\limsup_{S\to\infty}\max_s n_s<\infty$. We can also show that the variance of (ii) can be bounded above by $\left(\sum_{i=1}^n\sum_{k\neq i}^n\max_t(\psi_{ikt}-\beta_2 G_{ik})\right)^2 = o(n^2)$ by Assumption 4.2(iv). We conclude that $n^{-1}Z_n'\varepsilon\to^p 0$.

A.4 Central limit theorem

Let $r_S = n(S)^{-1}$. We have $r_S^{1/2}(W'W)(\hat{\beta}-\beta^{(S)}) = r_S^{1/2}W'\varepsilon_n$ and apply the Lindeberg Central Limit Theorem to the right-hand side term. We first define $a_s \equiv \sum_{i=1}^n w_{is}Y_i(0), b_{st}\equiv\sum_{i=1}^n w_{is}w_{it}(\psi_{it}-\beta_1), c_{st}\equiv\sum_{i=1}^n\sum_{j\neq i}^n w_{is}w_{jt}(\psi_{ijt}-\beta_2 G_{ij})$. Ignoring dependence of the variances on the conditioning set, let $\sigma_s^2=\operatorname{var}(\mathcal{X}_s|\mathcal{F}_n)$.

We can write

$$
\begin{aligned}
X'\varepsilon_n &= \sum_{s=1}^{S}\left(a_s\mathcal{X}_s + b_{ss}(\mathcal{X}_s^2-\sigma_s^2) + \sum_{t\neq s} b_{st}\mathcal{X}_s\mathcal{X}_t + c_{ss}(\mathcal{X}_s^2-\sigma_s^2) + \sum_{t\neq s} c_{st}\mathcal{X}_s\mathcal{X}_t\right) \\
&= \sum_{s=1}^{S}\left(a_s\mathcal{X}_s + (b_{ss}+c_{ss})(\mathcal{X}_s^2-\sigma_s^2) + \sum_{t=1}^{s-1}(b_{st}+b_{ts}+c_{st}+c_{ts})\mathcal{X}_s\mathcal{X}_t\right) \\
&= \sum_{s=1}^{S}\tilde{\mathcal{X}}_{1,s},
\end{aligned}
$$

where $\sum_s(b_{ss}\sigma_s^2 + c_{ss}\sigma_s^2) = 0$ because $\mathbb{E}[X'\varepsilon_n|\mathcal{F}_n] = 0$. Notice that the s-th term of the sum is a measurable functions $f(\mathcal{X}_1, ..., \mathcal{X}_s)$ so it follows that $\{\tilde{\mathcal{X}}_s\}_{s=1}^S$ is a martingale difference array with filtration $\tilde{\mathcal{F}}_s = \sigma(\mathcal{F}, \mathcal{X}_1, ..., \mathcal{X}_s)$. I.e. $\mathbb{E}[\tilde{\mathcal{X}}_{1,s}|\tilde{\mathcal{F}}_{s-1,n}] = 0$. It also follows that unconditionally $\mathbb{E}[\tilde{\mathcal{X}}_{1,s}\tilde{\mathcal{X}}_{1,t}|\mathcal{F}_n] = 0$.
Next, define $\bar{a}_s \equiv \sum_{i=1}^n\sum_{j\neq i}^n G_{ij}w_{js}Y_i(0)$, $\bar{b}_{st} \equiv \sum_{i=1}^n\sum_{j\neq i}^n G_{ij}w_{js}w_{it}(\psi_{it}-\beta_1)$, and finally $\bar{c}_{st} \equiv \sum_{i=1}^n\sum_{j\neq i}^n\sum_{k\neq i}^n G_{ij}w_{js}w_{kt}(\psi_{ikt}-\beta_2 G_{ik})$. Using the same arguments as above, we can write $Z_n'\varepsilon_n = \sum_{s=1}^S\tilde{\mathcal{X}}_{2,s}$. Therefore, we have $r_s^{1/2}W_n'\varepsilon_n = r_S^{1/2}\sum_{s=1}^S\tilde{\mathcal{X}}_s$ where $\tilde{\mathcal{X}}_s$ is a 2×1 vector martingale difference array. Using the results in Alj, Azrak, and Mélard 2014 together with the Cramer-Wold Device, we can check whether Lyapunov's condition is satisfied and a central limit theorem holds. That is, we verify whether the following two requirements are satisfied:
(i) $\lim_{S\to\infty} r_S^{1+\eta/2}\sum_{s=1}^S\mathbb{E}[|t'\tilde{\mathcal{X}}_s|^{2+\eta}|\mathcal{F}_n] = 0$ for all $t\in\mathbb{R}^2$, (ii) $r_S\sum_{s=1}^S \text{var}(t'\tilde{\mathcal{X}}_s|\mathcal{F}_n) - r_S\sum_{s=1}^S\text{var}(t'\tilde{\mathcal{X}}_s|\tilde{\mathcal{F}}_{s-1,n}) \to 0$.
(i) We need to show that $r_S^{1+\eta/2}\sum_s\mathbb{E}[|t_1\tilde{\mathcal{X}}_{1,s}|^{2+\eta}|\mathcal{F}] \to 0$ and $r_S^{1+\eta/2}\sum_s\mathbb{E}[|t_2\tilde{\mathcal{X}}_{2,s}|^{2+\eta}|\mathcal{F}] \to 0$. This follows from using Minkowski's inequality for the s-th term of the sum followed by the c_r-inequality. Consider the first. We will use the c_r-inequality to

bound $\mathbb{E}[|\tilde{\mathcal{X}}_{1,s}|^{2+\eta}|\mathcal{F}]$. Let's consider the term

$$
\begin{aligned}
r_S^{1+\eta/2}\sum_s \mathbb{E}[|\sum_{t=1}^{s-1} c_{st}\mathcal{X}_s\mathcal{X}_t|^{2+\eta}|\mathcal{F}] &= r_S^{1+\eta/2}\sum_s \mathbb{E}\left[\left|\sum_{t=1}^{s-1}\left(\sum_{i=1}^{n}\sum_{j\neq i}^{n} w_{is}w_{jt}(\psi_{ijt}-\beta_2 G_{ij})\mathcal{X}_s\mathcal{X}_t\right)\right|^{2+\eta}\middle|\mathcal{F}\right] \\
&\preceq r_S^{1+\eta/2}\sum_s \mathbb{E}\left[\left|\sum_{t=1}^{s-1}\left(\sum_{i=1}^{n}\sum_{j\neq i}^{n} w_{is}w_{jt}(\psi_{ijt}-\beta_2 G_{ij})\mathcal{X}_t\right)\right|^{2+\eta}\middle|\mathcal{F}\right] \\
&\leq r_S^{1+\eta/2}\sum_s \mathbb{E}\left[\left(\sum_{i=1}^{n} w_{is}\sum_{j\neq i}^{n}\max_{t\leq S}|\psi_{ijt}-\beta_2 G_{ij}|\sum_{t=1}^{s-1} w_{jt}|\mathcal{X}_t|\right)^{2+\eta}\middle|\mathcal{F}\right] \\
&\leq r_S^{1+\eta/2}\sum_s \mathbb{E}\left[\left(\sum_{t=1}^{s-1}\alpha_{s,t}\right)^{2+\eta}\left(\sum_{t=1}^{s-1}\alpha_{s,t}|\mathcal{X}_t|\right)^{2+\eta}\middle|\mathcal{F}\right] \\
&\leq r_S^{1+\eta/2}\sum_s \left(\sum_{t=1}^{s-1}\alpha_{s,t}\right)^{2+\eta}\sum_{t=1}^{s-1}\alpha_{s,t}\mathbb{E}\left[|\mathcal{X}_t|^{2+\eta}|\mathcal{F}\right] \\
&\preceq r_S^{1+\eta/2}\sum_s \left(\sum_{t=1}^{s-1}\alpha_{s,t}\right)^{2+\eta} \\
&\leq \sum_{s=1}^{S}\left(n^{-1/2}\sum_{i=1}^{n} w_{is}\sum_{j\neq i}^{n}\max_{t\leq S}|\psi_{ijt}-\beta_2 G_{ij}|\right)^{2+\eta},
\end{aligned}
$$

where $\alpha_{s,t} \equiv \sum_{i=1}^n w_{is}\sum_{j\neq i}^n M_{ij}w_{jt}$ with M_{ij} the maximum misspecification error between i and j and the fifth line uses Jensen's inequality pointwise. Assumption 4.3(ii) implies that the terms inside the parentheses converge to zero as $S \to \infty$ and to prove that the sequence of partial sums converges to zero as $S \to \infty$, we make use of the fact that it converges to zero for $\eta = 0$, so must also converge for $\eta > 0$.

Next, consider

$$\sum_{s=1}^{S} \mathbb{E}\left[\left| r_S^{1/2} \sum_{t\neq s} b_{st}\mathcal{X}_s\mathcal{X}_t \right|^{2+\eta} \middle| \mathcal{F}\right] = \sum_{s=1}^{S} \mathbb{E}\left[\left| r_S^{1/2} \sum_{t=1}^{s-1} \left(\sum_{i=1}^{n} w_{is}w_{it}(\psi_{it} - \beta_1^{(S)})\right) \mathcal{X}_s\mathcal{X}_t \right|^{2+\eta} \middle| \mathcal{F}\right]$$
$$\preceq r_S^{1+\eta/2} \sum_{s=1}^{S} \mathbb{E}\left[\left(\sum_{t=1}^{s-1} \alpha_{st}|\mathcal{X}_t|\right)^{2+\eta} \middle| \mathcal{F}\right]$$
$$\preceq r_S^{1+\eta/2} \sum_{s=1}^{S} \left(\sum_{t=1}^{s-1} \alpha_{st}\right)^{2+\eta}$$
$$\preceq r_S^{1+\eta/2} \sum_{s=1}^{S} \left(\sum_{t=1}^{s-1} \sum_{i=1}^{n} w_{is}w_{it}\right)^{2+\eta}$$
$$\leq \left(\frac{\max_{s\leq S} n_s}{n^{1/2}}\right)^{\eta} n^{-1} \sum_{s=1}^{S} n_s^2$$
$$= o(1),$$

where $\alpha_{st} \equiv \sum_{i=1}^{n} w_{is}w_{it}|\psi_{it} - \beta_1|$ and we again apply Jensen's inequality pointwise as well bounded misspecification $|\psi_{it} - \beta_1|$. The final line then makes use of eventually bounded maximum sector size.
Consider the term

$$\sum_{s=1}^{S} \mathbb{E}[|r_S^{1/2} a_s \mathcal{X}_s|^{2+\eta}|\mathcal{F}] = \sum_{s=1}^{S} \mathbb{E}\left[\left| r_S^{1/2} \left(\sum_{i=1}^{n} w_{is}Y_i(0)\right) \mathcal{X}_s \right|^{2+\eta} \middle| \mathcal{F}\right]$$
$$\preceq \sum_{s=1}^{S} \mathbb{E}\left[|r_S^{1/2} n_s \mathcal{X}_s|^{2+\eta} \middle| \mathcal{F}\right]$$
$$\preceq n^{-(1+\eta/2)} \sum_{s=1}^{S} n_s^2 n_s^{\eta}$$
$$= o(1).$$

[The terms $\sum_s (b_{ss} + c_s s)(\mathcal{X}_s^2 - \sigma_s^2)$ and $\sum_{t=1}^{s-1}(b_{ts} + c_{ts})\mathcal{X}_s\mathcal{X}_t$ should follow similar arguments to ones above.]

We now make the arguments for $r_S^{1+\eta/2}\sum_s \mathbb{E}[|\tilde{\mathcal{X}}_{2,s}|^{2+\eta}|\mathcal{F}] \to 0$. We will again bound the terms above using the c_r-inequality and consider each term separately converges to zero. Let's consider the term

$$\begin{aligned}
\sum_{s=1}^{S}\mathbb{E}\left[\left|r_S^{1/2}\sum_{t=1}^{s-1}\bar{c}_{st}\mathcal{X}_s\mathcal{X}_t\right|^{2+\eta}\middle|\mathcal{F}\right] &= \sum_{s=1}^{S}\mathbb{E}\left[\left|r_S^{1/2}\sum_{t\neq s}\left(\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{k\neq i}^{n}G_{ij}w_{js}w_{kt}(\psi_{ikt}-\beta_2^{(S)}G_{ik})\right)\mathcal{X}_s\mathcal{X}_t\right|^{2+\eta}\right. \\
&\preceq r_S^{1/2}\sum_{s=1}^{S}\left(\sum_{t=1}^{s-1}\alpha_{st}\right)^{2+\eta} \\
&\leq \sum_{s=1}^{S}\left(n^{-1/2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{k\neq i}^{n}G_{ij}w_{js}M_{ik}\right)^{2+\eta} \\
&= o(1),
\end{aligned}$$

where the final line uses that the statement holds for $\eta = 0$ together with Assumption 4.4(i) to restrict the rate of growth of the total misspecification error, so must hold for $\eta > 0$.

Let's consider the term

$$\begin{aligned}
\sum_{s=1}^{S}\mathbb{E}\left[\left|r_S^{1/2}\sum_{t\neq s}\bar{b}_{st}\mathcal{X}_s\mathcal{X}_t\right|^{2+\eta}\middle|\mathcal{F}\right] &= \sum_{s=1}^{S}\mathbb{E}\left[\left|r_S^{1/2}\sum_{t=1}^{s-1}\left(\sum_{i=1}^{n}\sum_{j\neq i}^{n}G_{ij}w_{js}w_{it}(\psi_{it}-\beta_1^{(S)})\right)\mathcal{X}_s\mathcal{X}_t\right|^{2+\eta}\middle|\mathcal{F}\right] \\
&\preceq r_S^{1+\eta/2}\sum_{s=1}^{S}\left(\sum_{t=1}^{s-1}\sum_{i=1}^{n}\sum_{j\neq i}^{n}G_{ij}w_{js}w_{it}|\psi_{it}-\beta_1^{(S)}|\right)^{2+\eta} \\
&\preceq \sum_{s=1}^{S}\left(n^{-1/2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}G_{ij}w_{js}\right)^{2+\eta} \\
&= o(1),
\end{aligned}$$

where the final line uses Assumption 4.3(iii). Finally consider the term

$$\sum_{s=1}^{S}\mathbb{E}[|r_S^{1/2}\bar{a}_s\mathcal{X}_s|^{2+\eta}|\mathcal{F}] = \sum_{s=1}^{S}\mathbb{E}\left[\left|r_S^{1/2}\left(\sum_{i=1}^{n}\sum_{j\neq i}^{n}G_{ij}w_{js}Y_i(0)\right)\mathcal{X}_s\right|^{2+\eta}\middle|\mathcal{F}\right]$$
$$\preceq \sum_{s=1}^{S}\left(n^{-1/2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}G_{ij}w_{js}\right)^{2+\eta}$$
$$= o(1).$$

[The terms $\sum_s \bar{b}_{ss}(\mathcal{X}_s^2-\sigma_s^2)$ and $\sum_s \bar{c}_{ss}(\mathcal{X}_s^2-\sigma_s^2)$ should follow similar arguments to ones above.] This completes the proof of Lyapunov's condition $r_S^{1+\eta/2}\sum_{s=1}^{S}\mathbb{E}[|t'\tilde{\mathcal{X}}_s|^{2+\eta}|\mathcal{F}_n] \to 0$ as $S\to\infty$ for any real vector $t\in\mathbb{R}^2$.

We now need to obtain conditions under which the conditional variance converges. Define $V_n(t) := \text{var}\left(r_S^{1/2}t'W'\varepsilon_n|\mathcal{F}\right) = \text{var}(r_S^{1/2}\sum_{s=1}^{S}t'\tilde{\mathcal{X}}_s|\mathcal{F}_n)$. Then $r_S\sum_{s=1}^{S}\text{var}(t'\tilde{\mathcal{X}}_s|\tilde{\mathcal{F}}_{s-1}) - V_n(t) = r_S t'\sum_{s=1}^{S}(\mathbb{E}[\tilde{\mathcal{X}}_s\tilde{\mathcal{X}}_s'|\tilde{\mathcal{F}}_{s-1}] - \mathbb{E}[\tilde{\mathcal{X}}_s\tilde{\mathcal{X}}_s'|\mathcal{F}])t$ with

$$\text{var}(\tilde{\mathcal{X}}_s|\mathcal{F}) = \begin{pmatrix}\mathbb{E}[\tilde{\mathcal{X}}_{1,s}^2|\mathcal{F}] & \mathbb{E}[\tilde{\mathcal{X}}_{1,s}\tilde{\mathcal{X}}_{2,s}|\mathcal{F}]\\ \mathbb{E}[\tilde{\mathcal{X}}_{1,s}\tilde{\mathcal{X}}_{2,s}|\mathcal{F}] & \mathbb{E}[\tilde{\mathcal{X}}_{2,s}^2|\mathcal{F}]\end{pmatrix}.$$

Since for any symmetric matrix A, $|t'At| \leq ||A||t't$ where $||A||$ is either the spectral or Frobenius matrix norm. It is sufficient to show that each element of A converges to zero.

For ease of reading, we drop the conditioning notation when referring to the 'unconditional' expectation that conditions only on $\mathcal{F}$. Consider the term

$$r_S\sum_{s=1}^{S}\mathbb{E}[\tilde{\mathcal{X}}_{1,s}^2|\mathcal{F}_{s-1}] - \mathbb{E}[\tilde{\mathcal{X}}_{1,s}^2] = 2r_S\sum_{s=2}^{S}\left(\sigma_s^2a_s^2 + \mathbb{E}[\tilde{\mathcal{X}}_s^3](b_{ss}+c_{ss})\right)\sum_{t=1}^{s-1}(b_{st}+b_{ts}+c_{st}+c_{ts})\mathcal{X}_t$$
$$+ r_S\sum_{s=2}^{S}\sigma_s^2\sum_{t=1}^{s-1}(b_{st}+b_{ts}+c_{st}+c_{ts})^2(\mathcal{X}_t^2-\sigma_t^2)$$
$$+ r_S\sum_{s=2}^{S}\sigma_s^2\sum_{t\neq u}^{s-1}(b_{st}+b_{ts}+c_{st}+c_{ts})(b_{su}+b_{us}+c_{su}+c_{us})\mathcal{X}_t\mathcal{X}_u.$$

Notice that the s-th term is zero when we take the expectation with respect to $\mathcal{F}_{s-1}$. Call the three terms above $2D_1 + D_2 + D_3$. Since $\mathbb{E}[D_j] = 0$, we will prove that var$(D_j) \to 0$. Observe that we can we can rewrite D_1 as

$$D_1 = r_S \sum_{t=1}^{S-1} \left(\sum_{s=t+1}^{S} (\sigma_s^2 a_s^2 + \mathbb{E}[\tilde{\mathcal{X}}_s^3](b_{ss} + c_{ss}))(b_{st} + b_{ts} + c_{st} + c_{ts}) \right) \mathcal{X}_t,$$

so that

$$\begin{aligned} \text{var}(D_1) &= n^{-2} \sum_{t=1}^{S-1} \sigma_t^2 \left[\sum_{s=t+1}^{S} (\sigma_s^2 a_s^2 + \mathbb{E}[\tilde{\mathcal{X}}_s^3](b_{ss} + c_{ss}))(b_{st} + b_{ts} + c_{st} + c_{ts}) \right]^2 \\ &\preceq n^{-2} \sum_{t=1}^{S-1} \sigma_t^2 \sum_{s=t+1}^{S} \sum_{s'=t+1}^{S} \mathbb{E}[\mathcal{X}_s^3]^2 (b_{ss} + c_{ss})(c_{st} + c_{ts})(b_{s's'} + c_{s's'})(c_{s't} + c_{ts'}), \end{aligned}$$

where the second line uses the fact that the order of the variance is determined by the dominant term $(b_{ss}+c_{ss})(c_{st}+c_{ts})$. We have $|b_{ss}+c_{ss}| \preceq n_s + \sum_{i=1}^n \sum_{j\neq i}^n w_{is} w_{js} |\psi_{ijs} - \beta_2 G_{ij}|$ while $|c_{st} + c_{ts}| \preceq \sum_{i=1}^n \sum_{j\neq i}^n w_{is} w_{jt} |\psi_{ijt} - \beta_2 G_{ij}|$. It follows that

$$\begin{aligned} \text{var}(D_1) \preceq & n^{-2} \sum_{t=1}^{S-1} \sum_{s=1}^{S} \sum_{s'=1}^{S} \left(\sum_{i=1}^{n} \sum_{j\neq i}^{n} w_{is} w_{js} |\psi_{ijs} - \beta_2 G_{ij}| \right) \left(\sum_{i=1}^{n} \sum_{j\neq i}^{n} w_{is'} w_{js'} |\psi_{ijs'} - \beta_2 G_{ij}| \right) \\ & \cdot \left(\sum_{i=1}^{n} \sum_{j\neq i}^{n} w_{is} w_{jt} |\psi_{ijt} - \beta_2 G_{ij}| \right) \left(\sum_{i=1}^{n} \sum_{j\neq i}^{n} w_{is'} w_{jt} |\psi_{ijt} - \beta_2 G_{ij}| \right) \\ \leq & \left(n^{-1/2} \sum_{i=1}^{n} \sum_{j\neq i}^{n} \max_{s \leq S} |\psi_{ijs} - \beta_2 G_{ij}| \right)^4 \to 0 \end{aligned}$$

by Assumption 4.4.

Consider the term D_3, which can be rewritten as

$$
\begin{aligned}
D_3 &= r_S \sum_{s=2}^{S} \sigma_s^2 \sum_{t\neq u}^{s-1} (b_{st}+b_{ts}+c_{st}+c_{ts})(b_{su}+b_{us}+c_{su}+c_{us})\mathcal{X}_u\mathcal{X}_t \\
&= r_S \sum_{s=3}^{S} \sigma_s^2 \sum_{t=2}^{s-1}\sum_{u=1}^{t-1} 2(b_{st}+b_{ts}+c_{st}+c_{ts})(b_{su}+b_{us}+c_{su}+c_{us})\mathcal{X}_u\mathcal{X}_t \\
&= 2r_S \sum_{t=2}^{S-1}\sum_{u=1}^{t-1}\sum_{s=u+2}^{S} \sigma_s^2 (b_{st}+b_{ts}+c_{st}+c_{ts})(b_{su}+b_{us}+c_{su}+c_{us})\mathcal{X}_u\mathcal{X}_t,
\end{aligned}
$$

which has variance

$$
\begin{aligned}
\operatorname{var}(D_3) &= 4n^{-2}\sum_{t=1}^{S-1}\sum_{u=1}^{t-1}\left(\sum_{s=u+2}^{S}\sigma_s^2(b_{st}+b_{ts}+c_{st}+c_{ts})(b_{su}+b_{us}+c_{su}+c_{us})\right)^2\sigma_u^2\sigma_t^2 \\
&\preceq n^{-2}\sum_{t=1}^{S-1}\sum_{u=1}^{t-1}\sum_{s=u+2}^{S}\sum_{s'=u+2}^{S}(c_{st}+c_{ts})(c_{su}+c_{us})(c_{s't}+c_{t's})(c_{s'u}+c_{u's}) \\
&\preceq n^{-2}\sum_{t=1}^{S-1}\sum_{u=1}^{t-1}\sum_{s=u+2}^{S}\sum_{s'=u+2}^{S}\left(\sum_{i=1}^{n}\sum_{j\neq i}^{n} w_{is}w_{it}M_{ij}\right)\left(\sum_{i=1}^{n}\sum_{j\neq i}^{n} w_{is}w_{iu}M_{ij}\right)\left(\sum_{i=1}^{n}\sum_{j\neq i}^{n} w_{is'}w_{it}M_{ij}\right)\Bigg(\sum_{i}
\\
&\leq \left(n^{-1/2}\sum_{i=1}^{n}\sum_{j\neq i}^{n} M_{ij}\right)^4 \to 0.
\end{aligned}
$$

Finally, the term D_2 is

$$
D_2 = r_S \sum_{t=1}^{S-1}\sum_{s=t+1}^{S}\sigma_s^2(b_{st}+b_{ts}+c_{st}+c_{ts})^2(\mathcal{X}_t^2-\sigma_t^2).
$$

The arguments follow similarly to D_3. We therefore conclude that the $(1,1)$ entry of $r_S t' \sum_{s=1}^{S}(\mathbb{E}[\tilde{\mathcal{X}}_s\tilde{\mathcal{X}}_s'|\tilde{\mathcal{F}}_{s-1}] - \mathbb{E}[\tilde{\mathcal{X}}_s\tilde{\mathcal{X}}_s'|\mathcal{F}])t$ converges to zero for every value of $t \in \mathbb{R}^2$. We next consider the $(2,2)$ entry.

Assumption A.1. *For* $n = n(S)$, $a_s \equiv \sum_{i=1}^{n} w_{is}Y_i(0)$, *and* $b_{st} \equiv \sum_{i=1}^{n} w_{is}w_{it}(\psi_{it} - \beta_1)$, $\lim_{S\to\infty} n^{-1}\sum_{s=1}^{S}\mathbb{E}\left[\left(a_s\mathcal{X}_s + b_{ss}(\mathcal{X}_s-\sigma_s^2)+\sum_{t\neq s} b_{st}\mathcal{X}_s\mathcal{X}_t\right)^2\middle|\mathcal{F}_n\right]$ *converges to* *a positive value,* V_1.

For $\bar{a}_s \equiv \sum_{i=1}^n \sum_{j\neq i}^n G_{ij} w_{js} Y_i(0), \bar{b}_{st} \equiv \sum_{i=1}^n \sum_{j\neq i}^n G_{ij} w_{js} w_{it}(\psi_{it} - \beta_1)$*, and* $\bar{c}_{st} \equiv \sum_{i=1}^n \sum_{j\neq i}^n \sum_{k\neq i}^n G_{ij} w_{js} w_{kt}(\psi_{ikt} - \beta_2 G_{ik})$*. Then* $\lim_{S\to\infty} n^{-1} \sum_{s=1}^S \mathbb{E}\left[\left(\bar{a}_s \mathcal{X}_s + \bar{b}_{ss}(\mathcal{X}_s - \sigma_s^2) + \sum_{t\neq s} \bar{b}_{st} \mathcal{X}_s \mathcal{X}_t\right)^2 + 2\bar{a}_s \bar{c}_{ss}(\mathcal{X}_s^2 - \sigma_s^2)\mathcal{X}_s \middle| \mathcal{F}\right]$ *converges to a positive value* V_2.

Proposition A.2. *Under Assumptions 4.2, 4.4, and Assumption A.1,* $n^{-1} \sum_{s=1}^S \mathbb{E}[\tilde{\mathcal{X}}_{1,s}^2 | \mathcal{F}_n] \to V_1$ *and* $n^{-1} \sum_{s=1}^S \mathbb{E}[\tilde{\mathcal{X}}_{2,s}^2 | \mathcal{F}] \to V_2$.

Proof. Recall that $\tilde{\mathcal{X}}_{1,s} = a_s \mathcal{X}_s + b_{ss}(\mathcal{X}_s^2 - \sigma_s^2) + \sum_{t\neq s} b_{st} \mathcal{X}_s \mathcal{X}_t + c_{ss}(\mathcal{X}_s^2 - \sigma_s^2) + \sum_{t\neq s} c_{st} \mathcal{X}_s \mathcal{X}_t$. We can show that

$$\begin{aligned}
\mathbb{E}[\tilde{\mathcal{X}}_{1,s}^2 | \mathcal{F}_n] = a_s^2 \mathbb{E}[\mathcal{X}_s^2 | \mathcal{F}_n] &+ 2a_s(b_{ss} + c_{ss})\mathbb{E}[(\mathcal{X}_s^2 - \sigma_s^2)\mathcal{X}_s | \mathcal{F}_n] \\
&+ (b_{ss}^2 + 2b_{ss}c_{ss} + c_{ss}^2)\mathbb{E}[(\mathcal{X}_s^2 - \sigma_s^2)^2 | \mathcal{F}_n] \\
&+ \sum_{t\neq s} (b_{st}^2 + 2b_{st}c_{st} + c_{st}^2)\mathbb{E}[\mathcal{X}_s^2 | \mathcal{F}_n]\mathbb{E}[\mathcal{X}_t^2 | \mathcal{F}_n].
\end{aligned}$$

Consider the term

$$\begin{aligned}
n^{-1} \sum_{s=1}^S a_s^2 \mathbb{E}[\mathcal{X}_s^2 | \mathcal{F}] &\preceq n^{-1} \sum_{s=1}^S \left(\sum_{i=1}^n w_{is} Y_i(0)\right)^2 \\
&\leq C_y^2 n^{-1} \sum_{s=1}^S n_s^2 = O(1).
\end{aligned}$$

Next,

$$n^{-1}\sum_{s=1}^{S}2a_s(b_{ss}+c_{ss})\mathbb{E}[(\mathcal{X}_s^2-\sigma_s^2)\mathcal{X}_s|\mathcal{F}] \preceq n^{-1}\sum_{s=1}^{S}\left(\sum_{i=1}^{n}w_{is}Y_i(0)\right)(\sum_{i=1}^{n}w_{is}^2(\psi_{is}-\beta_1)$$
$$+\sum_{i=1}^{n}\sum_{j\neq i}^{n}w_{is}w_{js}(\psi_{ijs}-\beta_2G_{ij}))$$
$$\preceq n^{-1}\sum_{s=1}^{S}n_s\left(\sum_{i=1}^{n}w_{is}^2+\sum_{i=1}^{n}\sum_{j\neq i}^{n}w_{is}w_{js}\max_{s\leq S}|\psi_{ijs}-\beta_2G_{ij}|\right)$$
$$\leq n^{-1}\sum_{s=1}^{S}n_s^2+n^{-1}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\max_{s\leq S}|\psi_{ijs}-\beta_2G_{ij}|$$
$$=O(1)+o(1).$$

The last line implies that $n^{-1}\sum_{s=1}^{S}2a_sc_{ss}\mathbb{E}[(\mathcal{X}_s^2-\sigma_s^2)\mathcal{X}_s|\mathcal{F}]=o(1)$.
Next, consider the term

$$n^{-1}\sum_{s=1}^{S}(b_{ss}+c_{ss})^2\mathbb{E}[(\mathcal{X}_s^2-\sigma_s^2)^2] \preceq n^{-1}\sum_{s=1}^{S}\left(\sum_{i=1}^{n}w_{is}^2(\psi_{is}-\beta_1)+\sum_{i=1}^{n}\sum_{j\neq i}^{n}w_{is}w_{js}(\psi_{ijs}-\beta_2G_{ij})\right)$$
$$\leq C_{\psi,1}^2n^{-1}\sum_{s=1}^{S}n_s^2+\left(n^{-1}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\max_{s\leq S}|\psi_{ijs}-\beta_2G_{ij}|\right)^2,$$

where the second term converges to zero. Next, we can show that $n^{-1}\sum_{s=1}^{S}\sum_{t\neq s}b_{st}^2\preceq n^{-1}\sum_{s=1}^{S}n_s^2$, $n^{-1}\sum_{s=1}^{S}\sum_{t\neq s}b_{st}c_{st}=o(1)$ under the assumption $\limsup\max_s n_s<\infty$ and total misspecification error being order smaller than n. Finally, $n^{-1}\sum_{s=1}^{S}\sum_{t\neq s}c_{st}^2=o(1)$ under Assumption 4.4(i). We conclude that $n^{-1}\sum_{s=1}^{S}\sum_{t\neq s}(b_{st}^2+2b_{st}c_{st}+$

$c_{st}^2)\mathbb{E}[(\mathcal{X}_t\mathcal{X}_s)|\mathcal{F}] = O(1)$. This leaves us with

$$\lim_{S\to\infty} n^{-1}\sum_{s=1}^{S}\mathbb{E}\left[\tilde{\mathcal{X}}_{1,s}^2\middle|\mathcal{F}\right] = \lim_{S\to\infty}\Bigg(n^{-1}\sum_{s=1}^{S}a_s^2\mathbb{E}[\mathcal{X}_s^2|\mathcal{F}] + n^{-1}\sum_{s=1}^{S}2a_{ss}b_{ss}\mathbb{E}[(\mathcal{X}_s^2-\sigma_s^2)\mathcal{X}_s|\mathcal{F}]$$
$$+ n^{-1}\sum_{s=1}^{S}b_{ss}^2\mathbb{E}[(\mathcal{X}_s^2-\sigma_s^2)^2|\mathcal{F}]$$
$$+ n^{-1}\sum_{s=1}^{S}\sum_{t\neq s}b_{st}^2\mathbb{E}[(\mathcal{X}_s\mathcal{X}_t)^2|\mathcal{F}]\Bigg).$$

Arguments for $n^{-1}\sum_{s=1}^{S}\mathbb{E}[\tilde{\mathcal{X}}_{2,s}^2|\mathcal{F}]$ follow similar arguments. □

B Consistency with control variables

This section can help us learn how to deal with control variables, $\{V_i\}_{i=1}^n$, should random assignment hold conditional on V. Such controls will be included in the set characterising our sample, $\mathcal{F}$. Define the projection models $X_i = Z_{ni}\delta + \ddot{X}_{ni}$ and $Z_{ni} = X_i\gamma + \ddot{Z}_{ni}$. Let $\tilde{X}_n = X - Z_n\hat{\delta} = (I_n - H_Z)X$ where $H_Z = Z_n(Z_n'Z_n)^{-1}Z_n'$. The OLS estimator of β_1 is

$$\begin{aligned}\hat{\beta}_1 &= \frac{\tilde{X}_n'Y}{\tilde{X}_n'\tilde{X}_n}\\ &= \beta_1 + \frac{\tilde{X}_n'(Y-\tilde{X}_n\beta_1)}{\tilde{X}_n'\tilde{X}_n}\\ &= \beta_1 + \frac{\ddot{X}_n'\varepsilon - (\hat{\delta}-\delta)Z_n'\varepsilon}{(\ddot{X}_n - Z_n(\hat{\delta}-\delta))'(\ddot{X}_n - Z_n(\hat{\delta}-\delta))}\end{aligned} \tag{31}$$

where the final equality uses the fact that $(I_n - H_Z)(Y - \tilde{X}_n\beta_1) = (I_n - H_Z)\varepsilon$. Similarly,

$$\begin{aligned}\hat{\beta}_2 &= \frac{\tilde{Z}_n' Y}{\tilde{Z}_n'\tilde{Z}_n} \\ &= \beta_2 + \frac{\ddot{Z}_n'\varepsilon - (\hat{\gamma}-\gamma)X'\varepsilon}{(\ddot{Z}_n - X(\hat{\gamma}-\gamma))'(\ddot{Z}_n - X(\hat{\gamma}-\gamma))}.\end{aligned} \tag{32}$$

Consider the projection models $X_i = Z_{ni}\delta + \ddot{X}_{ni}$ and $Z_{ni} = X_i\gamma + \ddot{Z}_{ni}$. We can write $\hat{\delta} - \delta = (Z_n'Z_n)^{-1}Z_n'\ddot{X}_n$ and $\hat{\gamma} - \gamma = (X'X)^{-1}X'\ddot{Z}_n$.

Proposition B.1. *Under Assumption 4.1, (i) $\hat{\delta} - \delta = o_p(1)$ and (ii) $\hat{\gamma} - \gamma = o_p(1)$.*

Proof of Proposition B.1. We want to show that $n^{-1}Z_n'\ddot{X} \to^p 0$. We have $n^{-1}Z_n'\ddot{X} = n^{-1}Z_n'X - \delta n^{-1}Z_n'Z_n$ together with Assumption 4.1

$$\begin{aligned}n^{-1}Z_n'\ddot{X} &= n^{-1}Z_n'X - E\left[n^{-1}Z_n'X \,|\, \mathcal{F}\right]\left(1 + \frac{n^{-1}Z_n'Z_n - E[n^{-1}Z_n'Z_n|\mathcal{F}]}{n^{-1}E[Z_n'Z_n|\mathcal{F}]}\right) \\ &= n^{-1}Z_n'X - E\left[n^{-1}Z_n'X \,|\, \mathcal{F}\right]\left(1 + \frac{o_p(1)}{Q_Z + o(1)}\right) \\ &= o_p(1)\end{aligned}$$

to conclude $\hat{\delta} - \delta = o_p(1)$.

Next, consider the estimand $\gamma^{(n)} = E[X'Z_n|\mathcal{F}]/E[X'X|\mathcal{F}]$. Focusing on the denominator we have $E[X'X|\mathcal{F}] = \sum_s(\sum_i w_{is}^2)E[\mathcal{X}_s^2|\mathcal{F}] = O(\sum_s n_s)$. We have already found that the numerator is $O(\sum_s n_s^2)$ so under Assumption 4.1(iv) and (v) we conclude that $\gamma^{(n)} = O(1)$. We can write $n^{-1}X'\ddot{Z}_n = n^{-1}X'Z_n - n^{-1}\gamma X'X$:

$$n^{-1}X'Z_n - n^{-1}E[X'Z_n|\mathcal{F}]\left(1 + \frac{n^{-1}X'X - n^{-1}E[X'X|\mathcal{F}]}{n^{-1}E[X'X|\mathcal{F}]}\right).$$

The numerator equals $n^{-1}\sum_s\left(\sum_i^n w_{is}^2\right)(\mathcal{X}_s^2 - \text{var}(\mathcal{X}_s|\mathcal{F})) + 2n^{-1}\sum_{s<t}\left(\sum_{i=1}^n w_{is}w_{it}\right)\mathcal{X}_s\mathcal{X}_t$ where both terms converge in probability to zero. Finally, using that $n^{-1}\sum_{i=1}^n Z_{ni}X_i - \mathbb{E}[n^{-1}\sum_{i=1}^n Z_{ni}X_i|\mathcal{F}] = o_p(1)$, we can conclude that $\hat{\gamma} - \gamma = (n^{-1}X'X)^{-1}n^{-1}X'\ddot{Z}_n = o_p(1)$. □