

Bounding Omitted Variable Bias Using Auxiliary Data: With an Application to Estimate Neighborhood Effects

Yujung Hwang *

June 20, 2025

Abstract

This paper develops a method for estimating a superset that contains the linear projection coefficients of an outcome variable y onto a full set of covariates (x, z, w) , when the available data are split across two independent sources. One dataset identifies the joint distribution of y and a subset of covariates (z, w) , while the other identifies the joint distribution of (x, z) . This framework is motivated by applications where administrative data lacks information on individual-level heterogeneities—such as preferences—but auxiliary data (e.g., survey) provide such detail. I illustrate the application of the method in estimating neighborhood effects while addressing selection into neighborhood choice. The micro Census data contain rich longitudinal residence information but limited detail on individual heterogeneity, while the auxiliary cross-sectional survey provides richer information on heterogeneity but lacks residential histories. I find that growing up in an ethnic enclave substantially increases the likelihood of living in an ethnic enclave later in life, but I find little evidence of neighborhood effects on most other cultural assimilation outcomes.

JEL Classification : C01, C80, J01, J15, J16, J61, Z10, Z13

Keywords : Omitted Variable Bias, Two-Sample Least Squares, Data Combination, Intergenerational Cultural Transmission, Neighborhood Effect

*Email : yujunghwang@gmail.com. This paper subsumes a working paper titled, “Neighborhood Effects on Intergenerational Cultural Transmission” (Hwang (2022)). I provide an open-source R package, bndovb (version 1.2), to implement the estimator for a simple case ($d_w = 0, d_x = 1$): <https://github.com/yujunghwang/bndovb>. The research was approved by the Homewood Institutional Review Board (HIRB00011748). This research was financially supported by The Yale MacMillan International Dissertation Fellowship and the JHU COVID-19 Accelerator grant. The permission of the Office for National Statistics to use the Longitudinal Study is gratefully acknowledged, as is the help provided by staff of the Centre for Longitudinal Study Information & User Support (CeLSIUS). CeLSIUS is funded by the ESRC under project ES/V003488/1. The author alone is responsible for the interpretation of the data. This work contains statistical data from ONS which is Crown Copyright. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets that may not exactly reproduce National Statistics aggregates. During the preparation of this manuscript, the author used Grammarly version 14.1240.0, Chat GPT version 4.0, Claude version 3.7 Sonnet in order to conduct a literature review, to improve clarity in writing, and to proofread mathematical proofs. After using these tools, the author reviewed and edited the content as needed. I thank many of my mentors and three anonymous referees for helpful comments. All errors in this paper are my own.

1 Introduction

Researchers often do not have all relevant variables in one dataset. Say, a correct regression model, called long regression, should regress y onto a full set of covariates (x, z, w) . However, the available dataset may contain only (y, z, w) . If a researcher regresses y onto (z, w) only, then the estimates obtained from this short regression – regression omitting a subset of relevant variables – suffer from omitted variable bias. This omitted variable bias due to data limitation is a common challenge in empirical research. For example, administrative data often lacks information on individual-level heterogeneities – such as preferences – but ignoring these key individual heterogeneities in a regression might result in an omitted variable bias if these heterogeneities are correlated with other included covariates.

This paper develops a data combination approach to address omitted variable bias in a general two-sample least squares framework. I consider two independent datasets: one that identifies the joint distribution of (y, z, w) , and another that identifies the joint distribution of (x, z) . Leveraging information from both sources, I study projections of a superset that contains the linear projection coefficients from the long regression of y on (x, z, w) . For example, if administrative data lack key individual-level heterogeneities, x , one can use an auxiliary survey that captures the missing variables x to partially identify the coefficients while addressing omitted variable bias.

The key contribution of this paper is to extend the framework to cases where covariates w are not observed in the auxiliary dataset—that is, when the dimension of w is greater than zero, $d_w > 0$. This case has received limited attention in the literature ([Cross and Manski \(2002\)](#), [Fan et al. \(2014\)](#), [Pacini \(2019\)](#), [Fan et al. \(2017\)](#), [D'Haultfoeuille et al. \(2022\)](#), [Fan et al. \(2025\)](#)). This generalization significantly broadens the applicability of the data combination approach to empirical settings where auxiliary data are only partially informative.

The special case where the auxiliary data contains all covariates from the long regression (i.e., $d_w = 0$) has been studied extensively by [Pacini \(2019\)](#) and [Fan et al. \(2025\)](#) using optimal transport techniques. Assuming $d_w = 0$, [Pacini \(2019\)](#) derived sharp identified sets for

the scalar x case ($d_x = 1$) and constructed a superset for the multivariate x case ($d_x > 1$) (Theorem 1 in [Pacini \(2019\)](#)). [Fan et al. \(2025\)](#) extended this result in Proposition 4.1 by characterizing the sharp set for when x is multivariate, showing in their Corollary 4.1 that it is nested within the superset of [Pacini \(2019\)](#). My framework nests the $d_w = 0$ case and yields a similar characterization to that of [Pacini \(2019\)](#). I discuss how to extend his framework to allow for measurement errors in covariates and compare the computational efficiency of my estimator to his in the Online Appendix.

The case $d_w > 0$, which is the primary focus of this paper, differs fundamentally from the $d_w = 0$ case. Specifically, the identified set of parameters in the $d_w > 0$ case¹ is not necessarily convex. In contrast, when $d_w = 0$, the identified set of parameters² is convex. Since only convex sets can be characterized by support functions, the support-function-based identification strategy used in [Pacini \(2019\)](#) and [Fan et al. \(2025\)](#) does not extend to the $d_w > 0$ case. Despite the loss of convexity, I show that projections of the superset onto each dimension are bounded intervals, and therefore, convex. I provide a plug-in estimator for these bounded intervals.

The key intuition behind my approach is that I can apply the Monotone Rearrangement Inequality (Cambanis-Simons-Stout inequality, [Cabanis et al. \(1976\)](#))³ element-wise to every unidentified second-order moment in the OLS estimator. This approach ignores dependence between unidentified moments (e.g., dependence between $E[x_1y]$ and $E[x_2y]$ when $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, or dependence between $E[xy]$ and $E[xw']$), so the Cartesian product of interval bound estimates forms only a superset for a vector of unidentified moments. Next, I define a set-valued generalization of the OLS estimator, where inputs include the superset of a vector of unidentified moments. The projection of the image of this set-valued OLS estimator gives a bounded interval for each parameter. This plug-in estimator is computationally light, which makes it practically useful. A common concern for identifying a superset is that it may be too large to

¹For example, the sharp set Θ_I from equation (3), page 11 of [Fan et al. \(2025\)](#), or the superset in my Proposition 2 is not convex.

²For example, [Fan et al. \(2025\)](#) Theorem 3.2 (ii) has shown that the sharp set described in their Proposition 4.1 is convex. Similarly, my Proposition 3 (2) shows that the superset when $d_w = 0$ is convex.

³The same inequality can be proven by using a different mathematical result, Optimal Transportation Theorem ([Galichon \(2018\)](#)).

be informative. In the empirical application, however, I show that the estimated supersets are reasonably tight.

For inference in the case of $d_w = 0$, I suggest using the numerical delta method developed by Hong and Li (2018), which is applicable to estimators that may not be Hadamard differentiable but are Hadamard directionally differentiable. For a tuning parameter choice for this numerical delta method, I suggest using the double bootstrap used in Hong and Li (2020). For the case of $d_w > 0$, developing inference methods presents additional challenges. The moment matrix in the OLS estimator may approach singularity in certain regions of the parameter space, necessitating a condition number restriction to ensure numerical stability. Due to these technical complexities, I defer the development of formal inference procedures for this case to future work.

I demonstrate the application of this method by estimating neighborhood effects; see Graham (2018) for a comprehensive overview on this literature. To my knowledge, this study represents the first attempt to estimate causal neighborhood effects using a data combination approach.⁴ A persistent challenge in neighborhood effects research is selection bias: individuals' neighborhood choices often correlate with unobservable characteristics that simultaneously affect outcomes of interest. In this empirical application, I examine how childhood neighborhood environments influence the cultural assimilation of second-generation immigrants. Parental cultural traits constitute a natural source of omitted variable bias in this context, as less-assimilated parents may preferentially select into ethnic enclaves, thereby confounding neighborhood effects with parental influence.⁵

For this analysis, I utilize two independent datasets: (1) a longitudinal administrative dataset tracking neighborhood residence histories and adulthood cultural outcomes for second-generation immigrants, and (2) a cross-sectional survey providing detailed information on

⁴Alternative research designs for estimating causal neighborhood effects include natural experiments (e.g., Moving-to-Opportunity randomized controlled trials, randomized social housing allocations; Algan et al. (2022), Chetty et al. (2016), Chyn (2018)) and sibling designs (Chetty and Hendren (2017)). My approach differs from these established methods as it neither requires natural experiments nor longitudinal household-level data tracking siblings over time. Instead, it assumes researchers can identify sources of omitted variable bias and access separate datasets containing information about these missing variables.

⁵My framework can accommodate additional potential omitted variables as well. In such cases, auxiliary data must contain either these omitted variables or appropriate proxies.

first-generation immigrants' cultural preferences and residential choices. Since the auxiliary survey data contains all relevant covariates, the estimator for the $d_w = 0$ case is directly applicable. However, for comparison, I deliberately omit several covariates from the auxiliary data to demonstrate the estimator for the $d_w > 0$ case as well. Throughout the implementation of both estimators, I address potential measurement errors in the proxies for parental cultural preferences.

The results from the $d_w = 0$ estimator reveal limited statistical evidence that childhood neighborhood significantly affects most cultural assimilation outcomes among second-generation immigrants. The sole adulthood outcome consistently influenced by childhood neighborhood exposure is the likelihood of continuing to reside in an ethnic enclave. For individuals raised in such neighborhoods, this likelihood increases by 39 to 52 percentage points (representing 69 to 92% of the mean probability), based on our interval estimates. This finding aligns with [Algan et al. \(2022\)](#), who similarly documented limited neighborhood influence on cultural assimilation in a different country (France) using quasi-random variation in social housing allocation rather than a data combination approach.

When comparing results across estimators, the $d_w > 0$ case yields slightly wider interval estimates than the $d_w = 0$ case, with the former consistently nesting the interval estimates of the latter. The difference between these two estimates reflects how much information is lost by deliberately omitting variables in auxiliary data. The $d_w > 0$ case estimator lacks confidence interval estimates, so it is difficult to make a statistical inference on these estimates.

Section 2 introduces the model assumptions. Section 3 presents the identification results, which are illustrated through a numerical example in Section 4. Section 5 defines the plug-in estimator. Section 6 applies our methodology to empirical data. Section 7 concludes. Monte Carlo simulation details are provided in the Online Appendix due to space constraints.

2 Model

Following standard convention in the literature, I treat all vectors—both random vectors and parameter vectors—as column vectors by default. For notational simplicity, when referring to joint distributions, I use expressions such as (y, x, z, w) without explicitly indicating vector orientation.

Consider the linear regression model in equation (1), where the parameter of interest is the vector $(\alpha', \beta', \gamma') \in \mathbb{R}^{d_x+d_z+d_w}$. Suppose that no single dataset identifies the full joint distribution of (y, x, z, w) . Instead, two independent datasets identify the joint distributions of (y, z, w) and (x, z) , respectively, as formalized in Assumption 1. The vector z contains covariates common to both datasets. In cases where the datasets do not share any common variables, z contains only a vector that equals one for all observations.

$$y = x'\alpha + z'\beta + w'\gamma + \epsilon, \quad E[(x', z', w')'\epsilon] = 0, \quad (1)$$

$$y \in \mathcal{Y} = \mathbb{R}, \quad x \in \mathcal{X} = \mathbb{R}^{d_x}, \quad z \in \mathcal{Z} = \mathbb{R}^{d_z}, \quad w \in \mathcal{W} = \mathbb{R}^{d_w}, \quad d_x \geq 1, d_z \geq 1, d_w \geq 0.$$

Assumption 1. *The main dataset identifies the joint distribution of (y, z, w) , and the auxiliary dataset identifies the joint distribution of (x, z) .*

Assumption 1 is trivially satisfied when both datasets contain the variables (y, z, w) and (x, z) without measurement error. However, the assumption may also hold when the data contain noisy but sufficiently informative measurements (i.e., proxies) that allow identification of the respective joint distributions. A large literature ([Allman et al. \(2009\)](#), [Hu \(2008\)](#), [Hu and Schennach \(2008\)](#), [Hu and Shiu \(2016\)](#), [Cunha et al. \(2010\)](#)) has established sufficient conditions for identification from proxies. These conditions, which are summarized in Online Appendix Section A, typically require multiple (usually three or more) conditionally independent proxies for the same latent variable, unless a strong parametric assumption (such as error-in-variable structure) is imposed.

Beyond Assumption 1, I make two other Assumptions: First, Assumption 2 requires that

the joint distribution of (y, x, z, w) has finite second moments. Second, Assumption 3 states that the moment matrix E , stated in Assumption 3, is well-conditioned with a known bound for its condition number. Assumption 3 ensures numerical stability when applying the OLS estimator and can be satisfied by removing highly multicollinear covariates from the model. When the moment matrix is singular due to perfect multicollinearity, the condition number becomes infinity. In practice, I use 2000 as an upper bound of the condition number, for example, in Section 4 illustrating a numerical example and in the empirical application in Section 6.

Assumption 2. *The variance-covariance matrix of (y, x, z, w) in equation (1) is finite.*

Assumption 3. *The moment matrix $E \equiv \begin{bmatrix} E[xx'] & E[xz'] & E[xw'] \\ E[zx'] & E[zz'] & E[zw'] \\ E[wx'] & E[wz'] & E[ww'] \end{bmatrix}$ is well-conditioned.*

Specifically, its condition number $\kappa(E)$ – the ratio of the largest to smallest eigenvalue – is bounded with a known bound, ensuring the moment matrix is invertible and numerical computations remain stable.

$$\kappa(E) \leq \kappa_0, \quad 0 < \kappa_0 < \infty. \quad (2)$$

Without additional assumptions—such as conditional independence or exclusion restrictions—the parameters (β', γ') are generally not point-identified using the main dataset alone (Ridder and Moffitt (2007)).

3 Identification

This section introduces the sharp identified set of parameters and compares it with the super-set analyzed in this paper. I begin by defining the sharp identified set of the moment vector $(E[xy]', E[xw_1]', \dots, E[xw_{d_w}]')$, denoted by $\Theta^{\text{sharp},m}$. The sharp identified set of the parameter vector $(\alpha', \beta', \gamma')'$, denoted $\Theta^{\text{sharp},p}$, is then defined as the image of a set-valued function $\mathcal{B}_{\text{OLS}} : \mathcal{K}(\mathbb{R}^{d_x+d_z+d_w}) \rightrightarrows \mathcal{K}(\mathbb{R}^{d_x+d_z+d_w})$. Here $\mathcal{K}(\mathbb{R}^n)$ denotes the collection of all compact

subsets of \mathbb{R}^n . Here, I abuse the notation and use $\kappa(m)$ to refer to the condition number of the moment matrix ((*) in equation 4) with A that arises from a moment vector m . The lower and upper bounds of each parameter can be derived by projecting $\Theta^{\text{sharp},p}$ onto each coordinate dimension.

Remark 1 (Sharp Set Identification, General Case $d_w \geq 0$). *Let $\mathcal{M}(f(y, z, w), f(x, z))$ denote the set of joint distributions over (y, x, z, w) whose marginals over (y, z, w) and (x, z) equal $f(y, z, w)$ and $f(x, z)$, respectively. Then the sharp identified set of the moment vector is*

$$\Theta^{\text{sharp},m} \equiv \left\{ \left[E_{\tilde{f}}[xy]', E_{\tilde{f}}[xw_1]', \dots, E_{\tilde{f}}[xw_{d_w}]' \right]' \middle| \tilde{f} \in \mathcal{M}(f(y, z, w), f(x, z)) \right\}.$$

The sharp identified set of the parameters is

$$\Theta^{\text{sharp},p} = \mathcal{B}_{\text{OLS}}(\Theta^{\text{sharp},m}), \quad (3)$$

$$\text{where } \mathcal{B}_{\text{OLS}}(M) \equiv \left\{ \underbrace{\begin{bmatrix} E[xx'] & E[xz'] & A' \\ E[zx'] & E[zz'] & E[zw'] \\ A & E[wz'] & E[ww'] \end{bmatrix}}_{(*)}^{-1} \begin{bmatrix} B \\ E[zy] \\ E[wy] \end{bmatrix} \middle| m \equiv \begin{bmatrix} B \\ A_1 \\ \vdots \\ A_{d_w} \end{bmatrix} \in M, \text{ s.t. } \kappa(m) \leq \kappa_0 \right\}. \quad (4)$$

Here, B and A_k (k -th column of matrix A) represent the unidentified moment, $E[xy]$ and $E[wx_k]$, respectively. κ_0 refers to the upper bound for the condition number in Assumption 3. The lower and upper bounds of each parameter $\{\alpha', \beta', \gamma'\}$ can be found by projecting the identified set $\Theta^{\text{sharp},p}$ into each dimension.

$$\left[\alpha_l^{\text{sharp}\prime}, \beta_l^{\text{sharp}\prime}, \gamma_l^{\text{sharp}\prime} \right]' = \left[\inf \text{Proj}_1(\Theta^{\text{sharp},p}), \dots, \inf \text{Proj}_{d_x+d_z+d_w}(\Theta^{\text{sharp},p}) \right]'. \quad (5)$$

$$\left[\alpha_u^{\text{sharp}\prime}, \beta_u^{\text{sharp}\prime}, \gamma_u^{\text{sharp}\prime} \right]' = \left[\sup \text{Proj}_1(\Theta^{\text{sharp},p}), \dots, \sup \text{Proj}_{d_x+d_z+d_w}(\Theta^{\text{sharp},p}) \right]'. \quad (6)$$

The proof of the remark is omitted, as it directly follows from the definition of a sharp

identified set as the smallest set consistent with the observed marginals subject to the condition number restriction. Next, I discuss the topological properties of sharp identified sets.

Proposition 1 (Topological Properties of Sharp Identified Sets). .

Under Assumption 1, 2, 3, the following statements are true.

- (1) $\Theta^{\text{sharp},m}$ is non-empty, compact, and convex.
- (2) $\Theta^{\text{sharp},p}$ is non-empty and compact. When $d_w = 0$, $\Theta^{\text{sharp},p}$ is convex, but when $d_w > 0$, $\Theta^{\text{sharp},p}$ may not be convex.
- (3) $\text{Proj } \Theta^{\text{sharp},p}$ is non-empty, compact, and convex. It is a bounded interval on \mathbb{R} .

Proof. In Online Appendix C. □

The sharp set is characterized as an image of a continuous set-valued function $\mathcal{B}_{OLS}(\cdot)$ of $\Theta^{\text{sharp},m}$ but this set-valued function is not an affine function of $\Theta^{\text{sharp},m}$ because of the moment matrix inverse. Therefore, the convexity of $\Theta^{\text{sharp},m}$ does not guarantee the convexity of $\Theta^{\text{sharp},p}$. When $d_w = 0$, matrix A disappears and the condition number constraint always holds on the entire $\Theta^{\text{sharp},m}$ from Assumption 3. The set-valued function $\mathcal{B}_{OLS}(\cdot)$ becomes an affine function of $m = [B] \in \Theta^{\text{sharp},m}$. Therefore, the convexity of $\Theta^{\text{sharp},m}$ guarantees the convexity of $\Theta^{\text{sharp},p}$. Fan et al. (2025)'s Theorem 3.2 has also shown that their sharp set (Θ_I in their equation (3)), analogous to $\Theta^{\text{sharp},p}$, is convex when $d_w = 0$.

Because of the potential non-convexity of $\Theta^{\text{sharp},p}$, we can not utilize support functions to characterize $\Theta^{\text{sharp},p}$ directly, which is an approach taken by Pacini (2019) and Fan et al. (2025) for $d_w = 0$ case. However, one can still characterize a convex set $\Theta^{\text{sharp},m}$ using support functions, and it is straightforward to identify $\Theta^{\text{sharp},p}$ given $\Theta^{\text{sharp},m}$, because we know the set-valued function $\mathcal{B}_{OLS}(\cdot)$.

Remark 2 (Support Function Characterization of $\Theta^{\text{sharp},m}$). *Let θ be a unit direction vector belonging to $(d_x + d_x d_w)$ -dimensional unit sphere $\mathbb{S}^{d_x + d_x d_w}$. Then the support functions for $\Theta^{\text{sharp},m}$ is simply*

$$h(\theta) = \sup_{\tilde{f} \in \mathcal{M}(f(y,z,w), f(x,z))} E_{\tilde{f}}[\langle \theta, (yx', w_1 x', \dots, w_{d_w} x')' \rangle]. \quad (7)$$

Then, $\Theta^{\text{sharp},m}$ is characterized by this support function.

$$\Theta^{\text{sharp},m} = \bigcap_{\theta \in \mathbb{S}^{d_x+d_w}} \{w \in \mathbb{R}^{d_x+d_w} \mid \langle \theta, w \rangle \leq h(\theta)\}. \quad (8)$$

Solving (7) numerically reduces to a linear programming once \tilde{f} is discretized. It is challenging to solve (7) when (x, z, w) is high-dimensional. In that case, one may use efficient algorithms such as the Sinkhorn-Knopp algorithm (Sinkhorn and Knopp, 1967; Peyré et al., 2019).

I next introduce a pair of supersets, $\Theta^{\text{superset},m}$ and $\Theta^{\text{superset},p}$, which are computationally convenient in high-dimensional settings. To construct them, I begin by stating the Monotone Rearrangement Inequality for a second-order moment $E[x_k y]$ for each $k \in \{1, \dots, d_x\}$. For brevity, I omit the analogous formulae for the bounds on $E[x_k w_l]$, which can be derived by replacing y with w_l for each $l \in \{1, \dots, d_w\}$.

Lemma 1 (Monotone Rearrangement Inequality). *Under Assumption 2 and given marginals $f(y, z, w)$ and $f(x, z)$, the sharp upper bound for $E[x_k y]$ for each $k \in \{1, \dots, d_x\}$ is:*

$$\mu_u^{x_k y} \equiv \sup_{\tilde{f} \in \mathcal{M}(f(y, z, w), f(x, z))} E[x_k y] = E \left[y \cdot Q_{x_k|z}(F_{y|z}(y|z) \mid z) \right], \quad (9)$$

where $F_{y|z}(y|z)$ is the **conditional Cumulative Distribution Function (CDF)** of y given z , and $Q_{x_k|z}(u|z)$ is the **conditional quantile function (the inverse CDF)** of x_k given z , for $u \in [0, 1]$. The sharp lower bound is:

$$\mu_l^{x_k y} \equiv \inf_{\tilde{f} \in \mathcal{M}(f(y, z, w), f(x, z))} E[x_k y] = E \left[y \cdot Q_{x_k|z}(1 - F_{y|z}(y|z) \mid z) \right]. \quad (10)$$

Similarly, one can define the sharp lower and upper bounds for $E[x_k w_l]$, denoted as $\mu_l^{x_k w_l}$ and $\mu_u^{x_k w_l}$, by replacing y with w_l .

Proof of Lemma 1. In Online Appendix C. □

We can make a few observations about the bound formula in equations 9 and 10: (i) The

sharp bounds in equations 9, 10 are finite under Assumption 2 because of the Cauchy-Schwarz inequality. This is stated in Corollary 1. (ii) The tightness of the bound for $E[x_k y]$ depends on the extent to which z and x_k are correlated. If the covariates in z are more strongly correlated with x_k , then the quantile function $Q_{x_k|z}(\cdot|z)$ may provide a much tighter bound on $E[x_k y]$ ⁶. Fan et al. (2017) has shown that the interval bound on $E[x_k y]$ becomes a singleton if at least one of the conditional marginal distributions $F(y|z), F(x_k|z)$ is degenerate for almost all z (Theorem 3.2 in Fan et al. (2017))⁷.

Corollary 1. *Under Assumption 2, the sharp bounds in Lemma 1 (equations 9 and 10) are finite.*

Proof of Corollary 1. In Online Appendix C. □

Next, I state the identified supersets and their topological properties.

Proposition 2 (Superset Identification, General Case $d_w \geq 0$). *Under Assumptions 1 and 2, the superset of moments $\Theta^{superset,m}$ is defined as a hyperrectangle, which is a Cartesian product of interval bounds on the moment $E[x_k w_l], E[x_k y]$, whose infimum and supremum are identified from applying the Monotone Rearrangement Inequality (Lemma 1).*

$$\Theta^{superset,m} = \Theta_{xy} \times \Theta_{xw_1} \times \cdots \times \Theta_{xw_{d_w}}, \quad (11)$$

where $\Theta_{xy} \equiv [\inf E[x_1 y], \sup E[x_1 y]] \times \cdots \times [\inf E[x_{d_x} y], \sup E[x_{d_x} y]],$

$$\Theta_{xw_l} \equiv [\inf E[x_1 w_l], \sup E[x_1 w_l]] \times \cdots \times [\inf E[x_{d_x} w_l], \sup E[x_{d_x} w_l]].$$

Under Assumption 3, the superset of the parameters, $\Theta^{superset,p}$ is

$$\Theta^{superset,p} = \mathcal{B}_{OLS}(\Theta^{superset,m}). \quad (12)$$

\mathcal{B}_{OLS} is defined in the previous Remark 1.

The lower and upper bounds of each parameter $\{\alpha', \beta', \gamma'\}$ can be found by projecting the iden-

⁶However, this does not necessarily imply the tighter bound on the parameter of interest, $(\alpha', \beta', \gamma')$.

⁷However, $F(y|z)$ would be difficult to be degenerate when there is an error term in an equation for an outcome variable y .

tified set $\Theta^{superset,p}$ into each dimension.

$$\left[\alpha_l^{superset}, \beta_l^{superset}, \gamma_l^{superset} \right]' = \left[\inf Proj_1(\Theta^{superset,p}), \dots, \inf Proj_{d_x+d_z+d_w}(\Theta^{superset,p}) \right]'. \quad (13)$$

$$\left[\alpha_u^{superset}, \beta_u^{superset}, \gamma_u^{superset} \right]' = \left[\sup Proj_1(\Theta^{superset,p}), \dots, \sup Proj_{d_x+d_z+d_w}(\Theta^{superset,p}) \right]'. \quad (14)$$

Proof. In Online Appendix C. □

Proposition 3 (Topological Properties of Identified Supersets). .

Under Assumption 1, 2, 3, the following statements are true.

(1) $\Theta^{superset,m}$ is non-empty, compact, and convex.

(2) $\Theta^{superset,p}$ is non-empty and compact. When $d_w = 0$, $\Theta^{superset,p}$ is convex, but when $d_w > 0$, it may not be convex.

(3) $Proj \Theta^{superset,p}$ is non-empty, compact, and convex. It is a bounded interval on \mathbb{R} .

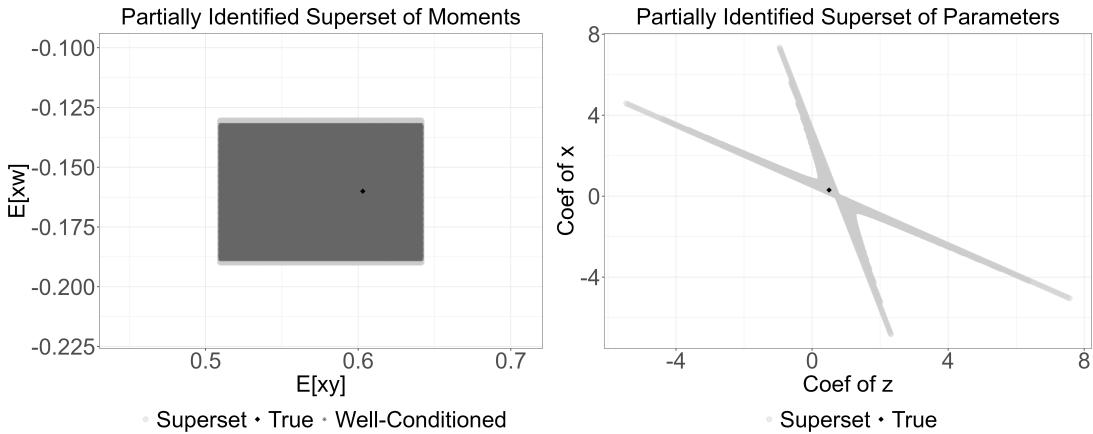
Proof. In Online Appendix C. □

Unlike the sharp set $\Theta^{\text{sharp},m}$, the superset $\Theta^{superset,m}$ does not consider potential dependence between moments, $(E[xy]', E[xw_1]', \dots, E[xw_{d_w}]')$, so $\Theta^{\text{sharp},m} \subset \Theta^{superset,m}$. The superset identification can be problematic when the model is misspecified, as explained in Li et al. (2020): that is, the two valid supersets of parameters in a misspecified model can be contradictory. This concern is less pronounced in this case, as only minimal assumptions (finite second moment of variance and covariance matrix and well-conditioned moment matrix) on the model were made.

4 Examples: Superset Identification

I illustrate the potential non-convexity of the identified set when $d_w > 0$ using a simple numerical example. In this setting, I assume $x \in \mathbb{R}$, $z \in \mathbb{R}$, $w \in \mathbb{R}$, and the main dataset contains

Figure 1: Set Identification with $d_w > 0$



(y, z, w) and the auxiliary dataset contains (z, x) . The chosen data-generating process ensures that the conditional distributions $y \mid z$, $x \mid z$, and $w \mid z$ are all normal, allowing for straightforward computation of the population lower and upper bounds of the moments $E[xy]$ and $E[xw]$ using Lemma 1.

Data Generating Process

$$z \sim \text{Normal}(0, 1). \quad (15)$$

$$x = 0.8z + \epsilon_1, \quad \epsilon_1 \sim \text{Normal}(0, 0.3). \quad (16)$$

$$w = -0.2z + \epsilon_2, \quad \epsilon_2 \sim \text{Normal}(0, 0.1). \quad (17)$$

$$y = 0.5z + 0.3x + 0.1w + \epsilon_3, \quad \epsilon_3 \sim \text{Normal}(0, 0.2). \quad (18)$$

The left panel of Figure 1 shows the identified superset of moments, $\Theta^{\text{superset},m}$, as a rectangular region with boundaries determined by monotone rearrangement inequality bounds on $E[xy]$ and $E[xw]$. Moments yielding well-conditioned matrices whose condition number is below 2000 appear in darker grey, forming a strict subset of $\Theta^{\text{superset},m}$ that contains the true population moment vector. Essentially, areas close to boundaries are excluded by the condition number restriction.

The right panel of Figure 1 shows the projection of $\mathcal{B}_{\text{OLS}}(\Theta^{\text{superset},m})$ onto the (x, z) coefficient plane, revealing a non-convex identified parameter set. This non-convexity means local

Figure 2: Set Identification with $d_w = 0$



optimization methods may fail to find valid bounds due to multiple local extrema, making global optimization methods like `DEoptim` ([Mullen et al. \(2011\)](#)) in R ideal for estimating the identified set's boundaries.

To contrast this with the $d_w = 0$ case, I use the same structural model for y but modify the data availability. Specifically, I re-label x and w as x_1 and x_2 , and assume that the main dataset contains (y, z) while the auxiliary dataset contains (z, x_1, x_2) . Thus, the joint distribution of (y, z, x_1, x_2) remains the same as that of (y, z, x, w) in the previous case, but only the informational structure across datasets differs.

The left panel of Figure 2 shows the identified superset of moments $(E[x_1y], E[x_2y])$, denoted $\Theta^{\text{superset},m}$. Note that the condition number constraint doesn't bind over all $\Theta^{\text{superset},m}$ in this case. Applying $\mathcal{B}_{\text{OLS}}(\cdot)$ yields $\Theta^{\text{superset},p}$, projected onto the (x_1, z) coefficient plane in the right panel.

The comparison with Figure 1 highlights the stark contrast in the topology of the identified parameter sets. While the $d_w > 0$ case yields a non-convex identified set, the $d_w = 0$ case results in a convex set with significantly tighter bounds. This contrast is particularly striking given that both scenarios are based on the same data-generating process; the only difference lies in the assignment of variables to the main and auxiliary datasets.

5 Estimation

I propose a plug-in estimator for a general case $d_w \geq 0$, where population moments in Proposition 2 are replaced with their sample counterparts. For a special case $d_w = 0$, I propose a much simpler estimator (Estimator 2) with a closed-form expression in Online Appendix Section B.

Estimator 1 (Estimator in a General Case $d_w \geq 0$). *Under Assumption 1, 2, 3, the set estimator for $\Theta^{superset,m}$ is obtained by applying the Monotone Rearrangement inequality.*

$$\widehat{\Theta}^{superset,m} = \widehat{\Theta}_{xy} \times \widehat{\Theta}_{xw_1} \times \cdots \times \widehat{\Theta}_{xw_{d_w}}, \quad (19)$$

$$\text{where } \widehat{\Theta}_{xy} \equiv [\widehat{\mu}_l^{x_1y}, \widehat{\mu}_u^{x_1y}] \times \cdots \times [\widehat{\mu}_l^{x_{d_x}y}, \widehat{\mu}_u^{x_{d_x}y}], \quad \widehat{\Theta}_{xw_k} \equiv [\widehat{\mu}_l^{x_1w_k}, \widehat{\mu}_u^{x_1w_k}] \times \cdots \times [\widehat{\mu}_l^{x_{d_x}w_k}, \widehat{\mu}_u^{x_{d_x}w_k}].$$

$$\widehat{\mu}_l^{x_kw_l} \equiv \frac{\sum_{i=1}^{N_m} w_{li} \widehat{Q}_{x_k|z}(1 - \widehat{F}_{w_l|z}(w_{li}|z_i)|z_i)}{N_m}, \quad \widehat{\mu}_u^{x_kw_l} \equiv \frac{\sum_{i=1}^{N_m} w_{li} \widehat{Q}_{x_k|z}(\widehat{F}_{w_l|z}(w_{li}|z_i)|z_i)}{N_m}.$$

$$\widehat{\mu}_l^{x_ky} \equiv \frac{\sum_{i=1}^{N_m} y_i \widehat{Q}_{x_k|z}(1 - \widehat{F}_{y|z}(y_i|z_i)|z_i)}{N_m}, \quad \widehat{\mu}_u^{x_ky} \equiv \frac{\sum_{i=1}^{N_m} y_i \widehat{Q}_{x_k|z}(\widehat{F}_{y|z}(y_i|z_i)|z_i)}{N_m}.$$

N_m is the sample size of the main data. $\widehat{Q}_{\cdot|z}(\cdot|z)$ is the estimate of **conditional Quantile Function** and $\widehat{F}_{\cdot|z}(\cdot|z)$ is the estimate of **conditional Cumulative Distribution Function**.

The set estimator for $\Theta^{superset,p}$ is

$$\widehat{\Theta}^{superset,p} = \widehat{\mathcal{B}}_{OLS}(\widehat{\Theta}^{superset,m}), \quad (20)$$

$$\text{where } \widehat{\mathcal{B}}_{OLS}(M) \equiv \left\{ \underbrace{\begin{bmatrix} \widehat{s}_{xx'} & \widehat{s}_{xz'} & A' \\ \widehat{s}_{zx'} & \widehat{s}_{zz'} & \widehat{s}_{zw'} \\ A & \widehat{s}_{wz'} & \widehat{s}_{ww'} \end{bmatrix}}_{(**)}^{-1} \begin{bmatrix} B \\ \widehat{s}_{zy} \\ \widehat{s}_{wy} \end{bmatrix} \middle| m \equiv \begin{bmatrix} B \\ A_1 \\ \vdots \\ A_{d_w} \end{bmatrix} \in M, \text{ s.t. } \kappa(m) \leq \kappa_0 \right\}. \quad (21)$$

The function $\widehat{\mathcal{B}}_{OLS}$ is the sample analogue of function \mathcal{B}_{OLS} , where population moments in \mathcal{B}_{OLS} are replaced with their sample counterparts. \widehat{s}_{\cdot} denote the sample analog of the population moment $E[\cdot]$. Here, B and A_k (k -th column of a matrix A) represent the sample unidentified moments \widehat{s}_{xy} and \widehat{s}_{wx_k} . κ_0 refers to the upper bound for the condition number in Assumption 3. The

lower and upper bound estimates of parameters are the infimum and supremum of the projection of the set $\widehat{\Theta}^{superset,p}$ into each dimension.

$$\left[\widehat{\alpha}_l^{superset}, \widehat{\beta}_l^{superset}, \widehat{\gamma}_l^{superset} \right]' = \left[\inf Proj_1(\widehat{\Theta}^{superset,p}), \dots, \inf Proj_{d_x+d_z+d_w}(\widehat{\Theta}^{superset,p}) \right]'. \quad (22)$$

$$\left[\widehat{\alpha}_u^{superset}, \widehat{\beta}_u^{superset}, \widehat{\gamma}_u^{superset} \right]' = \left[\sup Proj_1(\widehat{\Theta}^{superset,p}), \dots, \sup Proj_{d_x+d_z+d_w}(\widehat{\Theta}^{superset,p}) \right]'. \quad (23)$$

Estimator 1 for $(\widehat{\alpha}_l^{superset}, \widehat{\beta}_l^{superset}, \widehat{\gamma}_l^{superset})'$, $(\widehat{\alpha}_u^{superset}, \widehat{\beta}_u^{superset}, \widehat{\gamma}_u^{superset})'$ has a value function representation, which is useful for implementation.

Remark 3 (Value Function Representation). Let m be the vector of length $(d_x + d_x d_w)$ denoting unidentified moments $(E[xy]', E[xw_1]', \dots, E[xw_{d_w}'])$. That is, $m \equiv (m_1, \dots, m_{d_x+d_x d_w}) \in \mathbb{R}^{d_x+d_x d_w}$. Define the vector of sample moments

$\widehat{\theta} \equiv (\widehat{\mu}'_l, \widehat{\mu}'_u, \text{vec}(\widehat{s}_{xx'}), \text{vec}(\widehat{s}_{xz'}), \text{vec}(\widehat{s}_{zz'}), \text{vec}(\widehat{s}_{zw'}), \text{vec}(\widehat{s}_{ww'}), \text{vec}(\widehat{s}_{zy}), \text{vec}(\widehat{s}_{wy}))'$, where $\widehat{\mu}'_l \equiv (\widehat{\mu}_l^{xy}, \widehat{\mu}_l^{xw_1}, \dots, \widehat{\mu}_l^{xw_{d_w}})$, $\widehat{\mu}'_u \equiv (\widehat{\mu}_u^{xy}, \widehat{\mu}_u^{xw_1}, \dots, \widehat{\mu}_u^{xw_{d_w}})$, and $\text{vec}(\cdot)$ is a vectorization operator that converts matrices into row vectors. Let e_k denotes the k -th standard basis vector belonging to $\mathbb{R}^{d_x+d_z+d_w}$. Then, the Estimator 1 for $(\widehat{\alpha}_l^{superset}, \widehat{\beta}_l^{superset}, \widehat{\gamma}_l^{superset})'$ has the following value function representation. The estimator for the upper bounds can be defined by replacing ‘min’ with ‘max’.

$$\left[\widehat{\alpha}_l^{superset}, \widehat{\beta}_l^{superset}, \widehat{\gamma}_l^{superset} \right] = \left[\phi_1(\widehat{\theta}), \dots, \phi_{d_x+d_z+d_w}(\widehat{\theta}) \right], \quad (24)$$

$$\text{where } \phi_k(\widehat{\theta}) = \min e'_k \widehat{\mathcal{B}}_{OLS}(\{m\}), \quad (25)$$

$$s.t. \quad \widehat{\mu}_l^{x_1y} \leq m_1 \leq \widehat{\mu}_u^{x_1y},$$

$$\vdots$$

$$\widehat{\mu}_l^{x_{d_x}w_{d_w}} \leq m_{d_x+d_x d_w} \leq \widehat{\mu}_u^{x_{d_x}w_{d_w}},$$

$$\kappa(m) \leq \kappa_0.$$

The next proposition states the consistency results. The consistency proof requires that the estimators for the conditional CDF and the conditional quantile function jointly satisfy the uniform convergence condition. Also, it requires that the set-valued function $\mathcal{B}_{OLS}(\cdot)$ is continuous in the neighborhood of $\Theta^{\text{superset},m}$.

Assumption 4 (Uniform Convergence Assumption for Consistency). .

The estimators for the conditional CDF and the conditional quantile function satisfy the following uniform convergence condition⁸:

$$\lim_{N_a(N_m) \rightarrow \infty} \lim_{N_m \rightarrow \infty} \sup_{(y,z) \in \mathcal{Y} \times \mathcal{Z}} |(\widehat{Q}_{x_k|z}(\widehat{F}_{y|z}(y|z)|z) - Q_{x_k|z}(F_{y|z}(y|z)|z))| \xrightarrow{p} 0, \quad (26)$$

$$\lim_{N_a(N_m) \rightarrow \infty} \lim_{N_m \rightarrow \infty} \sup_{(w_l,z) \in \mathcal{W}_l \times \mathcal{Z}} |(\widehat{Q}_{x_k|z}(\widehat{F}_{w_l|z}(w_l|z)|z) - Q_{x_k|z}(F_{w_l|z}(w_l|z)|z))| \xrightarrow{p} 0, \quad (27)$$

$$\text{for } 1 \leq \forall k \leq d_x, \quad 1 \leq \forall l \leq d_w$$

where N_m is the sample size of the main data, $N_a(N_m)$ is the sample size of the auxiliary data with $N_a(N_m) \rightarrow \infty$ as $N_m \rightarrow \infty$, and $\mathcal{Y} \times \mathcal{Z}$ and $\mathcal{W}_l \times \mathcal{Z}$ denotes the support of (y, z) and (w_l, z) , respectively.

Proposition 4 (Consistency). *Under Assumptions 1, 2, 3, 4, the estimator 1 is consistent. That is, as $N_m \rightarrow \infty$ and subsequently $N_a(N_m) \rightarrow \infty$:*

$$(1) \mathbf{d}_H(\widehat{\Theta}^{\text{superset},m}, \Theta^{\text{superset},m}) \xrightarrow{p} 0.$$

$$(2) \mathbf{d}_H(\widehat{\Theta}^{\text{superset},p}, \Theta^{\text{superset},p}) \xrightarrow{p} 0.$$

$$(3) \|(\widehat{\alpha}_l^{\text{superset}'}, \widehat{\beta}_l^{\text{superset}'}, \widehat{\gamma}_l^{\text{superset}'}) - (\alpha_l^{\text{superset}'}, \beta_l^{\text{superset}'}, \gamma_l^{\text{superset}'})\|_2 \xrightarrow{p} 0.$$

$$\|(\widehat{\alpha}_u^{\text{superset}'}, \widehat{\beta}_u^{\text{superset}'}, \widehat{\gamma}_u^{\text{superset}'}) - (\alpha_u^{\text{superset}'}, \beta_u^{\text{superset}'}, \gamma_u^{\text{superset}'})\|_2 \xrightarrow{p} 0.$$

$\mathbf{d}_H(A, B)$ is the Hausdorff distance between sets A and B , defined as:

⁸This uniform convergence condition will hold, for example, when the estimators for the conditional CDF and the conditional quantile function converge uniformly to the true functions and when the conditional quantile function is continuous. If the support of z is discrete and finite, the empirical conditional CDF, $\widehat{F}_{y|z}(\tilde{y}|z) = \frac{1}{N} \sum_{i=1}^N 1\{y_i \leq \tilde{y}|z_i = z\}$, uniformly converges to the true conditional CDF, $F_{y|z}(y|z)$, almost surely by the Glivenko-Cantelli Theorem (Van der Vaart (2000), Theorem 19.1). When z includes continuous covariates, one may discretize them and compute the empirical conditional CDF, or one may use a parametric estimator for $F_{y|z}$ under the assumption that the assumed parametric distribution is a good approximation of the true distribution. Then, the conditional quantile function can be computed as the inverse of the empirical or parametrized conditional CDF.

$$\mathbf{d}_H(A, B) = \max\{\sup_{a \in A} \mathbf{d}(a, B), \sup_{b \in B} \mathbf{d}(A, b)\}, \text{ with } \mathbf{d}(a, B) = \inf_{b \in B} \|a - b\|_2.$$

Proof. In the Online Appendix Section C. □

The consistency result implies that finite sample bias in the estimators vanishes as sample size increases. While Chernozhukov et al. (2013) noted that sample analog estimators involving min and max operators exhibit finite sample bias, I deliberately choose not to correct for this bias in practice for two reasons. First, Jensen's inequality ensures that my estimators provide wider interval bounds than the population parameters on average: $E[\min(X, Y)] \leq \min(E(X), E(Y))$ and $\max(E(X), E(Y)) \leq E[\max(X, Y)]$. Consequently, the mean of the sample analog estimator of the form $[E[\min(X, Y)], E[\max(X, Y)]]$ is wider than the population interval bound $[\min(E(X), E(Y)), \max(E(X), E(Y))]$, making the inference conservatively valid without bias correction. Second, in practice, there is a risk of over-correcting the finite sample bias. In Online Appendix Section D.1, I compare my estimators with Pacini (2019)'s approach, which deploys a bootstrap method to correct for finite sample bias. This comparison demonstrates that Pacini (2019)'s method tends to overcorrect the bias, while my uncorrected estimator produces results closer to the population bounds. For further details on this comparison, please refer to the aforementioned Online Appendix section.

Before closing this section, let me discuss another common concern when using a two-sample estimator in practice; that is, the two samples used for estimation can be imbalanced. In that case, the assumption that both samples were generated from the same data generating process as in the population is violated.⁹ To address this concern, researchers can reweight the samples using post-stratification weights. The key assumptions to use post-stratification weights are stated in Assumption 5.¹⁰ First, for every stratification cell, the conditional distributions of variables (y, x, z, w) must be identical in population, main, and auxiliary samples, although the marginal of the stratification cell variable s can be different. Second, the support of the stratification cell variable must be the same in population, main and auxiliary

⁹Remember that the estimators need not only the conditional distribution estimates, $\hat{F}_{y|z}$, $\hat{F}_{w_l|z}$, $\hat{Q}_{x_k|z}$, but also the second-order moment estimates, such as $\hat{s}_{xx'}$, $\hat{s}_{xz'}$, $\hat{s}_{zz'}$. Therefore, the same conditional distributions of x given z , and y given z in both samples are not sufficient. We need a stronger assumption that marginal distributions from both samples are identical.

¹⁰Similar assumption has been used in Graham et al. (2016).

samples. Under Assumption 5, researchers can compute the weighted estimator using the post-stratification weight. See Johnson (2008) and Royal et al. (2019) for how to construct post-stratification weights in data.

Assumption 5 (Reweighting). *Let s be the stratification cell variable used to compute the post-stratification weights.*

1. (Conditional Distributional Equality) Assume $f_{pop}(y, x, z, w|s) = f_{main}(y, x, z, w|s) = f_{aux}(y, x, z, w|s)$ for $\forall s, y, x, z, w$.
2. (Common Support) $supp_{pop}(s) = supp_{main}(s) = supp_{aux}(s)$.

6 Empirical Application: Estimating Neighborhood Effects on Intergenerational Cultural Transmission

In this section, I estimate neighborhood effects on intergenerational cultural transmission among second-generation immigrants. Motivated by the literature (e.g., Bisin and Verdier (2000), Bisin and Verdier (2001)), I assume that second-generation immigrants form their cultural preferences under the joint influence of their parents and the neighborhoods in which they grow up. Accordingly, when estimating the effect of childhood neighborhood environment on adulthood outcomes, it is essential to control for parental cultural preferences to avoid omitted variable bias, as these preferences are likely to be correlated with residential location.

In practice, however, available data often lack reliable measurements of individual-level heterogeneity, such as cultural preferences—as is the case here. I apply the proposed estimators to two separate datasets: using the notation in equation 1, the main dataset, derived from longitudinal administrative records, includes (y, z, w) ; and the auxiliary dataset, based on a cross-sectional survey, includes (x, z) . In this setting, the outcome variable y is an adulthood outcome that reflects the second-generation immigrant’s cultural preference. The missing

variables in the main data, denoted by x , include a vector of discrete parental cultural preference indicators $\{1(u^P = k)\}_{k=2}^{\bar{u}}$, where the true latent preference type u^P can take on a finite number of values $\{1, \dots, \bar{u}\}$ with $\bar{u} > 1$.

The common covariate vector z includes a dummy variable $n \in \{0, 1\}$ indicating whether the individual grew up in an ethnic enclave, as well as other observable parental characteristics: mothers' religion, ethnicity¹¹, and both parents' cohort, education, immigration year group, and employment status during the individual's childhood (ages 0–10). The auxiliary dataset contains all relevant explanatory variables, and therefore, a much simpler Estimator 2 for $d_w = 0$ case is applicable. However, to demonstrate the application of Estimator 1 for the case $d_w > 0$, I intentionally treat parental religion as unobserved in the auxiliary data. This results in two missing maternal religion dummies (indicating whether the mother is Muslim/Sikh), implying $d_w = 2$.

This model implicitly assumes that the only relevant variables missing from the main dataset—but necessary to explain the outcome variables—are the parental cultural preference dummies, $\{1(u^P = k)\}_{k=2}^{\bar{u}}$.¹² In other words, the model assumes that the only unobserved parental trait affecting second-generation outcomes is cultural preference u^P , and that the observable vector of childhood parental characteristics, (z, w) , is comprehensive.¹³

The key parameter of interest is the coefficient in front of the childhood neighborhood dummy variable n . Denote this parameter as β_n . This parameter determines the effect of second-generation immigrants' childhood neighborhoods on their cultural attitudes. The main hypothesis that I consider is whether there is any significant neighborhood effect, that

¹¹In sample, the correlation between couple's religion and ethnicity was very high, leading to ill-conditioned moment matrix. Therefore, I included only mothers' religion and ethnicity in z and limited the use of fathers' religion and ethnicity to only when applying the Monotone Rearrangement inequality.

¹²In principle, one could allow for additional omitted parental characteristics (e.g., income or aspirations for children's economic success). However, I choose not to do so here, as incorporating too many omitted variables may lead to uninformative bounds. Importantly, the methodology employed remains applicable to a more general model, provided the auxiliary data contain all relevant sources of confounding.

¹³Parental income during early childhood is a potential omitted variable. However, the current specification of (z, w) is already extensive, and the omission of income is unlikely to meaningfully bias results. While parental income is available in the auxiliary survey data, it is not available in the administrative data. Using the auxiliary data, I examined whether household income explains ethnic enclave residency after controlling for all other covariates related to the female spouse. Because the auxiliary survey is not a household panel, male spouse variables are not observed. Even with this limitation, household income does not have significant explanatory power for ethnic enclave residency, suggesting that the effect of excluding parental income is minimal.

is, $H_0 : \beta_n = 0$. Note that the OLS estimand β_n is not equal to the average treatment effect (ATE) in general, but there is a special case when it is equal to the ATE: recently, Słoczyński (2022) has shown that the OLS estimand is equal to the ATE under a few assumptions.¹⁴ One of the assumptions required for this result is that the share of treated group is 0.5. In our sample, the share of second generation immigrants who grew up in an ethnic enclave is close to 0.5. So β_n in our case would be close to the ATE, if the other assumptions made in Słoczyński (2022) hold.

To estimate the sharp bounds on the second-order moment, I parametrize the conditional CDF and quantile function, because the conditioning variable vector z is high-dimensional and our sample size is not large enough to estimate the distributions nonparametrically.¹⁵ Specifically, I assume the parental cultural type u^P follows a multinomial logit model, conditioning on z , and the adulthood outcome variable y , which is discrete in this case, follows a logit model conditioning on z . When I apply the most general case estimator 1, I exclude maternal religion dummy variables from the conditioning variable z . The model fit of the multinomial logit model for u^P is good in our sample.¹⁶

6.1 Data

I use two different datasets for empirical analysis. The first data is the ONS Longitudinal Study (LS, hereafter, Office for National Statistics (2019)), which is a 1% random sample of

¹⁴This result is stated in Corollary 4 of Słoczyński (2022). The key assumptions to obtain this causal interpretation are Assumption 3, 4, 5 in Słoczyński (2022), and the condition that the share of the treated group is 0.5. Assumption 3 states that the average potential outcome is conditionally independent of treatment, conditional on other controlled covariates. This would hold in our case under the exogeneity assumption, $E[(x', z', w')'\epsilon] = 0$. Assumption 4 states that the means of potential outcomes are linear in the propensity score. Assumption 5 states that the conditional variances of the propensity score, conditional on treatment, are homogenous. Both Assumption 4 and Assumption 5 are less commonly used in the literature. Finally, the share of treated group in our sample is close to 0.5.

¹⁵If sample size allows, deploying a kernel estimator or sieve estimator could give more robust distributional estimates for y and u^P .

¹⁶McFadden's pseudo R^2 is 0.20 for the male sample and 0.24 for the female sample using the baseline control variables for z . When the coefficients are estimated via OLS, the R^2 of the model is between 0.07 and 0.32.

the census between 1971 and 2011 in England and Wales.¹⁷ The LS data is long enough to track the second-generation immigrants' neighborhoods and their adulthood outcomes. The second data is the Fourth National Survey of Ethnic Minorities (FNSEM, hereafter), 1993–94, which is a cross-sectional survey for ethnic minorities in England and Wales (Berthoud et al. (1997)). The FNSEM data includes detailed information about the first-generation immigrants' cultural preference and their neighborhood characteristics.

In the LS data, I use a sample of second-generation South Asian (Indian, Pakistani, and Bangladeshi) immigrants who were born before the year 1992 for analysis.¹⁸ South Asian immigrants are one of the largest ethnic minority groups in the UK, and they constitute about 5% of the population in the 2011 census. Many of them live in segregated neighborhoods. Second-generation immigrants are defined as people who were born in the UK and had at least one parent who was born abroad. I drop second-generation immigrants who were under 20 years of age in 2011 because I cannot observe adulthood outcomes for these individuals. I use a sample with complete information on parental characteristics.

I use seven adulthood outcomes for second-generation immigrants in the last Census year, that is, 2011, that indicate cultural preferences. These outcomes are female employment, female college graduation, whether married to a member of the same ethnicity, whether living in an ethnic enclave, whether married to a spouse with the same religion, whether having the same religion as their father, and whether having the same religion as their mother. The first two outcome variables reflect how much second-generation South Asian female immigrants deviate from traditional gender roles. The next two outcome variables are popular measures of interethnic attitudes and social integration. The last three outcome variables show the interreligious attitude and intergenerational religion transmission.

In the FNSEM data, I use a sample of first-generation South Asian immigrants who are

¹⁷To provide additional details about the LS data, the initial sample was drawn from the 1971 census on the basis of four birth dates, and study members' census records have been linked every 10 years up to the 2011 census. New LS members enter the study, on the basis of the four birth dates, through birth or immigration, and existing members leave through death or emigration, although their data are retained. Vital life events information (births, deaths and cancer registrations) are also linked to sample members' records. See [Shelton et al. \(2019\)](#) for more information about the dataset.

¹⁸In LS data, ethnicity was measured only since 1991. Therefore, the LS sample excludes any second-generation South Asian immigrants who never participated in the census after the year 1981.

married and are parents at the time of the survey (i.e., the years 1993 and 1994). I use this sample to learn about the cultural preferences of the parents of second-generation immigrants in the LS data under the assumption that the two data are comparable. First-generation immigrants are defined as people who were born outside of the UK. I use a subsample of the FNSEM data with complete parental characteristics information and at least one proxy variable for cultural preferences.

Despite the advantage of using two samples, potential sample imbalance between two datasets can be a concern. The LS data is a nationally representative sample, but the FNSEM is not. Table E.5 in Online Appendix shows that unweighted FNSEM data shows small but distinguishable differences in key demographic characteristics of the first-generation immigrants.

To address this sample imbalance, I reweight the FNSEM sample to balance the first-generation immigrant parents' key characteristics (sex, education, ethnicity, female employment, enclave residency)¹⁹. After reweighting the FNSEM data, most characteristics are indistinguishable between the two datasets; see Table E.4.²⁰ The only distinguishable characteristics are the father's age and father's employment status between ages 0 and 9: males in FNSEM data are 1.6 years younger and 14% less likely to work than the fathers of the LS sample, while their children are between ages 0 and 9.²¹ All maternal characteristics, including education level and employment status, are statistically indistinguishable in the two datasets.

To measure parental cultural preference in the FNSEM data, I use four groups of self-

¹⁹The FNSEM is not a household survey. Therefore, I assign an equal weight of 50% to male and female FNSEM samples, respectively. While computing the expected type of male FNSEM sample, I use their ethnicity and religion information, as spouse information is not available. In LS data, I estimate mothers' and fathers' cultural preference type probabilities separately and take an average to compute the probabilities of parents' representative cultural type. Note that I have to use the father's ethnicity and religion information in the LS data to compute the father's cultural preference type to utilize the FNSEM sample estimates.

²⁰I do not reweight the LS sample because the LS sample is a random sample of the population in the census and thus it is a representative sample. The FNSEM is a survey sample that was designed to oversample ethnic minorities.

²¹Gap in the father's employment rate can be further reduced by reweighting the FNSEM data along with father's employment rate. However, doing so will make the two samples unbalanced along other dimensions, such as parents' ethnicities and religions. It is well known that perfectly balancing two samples in every dimension through reweighting is difficult when there are many dimensions to reweight. I decided to prioritize balancing other dimensions (parents' ethnicity, religion, and mother's characteristics) since they seem more important in cultural transmission.

reported variables: non-Western religiosity, English fluency and usage, frequency of wearing ethnic clothes, and interracial marriage preference. Table E.1 summarizes the questionnaires for these variables. Note that I define non-Western religiosity as adherence to either Islam, Hinduism, or Sikhism.

I construct an ethnic enclave dummy variable n in both datasets using consistent geography over time – ward administrative units in the year 1991. Wards are small geographic units; there were 9527 wards (with any residents) in England and Wales, and the average number of residents per ward was 5500. I define a neighborhood as an ethnic enclave if its own ethnic group share in the neighborhood is above 10%.²² Finally, a childhood neighborhood is defined as a neighborhood lived before age 10, which is measured in a decennial census. Theoretically, one may attempt to distinguish the effect of childhood neighborhood of residence before age 10 from the effect of childhood neighborhood of residence after age 10. However, high persistence in neighborhood choice among our sample makes it difficult to separately estimate these effects: 86% of those who lived in an ethnic enclave between age 0 and 9 continue to live in an ethnic enclave between age 10 and 19 (Table E.12 in Online Appendix). Therefore, I report the neighborhood effect estimates using the residency before age 10.

6.2 Estimation Results

I preset the number of cultural preference types to three, which is the maximum number of types I can estimate with the data at hand. The choice of the number of types is constrained by available proxy variables. Many of the proxy variables available in the data are very coarse,

²²It is infeasible to use a continuous variable of its own ethnic group share instead of a binary ethnic enclave variable. To protect the confidentiality, the LS data restricts access to exact neighborhood geocodes or own ethnic group share. I asked an LS staff member to construct an ethnic enclave dummy variable for my project. The dummy variable is coarse, so it does not allow identifying each neighborhood.

To determine the optimal cutoff for defining an ethnic enclave, I first apply a k-means clustering algorithm to local area aggregate statistics from the census, grouping neighborhoods based on similar ethnic group shares. I then select the cutoff value from the FNSEM data's categorical variable for same-ethnicity neighborhood share that is closest to the threshold identified by the k-means algorithm. This value turns out to be 10%. The FNSEM data provides only a categorical variable for the share of same-ethnicity individuals in a ward, with the following cutoff intervals: 0–1.99%, 2–4.99%, 5–9.99%, 10–14.99%, 15–24.99%, 25–32.99%, and 33% or more.

with only 2–4 choice options (Table E.1). If all proxy variables exhibit a full-rank measurement structure matrix, defined in Assumption A.1 (3), then I can estimate at most four types. However, EM algorithm fails to estimate the measurement structure matrices for four types, which indicates a lack of variation in data to identify four types.

To satisfy Assumption A.1 (3) with three types, we need to show there are two proxies whose measurement structure matrices have at least rank three, $\text{rank}(M_k), \text{rank}(M_{k'}) \geq 3$, and there is another proxy whose measurement structure matrix rank is no less than 2. I show a sufficient condition of this; there are four proxy variables with matrix rank no less than three. To confirm this, I perform a Kleibergen and Paap (2006)'s matrix rank test on the matrix of form $M_{kk'}$, whose ij -th entry is $P(m_k = i\text{-th largest value} \text{ and } m_{k'} = j\text{-th largest value})$.²³ I perform this matrix rank test twice, each for proxy pairs, (*Religious Importance* and *English Fluency*), and (*Whether Interviewed in English* and *Wearing Ethnic Cloth*), and confirmed their matrix rank is strictly greater than 2. See Table E.2 in the Appendix. Therefore, the Assumption A.1 (3) holds if we assume three cultural preference types.

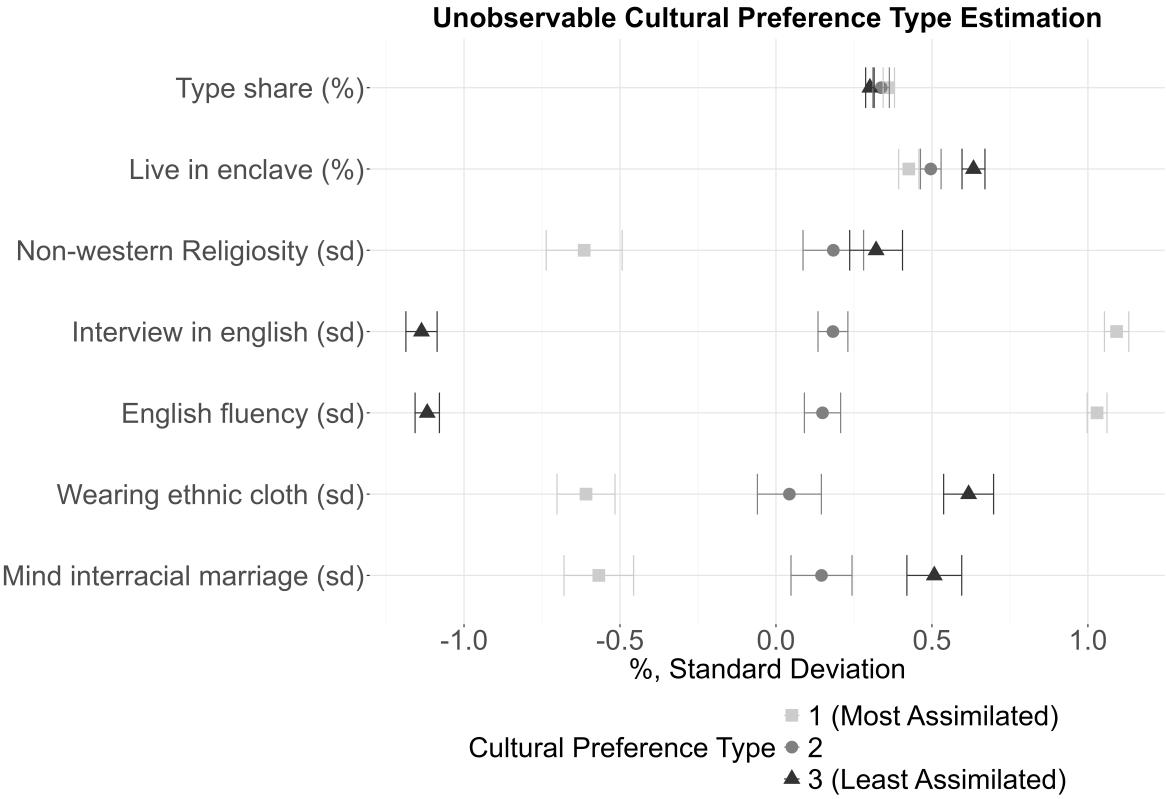
For identification, I make monotonicity assumption (4)(i) in Assumption A.1.²⁴ For the proxy variable satisfying the monotonicity relationship, I use a variable on whether the interview was done in English. I label the type whose interview was mostly done in another language as Type 3 (least assimilated) and the type whose interview was mostly done in English as Type 1 (most assimilated). Therefore, a higher type number corresponds to less cultural assimilation. Through robustness checks, I confirm that using a different proxy variable for the monotonicity assumption does not change the results (see Figure E.3 in the Online Appendix).

I estimate the cultural preference type probability for *every observation* using proxy variables in the FNSEM data. I apply the Arcidiacono and Jones (2003) EM-algorithm estimator that can be applied to a finite mixture model, as is the case here. Once I get the cultural pref-

²³To see why this is sufficient, consider that $\text{rank}(M_{kk'}) \leq \min(\text{rank}(M_k), \text{rank}(M_{k'}))$; This is because $M_{kk'} = M_k'D_xM_{k'}$, where D_x is the diagonal matrix, whose i -th diagonal entry is equal to $P(x = i\text{-th largest value})$. And applying the matrix rank rule, $\text{rank}(A, B) \leq \min(\text{rank}(A), \text{rank}(B))$, gives the result.

²⁴See Hu (2017) for an explanation on the underidentification when we do not utilize this kind of assumption.

Figure 3: Estimates of First-Generation Immigrants' Cultural Preference Type



Data Source: FNSEM

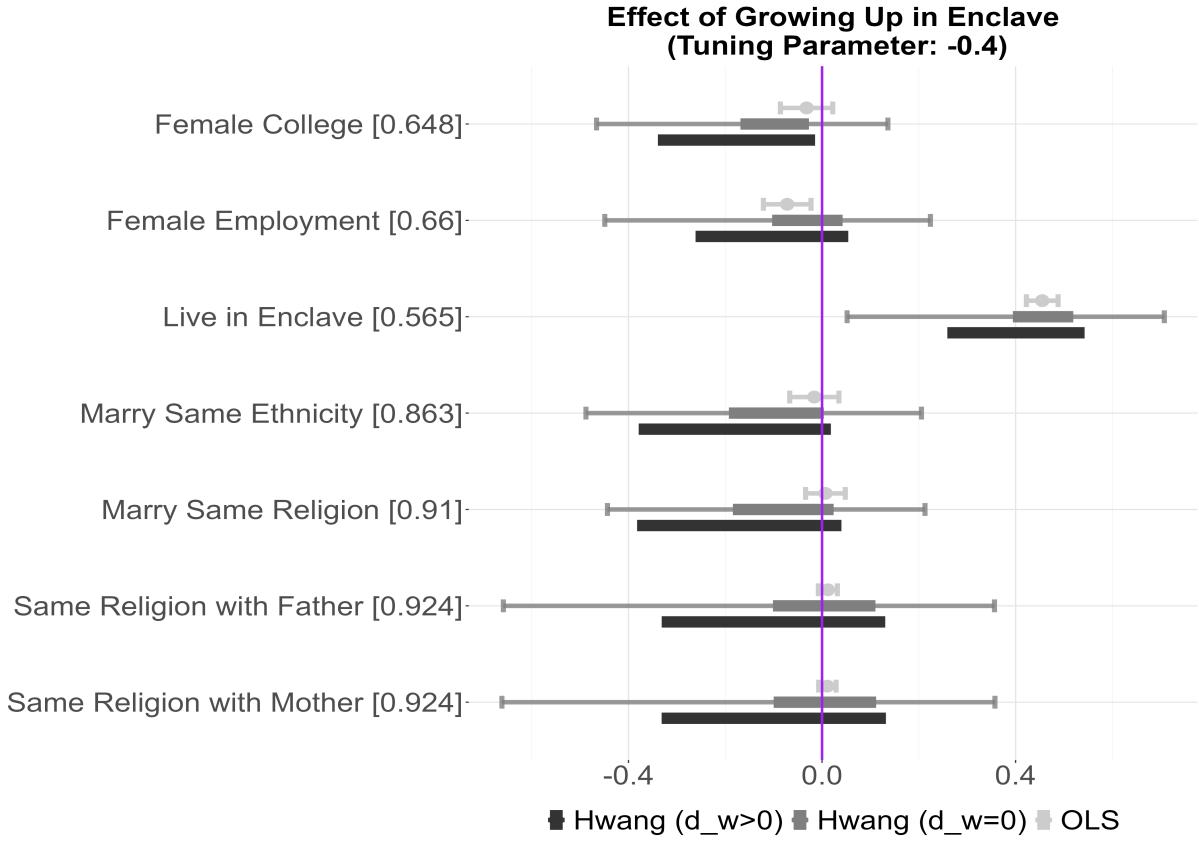
Note: The whisker bar denotes the 90% confidence interval around the average estimates. The units of 'Type share' and 'Live in enclave' are percentage points, and the units of other variables are standard deviations.

erence type probability estimates for every observation, I run the weighted multinomial logit regression to estimate the parameters for the conditional CDF $F(u^P|z)$ and its inverse. I also compute the average cultural type share.

Figure 3 shows the estimated cultural preference types of first-generation immigrant samples in FNSEM and their average characteristics. The shares of each type are close to one-third. The average characteristics of each cultural type are intuitive; the less assimilated groups are more likely to live in an ethnic enclave, have higher non-Western religiosity, conduct their interviews in languages other than English, show little English fluency, wear ethnic clothing more frequently, and dislike interracial marriage.

Next, I estimate the two-sample bounds on the regression coefficient β_n using the FNSEM estimates and I compare them with OLS estimates from a short regression without x . I apply both the Estimator 1 assuming $d_w > 0$ and Estimator 2 assuming $d_w = 0$. In this case,

Figure 4: Neighborhood Effects on Intergenerational Cultural Transmission



Data Source: FNSEM and ONS LS

Note: The graph presents the OLS point estimate from a short regression without x and two “Hwang” interval estimates of the neighborhood effect (β_n) under the assumptions of $d_w = 0$ and $d_w > 0$. The point estimates are displayed as dots and the interval estimates are shown as segment bands. The 90% confidence intervals for the “OLS” and “Hwang ($d_w = 0$)” estimates are drawn as whiskers. The 90% confidence intervals for “Hwang ($d_w = 0$)” estimates connect the lower end of the 90% CI for the lower bound estimate and the upper end of the 90% CI for the upper bound estimate and they are computed by Hong and Li (2018) numerical delta method described in Online Appendix section B. The tuning parameter for the numerical delta method is $N_m^{-0.4}$, where N_m is the sample size of the LS data.

Dependent variables are labeled on the left side of the figure. Numbers inside the brackets next to each dependent variable label are the averages of the dependent variables. The (z', w') controls include mothers’ religion, ethnicity and both parents’ cohort, education, immigration year group dummies, and employment status when the child was between age 0 and 10. For “Hwang ($d_w > 0$)” estimates, I assume the auxiliary data does not include two maternal religion dummy variables (whether a mother is a Muslim/Sikh).

there are two omitted variables, $1(u^P = 2)$, $1(u^P = 3)$, assuming three cultural preference types (that is, $d_x = 2$). When applying Estimator 2, the auxiliary data is assumed to include all right-hand side variables, so $d_w = 0$. When applying the Estimator 1, I additionally drop parental religion dummy variables²⁵ (whether a mother is a Muslim/Sikh) from auxiliary data, so $d_w = 2$ in this case.

²⁵I also assume father’s ethnicity and religion information is not available in the LS data while estimating father’s expected cultural preference type to apply monotone rearrangement inequality.

For the 90% confidence intervals of “Hwang ($d_w = 0$)” interval estimates, I display an interval connecting the lower end of the 90% confidence interval for the lower bound estimate and the upper end of the 90% confidence interval for the upper bound estimate.²⁶ The confidence interval for each lower and upper bound estimates are computed by Hong and Li (2018) numerical delta method described in Online Appendix Section B. The tuning parameter for the numerical delta method is set to $N_m^{-0.4}$, where N_m denotes the sample size of the LS data. This choice is used in the baseline results because it yields coverage rates closest to 90% across the considered tuning parameters (Table E.6) for the “Hwang ($d_w = 0$)” estimator. Estimation results using alternative tuning parameters (−0.49 and −0.3) are reported in Online Appendix Section F. The qualitative conclusions remain robust across these specifications.

Figure 4 summarizes the results: Table E.7 in the Appendix shows the numbers behind the figure; The point OLS estimates are displayed as dots and the interval estimates for “Hwang ($d_w > 0$)” and “Hwang ($d_w = 0$)” estimators are shown as segment bands. The 90% confidence intervals are drawn as whiskers. Dependent variables are listed on the left side of the figure. Numbers inside the brackets next to each dependent variable label are the averages of the dependent variables. All specifications include a comprehensive set of parental characteristics as controls. For both the OLS and Hwang ($d_w = 0$) estimates, these controls include: maternal religion and ethnicity, both parents’ birth cohorts, education levels, immigration year group indicators²⁷, and employment status during the child’s early development (ages 0-10). The Hwang ($d_w > 0$) estimates use the same control variables except for maternal religion, which are intentionally excluded to demonstrate the method’s application when some covariates are unavailable.

For interpretation, I take the “Hwang ($d_w = 0$)” estimates as baseline, as it is the only method that is equipped with inference and avoids the omitted variable bias. Figure 4 shows that there is a significant and sizable neighborhood effect for ethnic enclave residency. To

²⁶As discussed at the end of Online Appendix Section B, the coverage rate of this confidence interval for the parameter of interest is greater than $(1 - \tau)$ as shown in Imbens and Manski (2004), although it is difficult to tell precisely how greater the coverage rate of this interval is.

²⁷This is a binary indicator for whether arrival occurred before 1970. I retain observations with missing immigration year information (coded as 0) to maintain an adequate sample size, as excluding these cases would result in a sample too small for reliable analysis.

interpret the magnitude of the “Hwang ($d_w = 0$)” interval estimates, the second-generation South Asian immigrants who grew up in an ethnic enclave are 39 to 52 percentage points (representing 69 to 92% of the mean probability) more likely to live in an ethnic enclave in the adulthood²⁸. However, the statistical evidence on the neighborhood effect for other outcomes is weak: The OLS method suggests there might be a neighborhood effect on female employment with a 10% significance level, but the “Hwang ($d_w = 0$)” interval estimator does not support the statistical significance of the estimate. When I intentionally drop two maternal religion dummies in the auxiliary data, “Hwang ($d_w > 0$)” estimator delivers slightly wider interval estimates than that of “Hwang ($d_w = 0$)”, and they nest “Hwang($d_w = 0$)” interval estimates.

Due to limited sample sizes, I was unable to detect statistically significant heterogeneity in neighborhood effects across gender or age cohorts.²⁹ Figure E.1 in the Online Appendix presents neighborhood effect estimates by gender, while Figure E.2 displays estimates across age cohorts. The left panel of Figure E.2 shows results for individuals aged 20-29 in 2011 (born 1982-1991), while the right panel presents results for those aged 30-39 in 2011 (born 1972-1981). The corresponding numerical values for these figures are available in Tables E.8 and E.9 in the Online Appendix.

Although statistically indistinguishable, the neighborhood effect estimate on ethnic enclave residency turns out to be substantially larger for the young cohort than the old cohort. To interpret the magnitude of the “Hwang ($d_w = 0$)” estimates in Figure E.2, for an old cohort, growing up in an enclave increases ethnic enclave residency by 16 to 28 percentage points. For the young cohort, growing up in an ethnic enclave increases the probability of adulthood ethnic enclave residency by 50 to 63 percentage points. A larger magnitude observed for the younger cohort might be partially because they are younger and might still co-reside with their parents.

²⁸The ethnic enclave residency decision might be intertwined with the decision to live with parents. A large fraction of our sample members – 73% – were living with parents; See Table E.10 and Table E.11 in the Online Appendix. However, this does not change our interpretation very much. Essentially, the decision to live with parents is also endogenous and can be influenced by cultural preference.

²⁹Note that in this context, it is not possible to disentangle cohort effects from age effects.

To summarize, I find limited statistical evidence that childhood neighborhood changes second-generation immigrants' adulthood outcomes, after taking into account parental influence. The only adulthood outcome that childhood neighborhood meaningfully changes is ethnic enclave residency.

This finding contributes to the literature on environmental influences in the formation of social and political preferences ([Bisin and Verdier \(2000\)](#), [Bisin and Verdier \(2001\)](#), [Bisin et al. \(2004\)](#), [Merlino et al. \(2019\)](#), [Billings et al. \(2021\)](#), [Algan et al. \(2022\)](#)). The most closely related study is [Algan et al. \(2022\)](#), which studies cultural identity through baby name choices among Arab parents in France. Exploiting the quasi-random assignment of social housing units, they analyze the effect of neighbors on name choice. Their findings suggest little evidence that horizontal transmission via neighbors significantly influences cultural identity—a result that aligns with my own, despite differences in context and methodology – they had every variable in one dataset, there was no need to take a data combination approach.

7 Conclusion

This paper proposed a new two-sample estimator to bound the omitted variable bias using an auxiliary dataset. The key contribution of this paper lies in considering a general case where auxiliary data is only partially informative, and does not contain all right-hand side covariates ($d_w > 0$). I show that the identified set in this case may not be convex, which is fundamentally different from when auxiliary data contains all right-hand size covariates ($d_w = 0$), where the identified set is guaranteed to be convex.

This paper demonstrates the application of both estimators—for $d_w = 0$ and $d_w > 0$ cases—to an empirical question of substantial interest: estimating neighborhood effects in the absence of natural experiments when relevant information is distributed across two datasets. The results reveal that childhood neighborhood environments play a limited role in shaping the adulthood outcomes of second-generation immigrants once parental cultural preferences are properly accounted for. This finding aligns with [Algan et al. \(2013\)](#), who examined similar

questions in a different context using an alternative methodological approach.

References

- Algan, Y., A. Bisin, A. Manning, and T. Verdier (2013). *Cultural integration of immigrants in Europe*. Oxford University Press.
- Algan, Y., C. Malgouyres, T. Mayer, and M. Thoenig (2022). The economic incentives of cultural transmission: Spatial evidence from naming patterns across france. *The Economic Journal* 132(642), 437–470.
- Allman, E. S., C. Matias, and J. A. Rhodes (2009). Identifiability of parameters in latent structure models with many observed variables. *The Annals of Statistics*, 3099–3132.
- Arcidiacono, P. and J. B. Jones (2003). Finite mixture distributions, sequential likelihood and the em algorithm. *Econometrica* 71(3), 933–946.
- Berthoud, R. G. et al. (1997). Fourth National Survey of Ethnic Minorities, 1993-1994 [computer file]. Colchester, Essex: UK Data Archive [distributor], May 1997. SN: 3685. <http://dx.doi.org/10.5255/UKDA-SN-3685-1>.
- Billings, S. B., E. Chyn, and K. Haggag (2021). The long-run effects of school racial diversity on political identity. *American Economic Review: Insights* 3, 267–84.
- Bisin, A., G. Topa, and T. Verdier (2004). Religious intermarriage and socialization in the united states. *Journal of political Economy* 112(3), 615–664.
- Bisin, A. and T. Verdier (2000). “beyond the melting pot”: cultural transmission, marriage, and the evolution of ethnic and religious traits. *The Quarterly Journal of Economics* 115(3), 955–988.
- Bisin, A. and T. Verdier (2001). The economics of cultural transmission and the dynamics of preferences. *Journal of Economic theory* 97(2), 298–319.

- Bugni, F. A., I. A. Canay, and X. Shi (2017). Inference for subvectors and other functions of partially identified parameters in moment inequality models. *Quantitative Economics* 8(1), 1–38.
- Cambanis, S., G. Simons, and W. Stout (1976). Inequalities for $E_K(x, y)$ when the marginals are fixed. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 36(4), 285–294.
- Cavalli-Sforza, L. L. and M. W. Feldman (1981). *Cultural transmission and evolution: a quantitative approach*. Number 16. Princeton University Press.
- Chernozhukov, V., S. Lee, and A. M. Rosen (2013). Intersection bounds: Estimation and inference. *Econometrica* 81(2), 667–737.
- Chetty, R. and N. Hendren (2017). The impacts of neighborhoods on intergenerational mobility i: Childhood exposure effects. Technical report, National Bureau of Economic Research.
- Chetty, R., N. Hendren, and L. F. Katz (2016). The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review* 106(4), 855–902.
- Chyn, E. (2018). Moved to opportunity: The long-run effects of public housing demolition on children. *American Economic Review* 108(10), 3028–56.
- Cross, P. J. and C. F. Manski (2002). Regressions, short and long. *Econometrica* 70(1), 357–368.
- Cunha, F., J. J. Heckman, and S. M. Schennach (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica* 78(3), 883–931.
- D’Haultfoeuille, X., C. Gaillac, and A. Maurel (2022). Partially linear models under data combination. Technical report, National Bureau of Economic Research.
- Fan, Y., E. Guerre, and D. Zhu (2017). Partial identification of functionals of the joint distribution of “potential outcomes”. *Journal of Econometrics* 197(1), 42–59.

- Fan, Y., B. Pass, and X. Shi (2025). Partial identification in moment models with incomplete data via optimal transport. *arXiv preprint arXiv:2503.16098*.
- Fan, Y., R. Sherman, and M. Shum (2014). Identifying treatment effects under data combination. *Econometrica* 82(2), 811–822.
- Fang, Z. and A. Santos (2019). Inference on directionally differentiable functions. *The Review of Economic Studies* 86(1), 377–412.
- Galichon, A. (2018). *Optimal transport methods in economics*. Princeton University Press.
- Graham, B. S. (2018). Identifying and estimating neighborhood effects. *Journal of Economic Literature* 56(2), 450–500.
- Graham, B. S., C. C. d. X. Pinto, and D. Egel (2016). Efficient estimation of data combination models by the method of auxiliary-to-study tilting (ast). *Journal of Business & Economic Statistics* 34(2), 288–301.
- Hong, H. and J. Li (2018). The numerical delta method. *Journal of Econometrics* 206(2), 379–394.
- Hong, H. and J. Li (2020). The numerical bootstrap. *The Annals of Statistics* 48(1), 397–412.
- Hu, Y. (2008). Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution. *Journal of Econometrics* 144(1), 27–61.
- Hu, Y. (2017). The econometrics of unobservables: Applications of measurement error models in empirical industrial organization and labor economics. *Journal of Econometrics* 200, 154–168.
- Hu, Y. and S. M. Schennach (2008). Instrumental variable treatment of nonclassical measurement error models. *Econometrica* 76(1), 195–216.
- Hu, Y. and J.-L. Shiu (2016). Nonparametric identification using instrumental variables: sufficient conditions for completeness. Technical report, Working papers//the Johns Hopkins University, Department of Economics.

Hwang, Y. (2022). Neighborhood effects on intergenerational cultural transmission. *SSRN Working Paper*.

Hwang, Y. (2025). Conditional choice probability estimation with an imperfectly measured latent state. *Available at SSRN 3535098*.

Imbens, G. W. and C. F. Manski (2004). Confidence intervals for partially identified parameters. *Econometrica* 72(6), 1845–1857.

Johnson, D. R. (2008). Using weights in the analysis of survey data. *Presentation prepared for the Population Research Institute, Pennsylvania State University, November*.

Kaido, H., F. Molinari, and J. Stoye (2019). Confidence intervals for projections of partially identified parameters. *Econometrica* 87(4), 1397–1432.

Kasahara, H. and K. Shimotsu (2014). Non-parametric identification and estimation of the number of components in multivariate mixtures. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.

Kasahara, H. and K. Shimotsu (2015). Testing the number of components in normal mixture regression models. *Journal of the American Statistical Association* 110(512), 1632–1645.

Kleibergen, F. and R. Paap (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of econometrics* 133(1), 97–126.

Li, L., D. Kédagni, and I. Mourifié (2020). Discordant relaxations of misspecified models. *arXiv preprint arXiv:2012.11679*.

Magnus, J. R. and H. Neudecker (2019). *Matrix differential calculus with applications in statistics and econometrics*. John Wiley & Sons.

Masten, M. A. and A. Poirier (2020). Inference on breakdown frontiers. *Quantitative Economics* 11(1), 41–111.

- Merlino, L. P., M. F. Steinhardt, and L. Wren-Lewis (2019). More than just friends? school peers and adult interracial relationships. *Journal of Labor Economics* 37(3), 663–713.
- Mullen, K. M., D. Ardia, D. L. Gil, D. Windover, and J. Cline (2011). Deoptim: An r package for global optimization by differential evolution. *Journal of Statistical Software* 40, 1–26.
- Office for National Statistics (2019). released 11 June 2019, ONS SRS Metadata Catalogue, dataset, ONS Longitudinal Study - England and Wales. <https://doi.org/10.57906/z9xn-ng05>.
- Pacini, D. (2019). Two-sample least squares projection. *Econometric Reviews*, 38(1), 95–123.
- Peyré, G., M. Cuturi, et al. (2019). Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning* 11(5-6), 355–607.
- Ridder, G. and R. Moffitt (2007). The econometrics of data combination. *Handbook of econometrics* 6, 5469–5547.
- Royal, K. D. et al. (2019). Survey research methods: A guide for creating post-stratification weights to correct for sample bias. *Education in the Health Professions* 2(1), 48.
- Shelton, N., C. E. Marshall, R. Stuchbury, E. Grundy, A. Dennett, J. Tomlinson, O. Duke-Williams, O. Staff, and W. Xun (2019). Cohort profile: the office for national statistics longitudinal study (the ls). *International journal of epidemiology* 48(2), 383–384g.
- Sinkhorn, R. and P. Knopp (1967). Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics* 21(2), 343–348.
- Słoczyński, T. (2022). Interpreting ols estimands when treatment effects are heterogeneous: Smaller groups get larger weights. *The Review of Economics and Statistics*, 104(3): 501–509.
- Van der Vaart, A. W. (2000). *Asymptotic statistics*, Volume 3. Cambridge university press.

FOR ONLINE PUBLICATION: “Bounding Omitted Variable Bias
Using Auxiliary Data: With an Application to Estimate
Neighborhood Effects” by Yujung Hwang

A Lower-level assumptions that guarantee the Assumption 1

This section outlines lower-level conditions that imply Assumption 1. Multiple sets of sufficient conditions exist and have been studied in prior work, including [Allman et al. \(2009\)](#), [Hu \(2008\)](#), [Hu and Schennach \(2008\)](#), [Hu \(2017\)](#), and [Cunha et al. \(2010\)](#). Below, I present one such set of assumptions.³⁰ Assumption A.1 pertains to discrete error-ridden variables, while Assumption A.2 addresses continuous ones. For notational simplicity, I focus on the case where a scalar covariate x is measured with error in the auxiliary dataset using multiple proxy variables. If multiple variable in equation (1) are subject to measurement error, each must be associated with its own set of proxy variables satisfying either Assumption A.1 or Assumption A.2.

Assumption A.1 (Discrete Proxies). *Suppose the proxies $\{m_k\}_{k=1}^K$ are noisy measurements of x , and the following conditions hold:*

- (1) (*discreteness*) *x and its proxies $\{m_k\}_{k=1}^K$ are discrete.*
- (2) (*conditional independence*) *The proxy variables m_1, \dots, m_k and other covariates z are mutually independent conditional on x .*
- (3) (*rank condition*) *Let the cardinality of the support of x be $\bar{U} > 0$. Define the measurement structure matrix M_k whose ij -th component is the conditional probability of having k -th*

³⁰Alternative assumptions exist, particularly for discrete proxy variables. For example, in place of the matrix rank condition and monotonicity in Assumption A.1, different identification strategies can be used. For continuous proxy variables, Theorem 1 in [Cunha et al. \(2010\)](#) establishes identification using only two proxies under a parametric error-in-variables framework. These alternative conditions are omitted here for brevity. A more comprehensive discussion is available in the cited literature.

proxy m_k equal to j -th largest value when the latent variable x is i -th largest value. That is, $(M_k)_{ij} = [P(m_k = j\text{-th largest value}|x = i\text{-th largest value})]$. Assume there exist at least two proxy variables whose measurement structure matrix M_k has a full (row) rank and another proxy variable whose measurement structure matrix has a rank no less than 2.

(4) (monotonicity) One of the following assumptions must be true.

(i) There is a proxy m_k for x such that $P(m_k = j|x)$ is strictly decreasing/increasing in x for some j .

(ii) There is a proxy m_k for x such that $P(m_k = j|x = j) > P(m_k = j|x = j')$ for $j \neq j'$.

Assumption A.2 (Continuous Proxies). Assume there exist three proxy variables m_1, m_2, m_3 for a continuous covariate x , and the following conditions are satisfied:

(1) The (conditional) joint density of x, m_1, m_2, m_3 , conditional on covariates z , is bounded, and so are all their marginal and conditional densities.

(2) (conditional independence) The proxy variables m_1, \dots, m_k and other covariates z are mutually independent conditional on x .

(3) $f_{m_1|m_2,z}(m_1|m_2, z)$ and $f_{m_1|x}(m_1|x)$ form a bounded complete family of distributions indexed by (m_2, z) and m_1 , respectively.

(4) Whenever $x \neq \tilde{x}$, $f_{m_3|x}(m_3|x)$ and $f_{m_3|\tilde{x}}(m_3|\tilde{x})$ differ over a set of strictly positive probability.

(5) There exists a known functional Ψ , mapping a density to a vector, that has the property that $\Psi[f_{m_1|x}(\cdot|x)] = x$.

While the requirement of three proxies may seem restrictive, it is often satisfied in practice—especially when x represents a qualitative construct such as ability, belief, or preference. Survey instruments commonly include multiple questions targeting the same latent trait. For instance, [Hwang \(2025\)](#) catalogs proxy variables for common sources of unobserved heterogeneity in economics and shows that multiple proxies are often available. Also, see Section 6 for examples related to cultural preferences.

The conditional independence assumption implies that measurement errors in the proxies (i.e., $m_k - x$) are independent of other covariates in the same data, z . This is plausible when

errors reflect random variation in survey responses. The rank condition (Assumption A.1(3)) and its analogs in Assumption A.2(3)–(4) ensure that the proxies are sufficiently informative about x . In the discrete case, the rank condition is testable using existing methods (e.g., Kleibergen and Paap (2006), Kasahara and Shimotsu (2014), Kasahara and Shimotsu (2015)). In Section 6, I implement a rank test based on Kleibergen and Paap (2006). For alternative testing approaches, see the cited references.

The monotonicity condition in Assumption A.1(4) fixes the labeling of the latent discrete variable x , resolving under-identification in the presence of multiple equivalent labelings. Its analog in Assumption A.2(5) plays a similar role for continuous variables, ensuring identification up to location and scale. In practice, the choice of the proxy variable used for the monotonicity condition can be arbitrary. A robustness check using alternative proxies is recommended. As shown in Section 6, the results are stable across different choices of the monotonicity-enforcing proxy. For further discussion, see Hwang (2025) and references therein.

The next lemma formalizes that Assumptions A.1 and A.2 are sufficient for Assumption 1. These results follow from prior work, including Hu (2008), Hu (2017), Hu and Schennach (2008), and Cunha et al. (2010). The main difference here is the inclusion of additional covariates z .

Lemma A.1 (Sufficient Conditions for Assumption 1). (i) *When the omitted covariate x is discrete, Assumption A.1 guarantees Assumption 1.*

(ii) *When the omitted covariate x is continuous, Assumption A.2 guarantees Assumption 1.*

Proof. See Online Appendix Section C. □

B Discussion of $d_w = 0$ Case

This section discusses the special case $d_w = 0$. While this case has been widely studied in [Pacini \(2019\)](#) and [Fan et al. \(2025\)](#), I use a slightly different exposition from them. The special case $d_w = 0$ is much easier to handle in practice than $d_w > 0$ case, since it offers a closed-form expression for the lower bound and the upper bound of each parameter. I also discuss an inference method in the case of $d_w = 0$.

To state the identification result, I first define lower bound and upper bound operators. These operators compare and combine elements from two weighted sums in a specific way. Consider two vectors h^1 and h^2 , and a matrix A . For each row k of matrix A , we can compute two inner products: $A_k \cdot h^1 = \sum_{l=1}^L A_{kl} h_l^1$ and $A_k \cdot h^2 = \sum_{l=1}^L A_{kl} h_l^2$. Instead of directly using these inner products, the operators break them down into their individual terms ($A_{kl} h_l^1$ and $A_{kl} h_l^2$), compare corresponding terms, and then sum up the minimum (or maximum) of each pair.

Definition 1 (Lower and Upper Bound Operators). *The operator \mathcal{L} maps a real matrix A of size K -by- L ($K, L \geq 1$) and two real vectors h^1 and h^2 of length L to a real vector g of length K . That is, $\mathcal{L} : \mathbb{R}^{K \times L} \times \mathbb{R}^L \times \mathbb{R}^L \rightarrow \mathbb{R}^K$. Write $\mathcal{L}(A, h^1, h^2) = g$.*

For notation, let A_{kl} be the (k, l) -th element in matrix A , h_l^1, h_l^2 be the l -th element of a vector h^1, h^2 . The k -th element of g is defined as follows :

$$g_k = \sum_{l=1}^L \min(A_{kl} h_l^1, A_{kl} h_l^2). \quad (\text{B.1})$$

The upper bound operator \mathcal{U} is analogously defined by replacing \min with \max .

Next, I state the lower bound and upper bound identification results and their plug-in estimator. I characterize the projected interval bound on each parameter.³¹ Note that I do not require Assumption 3 unlike in the case of $d_w > 0$, because when $d_w = 0$, none of the unidentified moments appear in a moment matrix (**) in equation 21. Instead, I make a relaxed

³¹[Pacini \(2019\)](#) provides a similar formula for the projection: see Section 4.1 of [Pacini \(2019\)](#).

Assumption B.1 that the moment matrix (**) is non-singular, which is guaranteed when (x, z) does not contain any perfectly multicollinear covariate. Note that when x is a deterministic nonlinear function of z ³², then the bound estimates become a singleton.

Assumption B.1 (Non-Singularity). *The moment matrix $E \equiv \begin{bmatrix} E[xx'] & E[xz'] \\ E[zx'] & E[zz'] \end{bmatrix}$ is non-singular.*

Proposition B.1. *Assume $d_w = 0$. Under Assumptions 1, 2, and B.1, the lower bound of the parameter of interest (α', β') is given by*

$$\begin{bmatrix} \tilde{\alpha}_l^{superset} \\ \tilde{\beta}_l^{superset} \end{bmatrix} = \begin{bmatrix} \mathcal{L}(A, \inf E[xy], \sup E[xy]) + BE[zy] \\ \mathcal{L}(C, \inf E[xy], \sup E[xy]) + DE[zy] \end{bmatrix}, \quad \begin{bmatrix} A & B \\ C & D \end{bmatrix} \equiv \begin{bmatrix} E[xx'] & E[xz'] \\ E[zx'] & E[zz'] \end{bmatrix}^{-1}. \quad (\text{B.2})$$

Similarly, the upper bound of the parameter of interest (α', β') is given by replacing the lower bound operator with the upper bound operator. The bounds

$\tilde{\alpha}_l^{superset}, \tilde{\beta}_l^{superset}, \tilde{\alpha}_u^{superset}, \tilde{\beta}_u^{superset}$ are finite. Moreover, if $F(x|z)$ is degenerate for almost all z , then the bound estimates are singleton. That is, $\tilde{\alpha}_l^{superset} = \tilde{\alpha}_u^{superset}$ and $\tilde{\beta}_l^{superset} = \tilde{\beta}_u^{superset}$.

Proof. In the Online Appendix Section C. □

Estimator 2 (Estimator when there are one or more omitted variables and $d_w = 0$). *Under Assumptions 1, 2, B.1 and for $d_w = 0$, the lower bound estimator for the parameter of interest, (α, β) , is*

$$\begin{bmatrix} \widehat{\tilde{\alpha}}_l^{superset} \\ \widehat{\tilde{\beta}}_l^{superset} \end{bmatrix} = \begin{bmatrix} \mathcal{L}(\widehat{A}, \widehat{\mu}_l^{xy}, \widehat{\mu}_u^{xy}) + \widehat{B}\widehat{s}_{zy} \\ \mathcal{L}(\widehat{C}, \widehat{\mu}_l^{xy}, \widehat{\mu}_u^{xy}) + \widehat{D}\widehat{s}_{zy} \end{bmatrix}, \quad \begin{bmatrix} \widehat{A} & \widehat{B} \\ \widehat{C} & \widehat{D} \end{bmatrix} \equiv \begin{bmatrix} \widehat{s}_{xx'} & \widehat{s}_{xz'} \\ \widehat{s}_{zx'} & \widehat{s}_{zz'} \end{bmatrix}^{-1}. \quad (\text{B.3})$$

Similarly, the upper bound of the parameter of interest (α, β) is given by replacing the lower bound operator to upper bound operator (Definition 1). $\widehat{\mu}_l^{xy}, \widehat{\mu}_u^{xy}$ are vectors of length d_x and the k -th element of $\widehat{\mu}_l^{xy}, \widehat{\mu}_u^{xy}$ are defined as in Estimator 1. $\widehat{s}_.$ denote the sample analog of the population moment $E[\cdot]$.

³²When x is a linear function of z , there is a perfect multicollinearity, which violates the Assumption B.1.

Next, I state the consistency result for this plug-in estimator.

Proposition B.2 (Consistency for $d_w = 0$). *Assume $d_w = 0$. Then, under Assumptions 1, 2, B.1 and Assumption 4, the Estimator 2 is consistent. That is,*

$$\widehat{\alpha}_l^{superset} \xrightarrow{p} \tilde{\alpha}_l^{superset} \text{ as } N_m \rightarrow \infty, \text{ and subsequently } N_a(N_m) \rightarrow \infty. \quad (\text{B.4})$$

$$\widehat{\beta}_l^{superset} \xrightarrow{p} \tilde{\beta}_l^{superset} \text{ as } N_m \rightarrow \infty, \text{ and subsequently } N_a(N_m) \rightarrow \infty. \quad (\text{B.5})$$

$$\widehat{\alpha}_u^{superset} \xrightarrow{p} \tilde{\alpha}_u^{superset} \text{ as } N_m \rightarrow \infty, \text{ and subsequently } N_a(N_m) \rightarrow \infty. \quad (\text{B.6})$$

$$\widehat{\beta}_u^{superset} \xrightarrow{p} \tilde{\beta}_u^{superset} \text{ as } N_m \rightarrow \infty, \text{ and subsequently } N_a(N_m) \rightarrow \infty. \quad (\text{B.7})$$

Proof. In the Online Appendix Section C. □

For inference on the lower and upper bounds of the parameter of interest, I propose to use an alternative bootstrap method based on the numerical delta method. We can not use a standard bootstrap method because of the non-differentiability of the estimator³³. The standard bootstrap method fails when the estimator is not Hadamard differentiable (Fang and Santos (2019))³⁴. However, Fang and Santos (2019) has also shown that an alternative bootstrap method works if an estimator is Hadamard *directionally* differentiable at a population moment vector (Assumption 1 (ii), Theorem 3.2 in Fang and Santos (2019); Theorem 3.1 in Hong and Li (2018)). The next proposition states that Estimator 2 is Hadamard directionally differentiable at population moment vector θ_0 .

Proposition B.3 (Hadamard Directional Differentiability). .

Make Assumption 2 and B.1. Define a vector of population moments,

$\theta_0 \equiv (\mu_l^{xy'}, \mu_u^{xy'}, \text{vec}(E[xx']), \text{vec}(E[xz']), \text{vec}(E[zx']), \text{vec}(E[zy']))'$, where $\text{vec}(\cdot)$ is a vectorization operator that converts matrices into row vectors. Denote a sample counterpart to θ_0 as $\widehat{\theta}$.

³³I thank the anonymous referee for suggesting this method.

³⁴For the definition of Hadamard differentiability and Hadamard directional differentiability, see Definition 2.1 in Fang and Santos (2019) p.382.

Note that the Estimator 2 is a function of $\hat{\theta}$, denoted as $\phi(\hat{\theta})$.

$$\begin{aligned} \left[\hat{\alpha}_l^{superset}, \hat{\beta}_l^{superset} \right] &= \left[\phi_1(\hat{\theta}), \dots, \phi_{d_x+d_z}(\hat{\theta}) \right], \\ \text{where } \phi_k(\hat{\theta}) &= \sum_{l=1}^{d_x} \min(\hat{A}_{kl} \hat{\mu}_l^{x_l y}, \hat{A}_{kl} \hat{\mu}_u^{x_l y}) + \sum_{l=1}^{d_x} \hat{B}_{kl} \hat{s}_{z_k y}, \\ \text{and } \begin{bmatrix} \hat{A} & \hat{B} \\ \hat{C} & \hat{D} \end{bmatrix} &\equiv \begin{bmatrix} \hat{s}_{xx'} & \hat{s}_{xz'} \\ \hat{s}_{zx'} & \hat{s}_{zz'} \end{bmatrix}^{-1}. \end{aligned} \quad (\text{B.8})$$

For $(\hat{\alpha}_u^{superset}, \hat{\beta}_u^{superset})$, replace ‘min’ with ‘max’.

The function $\phi(\cdot)$ is Hadamard directionally differentiable in θ_0 .

Proof. In the Online Appendix Section C. □

I make Assumption B.2 to use the Hong and Li (2018) numerical delta method³⁵. First, I assume the ratio of the sample size of the main data and the sample size of the auxiliary data converges to a constant as the sample size from each sample increases. Unlike Hong and Li (2018), I assume some of the moments in $\hat{\theta}$ are estimated from separate data, which may have a different sample size. So this assumption is to avoid a singularity that arises when a subset of moments in the estimator converges at a much faster rate than others. Second, I assume that $\hat{\theta}$ defined in Assumption B.2 weakly converges to a Gaussian distribution as standard in bootstrap theory. The sample moments converge to the population moments by Central Limit Theorem. The sample size restriction, Assumption B.2 (a), guarantees that the auxiliary data moments converge to a limiting distribution at the same rate $\sqrt{N_m}$. So the Assumption B.2 (b) is essentially about the estimator for the conditional CDF $F_{y|z}, F_{w_l|z}$ and the conditional quantile function $Q_{x_k|z}$. Masten and Poirier (2020) has shown that when the support of z is discrete and finite, the empirical conditional CDF and conditional Quantile function converge uniformly in distribution to a Gaussian process at rate $\sqrt{N_m}$ (see Lemma 2 and Lemma C1 in Masten and Poirier (2020)). Then Assumption B.2 would hold. When the support of z is continuous or infinite, one can either discretize the support of z or use a parametric estimator

³⁵I thank Aureo de Paula for providing helpful comments on this assumption

for $F_{y|z}, F_{w_l|z}, Q_{x_k|z}$, which guarantees the convergence to a Gaussian process at rate $\sqrt{N_m}$.

Assumption B.2. (a) Let the sample size of the main data N_m and the sample size of the auxiliary data $N_a(N_m)$, which depends on the sample size of the main data N_m . Assume the ratio of the two sample sizes converges to a constant.

$$\lim_{N_m \rightarrow \infty} \frac{N_m}{N_a(N_m)} \rightarrow k \in \mathbb{R}$$

(b) Assume $\sqrt{N_m}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \Sigma)$ for some $N_m \rightarrow \infty, N_a(N_m) \rightarrow \infty$, where $\hat{\theta}$ is defined in Remark 3 and θ_0 is defined in Proposition B.3.

Note that Assumption B.2 (b) does not require the two samples to be independent. This is expressed in allowing Σ to be any variance-covariance matrix. As long as a vector $\hat{\theta}$ converges to a joint normal distribution, the numerical delta method applies. If the two samples are independent, then this implies that the moments estimated from main data only and the moments estimated from auxiliary data only must have zero covariance. Then, Σ will include zero block matrices corresponding to zero covariance between the moments $(\hat{s}_{xx'}, \hat{s}_{xz'})$ and (\hat{s}_{zy}) .

Next, I summarize the Hong and Li (2018) numerical delta method.

Algorithm 1. (*Numerical delta method, p.382 in Hong and Li (2018)*)

Under the Assumption B.2, the following bootstrap procedure provides an asymptotically valid inference. Let S be the number of bootstrap draws. Let λ_n be the tuning parameter such that $\lambda_n \sqrt{N_m} \rightarrow \infty$ and $\lambda_n \rightarrow 0$ as $\sqrt{N_m} \rightarrow \infty, \sqrt{N_a(N_m)} \rightarrow \infty$. Let $\hat{\theta}_n$ denote the estimator of θ_0 based on a sample size indexed by n , and let $\hat{\theta}_n^*$ denote the corresponding estimator based on a bootstrap sample of the same size.

1. Define $\mathbb{Z}_n^* = \sqrt{N_m}(\hat{\theta}_n^* - \hat{\theta}_n)$. Draw \mathbb{Z}_s from the distribution of \mathbb{Z}_n^* for $s = 1, \dots, S$.
2. For the given λ_n , evaluate for each s :

$$\widehat{\phi}'_n(\mathbb{Z}_s) \equiv \frac{\phi(\widehat{\theta}_n + \lambda_n \mathbb{Z}_s) - \phi(\widehat{\theta}_n)}{\lambda_n} \quad (\text{B.9})$$

$\phi(\cdot)$ refers to the estimator 2 defined as a function of moments $\widehat{\theta}$ in equation B.8.

3. Let the $c_{\tau/2}$ and $c_{1-\tau/2}$ be the $\tau/2$ and $1 - \tau/2$ be the empirical percentiles of $\widehat{\phi}'_n(\mathbb{Z}_s)$. Then, a $(1 - \tau)$ two-sided equal-tailed confidence interval for $\phi(\theta_0)$ is

$$CI^{equal-tail}(\phi(\widehat{\theta})) \equiv \left[\phi(\widehat{\theta}) - \frac{1}{\sqrt{N_m}} c_{1-\tau/2}, \phi(\widehat{\theta}) + \frac{1}{\sqrt{N_m}} c_{\tau/2} \right]. \quad (\text{B.10})$$

Similarly, a symmetric confidence interval can be constructed. Let $d_{1-\tau}$ be the $1 - \tau$ percentile of $|\widehat{\phi}'_n(\mathbb{Z}_n^*)|$. The symmetric confidence interval is

$$CI^{symmetric}(\phi(\widehat{\theta})) \equiv \left[\phi(\widehat{\theta}) - \frac{1}{\sqrt{N_m}} d_{1-\tau}, \phi(\widehat{\theta}) + \frac{1}{\sqrt{N_m}} d_{1-\tau} \right]. \quad (\text{B.11})$$

The numerical delta method is not computationally costly because computing the numerical derivative $\widehat{\phi}'_n$ requires computing $\phi(\cdot)$ function twice, where $\phi(\cdot)$ refers to the estimator 2 defined as a function of moments $\widehat{\theta}$ in equation B.8. Also, note that when $\lambda_n = N_m^{-1/2}$, then the numerical delta method is identical to the standard bootstrap.

There is no rule-of-thumb on how to choose the tuning parameter λ_n , and discussing which method is the best to find the optimal λ_n is beyond the scope of this paper. Following Hong and Li (2020), I suggest doing the double bootstrap to choose the optimal tuning parameter; first, one draws B_1 bootstrap samples and computes the bootstrap estimates $\phi(\widehat{\theta})^{(b1)}$. Conditional on each bootstrap sample, one draws again B_2 bootstrap sample and construct confidence intervals $CI^{equal-tail,(b1,b2)}$ or $CI^{symmetric,(b1,b2)}$ centered around $\phi(\widehat{\theta})^{(b1)}$ and count what fraction of confidence intervals cover $\phi(\widehat{\theta})$. The tuning parameter λ_n , which gives the coverage rate close to $(1 - \tau)$ can be chosen as an optimal tuning parameter; see Online Appendix Section D.4 for demonstration of choosing a tuning parameter.

Note that the constructed confidence interval is for the population lower bound and the upper bound (of the parameter of interest) and not the parameter itself. Denote the lower bound estimate as $\phi^{lb}(\hat{\theta})$ and the upper bound estimate as $\phi^{ub}(\hat{\theta})$. Then, one can construct the confidence interval for the parameter of interest as $[\inf CI^{equal-tailed}(\phi^{lb}(\hat{\theta})), \sup CI^{equal-tailed}(\phi^{ub}(\hat{\theta}))]$; this is the confidence interval that I present in the empirical application section 6. However, the coverage rate of this confidence interval for a parameter of interest is greater than $(1 - \tau)$ as shown in [Imbens and Manski \(2004\)](#). For alternative inference methods that provide a pre-specified coverage rate for a single partially identified parameter, see [Bugni et al. \(2017\)](#) and [Kaido et al. \(2019\)](#)³⁶.

C Proof

Proof of Lemma A.1. .

For discrete case, [Hu \(2008\)](#), [Hu \(2017\)](#) have shown that a similar assumption to Assumption A.1 guarantees that one can identify $f(x)$ and $f(m_1|x), \dots, f(m_k|x)$ from observing $f(m_1, m_2, m_3)$. The only difference in Assumption A.1 conditions is in the conditional independence assumption (2), which also requires mutual independence of other covariates z and proxies (m_1, \dots, m_k) conditional on a latent variable x . To see why this new condition suffices to show Lemma A.1 (i), consider

$$\begin{aligned} f(m_1, \dots, m_k|z) &= \int_x f(m_1, \dots, m_k|x, z)f(x|z)dx, \\ &= \int_x f(m_1|x, z) \cdots f(m_k|x, z)f(x|z)dx, \quad (\because \text{mutual independence condition (2)}) \\ &= \int_x f(m_1|x) \cdots f(m_k|x)f(x|z)dx. \quad (\because \text{mutual independence condition (2)}) \end{aligned}$$

Therefore, [Hu \(2008\)](#), [Hu \(2017\)](#) results apply here when conditioning everything on z , with the extended mutual independence assumption. That is, under the Assumption A.1, one can

³⁶I thank the referee for this suggestion.

identify $f(m_1|x), \dots, f(m_k|x), f(x|z)$ from $f(m_1, \dots, m_k|z)$, which is observed in auxiliary data. Then $f(x, z) = f(x|z)f(z)$ and $f(z)$ are observed in auxiliary data, so $f(x, z)$ is identified.

Similarly, Assumption A.2 extends conditions for Cunha et al. (2010) Theorem 2 (Hu and Schennach (2008) Theorem 1) to account for other covariates z . The only differences in the conditions are in (2) and (3). (2) is the extended mutual independence assumption to account for other covariates z . The condition (3) requires that $f(m_1|m_2, z)$ (instead of $f(m_1|m_2)$ in Hu and Schennach (2008) version) is complete. To see why this is necessary, consider the extended version of the equation (5) in Hu and Schennach (2008) for our context.

$$f(m_1, m_3|m_2, z) = \int_x f(m_1|x)f(m_3|x)f(x|m_2, z)dx. \quad (\text{C.1})$$

To apply the proof of Theorem 1 in Hu and Schennach (2008), consider the extended linear operator equation (of equations 7, 8, 9 in Hu and Schennach (2008)). Using our notations,

$$L_{m_1, m_3|m_2, z} = L_{m_1|x}\Delta_{m_3, x}L_{x|m_2, z}, \quad (\text{C.2})$$

$$L_{x|m_2, z} = L_{m_1|x}^{-1}L_{m_1|m_2, z}, \quad (\text{C.3})$$

$$L_{m_1, m_3|m_2, z}L_{m_1|m_2, z}^{-1} = L_{m_1|x}\Delta_{m_3, x}L_{m_1|x}^{-1}. \quad (\text{C.4})$$

The last equation shows the spectral decomposition of the operator $L_{m_1, m_3|m_2, z}L_{m_1|m_2, z}^{-1}$ and to make it well-defined, it requires that the inverse of a linear operator $L_{m_1|m_2, z}$ exists. Therefore, Assumption (3) (completeness of density $f(m_1|m_2, z)$) is necessary. Also, Assumption (3) states that $f(m_1|x)$ is complete, so the inverse of a linear operator $L_{m_1|x}$ is well-defined.

The remaining proof of Theorem 1 in Hu and Schennach (2008) goes through in this setup, and the result guarantees the unique solution $f(m_1|x), f(m_3|x), f(x|m_2, z)$ exists. From $f(x|m_2, z)$, we can identify $f(x, z|m_2)$ from the following equation because $f(z|m_2)$ is ob-

served in auxiliary data.

$$f(x|m_2, z) = \frac{f(x, z|m_2)}{f(z|m_2)}. \quad (\text{C.5})$$

Then $f(x, z) = \int_{m_2} f(x, z|m_2) f(m_2) dm_2$ is identified. This concludes the proof. \square

Proof of Proposition 1. .

Proof of (1)

$\Theta^{\text{sharp},m}$ is non-empty because it must include $(E[xy]', E[xw_1]', \dots, E[xw_k]')'$, as true $f(y, x, z, w) \in \mathcal{M}(f(y, z, w), f(x, z))$. $\Theta^{\text{sharp},m}$ is compact and convex because the coupling set $\mathcal{M}(\cdot)$ is compact and convex, and the mapping $\tilde{f} \mapsto E_{\tilde{f}}[\cdot]$ is linear.

Proof of (2)

Let $\Theta_{\kappa_0}^{\text{sharp},m} = \{m \in \Theta^{\text{sharp},m} | \kappa(m) \leq \kappa_0\}$ be the subset of $\Theta^{\text{sharp},m}$ that satisfies the condition number constraint. Under Assumption 3, the true moment vector $(E[xy]', E[xw_1]', \dots, E[xw_{d_w}]')'$ yields a well-conditioned moment matrix with $\kappa(m) \leq \kappa_0$. Therefore, $\Theta_{\kappa_0}^{\text{sharp},m}$ is non-empty.

Since the condition number $\kappa(m)$ is a continuous function of m , the set $\{m | \kappa(m) \leq \kappa_0\}$ is closed. The intersection of this closed set with the compact set $\Theta^{\text{sharp},m}$ results in $\Theta_{\kappa_0}^{\text{sharp},m}$, which is compact.

$\Theta^{\text{sharp},p}$ is non-empty because $(E[xy]', E[xw_1]', \dots, E[xw_{d_w}]')' \in \Theta_{\kappa_0}^{\text{sharp},m}$, and $\mathcal{B}_{OLS}((E[xy]', E[xw_1]', \dots, E[xw_{d_w}]')') = (\alpha', \beta', \gamma')' \in \Theta^{\text{sharp},p}$.

The moment matrix is non-singular over $\Theta_{\kappa_0}^{\text{sharp},m}$ due to condition number constraint. Therefore, the function mapping $\mathcal{B}_{OLS}(\cdot)$ is continuous on $\Theta_{\kappa_0}^{\text{sharp},m}$, and by the continuous mapping theorem, the compactness of $\Theta_{\kappa_0}^{\text{sharp},m}$ implies that $\Theta^{\text{sharp},p}$ is also compact.

When $d_w = 0$, the moment matrix $(*)$ does not include any unidentified moments and by Assumption 3, the moment matrix is well-conditioned over all of $\Theta^{\text{sharp},m}$. Therefore, $\Theta_{\kappa_0}^{\text{sharp},m} = \Theta^{\text{sharp},m}$. $\mathcal{B}_{OLS}(\cdot)$ is an affine transformation over a convex domain, $\Theta_{\kappa_0}^{\text{sharp},m}$ re-

sulting in a convex $\Theta^{\text{sharp},p}$.

However, when $d_w > 0$, the function $\mathcal{B}_{OLS}(\cdot)$ involves matrix inversion and is no longer affine. Consequently, the convexity of $\Theta^{\text{sharp},p}$ is not guaranteed, as the image of a convex set under a continuous but non-affine function need not be convex. In addition, the set $\Theta_{\kappa_0}^{\text{sharp},m}$ may not be convex due to the condition number constraint.

Proof of (3)

Note that the projection function is linear. Therefore, the non-emptiness and compactness of $\Theta^{\text{sharp},p}$ imply its projection is also non-empty and compact. Any compact set on \mathbb{R} is a bounded interval, which is convex.

This concludes the proof. □

Proof of Lemma 1.

Theorem 2 in [Cambanis et al. \(1976\)](#) guarantees that the identified set of $E[yx_k]$ for any $k \in \{1, \dots, d_x\}$ is the closed interval because their condition (i) holds in this case: The function $k(X, Y) = XY$ is symmetric and $E[X^2]$ and $E[Y^2]$ are finite by Assumption 2.

Next, the sharp upper and lower bounds can be found by the following inequalities.

$$\begin{aligned} \sup_{\tilde{f} \in \mathcal{M}(f(y,z,w), f(x,z))} E[x_k y] &= \sup_{\tilde{f} \in \mathcal{M}(f(y,z,w), f(x,z))} E[E[x_k y|z]] \leq E[\sup E[x_k y|z]] = E[E[y \cdot Q_{x_k|z}(F_{y|z}(y|z)|z)]]. \\ \inf_{\tilde{f} \in \mathcal{M}(f(y,z,w), f(x,z))} E[x_k y] &= \inf_{\tilde{f} \in \mathcal{M}(f(y,z,w), f(x,z))} E[E[x_k y|z]] \geq E[\inf E[x_k y|z]] = E[E[y \cdot Q_{x_k|z}(1 - F_{y|z}(y|z)|z)]]. \end{aligned}$$

The inequality above follows from Fatou's lemma. The last equality follows from applying the monotone rearrangement inequality to $\sup E[x_k y|z]$, $\inf E[x_k y|z]$. Note that we can apply the monotone rearrangement inequality because $y|z$, $x_k|z$ have finite second moments for $\forall z$ almost surely by Assumption 2.

For the inequality of another direction, simply note that there exists $f(y, x, z)$ in $\mathcal{M}(f(y, z, w), f(x, z))$

that gives the right-hand side. For supremum, choose f such that its marginal on $\mathcal{Y} \times \mathcal{X}_k \times \mathcal{Z}$ is $f(y, x_k, z) = f(y, Q_{x_k|z}(F_{y|z}(y|z)|z), z)$ and, for infimum, choose f such that its marginal on $\mathcal{Y} \times \mathcal{X}_k \times \mathcal{Z}$ is $f(y, x_k, z) = f(y, Q_{x_k|z}(1 - F_{y|z}(y|z)|z), z)$. That is, conditional on z , when y and x_k are perfectly positively assortatively matched, it gives the supremum, and when they are perfectly negatively assortatively matched, it gives the infimum. And these matchings are conditional on fixed marginals, $f(y|z)$, $f(x_k|z)$, $f(x_k)$. Therefore, by definition of *sup* and *inf*, we have

$$\sup_{\tilde{f} \in \mathcal{M}(f(y,z,w), f(x,z))} E[x_k y] \geq E[E[y \cdot Q_{x_k|z}(F_{y|z}(y|z)|z)]] \quad (\text{C.6})$$

$$\inf_{\tilde{f} \in \mathcal{M}(f(y,z,w), f(x,z))} E[x_k y] \leq E[E[y \cdot Q_{x_k|z}(1 - F_{y|z}(y|z)|z)]] \quad (\text{C.7})$$

This completes the proof. \square

Proof of Corollary 1. .

The second-order moment $E[x_k y]$ has trivial bounds from Cauchy-Schwarz inequality.

$$|E[x_k y]| \leq \sqrt{E[x_k^2]E[y^2]} < \infty \quad (\text{C.8})$$

The second inequality is due to Assumption 2. Since the bounds in 9, 10 are sharp, the trivial bounds must nest them.

$$[\inf E[x_k y], \sup E[x_k y]] \subset \left[-\sqrt{E[x_k^2]E[y^2]}, \sqrt{E[x_k^2]E[y^2]} \right]. \quad (\text{C.9})$$

This concludes the proof. \square

Proof of Proposition 2. .

We need to show $[E[xy]', E[xw_1]', \dots, E[xw_{d_w}]]' \in \Theta^{\text{superset},m}$, $(\alpha', \beta', \gamma')' \in \Theta^{\text{superset},p}$, $\alpha_k \in [\alpha_{lk}^{\text{superset}}, \alpha_{uk}^{\text{superset}}]$, $k \in \{1, \dots, d_x\}$, $\beta_k \in [\beta_{lk}^{\text{superset}}, \beta_{uk}^{\text{superset}}]$, $k \in \{1, \dots, d_z\}$,

$\gamma_k \in [\gamma_{lk}^{\text{superset}}, \gamma_{uk}^{\text{superset}}], k \in \{1, \dots, d_w\}$, so they are the valid supersets.

(1) Proof for $\Theta^{\text{superset},m}$

The Monotone Rearrangement Inequality guarantees that the interval bounds must include the population moment, for $1 \leq k \leq d_x, 1 \leq l \leq d_w$, $E[x_k y] \in [\mu_l^{x_k y}, \mu_u^{x_k y}], E[x_k w_l] \in [\mu_l^{x_k w_l}, \mu_u^{x_k w_l}]$, where $\mu_l^{x_k y}, \mu_u^{x_k y}, \mu_l^{x_k w_l}, \mu_u^{x_k w_l}$ are identified from Lemma 1.

Therefore, $[E[xy]', E[xw_1]', \dots, E[xw_{d_w}]]' \in \Theta^{\text{superset},m}$.

(2) Proof for $\Theta^{\text{superset},p}$

From (1), $[E[xy]', E[xw_1]', \dots, E[xw_{d_w}]]' \in \Theta^{\text{superset},m}$ and under Assumption 3, the condition number constraint is satisfied for $[E[xy]', E[xw_1]', \dots, E[xw_{d_w}]]'$.

Therefore, $\mathcal{B}_{OLS}([E[xy]', E[xw_1]', \dots, E[xw_{d_w}]]') = (\alpha', \beta', \gamma')' \in \Theta^{\text{superset},p}$.

(3) Proof for $(\alpha_l^{\text{superset}}, \beta_l^{\text{superset}}, \gamma_l^{\text{superset}}), (\alpha_u^{\text{superset}}, \beta_u^{\text{superset}}, \gamma_u^{\text{superset}})$

From (2), $(\alpha', \beta', \gamma')' \in \Theta^{\text{superset},p}$. By applying $\text{Proj}_k(\cdot)$ function ($1 \leq k \leq d_x + d_z + d_w$) to both sides,

$$(\alpha', \beta', \gamma')e_k \in [\inf \text{Proj}_k(\Theta^{\text{superset},p}), \sup \text{Proj}_k(\Theta^{\text{superset},p})].$$

e_k is the $(d_x + d_z + d_w) \times 1$ -dimensional standard normal vector whose k -th element is 1 and 0 otherwise. So $(\alpha', \beta', \gamma')e_k$ is the k -th coefficient in a vector $(\alpha', \beta', \gamma')$.

This concludes the proof. □

Proof of Proposition 3. .

Proof is analogous to the proof of Proposition of 1.

Proof of (1)

$\Theta^{\text{superset},m}$ is non-empty because it must include $(E[xy]', E[xw_1]', \dots, E[xw_k])'$, as true $f(y, x, z, w) \in \mathcal{M}(f(y, z, w), f(x, z))$. The hyperrectangle $\Theta^{\text{superset},m}$ is a cartesian product of closed, bounded

intervals (from Corollary 1) in Euclidean space, and therefore, it is compact. Also, $\Theta^{\text{superset},m}$ is convex, as it is a hyperrectangle.

Proof of (2)

Let $\Theta_{\kappa_0}^{\text{superset},m} = \{m \in \Theta^{\text{superset},m} | \kappa(m) \leq \kappa_0\}$ be the subset of $\Theta^{\text{superset},m}$ that satisfies the condition number constraint. Under Assumption 3, the true moment vector $(E[xy]', E[xw_1]', \dots, E[xw_{d_w}]')'$ yields a well-conditioned moment matrix with $\kappa(m) \leq \kappa_0$. Therefore, $\Theta_{\kappa_0}^{\text{superset},m}$ is non-empty.

Since the condition number $\kappa(m)$ is a continuous function of m , the set $\{m | \kappa(m) \leq \kappa_0\}$ is closed. The intersection of this closed set with the compact set $\Theta^{\text{superset},m}$ results in $\Theta_{\kappa_0}^{\text{superset},m}$, which is compact.

$\Theta^{\text{superset},p}$ is non-empty because $(E[xy]', E[xw_1]', \dots, E[xw_{d_w}]')' \in \Theta_{\kappa_0}^{\text{superset},m}$, and $\mathcal{B}_{OLS}((E[xy]', E[xw_1]', \dots, E[xw_{d_w}]')') = (\alpha', \beta', \gamma')' \in \Theta^{\text{superset},p}$.

The moment matrix is non-singular over $\Theta_{\kappa_0}^{\text{superset},m}$ due to condition number constraint. Therefore, the function mapping $\mathcal{B}_{OLS}(\cdot)$ is continuous on $\Theta_{\kappa_0}^{\text{superset},m}$, and by the continuous mapping theorem, the compactness of $\Theta_{\kappa_0}^{\text{superset},m}$ implies that $\Theta^{\text{superset},p}$ is also compact.

When $d_w = 0$, the moment matrix $(*)$ does not include any unidentified moments and by Assumption 3, the moment matrix is well-conditioned over all of $\Theta^{\text{superset},m}$. Therefore, $\Theta_{\kappa_0}^{\text{superset},m} = \Theta^{\text{superset},m}$. $\mathcal{B}_{OLS}(\cdot)$ is an affine transformation over a convex domain, $\Theta_{\kappa_0}^{\text{superset},m}$ resulting in a convex $\Theta^{\text{superset},p}$.

However, when $d_w > 0$, the function $\mathcal{B}_{OLS}(\cdot)$ involves matrix inversion and is no longer affine. Consequently, the convexity of $\Theta^{\text{superset},p}$ is not guaranteed, as the image of a convex set under a continuous but non-affine function need not be convex. In addition, the set $\Theta_{\kappa_0}^{\text{superset},m}$ may not be convex due to the condition number constraint.

Proof of (3)

Note that the projection function is linear. Therefore, the non-emptiness and compactness of $\Theta^{\text{superset},p}$ imply its projection is also non-empty and compact. Any compact set on \mathbb{R} is a bounded interval, which is convex.

This concludes the proof. \square

Proof of Proposition 4. .

Proof of (1)

As both $\widehat{\Theta}^{\text{superset},m}$ and $\Theta^{\text{superset},m}$ are hyperrectangles, it suffices to show the lower bound and upper bound interval estimates that define the hyperrectangles to converge in probability to the population lower and upper bounds. By Assumption 4, we have uniform convergence of $\widehat{Q}_{x_k|z}(\widehat{F}_{y|z}(y|z)|z)$ to $Q_{x_k|z}(F_{y|z}(y|z)|z)$ as $N_m \rightarrow \infty$ and subsequently $N_a(N_m) \rightarrow \infty$ for $\forall k \in \{1, \dots, d_x\}$.

For the lower bound estimator:

$$\|\widehat{\mu}_l^{xy} - \mu_l^{xy}\| = \left\| \frac{1}{N_m} \sum_{i=1}^{N_m} y_i \widehat{Q}_{x_k|z}(1 - \widehat{F}_{y|z}(y_i|z_i)|z_i) - E[y \cdot Q_{x_k|z}(1 - F_{y|z}(y|z)|z)] \right\| \quad (\text{C.10})$$

$$\leq \left\| \frac{1}{N_m} \sum_{i=1}^{N_m} y_i \widehat{Q}_{x_k|z}(1 - \widehat{F}_{y|z}(y_i|z_i)|z_i) - \frac{1}{N_m} \sum_{i=1}^{N_m} y_i Q_{x_k|z}(1 - F_{y|z}(y_i|z_i)|z_i) \right\| \quad (\text{C.11})$$

$$+ \left\| \frac{1}{N_m} \sum_{i=1}^{N_m} y_i Q_{x_k|z}(1 - F_{y|z}(y_i|z_i)|z_i) - E[y \cdot Q_{x_k|z}(1 - F_{y|z}(y|z)|z)] \right\|. \quad (\text{C.12})$$

The first term converges to zero in probability by the uniform convergence assumption. The second term converges to zero in probability by the Law of Large Numbers. This holds for all $1 \leq k \leq d_x$. Thus, $\widehat{\mu}_l^{xy} \xrightarrow{p} \mu_l^{xy}$ as $N_m \rightarrow \infty$ and subsequently $N_a(N_m) \rightarrow \infty$. Similarly, we can show $\widehat{\mu}_u^{xy} \xrightarrow{p} \mu_u^{xy}$, $\widehat{\mu}_l^{xw_l} \xrightarrow{p} \mu_l^{xw_l}$, $\widehat{\mu}_u^{xw_l} \xrightarrow{p} \mu_u^{xw_l}$ for all $1 \leq l \leq d_w$.

Proof of (2)

By the triangle inequality for the Hausdorff distance, we have:

$$\mathbf{d}_H(\widehat{\Theta}^{\text{superset},p}, \Theta^{\text{superset},p}) = \mathbf{d}_H(\widehat{\mathcal{B}}_{\text{OLS}}(\widehat{\Theta}^{\text{superset},m}), \mathcal{B}_{\text{OLS}}(\Theta^{\text{superset},m})) \quad (\text{C.13})$$

$$\leq \mathbf{d}_H(\widehat{\mathcal{B}}_{\text{OLS}}(\widehat{\Theta}^{\text{superset},m}), \widehat{\mathcal{B}}_{\text{OLS}}(\Theta^{\text{superset},m})) + \mathbf{d}_H(\widehat{\mathcal{B}}_{\text{OLS}}(\Theta^{\text{superset},m}), \mathcal{B}_{\text{OLS}}(\Theta^{\text{superset},m})). \quad (\text{C.14})$$

For the first term, by Assumption 3, $\widehat{\mathcal{B}}_{\text{OLS}}(\cdot)$ is continuous in a neighborhood of $\Theta_{\kappa_0}^{\text{superset},m}$, that is, a subset of $\Theta^{\text{superset},m}$ that implies a well-conditioned moment matrix. Given $\mathbf{d}_H(\widehat{\Theta}^{\text{superset},m}, \Theta^{\text{superset},m}) \xrightarrow{p} 0$ from part (1), the continuous mapping theorem implies:

$$\mathbf{d}_H(\widehat{\mathcal{B}}_{\text{OLS}}(\widehat{\Theta}^{\text{superset},m}), \widehat{\mathcal{B}}_{\text{OLS}}(\Theta^{\text{superset},m})) \xrightarrow{p} 0. \quad (\text{C.15})$$

For the second term to converge to 0 in probability, it suffices to show the uniform convergence of $\widehat{\mathcal{B}}_{\text{OLS}}(\cdot)$ to $\mathcal{B}_{\text{OLS}}(\cdot)$.

For any fixed $m = \begin{bmatrix} B \\ A_1 \\ \vdots \\ A_{d_w} \end{bmatrix} \in \Theta_{\kappa_0}^{\text{superset},m}$, let's denote:

$$\widehat{G} = \begin{bmatrix} \widehat{s}_{xx'} & \widehat{s}_{xz'} & A' \\ \widehat{s}_{zx'} & \widehat{s}_{zz'} & \widehat{s}_{zw'} \\ A & \widehat{s}_{wz'} & \widehat{s}_{ww'} \end{bmatrix}, \quad G = \begin{bmatrix} E[xx'] & E[xz'] & A' \\ E[zx'] & E[zz'] & E[zw'] \\ A & E[wz'] & E[ww'] \end{bmatrix}$$

$$\widehat{b} = \begin{bmatrix} B \\ \widehat{s}_{zy} \\ \widehat{s}_{wy} \end{bmatrix}, \quad b = \begin{bmatrix} B \\ E[zy] \\ E[wy] \end{bmatrix}.$$

Then $\widehat{\mathcal{B}}_{\text{OLS}}(m) = \widehat{G}^{-1}\widehat{b}$ and $\mathcal{B}_{\text{OLS}}(m) = G^{-1}b$.

Consider the following inequality.

$$\|\widehat{\mathcal{B}}_{\text{OLS}}(m) - \mathcal{B}_{\text{OLS}}(m)\|_2 = \|\widehat{G}^{-1}\widehat{b} - G^{-1}b\|_2 \quad (\text{C.16})$$

$$= \|\widehat{G}^{-1}\widehat{b} - \widehat{G}^{-1}b + \widehat{G}^{-1}b - G^{-1}b\|_2 \quad (\text{C.17})$$

$$\leq \|\hat{G}^{-1}(\hat{b} - b)\|_2 + \|(\hat{G}^{-1} - G^{-1})b\|_2. \quad (\text{C.18})$$

Regarding the first term, consider the submultiplicative property of matrix norms.

$$\|\hat{G}^{-1}(\hat{b} - b)\|_2 \leq \|\hat{G}^{-1}\|_2 \cdot \|\hat{b} - b\|_2. \quad (\text{C.19})$$

Note that the smallest eigenvalue of \hat{G} (denoted as λ_{min}) is bounded away from zero because of the condition number restriction, $\kappa(m) = \lambda_{max}/\lambda_{min} \leq \kappa_0 < \infty$ and because the largest eigenvalue is finite, $\lambda_{max} < \infty$ by finite second moment Assumption 2. Therefore, $\|\hat{G}^{-1}\|_2$ is bounded.

The difference $\|\hat{b} - b\|_2$ involves only the sample moments \hat{s}_{zy} and \hat{s}_{wy} versus their population counterparts, since B is the same in both. Due to Law of Large numbers, $\|\hat{b} - b\|_2 \xrightarrow{p} 0$.

Therefore, $\|\hat{G}^{-1}(\hat{b} - b)\|_2 \xrightarrow{p} 0$.

For the second term, consider the following relationship.

$$\hat{G}^{-1} - G^{-1} = \hat{G}^{-1}(G - \hat{G})G^{-1}. \quad (\text{C.20})$$

Then,

$$\|(\hat{G}^{-1} - G^{-1})b\|_2 = \|\hat{G}^{-1}(G - \hat{G})G^{-1}b\|_2 \quad (\text{C.21})$$

$$\leq \|\hat{G}^{-1}\|_2 \cdot \|G - \hat{G}\|_2 \cdot \|G^{-1}\|_2 \cdot \|b\|_2. \quad (\text{C.22})$$

I have shown above that $\|\hat{G}^{-1}\|_2$ is bounded. By similar logic, $\|G^{-1}\|_2$ is bounded as well.

The term $\|G - \hat{G}\|_2$ involves differences between sample moments and population moments.

By the Law of Large Numbers, these differences converge to zero in probability.

Since b includes B which comes from the compact set $\Theta_{\kappa_0}^{\text{superset},m}$, $\|b\|_2$ is bounded. Thus, $\|(\hat{G}^{-1} - G^{-1})b\|_2 \xrightarrow{p} 0$.

So far, I showed pointwise convergence. That is, for any fixed $m \in \Theta^{\text{superset},m}$, $\|\hat{\mathcal{B}}_{\text{OLS}}(m) - \mathcal{B}_{\text{OLS}}(m)\|_2 \xrightarrow{p} 0$.

Next, $\widehat{\mathcal{B}}_{\text{OLS}}$ and \mathcal{B}_{OLS} are equicontinuous functions because of the bounded eigenvalue, which is implied by the condition number restriction $\kappa(m) \leq \kappa_0 < \infty$. Remember that $\Theta^{\text{superset},m}$ as well as its subset $\Theta_{\kappa_0}^{\text{superset},m}$ are compact from Proposition 3 and its proof. Therefore, this proves the uniform convergence of $\widehat{\mathcal{B}}_{\text{OLS}}(m)$ to $\mathcal{B}_{\text{OLS}}(m)$.

$$\sup_{m \in \Theta^{\text{superset},m}} \|\widehat{\mathcal{B}}_{\text{OLS}}(m) - \mathcal{B}_{\text{OLS}}(m)\|_2 \xrightarrow{p} 0. \quad (\text{C.23})$$

This implies the second term in our decomposition converges to zero in probability.

$$\mathbf{d}_H(\widehat{\mathcal{B}}_{\text{OLS}}(\Theta^{\text{superset},m}), \mathcal{B}_{\text{OLS}}(\Theta^{\text{superset},m})) \xrightarrow{p} 0. \quad (\text{C.24})$$

To see this, consider any point $p \in \widehat{\mathcal{B}}_{\text{OLS}}(\Theta^{\text{superset},m})$. By definition, $p = \widehat{\mathcal{B}}_{\text{OLS}}(m)$ for some $m \in \Theta^{\text{superset},m}$. The distance from p to the set $\mathcal{B}_{\text{OLS}}(\Theta^{\text{superset},m})$ is at most $\|\widehat{\mathcal{B}}_{\text{OLS}}(m) - \mathcal{B}_{\text{OLS}}(m)\|_2$, since $\mathcal{B}_{\text{OLS}}(m) \in \mathcal{B}_{\text{OLS}}(\Theta^{\text{superset},m})$. Taking the supremum over all such p , we get:

$$\sup_{p \in \widehat{\mathcal{B}}_{\text{OLS}}(\Theta^{\text{superset},m})} \inf_{q \in \mathcal{B}_{\text{OLS}}(\Theta^{\text{superset},m})} \|p - q\|_2 \leq \sup_{m \in \Theta^{\text{superset},m}} \|\widehat{\mathcal{B}}_{\text{OLS}}(m) - \mathcal{B}_{\text{OLS}}(m)\|_2. \quad (\text{C.25})$$

A similar argument applies in the reverse direction. Therefore:

$$\mathbf{d}_H(\widehat{\mathcal{B}}_{\text{OLS}}(\Theta^{\text{superset},m}), \mathcal{B}_{\text{OLS}}(\Theta^{\text{superset},m})) \xrightarrow{p} 0. \quad (\text{C.26})$$

To sum up, as both terms converge to zero in probability as $N_m \rightarrow \infty$ and subsequently $N_a(N_m) \rightarrow \infty$, we conclude:

$$\mathbf{d}_H(\widehat{\Theta}^{\text{superset},p}, \Theta^{\text{superset},p}) \xrightarrow{p} 0. \quad (\text{C.27})$$

Proof of (3)

Since $\text{Proj}(\cdot)$ is a continuous function, when $d_H(\widehat{\Theta}^{\text{superset},p}, \Theta^{\text{superset},p}) \xrightarrow{p} 0$, $d_H(\text{Proj}_k(\widehat{\Theta}^{\text{superset},p}), \text{Proj}_k(\Theta^{\text{superset},p})) \xrightarrow{p} 0$ for $\forall k \in \{1, \dots, d_x + d_z + d_w\}$.

Since the projection sets converge in Hausdorff distance, their infima and suprema must also converge in probability. Therefore,

$$\|(\widehat{\alpha}_l^{\text{superset}'}, \widehat{\beta}_l^{\text{superset}'}, \widehat{\gamma}_l^{\text{superset}'}) - (\alpha_l^{\text{superset}'}, \beta_l^{\text{superset}'}, \gamma_l^{\text{superset}'})\|_2 \xrightarrow{p} 0.$$

$$\|(\widehat{\alpha}_u^{\text{superset}'}, \widehat{\beta}_u^{\text{superset}'}, \widehat{\gamma}_u^{\text{superset}'}) - (\alpha_u^{\text{superset}'}, \beta_u^{\text{superset}'}, \gamma_u^{\text{superset}'})\|_2 \xrightarrow{p} 0.$$

This concludes the proof. \square

Proof of Proposition B.1. .

Assume $d_w = 0$. Note that I do not make a condition number constraint Assumption 3 when

$d_w = 0$, as none of the unidentified moments appear in the moment matrix (*) in equation 4.

Then the equation 12 reduces to

$$\Theta^{\text{superset},p} = \left\{ \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} M \\ E[zy] \end{bmatrix} \mid M \in \Theta^{\text{superset},m} \right\} \quad (\text{C.28})$$

$$= \left\{ \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} m_1 \\ \vdots \\ m_{d_x} \\ E[zy] \end{bmatrix} \mid \mu_l^{x_1 y} \leq m_1 \leq \mu_u^{x_{d_x} y}, \dots, \mu_l^{x_{d_x} y} \leq m_{d_x} \leq \mu_u^{x_{d_x} y} \right\} \quad (\text{C.29})$$

$$= \left\{ \begin{bmatrix} A \begin{bmatrix} m_1 \\ \vdots \\ m_{d_x} \end{bmatrix} + BE[zy] \\ C \begin{bmatrix} m_1 \\ \vdots \\ m_{d_x} \end{bmatrix} + DE[zy] \end{bmatrix} \mid \mu_l^{x_1 y} \leq m_1 \leq \mu_u^{x_{d_x} y}, \dots, \mu_l^{x_{d_x} y} \leq m_{d_x} \leq \mu_u^{x_{d_x} y} \right\}. \quad (\text{C.30})$$

M represents the unidentified moment $E[xy]$. Note that this equation is linear in M . To

expand the k -th element in the image vector ($1 \leq k \leq d_x$),

$$A_{k.} \begin{bmatrix} m_1 \\ \vdots \\ m_{d_x} \end{bmatrix} + B_{k.} E[zy] \quad (\text{C.31})$$

By definition of infimum,

$$(\alpha_l^{superset})_k = \inf_{\substack{\mu_l^{x_1y} \leq m_1 \leq \mu_u^{x_1y} \\ \vdots \\ \mu_l^{x_{d_x}y} \leq m_{d_x} \leq \mu_u^{x_{d_x}y}}} \left\{ A_{k.} \begin{bmatrix} m_1 \\ \vdots \\ m_{d_x} \end{bmatrix} + B_{k.} E[zy] \right\} \quad (\text{C.32})$$

$$= \inf_{\substack{\mu_l^{x_1y} \leq m_1 \leq \mu_u^{x_1y} \\ \vdots \\ \mu_l^{x_{d_x}y} \leq m_{d_x} \leq \mu_u^{x_{d_x}y}}} \left\{ \sum_{l=1}^{d_x} A_{kl} m_l \right\} + B_{k.} E[zy] \quad (\text{C.33})$$

$$= \sum_{l=1}^{d_x} \inf_{\mu_l^{x_l y} \leq m_1 \leq \mu_u^{x_l y}} \left\{ A_{kl} m_l \right\} + B_{k.} E[zy] \quad (\text{C.34})$$

$$= \sum_{l=1}^{d_x} \min \left(A_{kl} \mu_l^{x_l y}, A_{kl} \mu_u^{x_l y} \right) + B_{k.} E[zy] \quad (\text{C.35})$$

$A_{k.}$ is the k -th row of a matrix A , $B_{k.}$ is the k -th row of a matrix B , and m_l is the l -th element of m .

In the second-to-last equality, I use the following property of a minimum operator. Let $x \equiv (x_1, \dots, x_{d_x})$ be a vector of length d_x , whose domain is denoted as $\chi \equiv \chi_1 \times \dots \times \chi_{d_x}$. Assuming each x_l can vary independently over its domain χ_l , then we have

$$\min_{x \in \chi} \left\{ \sum_{l=1}^{d_x} x_l \right\} = \sum_{l=1}^{d_x} \min_{x_l \in \chi_l} \{x_l\}. \quad (\text{C.36})$$

Due to linearity, the minimum is achieved when A_{kl} is multiplied by either the minimum or

maximum of the interval bound. So this implies the last equality and gives the lower bound equation B.2.

The upper bound can be proven using a similar technique, but using the maximum operator property instead.

$$\max_{x \in \chi} \left\{ \sum_{l=1}^{d_x} x_l \right\} = \sum_{l=1}^{d_x} \max_{x_l \in \chi_l} \{x_l\}. \quad (\text{C.37})$$

Similar results can be shown for β_l, β_u as well. Note that Proposition 3 implies $\Theta^{\text{superset}, p}$ is compact, and therefore the bounds $\alpha_l, \beta_l, \alpha_u, \beta_u$ are finite.

About singleton result, Theorem 3.2. in Fan et al. (2017) has shown that when $F(x_k|z)$ is degenerate for almost all z , then the sharp bound on the moments $E[x_k y]$ becomes singleton. So, this implies that the bound on the parameter α_k is a singleton too. Similar relationship can be shown for all elements of (α', β') . Note that Fan et al. (2017) stated another condition that leads to a singleton bound: that is, when $F(y|z)$ is degenerate for almost all z . However, this is not plausible in our case because of the presence of the error term u in the outcome equation. This concludes the proof. \square

Proof of Proposition B.2. .

We first establish the convergence of the moment bounds $\hat{\mu}_l^{xy}$ and $\hat{\mu}_u^{xy}$. By Assumption 4, we have uniform convergence of $\hat{Q}_{x_k|z}(\hat{F}_{y|z}(y|z)|z)$ to $Q_{x_k|z}(F_{y|z}(y|z)|z)$ as $N_m \rightarrow \infty$ and subsequently $N_a(N_m) \rightarrow \infty$ for $\forall k$.

For the lower bound estimator:

$$\|\hat{\mu}_l^{xy} - \mu_l^{xy}\| = \left\| \frac{1}{N_m} \sum_{i=1}^{N_m} y_i \hat{Q}_{x_k|z}(1 - \hat{F}_{y|z}(y_i|z_i)|z_i) - E[y \cdot Q_{x_k|z}(1 - F_{y|z}(y|z)|z)] \right\| \quad (\text{C.38})$$

$$\leq \left\| \frac{1}{N_m} \sum_{i=1}^{N_m} y_i \hat{Q}_{x_k|z}(1 - \hat{F}_{y|z}(y_i|z_i)|z_i) - \frac{1}{N_m} \sum_{i=1}^{N_m} y_i Q_{x_k|z}(1 - F_{y|z}(y_i|z_i)|z_i) \right\| \quad (\text{C.39})$$

$$+ \left\| \frac{1}{N_m} \sum_{i=1}^{N_m} y_i Q_{x_k|z}(1 - F_{y|z}(y_i|z_i)|z_i) - E[y \cdot Q_{x_k|z}(1 - F_{y|z}(y|z)|z)] \right\|. \quad (\text{C.40})$$

The first term converges to zero in probability by the uniform convergence assumption. The second term converges to zero in probability by the Law of Large Numbers. Thus, $\hat{\mu}_l^{xy} \xrightarrow{p} \mu_l^{xy}$ as $N_m \rightarrow \infty$ and subsequently $N_a(N_m) \rightarrow \infty$. Similarly, we can show $\hat{\mu}_u^{xy} \xrightarrow{p} \mu_u^{xy}$ for the upper bound.

Next, under Assumption B.1, $\begin{bmatrix} E[xx'] & E[xz'] \\ E[zx'] & E[zz'] \end{bmatrix}$ is a non-singular matrix. Therefore, $\|\hat{A} - A\|_2, \|\hat{B} - B\|_2, \|\hat{C} - C\|_2, \|\hat{D} - D\|_2 \xrightarrow{p} 0$, as $N_m \rightarrow \infty$ and subsequently $N_a(N_m) \rightarrow \infty$. Finally, the Estimator 2 are continuous functions of $\hat{A}, \hat{B}, \hat{C}, \hat{D}, \hat{\mu}_l^{xy}, \hat{\mu}_u^{xy}, \hat{s}_{zy}$. By the continuous mapping theorem and the Law of Large Numbers, we have:

$$\hat{\alpha}_l^{superset} \xrightarrow{p} \tilde{\alpha}_l^{superset}, \quad (\text{C.41})$$

$$\hat{\beta}_l^{superset} \xrightarrow{p} \tilde{\beta}_l^{superset}, \quad (\text{C.42})$$

$$\hat{\alpha}_u^{superset} \xrightarrow{p} \tilde{\alpha}_u^{superset}, \quad (\text{C.43})$$

$$\hat{\beta}_u^{superset} \xrightarrow{p} \tilde{\beta}_u^{superset}. \quad (\text{C.44})$$

This completes the proof. \square

Proof of Proposition B.3. .

To show $\phi(\cdot)$ is Hadamard directionally differentiable at θ_0 , it suffices to show each element $\phi_k(\theta_0)$ is Hadamard directionally differentiable at θ_0 .

The function $\phi_k(\cdot)$ involves a matrix inversion to compute A, B, C, D , matrix multiplication, vector summation, and min and max operators. Note that the matrix inverse function at any non-singular matrix is Fréchet differentiable: denote the matrix inverse function $h(M) = M^{-1}$. Then function $h(\cdot)$'s derivative in the direction of G is,

$$\lim_{t \rightarrow 0} \frac{h(M + tG) - h(M)}{t} = -M^{-1}GM^{-1}. \quad (\text{C.45})$$

(Magnus and Neudecker (2019), Section 8.4). Therefore, under Assumption 2 and Assumption B.1, A, B, C, D are differentiable functions of population moments $E[xx'], E[xz'], E[zz']$.

Also, matrix multiplication and vector summation functions are trivially Hadamard differentiable. The $\min(x_1, x_2) : \mathbb{R}^2 \rightarrow \mathbb{R}$ and $\max(x_1, x_2) : \mathbb{R}^2 \rightarrow \mathbb{R}$ functions, which cause non-differentiability of Estimator 2, are Hadamard directionally differentiable functions as known in the literature (Masten and Poirier (2020))³⁷.

To sum up, $\phi_k(\cdot)$ for every k , as well as $\phi(\cdot)$ are Hadamard directionally differentiable at θ_0 . This concludes the proof. \square

D Monte Carlo Simulation Details

In this section, I present four sets of Monte Carlo simulations. Each Monte Carlo simulation is summarized below.

1. In Online Appendix Section D.1, I compare Estimator 2 with Pacini (2019)'s estimator in Monte Carlo simulation and show that the main difference between these two estimators is in bias correction³⁸: Estimator 2 gives estimates that are close to Pacini (2019)'s estimates when bias correction step in Pacini (2019) is omitted (Figure D.1, Table D.1). I intentionally skip the bias correction for three reasons: First, the bias must disappear asymptotically, as shown in the proof of the consistency result. So the bias concern is relieved when the sample size is big. Moreover, omitting the bias correction makes the inference only conservative while remaining valid. Second, it is easy to overcorrect the bias in practice. I show that the Pacini (2019)'s estimates without bias correction step³⁹ are closer to the population bounds. Third, bootstrap-based bias correction is computationally costly. Table D.2 shows that omitting the bias correction step in Pacini (2019)'s estimator saves 99% of the computational time.
2. In Online Appendix Section D.2, I show that the Estimator 2 performs well even when the proxy variables are contaminated with large amounts of measurement errors. I

³⁷See page 90 in Masten and Poirier (2020) for the Hadamard directional derivatives of these functions.

³⁸However, note that I consider bias correction when computing the confidence interval, so utilizing the bias-corrected confidence interval complements inference.

³⁹I have used the Pacini (2019)'s code for implementing the bias correction.

vary the variance of the measurement error between 10% and 100% of the variance of the latent omitted variable and show that the simulated estimate distributions are very similar even when the measurement error variance is the largest (Figure D.2).

3. Online Appendix Section D.3 shows the performance of the most general Estimator 1 for the case where $d_w > 0$ (Figure D.3). Most lower-bound to upper-bound intervals cover the true parameters, confirming the validity of the estimator. The computational time for this case was reasonable: the average time to compute the estimator once was 18 seconds.
4. Online Appendix Section D.4 demonstrates choosing a tuning parameter from double bootstrapping to achieve the target nominal rate. Next, I compute the confidence intervals for the lower and upper bounds using the chosen tuning parameter (Figure D.4). The coverage rates of the confidence interval estimates for the lower and upper bounds are similar to those found in double bootstrapping. The coverage rate of the confidence interval of the form $[\inf CI^{equal-tailed}(\phi^{lb}(\hat{\theta})), \sup CI^{equal-tailed}(\phi^{ub}(\hat{\theta}))]$ for the parameter of interest is greater than $(1 - \tau)$ as claimed in [Imbens and Manski \(2004\)](#) (Table D.5).

D.1 Comparison with Pacini (2019)

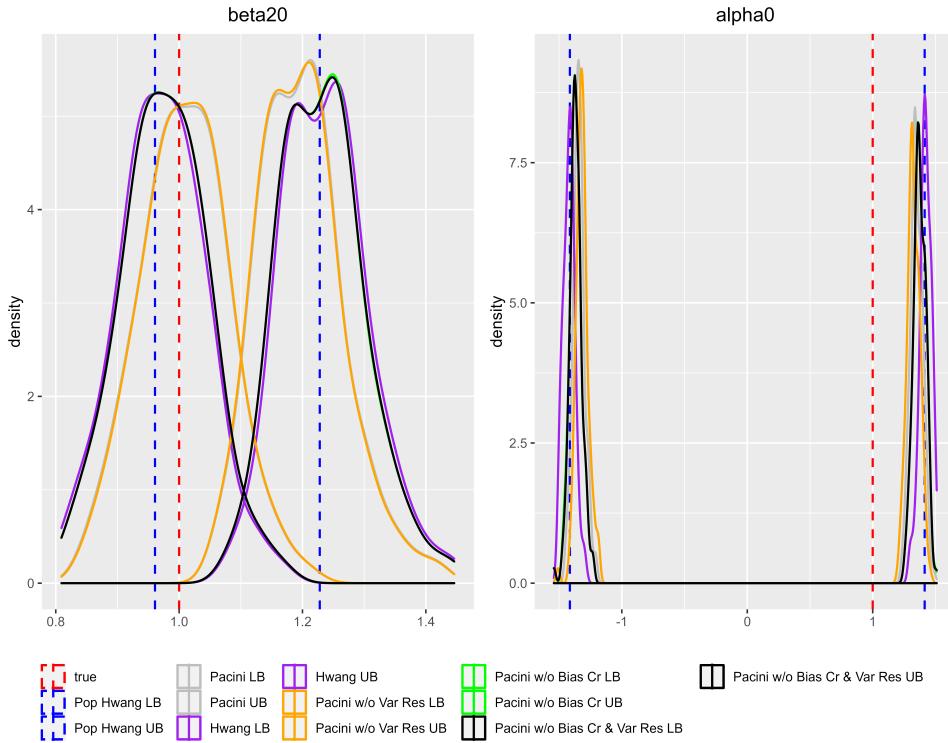
In this section, I compare Estimator 2 (“Hwang ($d_w = 0$)” with [Pacini \(2019\)](#)’s estimator. I use the same DGP and simulation sample size used in [Pacini \(2019\)](#)⁴⁰ and apply both Estimator 2 and [Pacini \(2019\)](#)’s estimator. To understand the differences between these two estimators, I additionally implement variations of [Pacini \(2019\)](#)’s estimator: in addition to the (i) original

⁴⁰I repeat the DGP used in [Pacini \(2019\)](#) here:

$$\begin{aligned} y &= \beta_{1o} + x\alpha_0 + z\beta_{2o} + u, \quad u \sim N(0, 1). \\ z &\sim N(0, 1). \\ x &= \delta z + \nu, \quad \nu \sim N(0, 1). \\ \theta_0 &\equiv (\alpha_o, \beta_{1o}, \beta_{2o}) = (1, 0, 1), \quad \delta = 0.0944. \end{aligned}$$

The sample sizes of the main and auxiliary datasets are both 1000. The number of Monte Carlo replications is 100.

Figure D.1: Comparison with Pacini (2019)



Note: To compute my bounds, I used the function ‘`bndovb`’ in the R package ‘`bndovb`’ (version 1.2).

Pacini (2019) estimator, I implement (ii) Pacini (2019) without a restriction that the variance-covariance matrix of the variables in the model (y, x, z) is positive semi-definite, (iii) Pacini (2019) without bias correction (Section 4.2 of Pacini (2019)), (iv) combination of (ii) and (iii).

Figure D.1 shows the simulation results. The red dashed line shows the true parameter used in the data generating process. The blue dashed lines show the population lower and upper bounds of Estimator 2, and the purple lines show Estimator 2’s lower and upper bound estimates. The gray, orange, green, black solid lines show the (i) original Pacini (2019), (ii) Pacini (2019) without variance restriction, (iii) Pacini (2019) without bias correction, (iv) Pacini (2019) without variance restriction and bias correction. Note that the population bounds of Estimator 2 must be the same as the population bound of (ii) Pacini (2019) without variance restriction. Table D.1 shows the summary statistics of these estimates.

Monte Carlo simulation results show that the important difference between Estimator 2 and Pacini (2019)’s is in bias correction. Pacini (2019)’s estimates without bias correction ((iii) and (iv)) are almost identical to those of Estimator 2. Also, Pacini (2019) turns out to over-

Table D.1: Summary Statistics of Estimates in Figure D.1

True Parameter	$\beta_{20} = 1$				$\alpha_0 = 1$			
	β_{20} LB		β_{20} UB		α_0 LB		α_0 UB	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Pop Hwang ($d_w = 0$)	0.96		1.23		-1.41		1.41	
Hwang ($d_w = 0$)	0.97	0.07	1.24	0.07	-1.42	0.05	1.42	0.05
Pacini	1.01	0.07	1.20	0.07	-1.34	0.05	1.35	0.05
Pacini w/o Var Res	1.01	0.07	1.20	0.07	-1.32	0.05	1.33	0.05
Pacini w/o Bias Cr	0.98	0.07	1.23	0.07	-1.37	0.05	1.37	0.05
Pacini w/o Bias Cr & Var Res	0.98	0.07	1.23	0.07	-1.37	0.05	1.37	0.05

Note: This table shows summary statistics (mean, standard errors) of estimates shown in Figure D.1.

Table D.2: Computation Time Comparison

	time (in seconds)
Pacini Duration	99232.62
Pacini w/o Variance Restriction Duration	96279.23
Pacini w/o Bias Correction Duration	960.91
Pacini w/o Variance Restriction and Bias Correction Duration	960.06
Hwang Duration	42.31

Note: Both estimators were computed on the same computer using the same software, the Dell Precision 7920 workstation with the R version 4.2.0 Patched and the platform x86_64-w64-mingw32 / x64 (64-bit). Both estimators did not use any parallelization in computation.

correct the bias under his bias correction step. Both Estimator 2 (purple line) and (ii) Pacini (2019) without variance restriction (orange line) share the same population bound (blue vertical line), but the mean of Estimator 2 is closer to the population bound than the mean of (ii) Pacini (2019) without variance restriction. In fact, (ii) Pacini (2019) bias correction makes the interval estimates much narrower than the population bound; that is, it over-corrects for the bias.

Bias correction is also the main source of the computational time difference. Estimator 2 took 42 seconds to compute 100 times, which is only 0.04% of Pacini (2019)'s computation time, 99,233 seconds (Table D.2). Removing the bias correction step in Pacini (2019)'s estimator saves 98,272 seconds, which is 99% of the total computational time needed to compute Pacini (2019)'s estimator. Omitting the variance restriction does not change much the computation time of Pacini (2019) estimator – 99,232.62 seconds (=27.56 hours) versus 96,279.23

seconds (=26.74 hours)⁴¹.

D.2 When Proxy Variable in Auxiliary Data Are Contaminated with Measurement Errors

In this section, I relax the assumption that the auxiliary dataset includes the covariates missing from the main dataset measured without error, and I show Estimator 2 still performs well in this setting with large amounts of measurement errors.

$$y = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \beta_3 x + \epsilon. \quad (\text{D.1})$$

The main data is assumed to include (y, z_1, z_2) but the auxiliary data includes $(z_1, z_2, \{m_k\}_{k=1}^K)$, where $\{m_k\}_{k=1}^K$ ($K \geq 3$) are noisy proxy variables for the unobservable x ; that is, auxiliary data no longer includes the latent omitted variable x . I assume $K = 3$ in this exercise, which is the minimum number of proxies necessary for nonparametric identification of the joint density for (z_1, z_2, x) in the auxiliary data (Hu (2008), Hu and Schennach (2008), and Cunha et al. (2010))⁴². To simulate noisy proxy variables, I assume the following data generating process:

$$m_k = x + \nu_k, \quad \nu_k \stackrel{i.i.d.}{\sim} N(0, \sigma_\nu^2).$$

To compute the closed-form population bounds and to compare them with estimates obtained from a finite sample, I assume joint normality between regressor variables, which gives

⁴¹Estimator 2 is still faster than Pacini (2019)'s estimator without computationally costly bias correction. There are two main reasons why the computation times are different. First, there are differences in implementation. Pacini (2019) adopted a support function characterization for an identified set in computation, because it is convex in the case of $d_w = 0$. In contrast, I utilized a closed-form expression for projected bounds without defining support functions (Estimator 2 for $d_w = 0$). Second, Pacini (2019) used a sieve estimation for the CDF and quantile function approximation, whereas I estimated them under the parametric normality assumption, which holds true for the data-generating process chosen for this exercise.

⁴²In fact, under the parametric error-in-variable assumption I make in this section, two proxy variables are sufficient. This is shown in Cunha et al. (2010). Three proxy variables are necessary in a general case.

a very simple covariance structure. Also, I generate z_1 conditional on z_2 and x , so that the short regression yields an omitted variable bias.

$$\begin{bmatrix} z_2 \\ x \\ u \\ \epsilon \end{bmatrix} \sim MVN \left(\mu \equiv \begin{bmatrix} \mu_{z_2} \\ \mu_x \\ \mu_u \\ \mu_\epsilon \end{bmatrix}, \Sigma \equiv \begin{bmatrix} \sigma_{z_2}^2 & 0 & 0 & 0 \\ 0 & \sigma_x^2 & 0 & 0 \\ 0 & 0 & \sigma_u^2 & 0 \\ 0 & 0 & 0 & \sigma_\epsilon^2 \end{bmatrix} \right).$$

$$z_1 = \gamma_0 + \gamma_1 z_2 + \gamma_2 x + u.$$

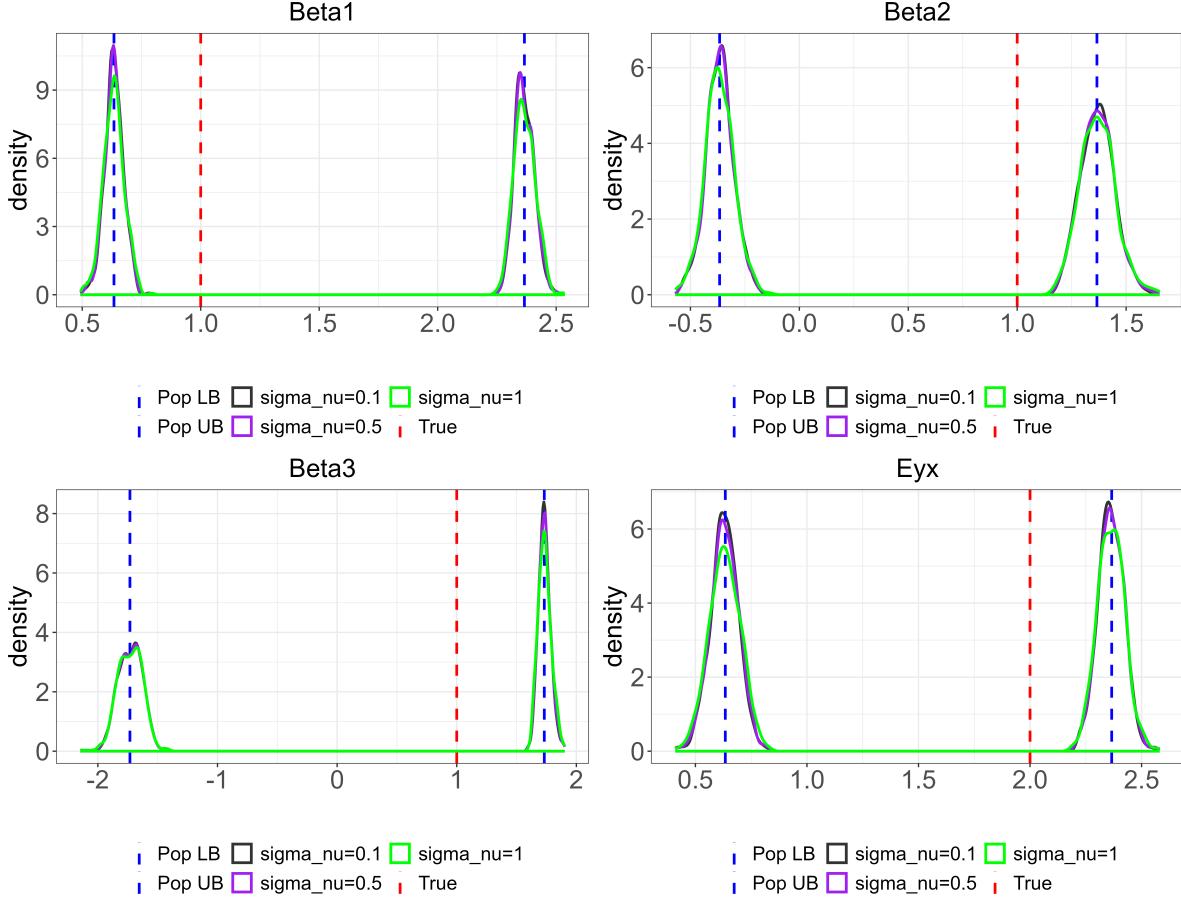
Then, the outcome variable and the regressors follow a joint normal distribution.

$$\begin{bmatrix} z_2 \\ x \\ z_1 \\ y \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \gamma_1 & \gamma_2 & 1 & 0 \\ (\beta_2 + \beta_1 \gamma_1) & (\beta_3 + \beta_1 \gamma_2) & \beta_1 & 1 \end{bmatrix}}_{\equiv A} \begin{bmatrix} z_2 \\ x \\ u \\ \epsilon \end{bmatrix} + \underbrace{\begin{bmatrix} 0 \\ 0 \\ \gamma_0 \\ \beta_0 \end{bmatrix}}_{\equiv B} \sim MVN(A\mu + B, A'\Sigma A).$$

For simulation, I set $\mu = (0, 0, 0, 0)$, $\Sigma = I$, $\beta = (0, 1, 1, 1)$, $\gamma = (0, 1, 1)$. To illustrate how measurement error in proxy variables affects the estimation, I vary the variance of the measurement error σ_ν^2 between 10% and 100% of the variance of the latent omitted variable x : when $\sigma_\nu^2 = Var(x) = 1$, the measurement errors are largest. When $\sigma_\nu^2 = 0.1 \times Var(x) = 0.1$, the measurement errors are smallest.

Figure D.2 reports the results of the Monte Carlo exercise from 500 simulations. These results suggest that the proposed Estimator 2 (for when $d_w = 0$) works well even when proxy variables are the most imprecise. A larger measurement error variance increases the standard errors of the bound estimates, but the overall distribution of the bound estimates is similar to when the measurement error variance is small (i.e., $\sigma_\nu = 0.1$).

Figure D.2: 500 times simulations, Sample Size $N_m = N_a = 3000$, $\beta = (0, 1, 1, 1)$, $\gamma = (0, 1, 1)$. Measurement Error SD $\{0.1, 0.5, 1\}$



Note : This figure shows 500 simulations of Estimator 2. Both main data and auxiliary data are of sample size 3000. “True” shows a true parameter used for the data generating process. “Pop LB” and “Pop UB” show the population lower and upper bounds of Estimator 2. “sigma_nu=0.1”, “sigma_nu=0.5”, “sigma_nu=1” show Estimator 2 estimates when the standard deviation of the measurement error is 0.1, 0.5, 1, respectively.

D.3 Monte Carlo Evidence of Estimator 1

This section presents the Monte Carlo evidence on the performance of the Estimator 1 for when there is a covariate w whose exact measurements or their proxies are not included in the auxiliary data ($d_w > 0$).

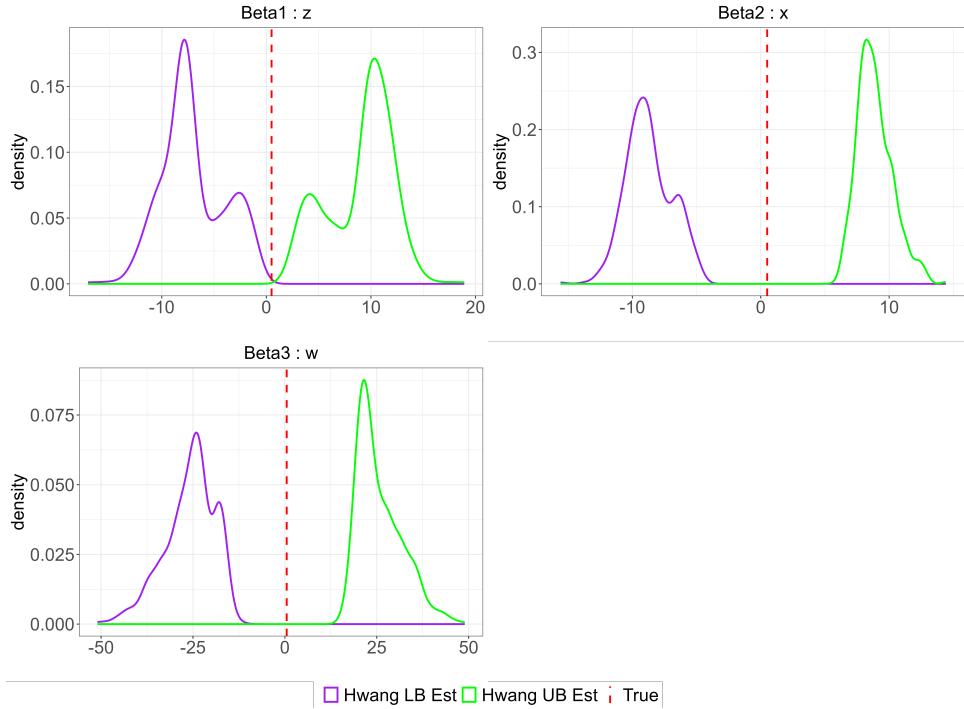
I use the following data-generating process:

$$y = \beta_0 + \beta_1 z + \beta_2 x + \beta_3 w + \epsilon, \quad \epsilon \sim N(0, 0.2). \quad (\text{D.2})$$

$$z \sim N(0, 1). \quad (\text{D.3})$$

$$w = 0.8z + \nu, \quad \nu \sim N(0, 0.3). \quad (\text{D.4})$$

Figure D.3: Monte Carlo Evidence of Estimator 1 for the Most General Case $d_w \geq 0$



Note: “Hwang LB Est” and “Hwang UB Est” show the lower bound and the upper bound estimates of Estimator 1 for when $d_w > 0$. “True” shows the true parameter used for the data generating process.

$$x = -0.2z + \omega, \quad \omega \sim N(0, 0.1). \quad (\text{D.5})$$

$$(\beta_0, \beta_1, \beta_2, \beta_3) = (0, 0.5, 0.3, 0.1). \quad (\text{D.6})$$

The main dataset includes (y, z, w) and the auxiliary dataset includes (z, x) . So z exists only in the main dataset. The sample sizes of the main and the auxiliary datasets are both 3000. The number of Monte Carlo replications is 500. Due to the normality assumption in the data-generating process, I parametrize the CDF, $F_{y|z}$, $F_{w|z}$, and the quantile function $Q_{x|z}$ assuming normality. I exclude the moment vector which gives the condition number (ratio of the largest to the smallest non-zero singular value of the matrix) of the matrix greater than 2000. Due to the non-convexity of the identified set of parameters (Figure 1), I use a global optimizer “DEoptim” (Mullen et al. (2011)).

Figure D.3 confirms that the interval connecting the lower bound and upper bound almost always includes the true parameter. In terms of computational time, it only took 18 seconds on average to compute the estimate once (Table D.3).

Table D.3: Computational Time for Estimator 1

(seconds)	
Mean	SD
18.51	1.97

Note: 500 Monte Carlo replications were parallelized on a Dell XPS 15 9510 laptop using 15 11th Gen Intel(R) Core(TM) i9-11900H @2.50GHz CPUs. I used the R version 4.2.2 and the platform was *x86_64 – w64–mingw32/x64*. The table reports the average time to compute the Estimator 1 once and the standard deviation of the computation time.

D.4 Computation of Confidence Intervals

In this section, I illustrate how to compute confidence intervals. I use the following data generating process (DGP).⁴³ Using the population bounds, I assess the coverage rate of the confidence interval estimates. I assume the main data includes (y, z) and the auxiliary data includes (x, z) . As $d_w = 0$ in this case, I apply Estimator 2.

$$y = \beta_{1o} + x\alpha_0 + z\beta_{2o} + u, \quad u \sim N(0, 1).$$

$$z \sim N(0, 1).$$

$$x = \delta z + \nu, \quad \nu \sim N(0, 1).$$

$$\theta_0 \equiv (\alpha_o, \beta_{1o}, \beta_{2o}) = (1, 0, 1), \quad \delta = 0.0944.$$

Table D.4 shows the coverage rates from the double bootstrap under different tuning parameters, $\{N_m^{-0.49}, N_m^{-0.4}, N_m^{-0.3}, N_m^{-0.2}, N_m^{-0.1}\}$. All but $N_m^{-0.1}$ provide similar coverage rates between 93% and 94%, which are quite close to the nominal rate 95%. The last tuning parameter $N_m^{-0.1}$ provides a coverage rate 99%, which is unnecessarily very high. Given similar coverage rates, it is better to choose a tuning parameter $N_m^{-0.49}$, which provides the tightest confidence interval.

Figure D.4 shows 500 simulations of bound and confidence interval estimates for lower

⁴³The DGP is same as the one used in Appendix Section D.1 and Pacini (2019).

Table D.4: Coverage Rates from Double Bootstrap with Different Tuning Parameters

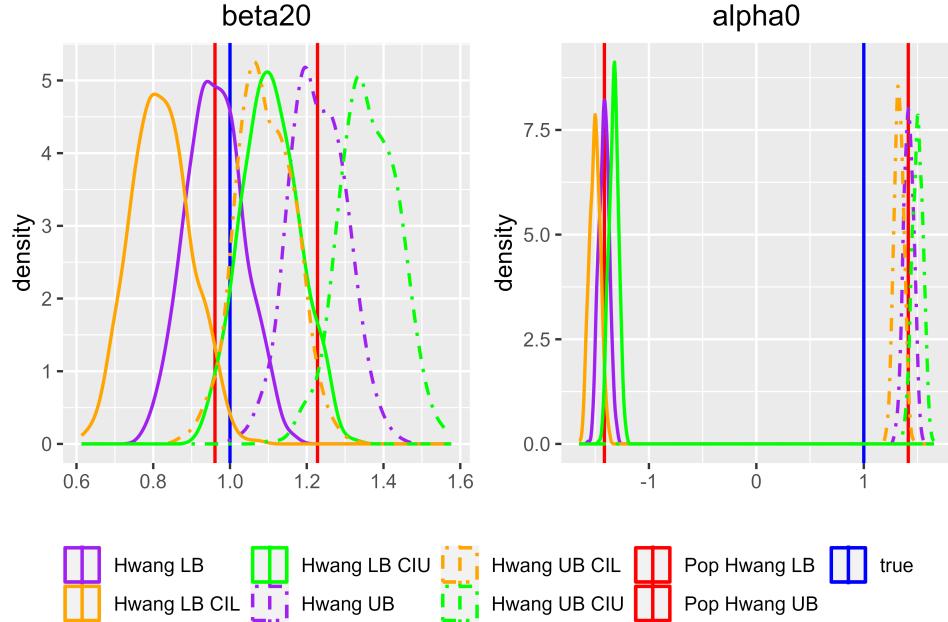
	Tuning Parameter				
	$N_m^{-0.49}$	$N_m^{-0.4}$	$N_m^{-0.3}$	$N_m^{-0.2}$	$N_m^{-0.1}$
$\inf \beta_{10}$	0.936	0.92	0.922	0.944	0.988
$\sup \beta_{10}$	0.944	0.936	0.938	0.954	0.988
$\inf \alpha_0$	0.906	0.898	0.894	0.916	0.996
$\sup \alpha_0$	0.942	0.94	0.93	0.946	0.994
$\inf \beta_{20}$	0.936	0.942	0.938	0.926	1
$\sup \beta_{20}$	0.948	0.946	0.936	0.928	1
overall	0.935	0.930	0.926	0.936	0.994

Note : This table presents the coverage rates from the double bootstrap with different tuning parameters. The number of bootstrap draws is (500, 500).

and upper bounds when the tuning parameter $N_m^{-0.49}$ is chosen. “true” shows the true parameter used for simulations, and “Pop Hwang LB” and “Pop Hwang UB” stand for population lower and upper bounds. “Hwang LB” and “Hwang UB” stand for Estimator 2’s lower and upper bound estimates. “Hwang LB CIL” and “Hwang LB CIU” show the (lower and upper) endpoints of the 95% confidence interval of the lower bound. “Hwang UB CIL” and “Hwang UB CIU” are the (lower and upper) endpoints of the 95% confidence interval of the upper bound.

Table D.5 shows the coverage rate of three different confidence intervals: (1) confidence interval for the population lower bound – how often [Hwang LB CIL, Hwang LB CIU] contains “Pop Hwang LB”, (2) confidence interval for the population upper bound – how often [Hwang UB CIL, Hwang UB CIU] contains “Pop Hwang UB” and (3) confidence interval for the parameter of interest, how often [Hwang LB CIL, Hwang UB CIU] contains “true”. The (3) confidence interval for the parameter of interest would be of ultimate interest in empirical works. The results confirm that the coverage rates for “Pop Hwang LB”, “Pop Hwang UB” are similar to the ones found from double bootstrapping in Table D.4. They are slightly below 95% as in Table D.4. However, the coverage rates for the parameter of interest are all above 95%. This is consistent with the finding in [Imbens and Manski \(2004\)](#) that the confidence interval of form $[\inf CI^{equal-tailed}(\phi^{lb}(\hat{\theta})), \sup CI^{equal-tailed}(\phi^{ub}(\hat{\theta}))]$ has the coverage rate above $(1 - \tau)$, where $(1 - \tau)$ is the target coverage rate of the confidence interval for population

Figure D.4: 500 times simulations, Sample Size $N_m = N_a = 1000$, Tuning Parameter $N_m^{-0.49}$



Note : This figure shows the simulation of Estimator 2 and its confidence interval estimates. “true” shows a parameter used for the data generating process. “Pop Hwang LB” and “Pop Hwang UB” show the population lower and upper bounds. “Hwang LB” and “Hwang UB” stand for Estimator 2’s lower and upper bound estimates. “Hwang LB CIL” and “Hwang LB CIU” show the (lower and upper) endpoints of the 95% confidence interval of the lower bound. “Hwang UB CIL” and “Hwang UB CIU” are the (lower and upper) endpoints of the 95% confidence interval of the upper bound. The tuning parameter to compute the confidence intervals is $N_m^{-0.49}$.

lower and upper bounds. This implies that even though the coverage rates for the population lower and upper bounds may be lower than the target coverage rate, the confidence interval of the form (3) could provide a much higher coverage rate of the parameter of interest.

Table D.5: Coverage Rate, Tuning Parameter $N_m^{-0.49}$

CI type	Parameter	Coverage Rate
[Hwang LB CIL, Hwang LB CIU]	Pop Hwang LB, $\inf \beta_{10}$	0.92
	Pop Hwang LB, $\inf \alpha_0$	0.94
	Pop Hwang LB, $\inf \beta_{20}$	0.94
[Hwang UB CIL, Hwang UB CIU]	Pop Hwang UB, $\sup \beta_{10}$	0.932
	Pop Hwang UB, $\sup \alpha_0$	0.932
	Pop Hwang UB, $\sup \beta_{20}$	0.924
[Hwang LB CIL, Hwang UB CIU]	true, β_{10}	0.952
	true, α_0	1
	true, β_{20}	0.992

Note : This figure shows the coverage rates for three different types of confidence interval estimates: (1) confidence interval for the population lower bound – how often [Hwang LB CIL, Hwang LB CIU] contains “Pop Hwang LB”, (2) confidence interval for the population upper bound – how often [Hwang UB CIL, Hwang UB CIU] contains “Pop Hwang UB” and (3) confidence interval for the parameter of interest, how often [Hwang LB CIL, Hwang UB CIU] contains “true”. The tuning parameter to compute the confidence intervals is $N_m^{-0.49}$.

E Additional Tables and Figures

Table E.1: Cultural Attitude Proxies in FNSEM used for estimation

Questionnaire	Response Scale
<i>Religion</i>	
• How important is your religion to the way you live your life ? Is it...	not at all important/not very important/fairly important/or very important?
<i>English Use</i>	
• INTERVIEWER CODE : whether interviewed in english	Interview wholly in english/Interview partly in English, partly in another language/Interview wholly in another language
• INTERVIEWER CODE : Assess respondent's english ability. If interview not conducted in english, attempt a conversation to assess ability	Speaks english fluently/Fairly well/Slightly/Not at all
<i>Clothing</i>	
• Do you ever wear Asian clothes such as sari, shalwar, kamiz, kurta or pyjama?	Yes/No
• (If Yes) Do you wear Asian clothes all the time or only sometimes?	All the times/sometimes
<i>Interracial Marriage</i>	
• Would you personally mind if a close relative were to marry a white person?	Yes/No
• Would you mind very much or just a little?	Very much/Little

Note : Religiosity variables were recoded to variables on non-Western religiosity using the religion variable.
Non-western religions include Islam, Hinduism, and Sikhism.

Table E.2: Kleibergen and Paap (2006) Matrix Rank Test, $H_0 : \text{rank}(M_{kk'}) = 2$

Proxies	rk-Statistic	Degrees of Freedom	P-Value
<i>Religious Importance and English Fluency</i>	21207.38	4	0.00
<i>Whether Interviewed in English and Wearing Ethnic Cloth</i>	14955.91	1	0.00

Table E.3: Second-Generation South Asian Immigrants Characteristics in Longitudinal Study data (Year 2011)

Variable	Obs	Mean	Std. Dev.
(Male) Year of birth	1676	1982.76	5.898
(Female) Year of birth	1467	1983.882	5.295
(Male) Married	1676	0.326	0.469
(Female) Married	1467	0.311	0.463
(Male) Hindu	1658	0.268	0.443
(Male) Muslim	1658	0.507	0.5
(Male) Sikh	1658	0.2	0.4
(Female) Hindu	1448	0.256	0.437
(Female) Muslim	1448	0.53	0.499
(Female) Sikh	1448	0.179	0.383
(Male) Degree or higher	1666	0.459	0.498
(Male) Others(voca/prof)	1666	0.019	0.137
(Male) Below/eqv A-level	1666	0.522	0.5
(Female) Degree or higher	1464	0.527	0.499
(Female) Others(voca/prof)	1464	0.01	0.101
(Female) Below/eqv A-level	1464	0.463	0.499
(Male) Employed	1647	0.713	0.452
(Female) Employed	1450	0.618	0.486
Living in Ethnic Enclave	3143	0.56	0.496

Data Source: ONS LS

Note : This table shows summary statistics of second-generation South Asian immigrant sample in Longitudinal Study in year 2011.

Table E.4: Comparison of Indian, Pakistani, and Bangladeshi First-Generation Immigrants' Characteristics in (weighted) FNSEM and LS data

Variable	Data	Obs	Mean	Std. Dev.	Male			Female			
					Difference	P value	Obs	Mean	Std. Dev.	Difference	P value
Year of Birth	FNSEM	1119	1950.673	11.705	-1.60	0.00	975	1955.931	10.453	-0.51	0.10
	LS	3053	1952.27	8.407			3047	1956.437	7.412		
Year came to the UK	FNSEM	1119	1971.65	8.921	-0.31	0.34	975	1976.283	8.331	0.09	0.77
	LS	2083	1971.958	8.73			2146	1976.191	8.006		
Indian	FNSEM	1119	0.554	0.497	0.00	1.00	975	0.536	0.499	-0.02	0.30
	LS	3143	0.554	0.497			3143	0.555	0.497		
Pakistani	FNSEM	1119	0.353	0.478	0.00	0.95	975	0.367	0.482	0.01	0.46
	LS	3143	0.354	0.478			3143	0.354	0.478		
Bangladeshi	FNSEM	1119	0.093	0.291	0.00	0.84	975	0.097	0.296	0.01	0.57
	LS	3143	0.091	0.288			3143	0.091	0.288		
Hindu	FNSEM	1119	0.255	0.436	-0.01	0.65	975	0.25	0.433	-0.01	0.42
	LS	3055	0.262	0.44			3054	0.263	0.44		
Muslim	FNSEM	1119	0.512	0.5	-0.01	0.73	975	0.528	0.499	0.01	0.70
	LS	3055	0.518	0.5			3054	0.521	0.5		
Sikh	FNSEM	1119	0.186	0.389	0.00	0.88	975	0.187	0.39	0.00	0.94
	LS	3055	0.188	0.391			3054	0.188	0.391		
Degree or higher	FNSEM	1119	0.22	0.414	0.01	0.73	975	0.135	0.342	0.00	0.81
	LS	3143	0.215	0.411			3143	0.138	0.345		
Other (voca/prof)	FNSEM	1119	0.17	0.376	0.00	0.94	975	0.135	0.342	-0.01	0.64
	LS	3143	0.169	0.375			3143	0.141	0.348		
Below/eqv A-level	FNSEM	1119	0.61	0.488	-0.01	0.77	975	0.73	0.444	0.01	0.58
	LS	3143	0.615	0.487	-0.14	0.00	3143	0.721	0.448		
Employed	FNSEM	1119	0.648	0.478	-0.14	0.00	975	0.287	0.453	0.00	0.82
	LS	2298	0.788	0.409			2315	0.283	0.45		
Total											
Living in Ethnic Enclave		FNSEM	2094	0.512	0.5	0.00	1.00				
		LS	2391	0.512	0.5						

Data Source: FNSEM and ONS LS

Note: This table compares the sample means of key variables in two datasets, the Fourth National Survey of Ethnic Minorities (FNSEM) and the Longitudinal Study (LS). The FNSEM sample is weighted using a post-stratification weight to balance sex, ethnicity, education, female employment, and enclave residency.

Table E.5: Comparison of Indian/Pakistani/Bangladeshi First Generations' Characteristics in (unweighted) FNSEM and LS data

Variable	Data	Male					Female				
		Obs	Mean	Std. Dev.	Difference	P value	Obs	Mean	Std. Dev.	Difference	P value
Year of Birth	FNSEM	1119	1950.38	11.93	-1.89	0.00	975	1956.22	10.30	-0.22	0.47
	LS	3053	1952.27	8.407			3047	1956.437	7.412		
Year came to the UK	FNSEM	1119	1971.43	9.07	-0.53	0.11	975	1977.07	8.15	0.88	0.00
	LS	2083	1971.958	8.73			2146	1976.191	8.006		
Indian	FNSEM	1119	0.49	0.50	-0.06	0.00	975	0.48	0.50	-0.08	0.00
	LS	3143	0.554	0.497			3143	0.555	0.497		
Pakistani	FNSEM	1119	0.33	0.47	-0.02	0.15	975	0.33	0.47	-0.02	0.17
	LS	3143	0.354	0.478			3143	0.354	0.478		
Bangladeshi	FNSEM	1119	0.18	0.39	0.09	0.00	975	0.19	0.40	0.10	0.00
	LS	3143	0.091	0.288			3143	0.091	0.288		
Hindu	FNSEM	1119	0.23	0.42	-0.03	0.04	975	0.22	0.42	-0.04	0.01
	LS	3055	0.262	0.44			3054	0.263	0.44		
Muslim	FNSEM	1119	0.57	0.50	0.05	0.00	975	0.58	0.49	0.06	0.00
	LS	3055	0.518	0.5			3054	0.521	0.5		
Sikh	FNSEM	1119	0.17	0.37	-0.02	0.18	975	0.17	0.37	-0.02	0.21
	LS	3055	0.188	0.391			3054	0.188	0.391		
Degree or higher	FNSEM	1119	0.17	0.38	-0.05	0.00	975	0.11	0.31	-0.03	0.02
	LS	3143	0.215	0.411			3143	0.138	0.345		
Other (voca/prof)	FNSEM	1119	0.10	0.31	-0.07	0.00	975	0.07	0.26	-0.07	0.00
	LS	3143	0.169	0.375			3143	0.141	0.348		
Below/eqv A-level	FNSEM	1119	0.72	0.45	0.11	0.00	975	0.82	0.38	0.10	0.00
	LS	3143	0.615	0.487			3143	0.721	0.448		
Employed	FNSEM	1119	0.60	0.49	-0.19	0.00	975	0.26	0.44	-0.02	0.18
	LS	2298	0.788	0.409			2315	0.283	0.45		
Total											
Living in Ethnic Enclave		FNSEM	2094	0.55	0.50	0.04	0.01				
		LS	2391	0.512	0.5						

Data Source: FNSEM and ONS LS

Note: This table compares the sample means of key variables in two datasets, the Fourth National Survey of Ethnic Minorities (FNSEM) and the Longitudinal Study (LS).

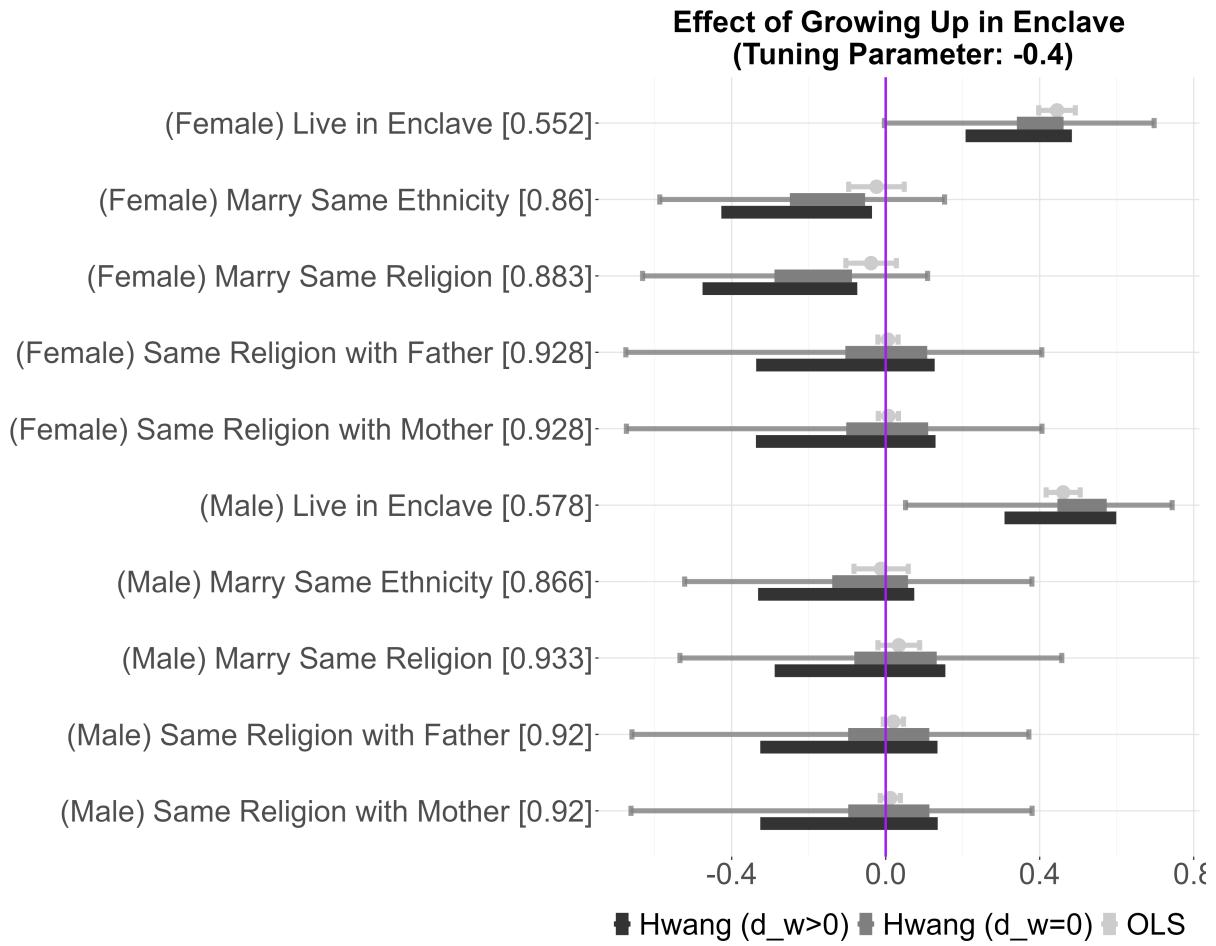
Table E.6: Coverage Rates for “Hwang ($d_w = 0$)” Estimator Under Different Tuning Parameters

		Tuning Parameters		
		Hwang ($d_w = 0$)		
		$N_m^{-0.49}$	$N_m^{-0.4}$	$N_m^{-0.3}$
Female Employment	LB	0.74	0.9	1
Female Employment	UB	0.79	0.93	1
Female College Graduation	LB	0.81	0.88	0.99
Female College Graduation	UB	0.82	0.96	1
Marry Same Ethnicity	LB	0.8	0.88	1
Marry Same Ethnicity	UB	0.79	0.94	1
Living in Enclave	LB	0.72	0.9	0.99
Living in Enclave	UB	0.87	0.93	0.99
Marry Same Religion	LB	0.77	0.89	1
Marry Same Religion	UB	0.8	0.95	1
Same Religion with Father	LB	0.7	0.88	0.96
Same Religion with Father	UB	0.82	0.94	0.99
Same Religion with Mother	LB	0.7	0.87	0.96
Same Religion with Mother	UB	0.82	0.94	1
Average		0.78	0.91	0.99

Data Source: FNSEM and ONS LS

Note: Numbers in each cell show the estimator’s (“Hwang ($d_w = 0$)”) coverage probabilities of the coefficient of neighborhood effect, β_n . The coverage probability is computed from the double bootstrap procedure described in Online Appendix Section B. For tuning parameters for the baseline results, I use $N_m^{-0.4}$ that gives the coverage rates for population bounds 91% on average.

Figure E.1: Evidence on Heterogeneous Effects by Gender



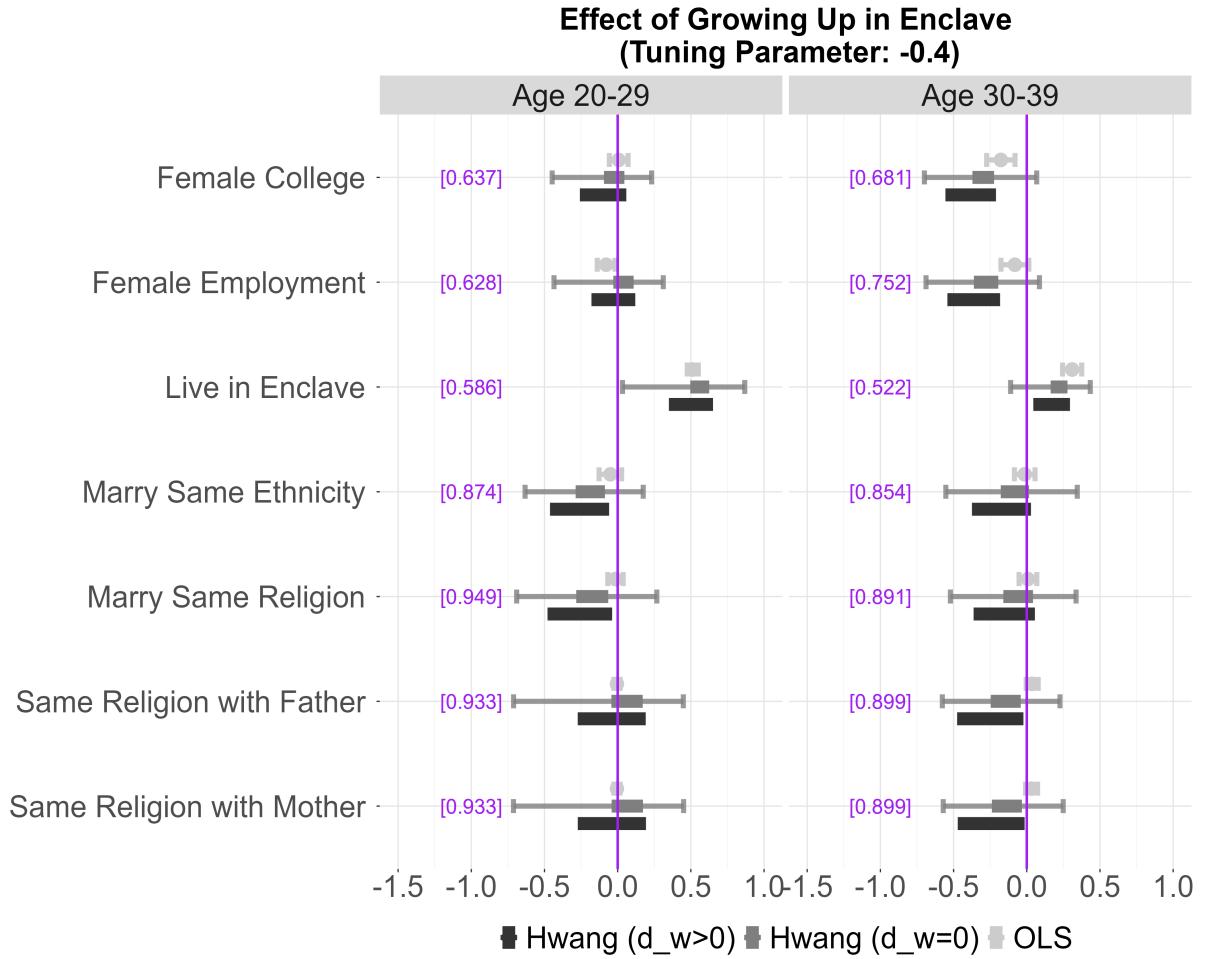
Data Source : FNSEM and ONS LS

Note: This figure shows the estimates in Figure 4 by gender.

The graph presents the OLS point estimate from a short regression without x and two “Hwang” interval estimates of the neighborhood effect (β_n) under the assumptions of $d_w = 0$ and $d_w > 0$. The point estimates are displayed as dots and the interval estimates are shown as segment bands. The 90% confidence intervals for the “OLS” and “Hwang ($d_w = 0$)” estimates are drawn as whiskers. The 90% confidence intervals for “Hwang ($d_w = 0$)” estimates connect the lower end of the 90% CI for the lower bound estimate and the upper end of the 90% CI for the upper bound estimate and they are computed by Hong and Li (2018) numerical delta method described in Online Appendix section B. The tuning parameter for the numerical delta method is $N_m^{-0.4}$, where N_m is the sample size of the LS data.

Dependent variables are labeled on the left side of the figure. Numbers inside the brackets next to each dependent variable label are the averages of the dependent variables. The (z', w') controls include mothers' religion, ethnicity and both parents' cohort, education, immigration year group dummies, and employment status when the child was between age 0 and 10. For “Hwang ($d_w > 0$)” estimates, I assume the auxiliary data does not include two maternal religion dummy variables (whether a mother is a Muslim/Sikh).

Figure E.2: Evidence on Heterogeneous Effects by Cohorts



Data Source : FNSEM and ONS LS

Note: This figure shows the estimates in Figure 4 by cohort; recall that the two cohorts are (i) those between 20 and 29 years of age in 2011 and (ii) those between 30 and 39 years of age in 2011.

The graph presents the OLS point estimate from a short regression without x and two “Hwang” interval estimates of the neighborhood effect (β_n) under the assumptions of $d_w = 0$ and $d_w > 0$. The point estimates are displayed as dots and the interval estimates are shown as segment bands. The 90% confidence intervals for the “OLS” and “Hwang ($d_w = 0$)” estimates are drawn as whiskers. The 90% confidence intervals for “Hwang ($d_w > 0$)” estimates connect the lower end of the 90% CI for the lower bound estimate and the upper end of the 90% CI for the upper bound estimate and they are computed by Hong and Li (2018) numerical delta method described in Online Appendix section B. The tuning parameter for the numerical delta method is $N_m^{-0.4}$, where N_m is the sample size of the LS data.

Dependent variables are labeled on the left side of the figure. Numbers inside the brackets next to each dependent variable label are the averages of the dependent variables. The (z', w') controls include mothers’ religion, ethnicity and both parents’ cohort, education, immigration year group dummies, and employment status when the child was between age 0 and 10. For “Hwang ($d_w > 0$)” estimates, I assume the auxiliary data does not include two maternal religion dummy variables (whether a mother is a Muslim/Sikh).

Table E.7: Numbers behind Figure 4

Dependent Variable		OLS	Hwang ($d_w = 0$)	Hwang ($d_w > 0$)	N
Female Employment [0.66]	Estimate	-0.07	[-0.1,0.04]	[-0.26,0.05]	984
	90% CI	[-0.12,-0.02]	[-0.45,0.22]		
Female College [0.648]	Estimate	-0.03	[-0.17,-0.03]	[-0.34,-0.01]	827
	90% CI	[-0.09,0.02]	[-0.47,0.14]		
Marry Same Ethnicity [0.863]	Estimate	-0.02	[-0.19,0]	[-0.38,0.02]	540
	90% CI	[-0.07,0.03]	[-0.49,0.21]		
Live in Enclave [0.565]	Estimate	0.46	[0.39,0.52]	[0.26,0.54]	2087
	90% CI	[0.42,0.49]	[0.05,0.71]		
Marry Same Religion [0.91]	Estimate	0.01	[-0.18,0.02]	[-0.38,0.04]	522
	90% CI	[-0.03,0.05]	[-0.44,0.21]		
Same Religion with Father [0.924]	Estimate	0.01	[-0.1,0.11]	[-0.33,0.13]	1990
	90% CI	[-0.01,0.03]	[-0.66,0.36]		
Same Religion with Mother [0.924]	Estimate	0.01	[-0.1,0.11]	[-0.33,0.13]	1990
	90% CI	[-0.01,0.03]	[-0.66,0.36]		

Data Source: FNSEM and ONS LS

Note: Numbers inside the brackets next to each dependent variable label are the averages of the dependent variables.

The controls (z, w) include mothers' religion, ethnicity and both parents' cohort, education, immigration year group dummies, and employment status when the child was between age 0 and 10. For "Hwang ($d_w > 0$)" estimates, I assume the auxiliary data does not include two maternal religion dummy variables (whether a mother is a Muslim/Sikh).

The 90% confidence intervals for "Hwang ($d_w = 0$)" estimates are computed by Hong and Li (2018) numerical delta method described in Online Appendix section B. The tuning parameter for the numerical delta method is $N_m^{-0.4}$, where N_m is the sample size of the LS data.

Table E.8: Numbers behind Figure E.1

Dependent Variable		OLS	Hwang ($d_w = 0$)	Hwang ($d_w > 0$)	N
(Female) Marry Same Ethnicity [0.86]	Estimate	-0.02	[-0.25,-0.05]	[-0.43,-0.04]	250
	90% CI	[-0.1,0.05]	[-0.59,0.15]		
(Male) Marry Same Ethnicity [0.866]	Estimate	-0.01	[-0.14,0.06]	[-0.33,0.07]	290
	90% CI	[-0.08,0.06]	[-0.52,0.38]		
(Female) Live in Enclave [0.552]	Estimate	0.44	[0.34,0.46]	[0.21,0.48]	997
	90% CI	[0.4,0.49]	[0,0.7]		
(Male) Live in Enclave [0.578]	Estimate	0.46	[0.45,0.57]	[0.31,0.6]	1090
	90% CI	[0.42,0.51]	[0.05,0.74]		
(Female) Marry Same Religion [0.883]	Estimate	-0.04	[-0.29,-0.09]	[-0.48,-0.07]	240
	90% CI	[-0.1,0.03]	[-0.63,0.11]		
(Male) Marry Same Religion [0.933]	Estimate	0.03	[-0.08,0.13]	[-0.29,0.16]	282
	90% CI	[-0.02,0.09]	[-0.53,0.46]		
(Female) Same Religion with Father [0.928]	Estimate	0.01	[-0.1,0.11]	[-0.34,0.13]	949
	90% CI	[-0.02,0.03]	[-0.67,0.41]		
(Male) Same Religion with Father [0.92]	Estimate	0.02	[-0.1,0.11]	[-0.33,0.13]	1041
	90% CI	[-0.01,0.05]	[-0.66,0.37]		
(Female) Same Religion with Mother [0.928]	Estimate	0.01	[-0.1,0.11]	[-0.34,0.13]	949
	90% CI	[-0.02,0.03]	[-0.67,0.41]		
(Male) Same Religion with Mother [0.92]	Estimate	0.01	[-0.1,0.11]	[-0.33,0.14]	1041
	90% CI	[-0.01,0.04]	[-0.66,0.38]		

Data Source: FNSEM and ONS LS

Note: Numbers inside the brackets next to each dependent variable label are the averages of the dependent variables.

The controls (z, w) include mothers' religion, ethnicity and both parents' cohort, education, immigration year group dummies, and employment status when the child was between age 0 and 10. For "Hwang ($d_w > 0$)" estimates, I assume the auxiliary data does not include two maternal religion dummy variables (whether a mother is a Muslim/Sikh).

The 90% confidence intervals for "Hwang ($d_w = 0$)" estimates are computed by Hong and Li (2018) numerical delta method described in Online Appendix section B. The tuning parameter for the numerical delta method is $N_m^{-0.4}$, where N_m is the sample size of the LS data.

Table E.9: Numbers behind Figure E.2

Dependent Variable		OLS	Hwang ($d_w = 0$)	Hwang ($d_w > 0$)	N
(20-29) Female Employment [0.628]	Estimate	-0.08	[-0.03,0.11]	[-0.18,0.12]	720
	90% CI	[-0.14,-0.02]	[-0.44,0.31]		
(20-29) Female College [0.637]	Estimate	0.01	[-0.09,0.05]	[-0.26,0.06]	562
	90% CI	[-0.06,0.07]	[-0.45,0.23]		
(20-29) Marry Same Ethnicity [0.874]	Estimate	-0.05	[-0.29,-0.09]	[-0.46,-0.06]	190
	90% CI	[-0.13,0.03]	[-0.63,0.17]		
(20-29) Live in Enclave [0.586]	Estimate	0.51	[0.5,0.63]	[0.35,0.65]	1432
	90% CI	[0.48,0.55]	[0.03,0.87]		
(20-29) Marry Same Religion [0.949]	Estimate	-0.01	[-0.28,-0.06]	[-0.48,-0.04]	178
	90% CI	[-0.07,0.04]	[-0.69,0.27]		
(20-29) Same Religion with Father [0.933]	Estimate	0	[-0.04,0.17]	[-0.27,0.19]	1349
	90% CI	[-0.03,0.02]	[-0.71,0.45]		
(20-29) Same Religion with Mother [0.933]	Estimate	0	[-0.04,0.17]	[-0.27,0.19]	1349
	90% CI	[-0.03,0.02]	[-0.71,0.45]		
(30-39) Female Employment [0.752]	Estimate	-0.08	[-0.36,-0.19]	[-0.54,-0.18]	250
	90% CI	[-0.18,0.01]	[-0.69,0.09]		
(30-39) Female College [0.681]	Estimate	-0.18	[-0.37,-0.22]	[-0.56,-0.21]	251
	90% CI	[-0.27,-0.08]	[-0.7,0.07]		
(30-39) Marry Same Ethnicity [0.854]	Estimate	-0.01	[-0.18,0.02]	[-0.37,0.03]	328
	90% CI	[-0.08,0.06]	[-0.55,0.35]		
(30-39) Live in Enclave [0.522]	Estimate	0.31	[0.16,0.28]	[0.04,0.3]	605
	90% CI	[0.24,0.38]	[-0.11,0.43]		
(30-39) Marry Same Religion [0.891]	Estimate	0.01	[-0.16,0.04]	[-0.36,0.06]	322
	90% CI	[-0.05,0.07]	[-0.52,0.34]		
(30-39) Same Religion with Father [0.899]	Estimate	0.04	[-0.25,-0.04]	[-0.48,-0.02]	594
	90% CI	[0,0.08]	[-0.58,0.23]		
(30-39) Same Religion with Mother [0.899]	Estimate	0.03	[-0.24,-0.03]	[-0.47,-0.01]	594
	90% CI	[-0.01,0.07]	[-0.57,0.25]		

Data Source: FNSEM and ONS LS

Note: Numbers inside the brackets next to each dependent variable label are the averages of the dependent variables.

The controls (z, w) include mothers' religion, ethnicity and both parents' cohort, education, immigration year group dummies, and employment status when the child was between age 0 and 10. For "Hwang ($d_w > 0$)" estimates, I assume the auxiliary data does not include two maternal religion dummy variables (whether a mother is a Muslim/Sikh).

The 90% confidence intervals for "Hwang ($d_w = 0$)" estimates are computed by Hong and Li (2018) numerical delta method described in Online Appendix section B.⁸¹ The tuning parameter for the numerical delta method is $N_m^{-0.4}$, where N_m is the sample size of the LS data.

Table E.10: Co-residence rate with parents by marital status

	Parent Not at Home	Parent at Home	Total
Not Married	11%	89%	100%
Married	60%	40%	100%
Total	27%	73%	100%

Source: ONS LS

Note: This table shows the co-residence rate with parents for second-generation South Asian immigrants.

Table E.11: Co-residence rate with parents by age group

	Parent Not at Home	Parent at Home	Total
20-29 y.o.	18%	82%	100%
30-39 y.o.	45%	55%	100%
40 over	41%	59%	100%
Total	27%	73%	100%

Source: ONS LS

Note: This table shows the co-residence rate with parents for second-generation South Asian immigrants.

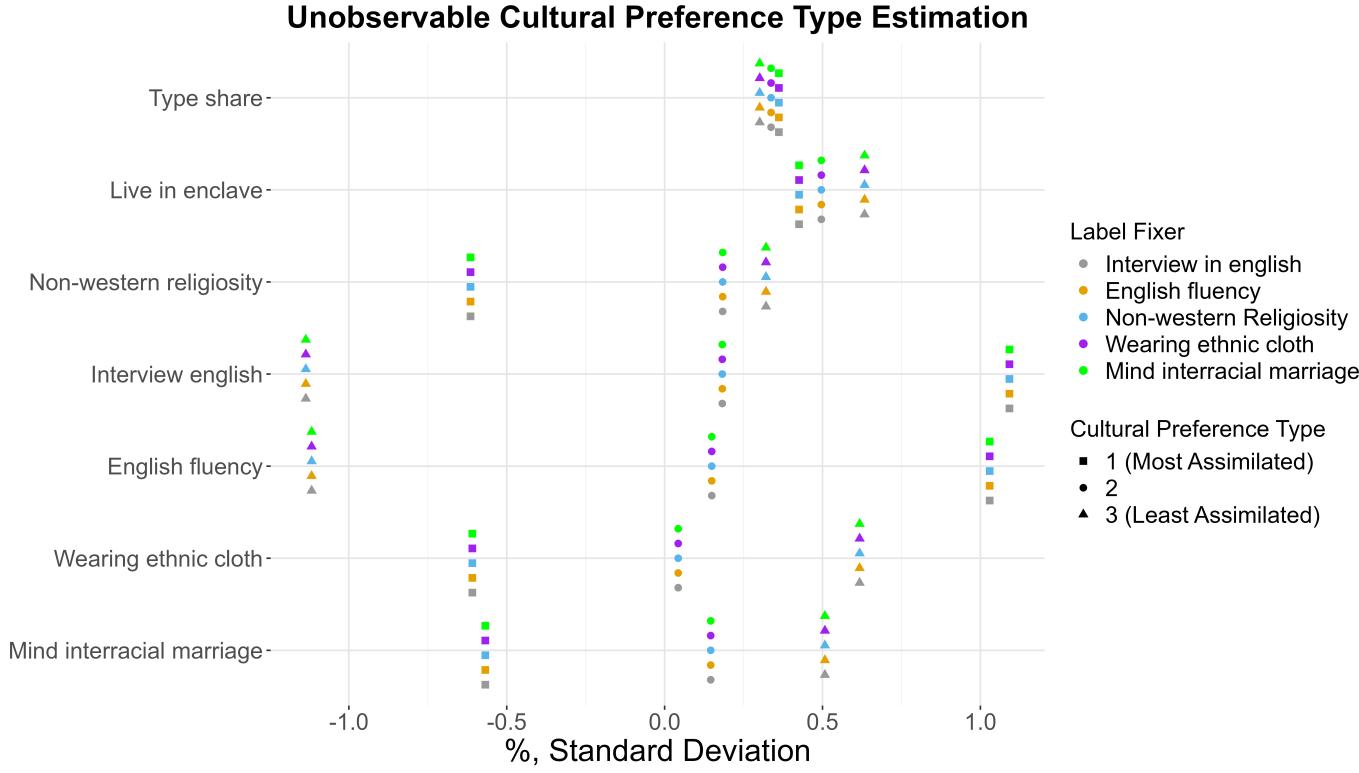
Table E.12: Persistence of Enclave Residency

		Living in Enclave Age 10-19		
		No	Yes	Total
Living in Enclave Age 0-9	No	76%	24%	100%
	Yes	14%	86%	100%

Data Source: ONS LS

Note: This table shows the persistence in ethnic enclave residency in LS data, 1971-2011. The cell numbers are the conditional probability of ethnic enclave residency between age 10-19 conditional on ethnic enclave residency between age 0-9.

Figure E.3: Robustness of Type Estimation to Choice of a Label Fixer



Data Source: FNSEM

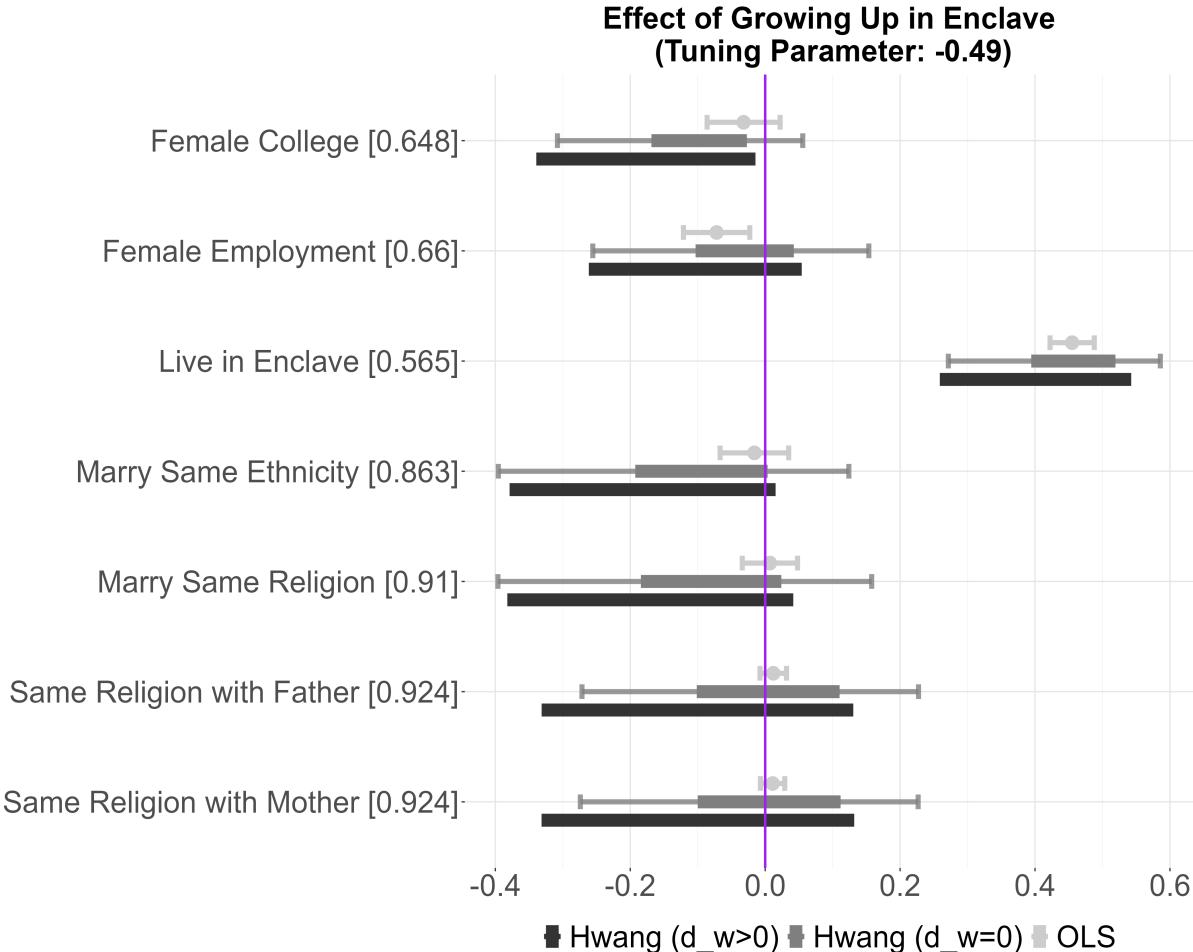
Note: This figure shows estimates when a different variable is used for fixing a label. The figure assures that the different choices of a label fixer do not affect the estimation results.

F Robustness Checks with Different Tuning Parameters

This section presents the results corresponding to Section 6 under alternative tuning parameters. Overall, the qualitative conclusions remain unchanged. I find little statistical evidence of neighborhood effects—except for ethnic enclave residency—in the full sample.

F.1 Tuning Parameter: -0.49

Figure F.1: Robustness Check: Neighborhood Effects on Intergenerational Cultural Transmission

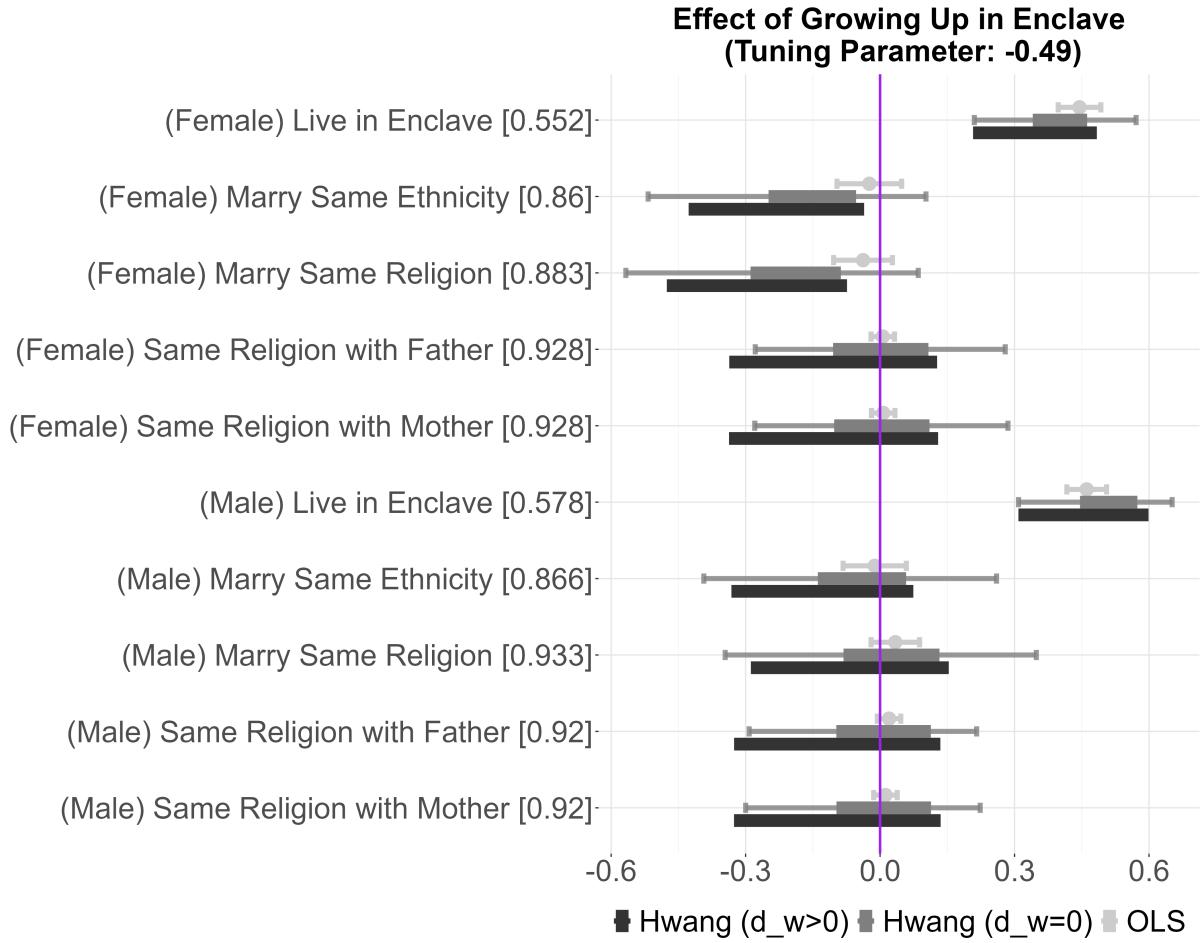


Data Source: FNSEM and ONS LS

Note: The graph presents the OLS point estimate from a short regression without x and two “Hwang” interval estimates of the neighborhood effect (β_n) under the assumptions of $d_w = 0$ and $d_w > 0$. The point estimates are displayed as dots and the interval estimates are shown as segment bands. The 90% confidence intervals for the “OLS” and “Hwang ($d_w = 0$)” estimates are drawn as whiskers. The 90% confidence intervals for “Hwang ($d_w = 0$)” estimates connect the lower end of the 90% CI for the lower bound estimate and the upper end of the 90% CI for the upper bound estimate and they are computed by Hong and Li (2018) numerical delta method described in Online Appendix section B. The tuning parameter for the numerical delta method is $N_m^{-0.49}$, where N_m is the sample size of the LS data.

Dependent variables are labeled on the left side of the figure. Numbers inside the brackets next to each dependent variable label are the averages of the dependent variables. The (z', w') controls include mothers’ religion, ethnicity and both parents’ cohort, education, immigration year group dummies, and employment status when the child was between age 0 and 10. For “Hwang ($d_w > 0$)” estimates, I assume the auxiliary data does not include two maternal religion dummy variables (whether a mother is a Muslim/Sikh).

Figure F.2: Evidence on Heterogeneous Effects by Gender



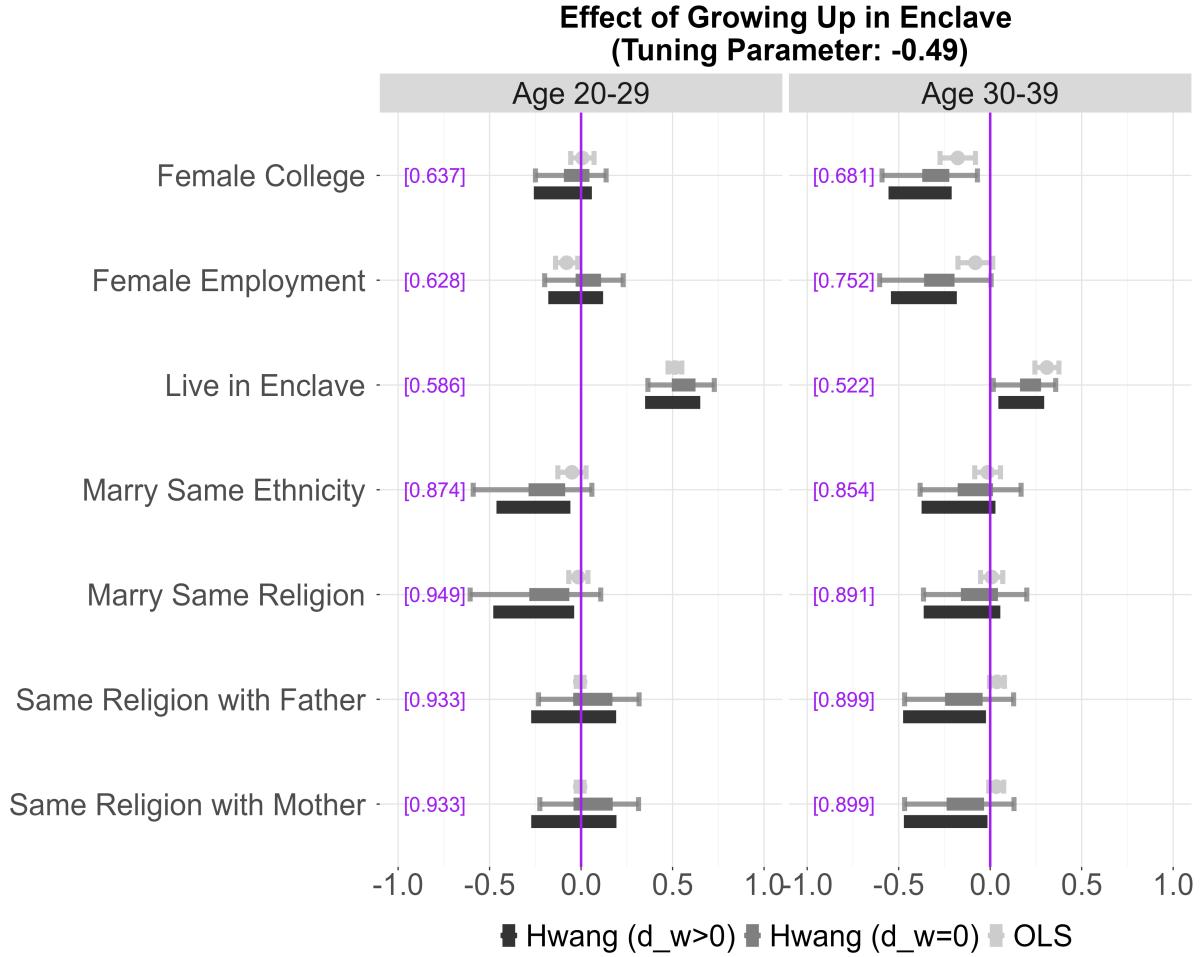
Data Source : FNSEM and ONS LS

Note: This figure shows the estimates in Figure 4 by gender.

The graph presents the OLS point estimate from a short regression without x and two “Hwang” interval estimates of the neighborhood effect (β_n) under the assumptions of $d_w = 0$ and $d_w > 0$. The point estimates are displayed as dots and the interval estimates are shown as segment bands. The 90% confidence intervals for the “OLS” and “Hwang ($d_w = 0$)” estimates are drawn as whiskers. The 90% confidence intervals for “Hwang ($d_w = 0$)” estimates connect the lower end of the 90% CI for the lower bound estimate and the upper end of the 90% CI for the upper bound estimate and they are computed by [Hong and Li \(2018\)](#) numerical delta method described in Online Appendix section B. The tuning parameter for the numerical delta method is $N_m^{-0.49}$, where N_m is the sample size of the LS data.

Dependent variables are labeled on the left side of the figure. Numbers inside the brackets next to each dependent variable label are the averages of the dependent variables. The (z', w') controls include mothers' religion, ethnicity and both parents' cohort, education, immigration year group dummies, and employment status when the child was between age 0 and 10. For “Hwang ($d_w > 0$)” estimates, I assume the auxiliary data does not include two maternal religion dummy variables (whether a mother is a Muslim/Sikh).

Figure F.3: Evidence on Heterogeneous Effects by Cohorts



Data Source : FNSEM and ONS LS

Note: This figure shows the estimates in Figure 4 by cohort; recall that the two cohorts are (i) those between 20 and 29 years of age in 2011 and (ii) those between 30 and 39 years of age in 2011.

The graph presents the OLS point estimate from a short regression without x and two “Hwang” interval estimates of the neighborhood effect (β_n) under the assumptions of $d_w = 0$ and $d_w > 0$. The point estimates are displayed as dots and the interval estimates are shown as segment bands. The 90% confidence intervals for the “OLS” and “Hwang ($d_w = 0$)” estimates are drawn as whiskers. The 90% confidence intervals for “Hwang ($d_w > 0$)” estimates connect the lower end of the 90% CI for the lower bound estimate and the upper end of the 90% CI for the upper bound estimate and they are computed by Hong and Li (2018) numerical delta method described in Online Appendix section B. The tuning parameter for the numerical delta method is $N_m^{-0.49}$, where N_m is the sample size of the LS data.

Dependent variables are labeled on the left side of the figure. Numbers inside the brackets next to each dependent variable label are the averages of the dependent variables. The (z', w') controls include mothers’ religion, ethnicity and both parents’ cohort, education, immigration year group dummies, and employment status when the child was between age 0 and 10. For “Hwang ($d_w > 0$)” estimates, I assume the auxiliary data does not include two maternal religion dummy variables (whether a mother is a Muslim/Sikh).

Table F.1: Numbers behind Figure F.1

Dependent Variable		OLS	Hwang ($d_w = 0$)	Hwang ($d_w > 0$)	N
Female Employment [0.66]	Estimate	-0.07	[-0.1,0.04]	[-0.26,0.05]	984
	90% CI	[-0.12,-0.02]	[-0.26,0.15]		
Female College [0.648]	Estimate	-0.03	[-0.17,-0.03]	[-0.34,-0.01]	827
	90% CI	[-0.09,0.02]	[-0.31,0.06]		
Marry Same Ethnicity [0.863]	Estimate	-0.02	[-0.19,0]	[-0.38,0.02]	540
	90% CI	[-0.07,0.03]	[-0.4,0.12]		
Live in Enclave [0.565]	Estimate	0.46	[0.39,0.52]	[0.26,0.54]	2087
	90% CI	[0.42,0.49]	[0.27,0.59]		
Marry Same Religion [0.91]	Estimate	0.01	[-0.18,0.02]	[-0.38,0.04]	522
	90% CI	[-0.03,0.05]	[-0.4,0.16]		
Same Religion with Father [0.924]	Estimate	0.01	[-0.1,0.11]	[-0.33,0.13]	1990
	90% CI	[-0.01,0.03]	[-0.27,0.23]		
Same Religion with Mother [0.924]	Estimate	0.01	[-0.1,0.11]	[-0.33,0.13]	1990
	90% CI	[-0.01,0.03]	[-0.27,0.23]		

Data Source: FNSEM and ONS LS

Note: Numbers inside the brackets next to each dependent variable label are the averages of the dependent variables.

The controls (z, w) include mothers' religion, ethnicity and both parents' cohort, education, immigration year group dummies, and employment status when the child was between age 0 and 10. For "Hwang ($d_w > 0$)" estimates, I assume the auxiliary data does not include two maternal religion dummy variables (whether a mother is a Muslim/Sikh).

The 90% confidence intervals for "Hwang ($d_w = 0$)" estimates are computed by Hong and Li (2018) numerical delta method described in Online Appendix section B. The tuning parameter for the numerical delta method is $N_m^{-0.49}$, where N_m is the sample size of the LS data.

Table F.2: Numbers behind Figure F.2

Dependent Variable		OLS	Hwang ($d_w = 0$)	Hwang ($d_w > 0$)	N
(Female) Marry Same Ethnicity [0.86]	Estimate	-0.02	[-0.25,-0.05]	[-0.43,-0.04]	250
	90% CI	[-0.1,0.05]	[-0.52,0.1]		
(Male) Marry Same Ethnicity [0.866]	Estimate	-0.01	[-0.14,0.06]	[-0.33,0.07]	290
	90% CI	[-0.08,0.06]	[-0.39,0.26]		
(Female) Live in Enclave [0.552]	Estimate	0.44	[0.34,0.46]	[0.21,0.48]	997
	90% CI	[0.4,0.49]	[0.21,0.57]		
(Male) Live in Enclave [0.578]	Estimate	0.46	[0.45,0.57]	[0.31,0.6]	1090
	90% CI	[0.42,0.51]	[0.31,0.65]		
(Female) Marry Same Religion [0.883]	Estimate	-0.04	[-0.29,-0.09]	[-0.48,-0.07]	240
	90% CI	[-0.1,0.03]	[-0.57,0.09]		
(Male) Marry Same Religion [0.933]	Estimate	0.03	[-0.08,0.13]	[-0.29,0.15]	282
	90% CI	[-0.02,0.09]	[-0.35,0.35]		
(Female) Same Religion with Father [0.928]	Estimate	0.01	[-0.1,0.11]	[-0.34,0.13]	949
	90% CI	[-0.02,0.03]	[-0.28,0.28]		
(Male) Same Religion with Father [0.92]	Estimate	0.02	[-0.1,0.11]	[-0.33,0.13]	1041
	90% CI	[-0.01,0.05]	[-0.29,0.22]		
(Female) Same Religion with Mother [0.928]	Estimate	0.01	[-0.1,0.11]	[-0.34,0.13]	949
	90% CI	[-0.02,0.03]	[-0.28,0.29]		
(Male) Same Religion with Mother [0.92]	Estimate	0.01	[-0.1,0.11]	[-0.33,0.14]	1041
	90% CI	[-0.01,0.04]	[-0.3,0.22]		

Data Source: FNSEM and ONS LS

Note: Numbers inside the brackets next to each dependent variable label are the averages of the dependent variables.

The controls (z, w) include mothers' religion, ethnicity and both parents' cohort, education, immigration year group dummies, and employment status when the child was between age 0 and 10. For "Hwang ($d_w > 0$)" estimates, I assume the auxiliary data does not include two maternal religion dummy variables (whether a mother is a Muslim/Sikh).

The 90% confidence intervals for "Hwang ($d_w = 0$)" estimates are computed by Hong and Li (2018) numerical delta method described in Online Appendix section B. The tuning parameter for the numerical delta method is $N_m^{-0.49}$, where N_m is the sample size of the LS data.

Table F.3: Numbers behind Figure F.3

Dependent Variable		OLS	Hwang ($d_w = 0$)	Hwang ($d_w > 0$)	N
(20-29) Female Employment [0.628]	Estimate	-0.08	[-0.03,0.11]	[-0.18,0.12]	720
	90% CI	[-0.14,-0.02]	[-0.2,0.23]		
(20-29) Female College [0.637]	Estimate	0.01	[-0.09,0.05]	[-0.26,0.06]	562
	90% CI	[-0.06,0.07]	[-0.25,0.14]		
(20-29) Marry Same Ethnicity [0.874]	Estimate	-0.05	[-0.29,-0.09]	[-0.46,-0.06]	190
	90% CI	[-0.13,0.03]	[-0.59,0.06]		
(20-29) Live in Enclave [0.586]	Estimate	0.51	[0.5,0.63]	[0.35,0.65]	1432
	90% CI	[0.48,0.55]	[0.37,0.73]		
(20-29) Marry Same Religion [0.949]	Estimate	-0.01	[-0.28,-0.06]	[-0.48,-0.04]	178
	90% CI	[-0.07,0.04]	[-0.61,0.11]		
(20-29) Same Religion with Father [0.933]	Estimate	0	[-0.04,0.17]	[-0.27,0.19]	1349
	90% CI	[-0.03,0.02]	[-0.23,0.32]		
(20-29) Same Religion with Mother [0.933]	Estimate	0	[-0.04,0.17]	[-0.27,0.19]	1349
	90% CI	[-0.03,0.02]	[-0.23,0.31]		
(30-39) Female Employment [0.752]	Estimate	-0.08	[-0.36,-0.19]	[-0.54,-0.18]	250
	90% CI	[-0.18,0.01]	[-0.6,0.01]		
(30-39) Female College [0.681]	Estimate	-0.18	[-0.37,-0.22]	[-0.56,-0.21]	251
	90% CI	[-0.27,-0.08]	[-0.59,-0.07]		
(30-39) Marry Same Ethnicity [0.854]	Estimate	-0.01	[-0.18,0.02]	[-0.37,0.03]	328
	90% CI	[-0.08,0.06]	[-0.38,0.17]		
(30-39) Live in Enclave [0.522]	Estimate	0.31	[0.16,0.28]	[0.04,0.3]	605
	90% CI	[0.24,0.38]	[0.02,0.36]		
(30-39) Marry Same Religion [0.891]	Estimate	0.01	[-0.16,0.04]	[-0.36,0.06]	322
	90% CI	[-0.05,0.07]	[-0.36,0.2]		
(30-39) Same Religion with Father [0.899]	Estimate	0.04	[-0.25,-0.04]	[-0.48,-0.02]	594
	90% CI	[0,0.08]	[-0.47,0.13]		
(30-39) Same Religion with Mother [0.899]	Estimate	0.03	[-0.24,-0.03]	[-0.47,-0.01]	594
	90% CI	[-0.01,0.07]	[-0.47,0.13]		

Data Source: FNSEM and ONS LS

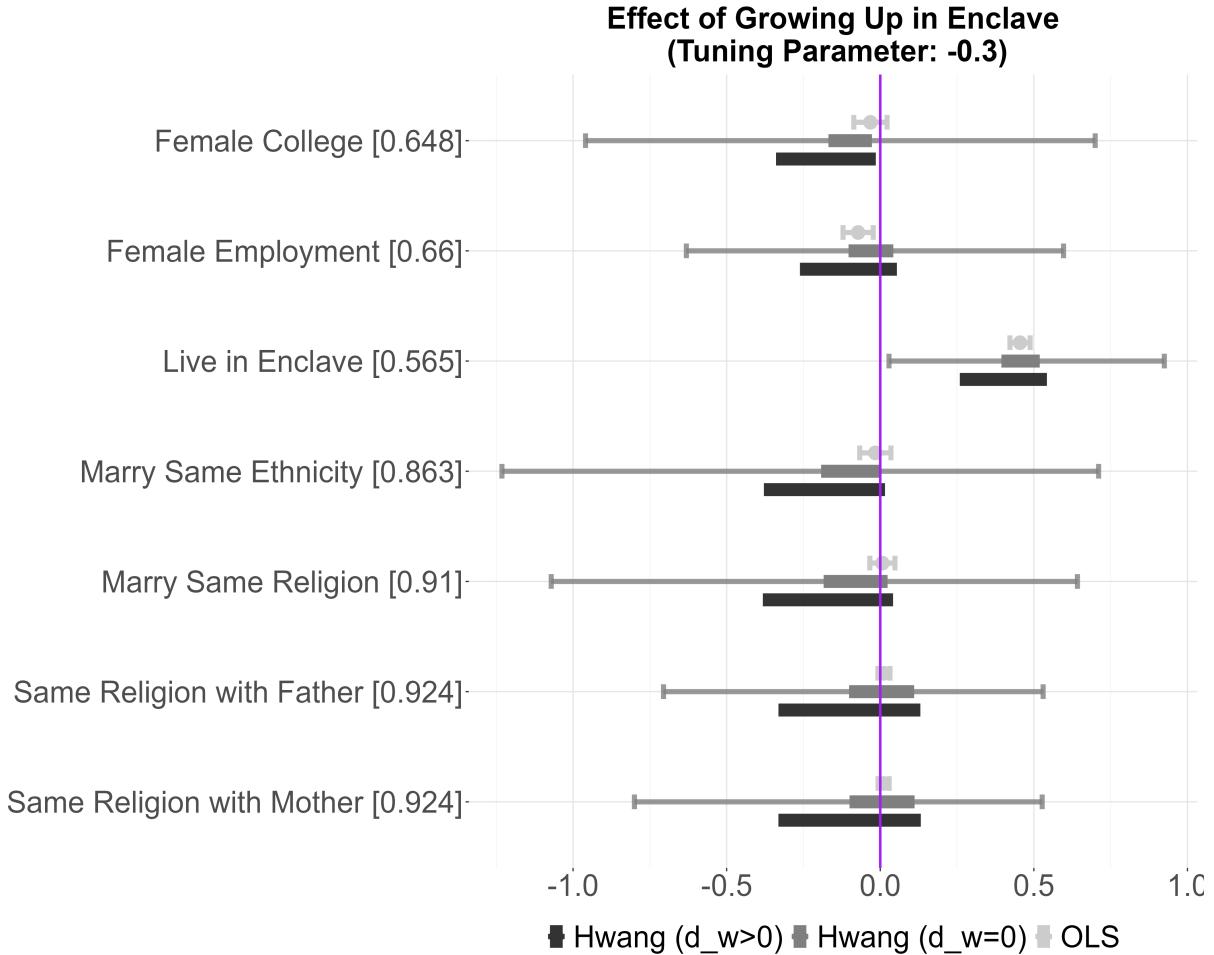
Note: Numbers inside the brackets next to each dependent variable label are the averages of the dependent variables.

The controls (z, w) include mothers' religion, ethnicity and both parents' cohort, education, immigration year group dummies, and employment status when the child was between age 0 and 10. For "Hwang ($d_w > 0$)" estimates, I assume the auxiliary data does not include two maternal religion dummy variables (whether a mother is a Muslim/Sikh).

The 90% confidence intervals for "Hwang ($d_w = 0$)" estimates are computed by Hong and Li (2018) numerical delta method described in Online Appendix section B.⁸⁹ The tuning parameter for the numerical delta method is $N_m^{-0.49}$, where N_m is the sample size of the LS data.

F.2 Tuning Parameter: -0.3

Figure F.4: Robustness Check: Neighborhood Effects on Intergenerational Cultural Transmission

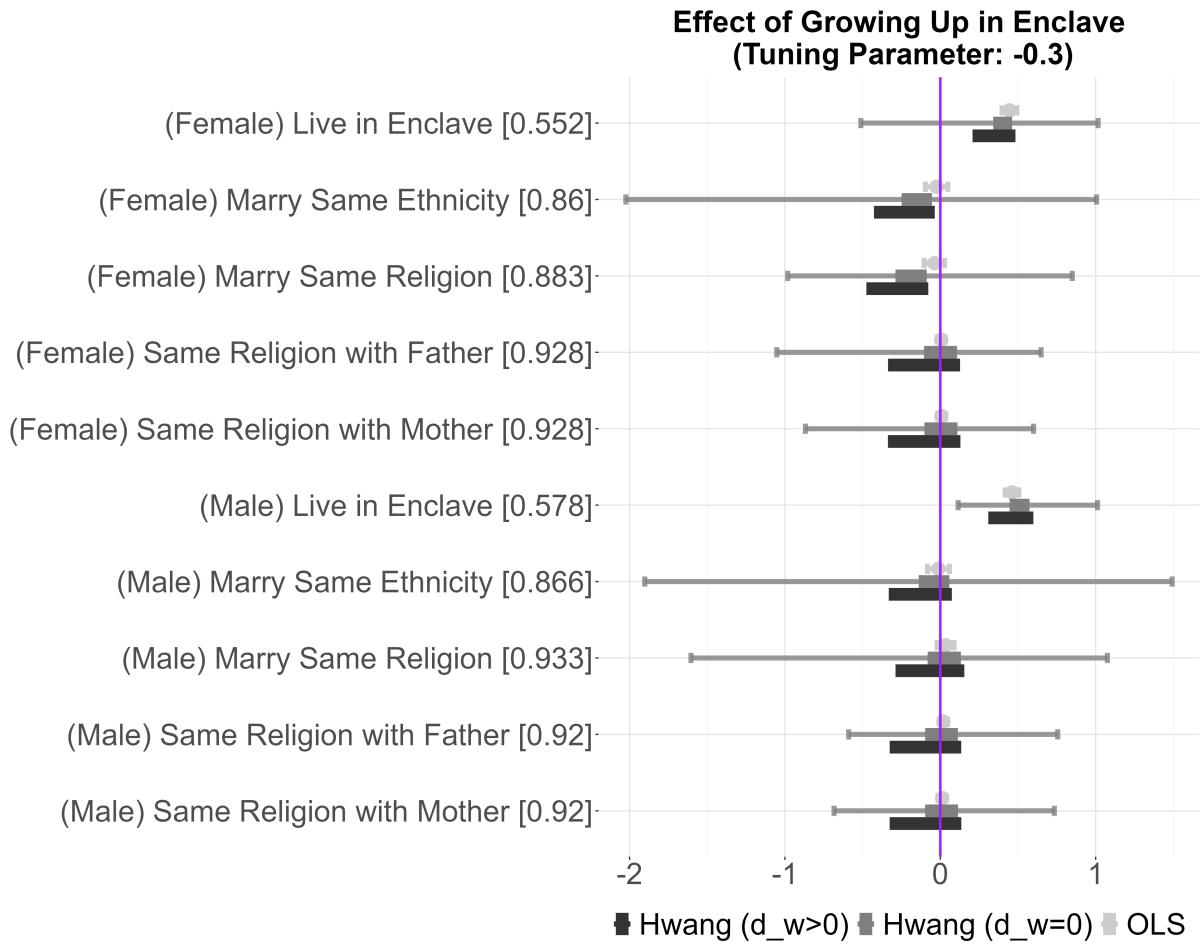


Data Source: FNSEM and ONS LS

Note: The graph presents the OLS point estimate from a short regression without x and two “Hwang” interval estimates of the neighborhood effect (β_n) under the assumptions of $d_w = 0$ and $d_w > 0$. The point estimates are displayed as dots and the interval estimates are shown as segment bands. The 90% confidence intervals for the “OLS” and “Hwang ($d_w = 0$)” estimates are drawn as whiskers. The 90% confidence intervals for “Hwang ($d_w = 0$)” estimates connect the lower end of the 90% CI for the lower bound estimate and the upper end of the 90% CI for the upper bound estimate and they are computed by Hong and Li (2018) numerical delta method described in Online Appendix section B. The tuning parameter for the numerical delta method is $N_m^{-0.3}$, where N_m is the sample size of the LS data.

Dependent variables are labeled on the left side of the figure. Numbers inside the brackets next to each dependent variable label are the averages of the dependent variables. The (z', w') controls include mothers’ religion, ethnicity and both parents’ cohort, education, immigration year group dummies, and employment status when the child was between age 0 and 10. For “Hwang ($d_w > 0$)” estimates, I assume the auxiliary data does not include two maternal religion dummy variables (whether a mother is a Muslim/Sikh).

Figure F.5: Evidence on Heterogeneous Effects by Gender



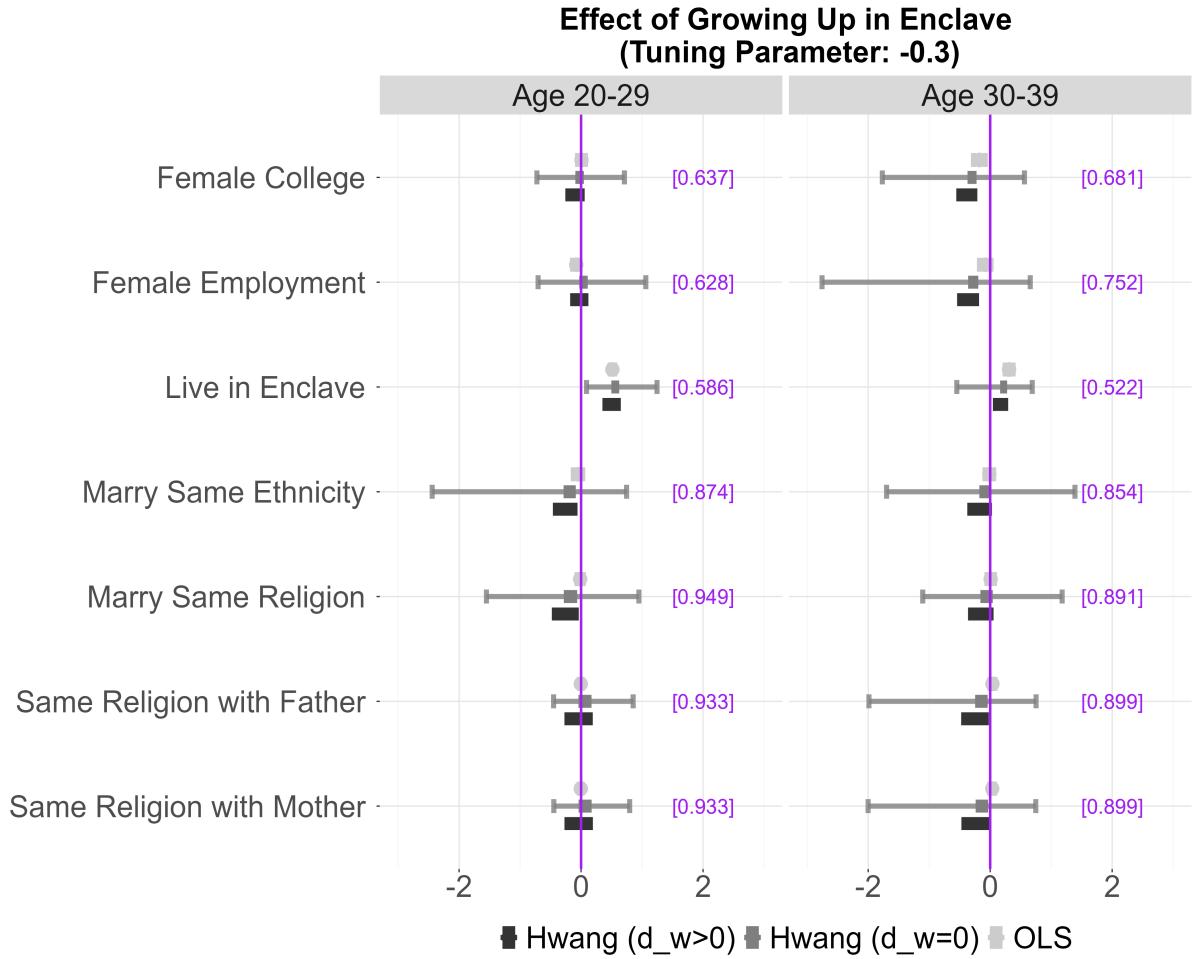
Data Source : FNSEM and ONS LS

Note: This figure shows the estimates in Figure 4 by gender.

The graph presents the OLS point estimate from a short regression without x and two “Hwang” interval estimates of the neighborhood effect (β_n) under the assumptions of $d_w = 0$ and $d_w > 0$. The point estimates are displayed as dots and the interval estimates are shown as segment bands. The 90% confidence intervals for the “OLS” and “Hwang ($d_w = 0$)” estimates are drawn as whiskers. The 90% confidence intervals for “Hwang ($d_w = 0$)” estimates connect the lower end of the 90% CI for the lower bound estimate and the upper end of the 90% CI for the upper bound estimate and they are computed by Hong and Li (2018) numerical delta method described in Online Appendix section B. The tuning parameter for the numerical delta method is $N_m^{-0.3}$, where N_m is the sample size of the LS data.

Dependent variables are labeled on the left side of the figure. Numbers inside the brackets next to each dependent variable label are the averages of the dependent variables. The (z', w') controls include mothers' religion, ethnicity and both parents' cohort, education, immigration year group dummies, and employment status when the child was between age 0 and 10. For “Hwang ($d_w > 0$)” estimates, I assume the auxiliary data does not include two maternal religion dummy variables (whether a mother is a Muslim/Sikh).

Figure F.6: Evidence on Heterogeneous Effects by Cohorts



Data Source : FNSEM and ONS LS

Note: This figure shows the estimates in Figure 4 by cohort; recall that the two cohorts are (i) those between 20 and 29 years of age in 2011 and (ii) those between 30 and 39 years of age in 2011.

The graph presents the OLS point estimate from a short regression without x and two “Hwang” interval estimates of the neighborhood effect (β_n) under the assumptions of $d_w = 0$ and $d_w > 0$. The point estimates are displayed as dots and the interval estimates are shown as segment bands. The 90% confidence intervals for the “OLS” and “Hwang ($d_w = 0$)” estimates are drawn as whiskers. The 90% confidence intervals for “Hwang ($d_w > 0$)” estimates connect the lower end of the 90% CI for the lower bound estimate and the upper end of the 90% CI for the upper bound estimate and they are computed by Hong and Li (2018) numerical delta method described in Online Appendix section B. The tuning parameter for the numerical delta method is $N_m^{-0.3}$, where N_m is the sample size of the LS data.

Dependent variables are labeled on the left side of the figure. Numbers inside the brackets next to each dependent variable label are the averages of the dependent variables. The (z', w') controls include mothers’ religion, ethnicity and both parents’ cohort, education, immigration year group dummies, and employment status when the child was between age 0 and 10. For “Hwang ($d_w > 0$)” estimates, I assume the auxiliary data does not include two maternal religion dummy variables (whether a mother is a Muslim/Sikh).

Table F.4: Numbers behind Figure F.4

Dependent Variable		OLS	Hwang ($d_w = 0$)	Hwang ($d_w > 0$)	N
Female Employment [0.66]	Estimate	-0.07	[-0.1,0.04]	[-0.26,0.05]	984
	90% CI	[-0.12,-0.02]	[-0.63,0.6]		
Female College [0.648]	Estimate	-0.03	[-0.17,-0.03]	[-0.34,-0.01]	827
	90% CI	[-0.09,0.02]	[-0.96,0.7]		
Marry Same Ethnicity [0.863]	Estimate	-0.02	[-0.19,0]	[-0.38,0.02]	540
	90% CI	[-0.07,0.03]	[-1.23,0.71]		
Live in Enclave [0.565]	Estimate	0.46	[0.39,0.52]	[0.26,0.54]	2087
	90% CI	[0.42,0.49]	[0.03,0.92]		
Marry Same Religion [0.91]	Estimate	0.01	[-0.18,0.02]	[-0.38,0.04]	522
	90% CI	[-0.03,0.05]	[-1.07,0.64]		
Same Religion with Father [0.924]	Estimate	0.01	[-0.1,0.11]	[-0.33,0.13]	1990
	90% CI	[-0.01,0.03]	[-0.71,0.53]		
Same Religion with Mother [0.924]	Estimate	0.01	[-0.1,0.11]	[-0.33,0.13]	1990
	90% CI	[-0.01,0.03]	[-0.8,0.53]		

Data Source: FNSEM and ONS LS

Note: Numbers inside the brackets next to each dependent variable label are the averages of the dependent variables.

The controls (z, w) include mothers' religion, ethnicity and both parents' cohort, education, immigration year group dummies, and employment status when the child was between age 0 and 10. For "Hwang ($d_w > 0$)" estimates, I assume the auxiliary data does not include two maternal religion dummy variables (whether a mother is a Muslim/Sikh).

The 90% confidence intervals for "Hwang ($d_w = 0$)" estimates are computed by Hong and Li (2018) numerical delta method described in Online Appendix section B. The tuning parameter for the numerical delta method is $N_m^{-0.3}$, where N_m is the sample size of the LS data.

Table F.5: Numbers behind Figure F.5

Dependent Variable		OLS	Hwang ($d_w = 0$)	Hwang ($d_w > 0$)	N
(Female) Marry Same Ethnicity [0.86]	Estimate	-0.02	[-0.25,-0.05]	[-0.43,-0.04]	250
	90% CI	[-0.1,0.05]	[-2.02,1.01]		
(Male) Marry Same Ethnicity [0.866]	Estimate	-0.01	[-0.14,0.06]	[-0.33,0.07]	290
	90% CI	[-0.08,0.06]	[-1.9,1.49]		
(Female) Live in Enclave [0.552]	Estimate	0.44	[0.34,0.46]	[0.21,0.48]	997
	90% CI	[0.4,0.49]	[-0.51,1.02]		
(Male) Live in Enclave [0.578]	Estimate	0.46	[0.45,0.57]	[0.31,0.6]	1090
	90% CI	[0.42,0.51]	[0.11,1.01]		
(Female) Marry Same Religion [0.883]	Estimate	-0.04	[-0.29,-0.09]	[-0.48,-0.08]	240
	90% CI	[-0.1,0.03]	[-0.98,0.85]		
(Male) Marry Same Religion [0.933]	Estimate	0.03	[-0.08,0.13]	[-0.29,0.15]	282
	90% CI	[-0.02,0.09]	[-1.6,1.08]		
(Female) Same Religion with Father [0.928]	Estimate	0.01	[-0.1,0.11]	[-0.34,0.13]	949
	90% CI	[-0.02,0.03]	[-1.05,0.65]		
(Male) Same Religion with Father [0.92]	Estimate	0.02	[-0.1,0.11]	[-0.33,0.13]	1041
	90% CI	[-0.01,0.05]	[-0.59,0.76]		
(Female) Same Religion with Mother [0.928]	Estimate	0.01	[-0.1,0.11]	[-0.34,0.13]	949
	90% CI	[-0.02,0.03]	[-0.87,0.6]		
(Male) Same Religion with Mother [0.92]	Estimate	0.01	[-0.1,0.11]	[-0.33,0.14]	1041
	90% CI	[-0.01,0.04]	[-0.68,0.73]		

Data Source: FNSEM and ONS LS

Note: Numbers inside the brackets next to each dependent variable label are the averages of the dependent variables.

The controls (z, w) include mothers' religion, ethnicity and both parents' cohort, education, immigration year group dummies, and employment status when the child was between age 0 and 10. For "Hwang ($d_w > 0$)" estimates, I assume the auxiliary data does not include two maternal religion dummy variables (whether a mother is a Muslim/Sikh).

The 90% confidence intervals for "Hwang ($d_w = 0$)" estimates are computed by Hong and Li (2018) numerical delta method described in Online Appendix section B. The tuning parameter for the numerical delta method is $N_m^{-0.3}$, where N_m is the sample size of the LS data.

Table F.6: Numbers behind Figure F.6

Dependent Variable		OLS	Hwang ($d_w = 0$)	Hwang ($d_w > 0$)	N
(20-29) Female Employment [0.628]	Estimate	-0.08	[-0.03,0.11]	[-0.18,0.12]	720
	90% CI	[-0.14,-0.02]	[-0.7,1.06]		
(20-29) Female College [0.637]	Estimate	0.01	[-0.09,0.05]	[-0.26,0.06]	562
	90% CI	[-0.06,0.07]	[-0.72,0.71]		
(20-29) Marry Same Ethnicity [0.874]	Estimate	-0.05	[-0.29,-0.09]	[-0.46,-0.06]	190
	90% CI	[-0.13,0.03]	[-2.44,0.75]		
(20-29) Live in Enclave [0.586]	Estimate	0.51	[0.5,0.63]	[0.35,0.65]	1432
	90% CI	[0.48,0.55]	[0.09,1.24]		
(20-29) Marry Same Religion [0.949]	Estimate	-0.01	[-0.28,-0.06]	[-0.48,-0.04]	178
	90% CI	[-0.07,0.04]	[-1.55,0.95]		
(20-29) Same Religion with Father [0.933]	Estimate	0	[-0.04,0.17]	[-0.27,0.19]	1349
	90% CI	[-0.03,0.02]	[-0.45,0.85]		
(20-29) Same Religion with Mother [0.933]	Estimate	0	[-0.04,0.17]	[-0.27,0.19]	1349
	90% CI	[-0.03,0.02]	[-0.45,0.8]		
(30-39) Female Employment [0.752]	Estimate	-0.08	[-0.36,-0.19]	[-0.54,-0.18]	250
	90% CI	[-0.18,0.01]	[-2.76,0.66]		
(30-39) Female College [0.681]	Estimate	-0.18	[-0.37,-0.22]	[-0.56,-0.21]	251
	90% CI	[-0.27,-0.08]	[-1.77,0.56]		
(30-39) Marry Same Ethnicity [0.854]	Estimate	-0.01	[-0.18,0.02]	[-0.38,0.03]	328
	90% CI	[-0.08,0.06]	[-1.7,1.39]		
(30-39) Live in Enclave [0.522]	Estimate	0.31	[0.16,0.28]	[0.04,0.3]	605
	90% CI	[0.24,0.38]	[-0.55,0.69]		
(30-39) Marry Same Religion [0.891]	Estimate	0.01	[-0.16,0.04]	[-0.36,0.06]	322
	90% CI	[-0.05,0.07]	[-1.11,1.18]		
(30-39) Same Religion with Father [0.899]	Estimate	0.04	[-0.25,-0.04]	[-0.48,-0.02]	594
	90% CI	[0,0.08]	[-1.99,0.75]		
(30-39) Same Religion with Mother [0.899]	Estimate	0.03	[-0.24,-0.03]	[-0.47,-0.01]	594
	90% CI	[-0.01,0.07]	[-2,0.75]		

Data Source: FNSEM and ONS LS

Note: Numbers inside the brackets next to each dependent variable label are the averages of the dependent variables.

The controls (z, w) include mothers' religion, ethnicity and both parents' cohort, education, immigration year group dummies, and employment status when the child was between age 0 and 10. For "Hwang ($d_w > 0$)" estimates, I assume the auxiliary data does not include two maternal religion dummy variables (whether a mother is a Muslim/Sikh).

The 90% confidence intervals for "Hwang ($d_w = 0$)" estimates are computed by Hong and Li (2018) numerical delta method described in Online Appendix section B.⁹⁵ The tuning parameter for the numerical delta method is $N_m^{-0.3}$, where N_m is the sample size of the LS data.

G An Economic Model to Justify the Main Regression Model

This section presents an intergenerational cultural transmission model which justifies the main regression model in equation (1). I closely follow the model setup of [Bisin and Verdier \(2001\)](#).

The model is an overlapping generation model with a single parent and a single child. Each agent lives for two periods and then dies; the first period is childhood and the second period is adulthood. An adult raises one child and each parent is characterized as having a heterogeneous cultural trait $u^P \in \mathcal{U} \subset \mathbb{R}^+$ and some observable characteristics vector $Z^P \in \mathbb{R}^N$.

At the beginning of adulthood, a parent first chooses a neighborhood among two neighborhoods, $n \in \{0, 1\}$. I label these neighborhoods as “Non-Ethnic Enclave” and “Ethnic Enclave.” Next, while living in the chosen neighborhood, parents may exert a costly effort, $\tau_i \in \mathbb{R}^+$, to socialize the child. Parent receives utility from (i) the amenity value of the neighborhood, which depends on the parent’s own cultural trait u^P and other demographics Z^P , (ii) the effort cost to socialize the child $C\tau$, and (iii) the child’s socialization outcome, u^C , which is realized at the end of childhood.

Parent’s Problem, conditional on neighborhood

$$V_n(u, Z^P) \equiv \max_{\tau} \underbrace{f(u^P, n, Z^P)}_{(i)} - \underbrace{C\tau}_{(ii)} + \underbrace{\gamma E[u^C | u^P, \tau, n]}_{(iii)}. \quad (\text{G.1})$$

The first term (i) captures both the utility gain for parents from living in a culturally similar neighborhood and the economic cost associated with different rents in each neighborhood. The second term (ii) is the parental effort cost to socialize the child. The third term (iii) captures the parent’s valuation of the child’s expected outcome $E[u^C | u^P, \tau, n]$, which depends on the socialization mechanism. The child’s cultural preference, u^C , is determined at the end of childhood by the parental cultural trait u^P , endogenous parental effort τ , neighborhood n during childhood, and some random component $\epsilon \sim F_\epsilon(\cdot)$.

Socialization Mechanism

$$u^C = \underbrace{\alpha(\tau)u^P}_{\text{vertical transmission}} + \underbrace{\beta n}_{\text{oblique transmission}} + \epsilon, \quad \epsilon \sim F_\epsilon(\cdot), \quad \beta > 0. \quad (\text{G.2})$$

Parents are assumed to know the socialization mechanism (G.2) when making decisions on socialization effort τ and a neighborhood choice n . $\alpha(\cdot)$ corresponds to the “production function” of vertical transmission⁴⁴ in Bisin and Verdier (2001). The above socialization mechanism assumes that parental efforts and the neighborhood are perfect substitutes for producing a child’s cultural preference.⁴⁵ In addition to the vertical transmission, there is an independent effect from neighborhood capturing oblique transmissions, such as learning from peers and adult neighbors.

Next, I make the following technical assumption on $\alpha(\cdot)$, which implies that higher parental effort to transmit their own cultural trait increases the chance that the child inherits a similar cultural trait.

Assumption G.1 (Socialization Mechanism).

α is a continuous function from \mathbb{R}^+ to $[0, 1] \subset \mathbb{R}^+$ that is strictly increasing in τ .

The optimal parental socialization effort level follows from a first-order condition.

⁴⁴Cavalli-Sforza and Feldman (1981) suggests three mechanisms for intergenerational cultural transmission. Vertical transmission is the influence of the parent’s preference on the child’s preference. Oblique transmission is the transmission from unrelated adults to the child’s preference. Horizontal transmission is the transmission from peers’ preferences to the child’s preferences. I include horizontal transmission in oblique transmissions in this paper as I do not have data to distinguish the effect of peers in the neighborhood from other adults in the neighborhood.

⁴⁵This assumption implies a simple regression model in the end. Relaxing this assumption would make the estimation more difficult – adding potential interaction effect between parental effect and the neighborhood is not easy given the current sample size.

Optimal Parental Socializing Effort

$$\tau^* : -C + \gamma \left\{ \frac{\partial \alpha(\tau^*)}{\partial \tau} u^P \right\} = 0. \quad (\text{G.3})$$

$$\tau^* = \tau(u^P). \quad (\text{G.4})$$

Using the optimal parental effort τ^* gives an additively separable equation for the child's cultural preference u^C .

Parent's Problem, Choosing Neighborhood

$$V(u^P, Z^P) \equiv \max\{V_0(u^P, Z^P), V_1(u^P, Z^P)\}. \quad (\text{G.5})$$

Due to endogenous neighborhood choice, parental preference u^P is correlated with neighborhood choice n . Therefore, without controlling for u^P , there will be omitted variable bias in the childhood neighborhood effect on a child's adulthood outcomes.

Child's Adulthood Outcomes

$$y^C = h(u^C, Z^P) + \nu, \quad \nu \sim F_\nu(.). \quad (\text{G.6})$$

$$= h(\tau(u^P)u^P + \beta n + \epsilon, Z^P) + \nu. \quad (\text{G.7})$$

Finally, the child's adulthood outcome y^C is determined by the child's adulthood cultural preference u^C , parental observable characteristics Z^P , and a stochastic error term ν . A linear approximation of equation G.6 yields the regression model in equation 1, where the key omitted variable from the main dataset—using their notation—is $x = \{u^P\}$, and the common covariate vector is $z = \{n, Z^P\}$.