NBER WORKING PAPER SERIES

COMPLEMENTARITY AND AGGREGATE IMPLICATIONS OF ASSORTATIVE MATCHING: A NONPARAMETRIC ANALYSIS

Bryan S. Graham
Guido W. Imbens
Geert Ridder

Working Paper 14860
http://www.nber.org/papers/w14860

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2009

Financial support for this research was generously provided through NSF grant SES 0136789, SES 0452590, and SES 0820361. We thank participants in seminars at Harvard-MIT, Princeton, UC Berkeley, NBER Labor Studies, Cemmap, and Brown University for comments. We thank Cristine Pinto for excellent research assistance. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

Complementarity and Aggregate Implications of Assortative Matching: A Nonparametric Analysis
Bryan S. Graham, Guido W. Imbens, and Geert Ridder
NBER Working Paper No. 14860
April 2009
JEL No. C14,C21,C52

ABSTRACT

This paper presents methods for evaluating the effects of reallocating an indivisible input across production units, taking into account resource constraints by keeping the marginal distribution of the input fixed. When the production technology is nonseparable, such reallocations, although leaving the marginal distribution of the reallocated input unchanged by construction, may nonetheless alter average output. Examples include reallocations of teachers across classrooms composed of students of varying mean ability. We focus on the effects of reallocating one input, while holding the assignment of another, potentially complementary, input fixed. We introduce a class of such reallocations -- correlated matching rules -- that includes the *status quo* allocation, a random allocation, and both the perfect positive and negative assortative matching allocations as special cases. We also characterize the effects of local (relative to the *status quo*) reallocations. For estimation we use a two-step approach. In the first step we nonparametrically estimate the production function. In the second step we average the estimated production function over the distribution of inputs induced by the new assignment rule. These methods build upon the partial mean literature, but require extensions involving boundary issues. We derive the large sample properties of our proposed estimators and assess their small sample properties via a limited set of Monte Carlo experiments.

Bryan S. Graham
Department of Economics
University of California, Berkeley
508-1 Evans Hall #3880
Berkeley, CA 94720-3880
and NBER
bgraham@econ.berkeley.edu

Guido W. Imbens
Department of Economics
Littauer Center
Harvard University
1805 Cambridge Street
Cambridge, MA 02138
and NBER
imbens@fas.harvard.edu

Geert Ridder
Department of Economics
University of Southern California
Kaprielian Hall
Los Angeles, CA 90089
ridder@usc.edu

1 Introduction

Consider a production function depending on a number of inputs. We are interested in the effect of a particular input on output, and specifically in the average effects of policies that change the allocation of this input across production units. For each production unit output may be monotone in this input, but at different rates. If the input is indivisible, and its aggregate stock fixed, it is impossible to simultaneously raise the input level for all production units. In such cases it may be of interest to consider the output effects of reallocations of the input across production units. Here we investigate econometric methods for assessing the effect of such reallocations on average output. We will call the average causal effects of such policies Aggregate Redistributional Effects (AREs). A key feature of the reallocations we consider is that, although they potentially alter input levels for each firm, they keep the marginal distribution of the input across the population of firms fixed.

The first contribution of our paper is to introduce a framework for considering such reallocations, and to define novel estimands that capture their key features. These estimands include the effects of focal reallocations, and a semiparametric class of reallocations, as well as the effect of a local reallocation. One focal reallocation redistributes the input across production units such that it has perfect rank correlation with a second input. We refer to this as the positive assortative matching allocation. We also consider a negative assortative matching allocation where the primary input is redistributed to have perfect negative rank correlation with the second input. A third allocation involves randomly assigning the input across firms. This allocation, by construction, ensures independence of the two inputs. A fourth allocation simply maintains the *status quo* assignment of the input. More generally, we consider a two parameter family of feasible reallocations that include these four focal allocations as special cases. Reallocations in this family may depend on the distribution of a second input or firm characteristic. This characteristic may be correlated with the firm-specific return to the input to be reallocated. Our family of reallocations, called correlated matching rules, includes each of the four focal allocations as special cases. In particular the family traces a path from the positive to negative assortative matching allocations. Each reallocation along this path keeps the marginal distribution of the two inputs fixed, but it induces a different level of correlation between the two inputs. Each of the reallocations we consider are members of a general class of reallocation rules that keep the marginal distributions of both inputs fixed. We also provide a local measure of complementarity that requires much weaker conditions on the support of the input distribution. This estimand measures whether a small step away from the status quo, towards perfect assortative matching allocation raises average output.

The second contribution of our paper is to derive statistical methods for estimation and inference for the proposed estimands. We derive an estimator for average output under all correlated matching allocations, and for the local complementarity measure. Our estimator requires that the first input is exogenous conditional on the second input and additional firm characteristics. Except for the case of perfect negative and positive rank correlation the estimator has the usual parametric convergence rate. For the two extremes the rate of convergence is slower, comparable to that of estimating a regression function with a scalar covariate at a point. In all cases we drive the asymptotic distribution of the estimator. In the first step of the estimation procedure we use a nonparametric estimator for the production function. We modify existing kernel estimators to deal with boundary issues that arise in our setting.

Our focus on reallocation rules that keep the marginal distribution of the inputs fixed is appropriate in applications where the input is indivisible, such as in the allocation of teachers to classes, or managers to production units. In other settings it may be more appropriate to consider allocation rules that leave the total amount of the input constant by fixing its average level. Such rules would require some modification of the methods considered in this paper.

Our methods may be useful in a variety of settings. One class of examples concerns complementarity of inputs in production functions (e.g. Athey and Stern, 1998). If the first and second inputs are everywhere complements, then the difference in average output between the positive and negative assortative matching allocations provides a nonparametric measure of the degree of complementarity. This measure is invariant to monotone transformations of the inputs. If the production function is not supermodular, the interpretation of this difference is not straightforward, although it still might be viewed as some sort of 'global' measure of input complementarity.

A second example concerns educational production functions. Card and Krueger (1992) study the relation between adult wages and teacher quality. Teacher quality may improve outcomes for all students, but average outcomes may be higher or lower depending on whether, given a fixed supply of teachers, the best teachers are assigned to the least prepared students or vice versa. Parents concerned solely with outcomes for their own children may be most interested in the effect of raising teacher quality on expected outcomes. A school board, however, may be more interested in maximizing expected outcomes given a fixed set of classes and and a fixed set of teachers, by optimally matching teachers to classes.

A third class of examples arises in settings with social interaction (c.f., Manski 1993; Brock and Durlauf 2001). Sacerdote (2001) studies peer effects in college by looking at the relation between individual outcomes and roommate characteristics. From the perspective of the individual student it may again be of interest whether having a roommate with different characteristics would, in expectation, lead to a different outcome. This is what Manski (1993) calls an exogenous or contextual effect. The college, however, may be interested in a different effect, namely the effect on average outcomes of changing the procedures for assigning roommates. While it may be very difficult for a college to quickly change the distribution of characteristics in incoming classes, it may be under its control to change the way roommates are assigned. In Graham, Imbens and Ridder (2006b) we study the peer effect setting further, developing methods appropriate for social groups of arbitrary size when agents are binary-typed. Our focus in that work is on the outcome and inequality effects of segregation.

If production functions are additive in inputs, the questions posed above have trivial answers: average outcomes are invariant to input reallocations. Although reallocations may raise outcomes for some units in that case, they will necessarily lower them by an offsetting amount for others. Reallocations are zero-sum games in this additive setting. With additive and linear functions, even more general assignment rules that allow the marginal input distribution to change, while keeping its average level fixed, do not affect average outcomes. In order for these questions to have non-trivial answers, one therefore needs to explicitly recognize, and allow for, non-additivity and non-linearity of a production function in its inputs. For this reason our approach is fully nonparametric.

The current paper builds on the larger treatment effect and program evaluation literature.[1] More directly, it is complementary to the small literature on the effect of treatment assignment rules (Manski, 2004; Dehejia, 2004; Hirano and Porter, 2005). Our focus is different from

[1]For recent surveys see Angrist and Krueger (2001), Heckman, Lalonde and Smith (2000), and Imbens and Wooldridge (2009).

that in the Manski, Dehejia, and Hirano-Porter studies. First, we allow for continuous rather than discrete or binary treatments. Second, our assignment policies take into account resource constraints (by leaving unchanged he marginal distribution of the treatment), whereas in the previous papers treatment assignment for one unit is not restricted by the assignments for other units. Our policies are redistributions. In the current paper we focus on estimation and inference for specific assignment rules. It is also interesting to consider optimal rules as in the Manski, Dehejia and Hirano-Porter studies. The class of feasible reallocations/redistributions includes all joint distributions of the two inputs with fixed marginal distributions. When the inputs are continuously-valued, as we assume in the current paper, this class of potential rules is very large. Characterizing the optimal allocation within this class is therefore a non-trivial problem. When both inputs are discretely-valued the problem with finding the optimal allocation is tractable as the joint distribution of the inputs is characterized by a finite number of parameters. In Graham, Imbens and Ridder (2006a) we consider optimal allocation rules when both inputs are discrete, allowing for general complementarity or substitutability of the inputs.

Our paper is also related to recent work on identification and estimation of models of social interactions (e.g., Manski, 1993; Brock and Durlauf, 2001; Graham, 2008; Moffitt, 2001). We do not focus on directly characterizing the within-group structure of social interactions, an important theme of this literature. Rather our goal is simply to estimate the average relationship between group composition and outcomes. The average we estimate may reflect endogenous behavioral responses by agents to changes in group composition, or even equal an average over multiple equilibria. Viewed in this light our approach is reduced form in nature. However it is sufficient for, say, an university administrator to characterize the outcome effects of alternative roommate assignment procedures, as long as the average response to group composition remains unchanged across such procedures.

The econometric approach taken here builds on the partial mean literature (e.g., Newey, 1994; Linton and Nielsen, 1995). In this literature one first estimates a regression function nonparametrically. In the second stage the regression function is averaged, possibly after some weighting with a known or estimable weight function, over some of the regressors. Similarly here we first estimate a the production function nonparametrically as the conditional mean of the outcome given the observed inputs. In the second stage the averaging is over the distribution of the regressors induced by the new assignment rule. This typically involves the original marginal distribution fo some of the regressors, but a different conditional distribution for others. Complications arise because this conditional covariate distribution may be degenerate, which will affect the rate of convergence for the estimator. In addition the conditional covariate distribution itself may require nonparametric estimation through its dependence on the assignment rule. For the policies we consider the assignment rule will involve distribution functions and their inverses similar to the way these enter in the changes-in-changes model of Athey and Imbens (2006).

The next section lays out our basic model and approach to identification. Section 3 then defines and motivates the estimands we seek to estimate. Section 4 presents of our estimators, and derives their large-sample properties, for the case where inputs are continuously-valued. Section 5 presents the results from a small Monte Carlo exercise.

2 Model

In this section we present the basic set up and identifying assumptions. For clarity of exposition we use the production function terminology; although our methods are appropriate for a wide range of applications, as emphasized in the introduction. For production unit or firm i, for $i = 1, \ldots, N$, the production function relates a triple of observed inputs, (W_i, X_i, V_i), and an unobserved input ε_i, to an output Y_i:

$$Y_i = k(W_i, X_i, V_i, \varepsilon_i). \tag{2.1}$$

The inputs W_i and X_i, and the output Y_i are scalars. The third observed input V_i and the unobserved input ε_i can both be vectors. We are interested in reallocating the input W across production units. We focus upon reallocations which hold the marginal distribution of W fixed. As such they are appropriate for settings where W is a plausibly indivisible input, such as a manager or teacher, with a certain level of experience and expertise. The presumption is also that the aggregate stock of W is difficult to augment. In addition to W there are two other (observed) firm characteristics that may affect output: X and V, where X is a scalar and V is a vector of dimension L_V. The first characteristic X could be a measure of, say, the quality of the long-run capital stock, with V being other characteristics of the firm such as location and age. These characteristics may themselves be inputs that can be varied, but this is not necessary for the arguments that follow. In particular the exogeneity assumption that we make for the first input need not hold for these characteristics.

We observe for each production unit, indexed by $i = 1, \ldots, N$, the level of the input, W_i, the characteristics X_i and V_i, and the realized output level, Y_i. In the educational example the unit of observation would be a classroom. The variable input W would be teacher quality, and X would be a measure of quality of the class, e.g., average test scores in prior years. The second characteristic V could include other measures of the class, e.g., its age or gender composition, as elements. In the roommate example the unit would be the individual, with W the quality of the roommate (measured by, for example, a high school test score), and the characteristic X would be own quality. The second set of characteristics V could be other characteristics of the dorm or of either of the two roommates such as smoking habits (which may be used by university administrators in the assignment of roommates).

Our key identifying assumption is that conditional on firm characteristics (X, V) the assignment of W, the level of the input to be reallocated, is exogenous:

Assumption 2.1 (Exogeneity)

$$\varepsilon \perp W \,\Big|\, X, V.$$

Let

$$g(w, x, v) = \mathbb{E}[Y|W = w, X = x, V = v], \tag{2.2}$$

and

$$\sigma^2(w, x, v) = \mathbb{V}[Y|W = w, X = x, V = v], \tag{2.3}$$

denote the expectation and the variance of the output conditional on input level w and characteristics x and v. We often refer to the derivative of $g(w, x, v)$ with respect to w, which will be denoted by

$$g_W(w, x, v) = \frac{\partial}{\partial w} g(w, x, v), \tag{2.4}$$

Under exogeneity we have – among firms with identical values of X and V – an equality between the counterfactual average output that we would observe if all firms in this subpopulation were assigned $W = w$, and the average output we observe for the subset of firms within this subpopulation that are in fact assigned $W = w$. Alternatively, the exogeneity assumption implies that the difference in $g(w, x, v)$ evaluated at two values of w, w_0 and w_1, has a causal interpretation as the average effect of assigning $W = w_1$ rather than $W = w_0$:

$$g(w_1, x, v) - g(w_0, x, v) = \mathbb{E}\left[k(w_1, X, V, \varepsilon) - k(w_0, X, V, \varepsilon) \right| X = x, V = v] .$$

Assumption 2.1 is often controversial. It holds under conditional random assignment of W to units; as would occur in a randomized experiment. However randomized allocation mechanisms are also used by administrators in some institutional settings. For example some universities match freshman roommates randomly conditional on responses to housing questionnaires (e.g., Sacerdote 2001). This assignment mechanism is consistent with Assumption 2.1. In other settings, particularly where assignment is bureaucratic, as may be true in some educational settings, a plausible set of conditioning variables may be available. In this paper we focus upon identification and estimation under Assumption 2.1. In principle, however, the methods could be extended to accommodate other approaches to identification based upon, for example, nonparametric instrumental variables methods (e.g., Matzkin, 2008, Imbens and Newey, 2009).

Much of the treatment effect literature (e.g., Angrist and Krueger, 2000; Heckman, Lalonde and Smith, 2000; Manski, 1990; Imbens and Wooldridge, 2009) has focused on the average effect of an increase in the value of the treatment. In particular, in the binary treatment case ($w \in \{0, 1\}$) interest has centered on the average treatment effect

$$\mathbb{E}[g(1, X, V) - g(0, X, V)].$$

With continuous inputs one may be interested in the full average output function $g(w, x, v)$ (Imbens, 2000; Flores, 2005) or in its derivative with respect to the input,

$$g_W(w, x, v),$$

at a point, or a weighted average,

$$\mathbb{E}\left[\omega(W, X, V) \cdot g_W(W, X, V)\right],$$

See Powell, Stock and Stoker (1989) or Hardle and Stoker, (1989) for estimands of this type.

Here we are interested in a fundamentally different class of estimands, one which has received little attention in the econometrics literature. We focus on policies that redistribute the input W, according to a rule based on the X characteristic of the unit. For example upon assignment mechanisms that match teachers of varying experience to classes of students based on average ability in the classes. One might assign those teachers with the most experience (highest values of W) to those classrooms with the highest ability students (highest values of X) and so on. In that case average outcomes would reflect perfect rank correlation between W and X. Alternatively, we could be interested in the average outcome if we were to assign W to be negatively perfectly rank correlated with X. A third possibility is to assign W so that it is independent of X. We are interested in the effect of such policies on the average value of the output. We refer to such effects in general as Aggregate Redistributional Effects (AREs). The three reallocations mentioned are a special case of a general set of reallocation rules that fix the marginal distributions of W and X, but allow for correlation in their joint distribution. For perfect assortative matching the correlation is 1, for negative perfect assortative matching -1, and for random allocation 0. By using a bivariate normal cupola we can trace out the path between these extremes.

We wish to emphasize that there are at least two limitations to our approach. First, we focus on comparing specific assignment rules, rather than searching for the optimal assignment rule. The latter problem is a particularly demanding problem in the current setting with continuously-valued inputs as the optimal assignment for each unit depends both on the characteristics of that unit as well as on the marginal distribution of characteristics in the population. When the inputs are discretely-valued both the problems of inference for a specific rule as well as the problem of finding the optimal rule become considerably more tractable. In that case any rule, corresponding to a joint distribution of the inputs, is characterized by a finite number of parameters. Maximizing estimated average output over all rules evaluated will then generally lead to the optimal rule. Graham, Imbens and Ridder (2006a) and, motivated by an early version of the current paper, Bhattacharya (2008), provide a discussion for the case with discrete covariates.

A second limitation is that the class of assignment rules we consider leaves all aspects of the marginal distribution of the inputs unchanged. This latter restriction is perfectly appropriate in cases where the inputs are indivisible, as, for example, in the social interactions and educational examples. In other cases one need not be restricted to such assignment rules. A richer class of estimands would allow for assignment rules that maintain some aspects of the marginal distribution of inputs but not others. An interesting class consists of assignment rules that maintain the average (and thus total) level of the input, but allow for its arbitrary distribution across units. This can be interpreted as assignment rules that "balance the budget". In such cases one might assign the maximum level of the input to some subpopulation and the minimum level of the input to the remainder of the population. Finally, one may wish to consider arbitrary decision rules where each unit can be assigned any level of the input within a set. In that case interesting questions include both the optimal assignment rule as a function of unit-level characteristics as well as average outcomes of specific assignment rules. In the binary treatment case such problems have been studied by Dehejia (2005), Manski (2004), and Hirano and Porter (2005).

3 Aggregate Redistributional Effects

Let $f_{W|X,V}(w|x,v)$ denote the conditional distribution of W given (X,V) in the data, and let $\tilde{f}_{W|X,V}(w|x,v)$ denote a potentially different conditional distribution. We will allow $\tilde{f}_{W|X,V}(w|x,v)$ to correspond to any distribution such that the implied marginal distribution for W_i remains unchanged, or

$$\int \tilde{f}_{W|X,V}(w|x,v) f_{X,V}(x,v) \mathrm{d}w\mathrm{d}x\mathrm{d}v = \int f_{W|X,V}(w|x,v) f_{X,V}(x,v) \mathrm{d}w\mathrm{d}x\mathrm{d}v.$$

This includes degenerate conditional distributions. In general we are interested in the average outcome that would result from the current distribution of (X, V, ε), if the distribution of W given (X, V) were changed from its current distribution, $f_{W|X,V}(w|x, v)$ to $\tilde{f}_{W|X,V}(w|x, v)$. We denote the expected output given such a reallocation by

$$\beta^{\text{are}}_{\tilde{f}} = \int g(w, x, v)\tilde{f}_{W|X,V}(w|x, v)f_{X,V}(x, v)\mathrm{d}w\mathrm{d}x\mathrm{d}v. \tag{3.5}$$

In the next two sections we discuss some specific choices for $\tilde{f}(\cdot)$.

3.1 Positive and Negative Assortive Matching Allocations

The first estimand we consider is expected average outcome given perfect assortative matching of W on X conditional on V:

$$\beta^{\text{pam}} = \mathbb{E}\left[g\left(F^{-1}_{W|V}(F_{X|V}(X|V)|V), X, V\right)\right], \tag{3.6}$$

where $F_{X|V}(X|V)$ denotes the conditional CDF of X given V, and $F^{-1}_{W|V}(q|V)$ is the q-th quantile (for $q \in [0, 1]$) associated with the conditional distribution of W given V (i.e., $F^{-1}_{W|V}(q|V)$ is a conditional quantile function). Therefore $F^{-1}_{W|V}(F_{X|V}(X|V)|V)$ computes a unit's location on the conditional CDF of X given V and reassigns it the corresponding quantile of the conditional distribution of W given V. Thus, among units with the same realization of V, those with the highest value of X are reassigned the highest value of W, and so on.

In order for β^{pam} to be well defined, we need some conditions on the joint distribution of (Y, W, X, V). We do not state these conditions here explicitly. When we discuss estimation, in Section 4, we provide conditions for consistent estimation., including compact support and smooth distributions for (W, X), and moment conditions for the conditional distribution of Y given (W, X). These conditions imply that β^{pam} is well defined.

The focus on reallocations within subpopulations defined by V, as opposed to population-wide reallocations, is motivated by the fact that the average outcome effects of such reallocations solely reflect complementarity or substitutability between W and X. To see why this is the case consider the alternative estimand

$$\beta^{\text{pam-pop}} = \mathbb{E}\left[g\left(F^{-1}_{W}\left(F_X(X)\right), X, V\right)\right]. \tag{3.7}$$

This gives average output associated with population-wide perfect assortative matching of W on X. If, for example, X and V are correlated, then this reallocation, in addition to altering the joint distribution of W and X, will alter the joint distribution of W and V. Say V is also a scalar and is positively correlated with X. Population-wide positive assortative matching will induce perfect rank correlated between W and X, but it will also affect the degree of correlation between W and V. This complicates the interpretation of the estimand when $g(w, x, v)$ is non-separable in w and v, as well as in w and x.

An example helps to clarify the issues involved. Let W denote an observable measure of teacher quality, X mean (beginning-of-year) achievement in a classroom, and V the fraction of the classroom that is female. If begining-of-year achievement varies with gender, (say, with classes with a higher fraction of girls having higher average achievement) then X and V will be correlated. A reallocation that assigns high quality teachers to high achievement classrooms, will also tend to assign such teachers to classrooms will an above average fraction of females.

Average achievement increases observed after implementing such a reallocation may reflect complementarity between teacher quality and begining-of-year student achievement or it may be that the effects of changes in teacher quality vary with gender and that, conditional on gender, their is no complementarity between teacher quality and achievement. By focusing on reallocations of teachers across classrooms with similar gender mixes, but varying baseline achievement, (3.6) provides a more direct avenue to learning about complementarity between W and X.[2]

Both (3.6) and (3.7) may be policy relevant, depending on the circumstances, and both are identified under Assumption 2.1 and additional support conditions (which we make explicit below). Under the additional assumption that

$$g(w, x, v) = g_1(w, x) + g_2(v),$$

the estimands, although associated with different reallocations, also have the same basic interpretation. In the current paper we focus upon (3.6), although it is conceptually straightforward to extend our results to (3.7).

Our second estimand is the expected average outcome given negative assortative matching:

$$\beta^{\text{nam}} = \mathbb{E}\left[g\left(F^{-1}_{W|V}\left(1 - F_{X|V}(X|V)|V\right), X, V\right)\right]. \tag{3.8}$$

If, within subpopulations homogenous in V, the two inputs W and X are everywhere complements, then the difference $\beta^{\text{pam}} - \beta^{\text{nam}}$ provides a measure of the strength of input complementarity. When $g(\cdot)$ is not supermodular, the interpretation of this difference is not straightforward. In Section 3.1 below we present a measure of 'local' (relative to the status quo allocation) complementarity between X and W.

3.2 Correlated Matching Allocations

The perfect positive and negative assortative allocations are focal allocations, being emphasized in the economic theory literature (e.g., Becker and Murphy, 2000; Legros and Newman, 2004). There are many more possible allocations. Two others that are of particular importance are the *status quo* allocation, and the random matching allocation. Average output under the *status quo* allocation is given by

$$\beta^{\text{sq}} = \mathbb{E}[Y] = \mathbb{E}[g(W, X, V)].$$

Average output under the random matching allocation is given by

$$\beta^{\text{rm}} = \int_v \left[\int_x \int_w g(w, x, v) f_{W|V}(w|v) f_{X|V}(x|v)\right] f_V(v) \mathrm{d}w\mathrm{d}x\mathrm{d}v.$$

This last estimand gives average output when W and X are independently assigned within subpopulations indexed by V.

These allocations are just four among the class of feasible allocations. This class is comprised of all joint distributions of inputs consistent with fixed marginal distributions (within subpopulations homogenous in V). As noted in the introduction, if the inputs are continuously distributed this class of joint distributions is very large. For this reason we only consider a

[2] We make the connection to complementarity more explicit in Section 3.3.

subset of these joint distributions. To be specific, we concentrate on a family of the feasible allocations, indexed by two parameters, τ and ρ, that includes as special cases the negative and positive assortative matching allocations, the independent allocation, and the status quo allocation. Let $\beta^{\text{cm}}(\tau, \rho)$ denote average output under the allocation indexed by τ and ρ. By changing the two parameters we trace out a "path" in two directions: further from or closer to the status quo allocation, and further from, or closer to, the perfect sorting allocations. Borrowing a term from the literature on cupolas, we call this class of feasible allocations "comprehensive," because it contains all four focal allocations as a special case. For ease of exposition we focus in the remainder of the paper on the case with no covariates beyond W and X, and so drop the argument V in the production function.

For the purposes of estimation, the correlated matching allocations are redefined using a truncated bivariate normal cupola. The truncation ensures that the denominator in the weights of the correlated matching ARE are bounded from 0, so that we do not require trimming. The bivariate standard normal probability density function (pdf) is

$$\phi(x_1, x_2; \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(x_1^2 - 2\rho x_1 x_2 + x_2^2)}, \qquad -\infty < x_1, x_2 < \infty$$

with a corresponding joint cumulative distribution function (cdf) denoted by $\Phi(x_1, x_2; \rho)$. Observe that

$$\Pr(-c < x_1 \leq c, -c < x_2 \leq c) = \Phi(c, c; \rho) - \Phi(c, -c; \rho) - [\Phi(-c, c; \rho) - \Phi(-c, -c; \rho)],$$

so that the truncated standard bivariate normal pdf is given by

$$\phi_c(x_1, x_2; \rho) = \frac{\phi(x_1, x_2; \rho)}{\Phi(c, c; \rho) - \Phi(c, -c; \rho) - [\Phi(-c, c; \rho) - \Phi(-c, -c; \rho)]}, \qquad -c < x_1, x_2 \leq c.$$

Denote the truncated bivariate cdf by Φ_c.

The truncated normal bivariate CDF gives a comprehensive cupola, because the corresponding joint cdf

$$H_{W,X}(w, x) = \Phi_c\left(\Phi_c^{-1}(F_W(w)), \Phi_c^{-1}(F_X(x)); \rho\right)$$

has marginal cdf's equal to $H_{W,X|V}(w, \infty|v) = F_W(w)$ and $H_{W,X}(\infty, x) = F_X(x)$, it reaches the upper and lower Fréchet bounds on the joint cdf for $\rho = 1$ and $\rho = -1$, respectively, and it has independent W, X as a special case for $\rho = 0$.

To define $\beta^{\text{cm}}(\rho, \tau)$, we note that joint pdf associated with $H_{W,X}(w, x)$ equals

$$h_{W,X}(w, x) = \phi_c\left(\Phi_c^{-1}(F_W(w)), \Phi_c^{-1}(F_X(x)); \rho\right) \frac{f_W(w) f_X(x)}{\phi_c\left(\Phi_c^{-1}(F_W(w))\right) \phi_c\left(\Phi_c^{-1}(F_X(x))\right)}.$$

Then we define $\beta^{\text{cm}}(\rho, 0)$ in terms of the truncated normal, as

$$\beta^{\text{cm}}(\rho, 0) = \int_{w,x} g(w, x) \frac{\phi_c\left(\Phi_c^{-1}(F_W(w)), \Phi_c^{-1}(F_X(x)); \rho\right)}{\phi_c\left(\Phi_c^{-1}(F_W(w))\right) \phi_c\left(\Phi_c^{-1}(F_X(x))\right)} f_W(w) f_X(x) \mathrm{d}w \mathrm{d}x. \tag{3.9}$$

Average output under the correlated matching allocation is given by

$$\beta^{\text{cm}}(\rho, \tau) = \tau \cdot \mathbb{E}[Y] + (1 - \tau) \cdot \beta^{\text{cm}}(\rho, 0) \tag{3.10}$$

$$= \tau \cdot \mathbb{E}[Y_i] + (1-\tau)$$
$$\times \int_{x,w} g(w,x) \frac{\phi_c\left(\Phi_c^{-1}(F_W(w)), \Phi_c^{-1}(F_X(x)); \rho\right)}{\phi_c\left(\Phi_c^{-1}(F_W(w))\right) \phi_c\left(\Phi_c^{-1}(F_X(x))\right)} f_W(w) f_X(x) \mathrm{d}w\mathrm{d}x,$$

for $\tau \in [0,1]$ and $\rho \in (-1,1)$.

The case with $\tau = 1$ corresponds to the *status quo*:

$$\beta^{\mathrm{sq}} = \beta^{\mathrm{cm}}(\rho, 1).$$

The case with $\tau = \rho = 0$ corresponds to random matching allocation of inputs:

$$\beta^{\mathrm{rm}} = \beta^{\mathrm{cm}}(0,0) = \int_x \int_w g(w,x) \mathrm{d}F_W(w) \mathrm{d}F_X(x).$$

The cases with $(\tau = 0, \rho \to 1)$ and $(\tau = 0, \rho \to -1)$ correspond respectively to the perfect positive and negative assortative matching allocations:

$$\beta^{\mathrm{pam}} = \lim_{\rho \to 1} \beta^{\mathrm{cm}}(\rho, 0), \qquad \text{and} \quad \beta^{\mathrm{nam}} = \lim_{\rho \to -1} \beta^{\mathrm{cm}}(\rho, 0).$$

More generally, with $\tau = 0$ we allocate the inputs using a normal copula in a way that allows for arbitrary correlation between W and X indexed by the parameter ρ. It would be conceptually straightforward to use other copulas.

3.3 Local Measures of Complementarity

A potential disadvantage of the correlated matching reallocation family of estimands $\beta^{\mathrm{cm}}(\rho, \tau)$, including the focal allocations β^{pam} and β^{nam} is that the support requirements that allow for precise estimation may be difficult to satisfy in practice. This is particularly relevant for allocations 'distant' from the status quo. For example, if the *status quo* is characterized by a high degree of correlation between the inputs, evaluating the effect of allocations with a small, or even negative, correlation between inputs, such as random matching, or negative assortative matching, can be difficult because such allocations rely on knowledge of the production funhction at pairs of input values (W, X) that are infrequently seen in the data. For this reason a measure of local (close to the status quo) complementarity between W and X would be valuable. To this end we next characterize the expected effect on output associated with a 'small' increase toward either positive or negative assortative matching. Such estimands may also be informative regarding the effects of "modest" policies that stay close to the *status quo*. The resulting estimand forms the basis of a simple test for local efficiency of the status quo allocation. We derive this local measure by considering matching on a family of transformations of X_i and W_i, indexed by a scalar parameter λ, where for some values of λ the matching is on W_i (corresponding to the status quo), and for other values of λ the matching is on X_i or $-X_i$, corresponding to positive and negative assortative matching respectively. We then focus on the derivative of the expected outcomes from matching on this family of transformations, evaluated at the value of λ that corresponds to the status quo.

For technical reasons, and to be consistent with the subsequent formal statistical analysis in Section 4 of the previously discussed estimands β^{pam} and β^{nam}, we assume that the support of X_i is the interval $[x_l, x_u]$, with midpoint $x_m = (x_u + x_l)/2$, and similarly that the support of W_i is the interval $[w_l, w_u]$, with midpoint $w_m = (w_u + w_l)/2$. Without loss of generality we will assume that $x_l = 0$, $x_m = 1/2$, $x_u = 1$, $w_l = 0$, $w_m = 1/2$, and $w_u = 1$. To focus on the key conceptual issues we continue to ignore the presence of additional covariates V_i. First define a smooth function $d(w)$ that goes to zero at the boundary of the support of W_i:

$$d(w) = 1_{w>w_m} \cdot (w_u - w) + 1_{w \leq w_m} \cdot (w - w_l).$$

We implement our local reallocation as follows: for $\lambda \in [-1, 1]$, define the random variable U_λ as a transformation of (X, W):

$$U_\lambda = \lambda \cdot X \cdot d(W)^{1-|\lambda|} + (\sqrt{1-\lambda^2}) \cdot W.$$

This gives us a parametric transformation of (W, X) that moves smoothly between $W = U_0$ and $X = U_1$. Now we consider reallocations based on positive assortative matching on U_λ, for a range of values of λ, as a smooth way of moving from the *status quo* (matching on W) to positive assortative matching (matching on X). For general λ the average output associated with positive assortative matching on U_λ is given by the local reallocation

$$\beta^{\mathrm{lr}}(\lambda) = \mathbb{E}[g(F_W^{-1}(F_{U_\lambda}(U_\lambda)), X)]. \tag{3.11}$$

For $\lambda = 0$ and $\lambda = 1$ we have $U_\lambda = W$ and $U_\lambda = X$ respectively, and hence $\beta^{\mathrm{lr}}(0) = \beta^{\mathrm{sq}}$ and $\beta^{\mathrm{lr}}(1) = \beta^{\mathrm{pam}}$. Perfect negative assortative matching is also nested in this framework since

$$\Pr(-X \leq -x) = \Pr(X \geq x) = 1 - F_X(x),$$

and hence for $\lambda = -1$ we have $\beta^{\mathrm{lr}}(-1) = \beta^{\mathrm{nam}}$. Values of λ close to zero induce reallocations of W that are 'local' to the status quo, with $\lambda > 0$ and $\lambda < 0$ generating shifts toward positive and negative assortative matching respectively.

We focus on the effect of a small reallocation as our local measure of complementarity:

$$\beta^{\mathrm{lc}} = \frac{\partial \beta^{\mathrm{lr}}}{\partial \lambda}(0). \tag{3.12}$$

This local complementarity measure has two interesting alternative representations which are given in the following theorem. Before stating this result we introduce one assumption. This assumption is stronger than needed for this theorem, but its full force will be used later. The required values of the parameters in this assumption, p and q, will be specified in the theorems.

Assumption 3.1 (Distribution of Data)
(i) $(Y_1, W_1, X_1), (Y_2, W_2, X_2), \ldots, (Y_N, W_N, X_N)$ *are independent and identically distributed,*
(ii) The support of W *is* $\mathbb{W} = [w_l, w_u]$*, a compact subset of* $\mathbb{R}$*,*
(iii) the support of X *is* $\mathbb{X} = [x_l, x_u]$*, a compact subset of* $\mathbb{R}$*,*
(iv) the joint probability density function of W *and* X *is bounded and bounded away from zero, and* q *times continuously differentiable on* $\mathbb{W} \times \mathbb{X}$*,*
(v) $g(w, x)$ *is* q *times continuously differentiable with respect to* w *and* x *on* $\mathbb{W} \times \mathbb{X}$*,*
(vi) $\mathbb{E}[|Y_i|^p | X_i = x]$ *is bounded.*

The first representation is as the expected value of the conditional (on W) covariance of X and the returns to W, $g_W(w, x) = \frac{\partial g}{\partial w}(w, x)$, weighted by $d(W)$. The second representation is as a weighted average of the cross-derivative $\frac{\partial^2 g}{\partial w \partial x}(w, x)$. Formally:

Theorem 3.1 *Suppose Assumption 3.1 holds with $q \geq 2$. Then, β^{lc} has two equivalent representations:*

$$\beta^{\text{lc}} = \mathbb{E}\left[d(W) \cdot \text{Cov}\left(g_W\left(W, X\right), X | W\right)\right], \tag{3.13}$$

and,

$$\beta^{\text{lc}} = \mathbb{E}\left[\delta(W, X) \cdot \frac{\partial^2 g}{\partial w \partial x}\left(W, X\right)\right], \tag{3.14}$$

where the weight function $\delta(w, x)$ is non-negative and has the form

$$\delta(w, x) = d(w) \cdot \frac{F_{X|W}(x|w) \cdot (1 - F_{X|W}(x|w))}{f_{X|W}(x|w)} \cdot \Big(\mathbb{E}\left[X | X > x, W = w\right] - \mathbb{E}\left[X | X \leq x, W = w\right]\Big).$$

The proofs for the Theorems given in the body of the text are presented in Appendix C.

Representation (3.13), as we demonstrate below, suggests a straightforward nonparametric approach to estimating β^{lc}. Representation (3.14) is valuable for interpretation. Equation (3.14) demonstrates that a test of $H_0 : \beta^{\text{lc}} = 0$ is a test of the the null hypothesis of no complementarity or substitutability between W and X. If $\beta^{\text{lc}} > 0$, then in the 'vicinity of the *status quo*' W and X are complements; if $\beta^{\text{lc}} < 0$, they are substitutes. The precise meaning of the "vicinity of the *status quo*" is implicit in the form of the weight function $\delta(w, x)$.

Deviations of β^{lc} from zero imply that the *status quo* allocation does not maximize average outcomes. For $\beta^{\text{lc}} > 0$ a shift toward positive assortative matching will raise average outcomes, while for $\beta^{\text{lc}} < 0$ a shift toward negative assortative matching will do so. Theorem 3.1 therefore provides the basis of a test of the null hypothesis that the *status quo* allocation is locally efficient.

4 Estimation and inference with continuously-valued inputs

In this section we discuss estimation and inference. For ease of exposition we focus on the case without additional exogenous covariates. Allowing for these would complicate the notation, without adding much insight. The estimators are all weighted averages of (derivatives of) nonparametric estimators for the regression function. These are what Newey (1994) calls full and partial means and derivatives. First, in Section 4.1 we describe the nonparametric estimators for the regression functions. In order to deal with boundary issues we use develop a new nonparametric kernel estimator. Note that in Newey (1994) fixed trimming methods are used to deal with these boundary issues. These are less attractive here because they change the nature of the estimands. Next, in Section 3.1 we present estimators for the first pair of estimands, β^{pam} and β^{nam}. In Section 4.3 we discuss estimation and inference for β^{cm} (including β^{rm}), and in Section 4.4 we discuss β^{lc}. Estimation of and inference for the *status quo* allocation β^{sq} is straightforward, as this estimand is a simple expectation, estimated by a sample average.

4.1 Estimating the Production and Distribution Functions

For the two distributions functions we use the empirical distribution functions:

$$\hat{F}_W(w) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{W_i \leq w}, \qquad \text{and} \quad \hat{F}_X(x) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{X_i \leq x}.$$

For the inverse distribution functions we use the definition:

$$\hat{F}_W^{-1}(q) = \inf_{w \in \mathbb{W}} \mathbf{1}_{\hat{F}_W(w) \geq q}, \qquad \text{and} \quad \hat{F}_X^{-1}(q) = \inf_{x \in \mathbb{X}} \mathbf{1}_{\hat{F}_X(x) \geq q}.$$

The estimands we consider in this paper depend on the regression function $g(w,x)$ (in the case of β^{pam}, β^{nam}, and β^{cm}), or its derivative in the case of β^{lc}. The latter also depends on the regression function $m(w)$, defined as

$$m(w) = \mathbb{E}[X|W = w]. \tag{4.15}$$

In order to estimate these objects, we need estimators for the regression functions $m(w)$ and $g(w,x)$, and the derivative $g_W(w,x)$. Write the regression function as

$$g(w,x) = \mathbb{E}\left[Y|W = w, X = x\right] = \frac{h_2(w,x)}{h_1(w,x)},$$

where

$$h_1(w,x) = f_{WX}(w,x), \qquad \text{and} \quad h_2(w,x) = g(w,x) \cdot f_{WX}(w,x).$$

To simplify the following discussion, we rewrite $h_1(w,x)$ and $h_2(w,x)$ as

$$h_m(w,x) = \mathbb{E}\left[\tilde{Y}_m|W = w, X = x\right] \cdot f_{WX}(w,x), \tag{4.16}$$

for $m = 1,2$, where $\tilde{Y} = (\tilde{Y}_1\ \tilde{Y}_2)'$, with $\tilde{Y}_1 = 1$, $\tilde{Y}_2 = Y_i$.

We focus on estimators for $h_m(w,x)$, and use those to estimate $g(w,x)$ and its derivatives. The standard Nadaraya-Watson (NW) estimator for $h_m(w,x)$ is, for some bivariate kernel $K(\cdot,\cdot)$,

$$\hat{h}_{\text{nw},m}(w,x) = \frac{1}{N \cdot b^2} \sum_{i=1}^{N} \tilde{Y}_{im} \cdot K\left(\frac{W_i - w}{b}, \frac{X_i - x}{b}\right). \tag{4.17}$$

We denote the resulting nonparametric estimator by $\hat{g}(w,x)$. We estimate the derivative of $g(w,x)$ with respect to w by taking the derivative of the NW estimator of $g(w,x)$.

Because the support of (W_i, X_i) is assumed to be bounded, we have to deal with boundary bias of the kernel estimators. Because we also need bias reduction by using higher order kernels we adopt the Nearest Interior Point (NIP) estimator of Imbens and Ridder (2009). This estimator divides, for given bandwidth b, the support of (W,X) into an internal region and a boundary region. On the internal region the uniform convergence of the standard NW kernel estimators holds, but the estimators must be modified on the boundary region of the support. The NIP estimator coincides with the usual NW kernel estimator on the internal set, but it is equal to a polynomial on the boundary set. The coefficients of this polynomial are those of a Taylor series expansion in a point of the internal set.

To obtain a compact expression for the NIP estimator we adopt the following notation. The vector $z = (w\ x)'$ has $L = 2$ components. Some of the results below are stated for general L, although we only use the case with $L = 2$. Let $\mathbb{Z} = \mathbb{W} \times \mathbb{X}$ denote the (compact) support of Z. Let λ denote an L vector of nonnegative integers, with $|\lambda| = \sum_{l=1}^{L} \lambda_l$, and $\lambda! = \prod_{l=1}^{L} \lambda_l!$. For L

vectors of nonnegative integers λ and μ let $\mu \leq \lambda$ be equivalent to $\mu_l \leq \lambda_l$ for all $l = 1, \ldots, L$, and define

$$\begin{pmatrix} \lambda \\ \mu \end{pmatrix} = \frac{\lambda!}{\mu!(\lambda-\mu)!} = \prod_{l=1}^{L} \frac{\lambda_l!}{\mu_l!(\lambda_l-\mu_l)!} = \prod_{l=1}^{L} \begin{pmatrix} \lambda_l \\ \mu_l \end{pmatrix}.$$

For L vectors λ and z, let $z^\lambda = \prod_{l=1}^{L} z_l^{\lambda_l}$. As shorthand for partial derivatives of some function g we use $g^{(\lambda)}(z)$:

$$g^{(\lambda)}(z) = \frac{\partial g^{|\lambda|}}{\partial z^\lambda}(z).$$

The definition of the internal region depends on the support of the kernel. Let $K : \mathbb{R}^L \mapsto \mathbb{R}$ denote the kernel function. We will assume that $K(u) = 0$ for $u \notin \mathbb{U}$ with $\mathbb{U}$ compact, and $K(u)$ bounded. For the bandwidth b define the internal set of the support $\mathbb{Z}$ as the subset of $\mathbb{Z}$ such that all $\tilde{z}$ with a distance of up to b times the support of the kernel from z are also in $\mathbb{Z}$

$$\mathbb{Z}_b^I = \left\{ z \in \mathbb{Z} \middle| \left\{ \tilde{z} \in \mathbb{R}^L \middle| \frac{z-\tilde{z}}{b} \in \mathbb{U} \right\} \subset \mathbb{Z} \right\}. \tag{4.18}$$

This is a compact subset of the interior of $\mathbb{Z}$ that contains all points that are sufficiently far away from the boundary that the standard kernel density estimator at those points is not affected by any potential discontinuity of the density at the boundary. If $\mathbb{U} = [-1,1]^L$ and $\mathbb{Z} = \bigotimes_{l=1}^{L} [z_{ll}, z_{ul}]$, we have $\mathbb{Z}_b^I = \bigotimes_{l=1}^{L} [z_{ll} + b, z_{ul} - b]$.[3] The complement of the interior region is the boundary region:

$$\mathbb{Z}_b^B = \mathbb{Z}/\mathbb{Z}_b^I = \left\{ z \in \mathbb{Z} \middle| \exists \tilde{z} \notin \mathbb{Z} \text{ s.t. } \frac{z-\tilde{z}}{b} \in \mathbb{U} \right\}. \tag{4.19}$$

Next, we need to develop some notation for Taylor series approximations. Define for a given, q times differentiable function $g : \mathbb{Z} \mapsto \mathbb{R}$, a point $r \in \mathbb{R}^L$ and an integer $s \leq q$, the $(s-1)$-th order polynomial function $t : \mathbb{Z} \mapsto \mathbb{R}$ based on the Taylor series, expansion of order $s-1$, of $g(z)$ around the point $r \in \mathbb{Z}$:

$$t(z; g, r, s) = \sum_{j=0}^{s-1} \sum_{|\lambda|=j} \frac{1}{\lambda!} \cdot g^{(\lambda)}(r) \cdot (z-r)^\lambda. \tag{4.20}$$

Because the function $g(z)$ is $q \geq s$ times continuously differentiable on $\mathbb{Z}$, the remainder term in the Taylor series expansion is

$$g(z) - t(z, g, r, s) = \sum_{|\lambda|=s} \frac{1}{\lambda!} g^{(\lambda)}(\overline{r}(s)) \cdot (z-r)^\lambda.$$

with $\overline{r}(z)$ intermediate between z and r. Because $\mathbb{Z}$ is compact, and the the s-th order continuous, the sth order derivative must be bounded, and therefore this remainder term is bounded

[3]The set $[-1,1]^L$ is the set of L vectors with components that are between -1 and 1. The set $\bigotimes_{l=1}^{L} [z_{ll}, z_{ul}]$ is the set of L vectors with l-th component between z_{ll} and z_{ul}.

by $C|z-r|^s$. For the NIP estimator we use this Taylor series expansion around a point that depends on z and the bandwidth. Specifically, we take the expansion around $r_b(z)$, the projection on the internal region

$$r_b(z) = \text{argmin}_{r \in \mathbb{Z}_b^I} \|z - r\| \tag{4.21}$$

With this preliminary discussion, the NIP estimator of order s of $h_m(z)$ can be defined as:

$$\hat{h}_{m,\text{nip},s}(z) = \sum_{j=0}^{p-1} \sum_{|\lambda|=j} \frac{1}{\lambda!} \cdot \hat{h}_{m,\text{nw}}^{(\mu)}(r_b(z))(z - r_b(z))^{\mu} \tag{4.22}$$

with $\hat{h}_{m,\text{nw}}^{(\lambda)}$ the λ-th derivative of the kernel estimator $\hat{h}_{m,\text{nw}}$. For values of z in the internal region $\mathbb{Z}_b^I$, the NIP estimator is identical to the NW kernel estimator, $\hat{h}_{m,\text{nip},s}(z) = \hat{h}_{m,nw}(z)$. It is only in the boundary region that a $s-1$-th order Taylor series expansion is used to address the poor properties of the NS estimator in that region.

Now the NIP estimator for $g(w,x)$ is

$$\hat{g}_{\text{nip},s}(w,x) = \frac{\hat{h}_{2,\text{nip},s}(w,x)}{\hat{h}_{1,\text{nip},s}(w,x)}, \tag{4.23}$$

and the NIP estimator for the first derivative of $g(w,x)$ with respect to w is

$$\widehat{\frac{\partial g_{\text{nip},s}}{\partial w}}(w,x) = \frac{\frac{\partial}{\partial w}\hat{h}_{2,\text{nip},s}(w,x)}{\hat{h}_{1,\text{nip},s}(w,x)} - \frac{\hat{h}_{2,\text{nip},s}(w,x) \cdot \frac{\partial}{\partial w}\hat{h}_{1,\text{nip},s}(w,x)}{\left(\hat{h}_{1,\text{nip},s}(w,x)\right)^2}. \tag{4.24}$$

Unlike the NW kernel estimator, the NIP estimator is uniformly consistent. Its properties are discussed in more detail in Imbens and Ridder (2009). A formal statement of the relevant properties for our discussion is given in Lemmas A.9, A.10, and A.11, and Theorems A.1, A.2, and A.3 in Appendix A.

In the remainder of the paper we drop the subscripts from the estimator of the regression function. Unless specifically mentioned, $\hat{g}(w,x)$ will be used to denote $\hat{g}_{\text{nip},s}(w,x)$, for s equal to the order of the kernel, with its value stated in the Lemmas and Theorems.

Next we introduce two more assumptions. Assumption 4.1 describes the properties of the kernel function, and Assumption 4.2 gives the rate on the bandwidth. Before stating the next assumption we need to introduce a class of restrictions on kernel functions. The restrictions govern the rate at which the kernel, which is assumed to have compact support), goes to zero on the boundary of its support. This property allows us to deal with some of the boundary issues. Such properties have previously been used in, for example, Powell, Stock and Stoker (1989).

Definition 4.1 (DERIVATIVE ORDER OF A KERNEL) *A kernel function $K : \mathbb{U} \mapsto \mathbb{R}$ is of derivative order d, if, for all u in the boundary of the set $\mathbb{U}$, and all $|\lambda| \leq d-1$,*

$$\lim_{v \to u} \frac{\partial^{\lambda}}{\partial u^{\lambda}} K(v) = 0.$$

Assumption 4.1 (Kernel)
(i) $K : \mathbb{R}^L \mapsto \mathbb{R}$, *with* $K(u) = \prod_{l=1}^{L} \mathcal{K}(u_l)$,
(ii) $K(u) = 0$ *for* $u \notin \mathbb{U}$, *with* $\mathbb{U} = [-1, 1]^L$,
(iii) $K(\cdot)$ *is* r *times continuously differentiable, with the* r*-th derivative bounded on the interior of* $\mathbb{U}$,
(iv) $K(\cdot)$ *is a kernel of order* s*, so that* $\int_{\mathbb{U}} K(u)du = 1$ *and* $\int_{\mathbb{U}} u^\lambda K(u)du = 0$ *for all* λ *such that* $0 < |\lambda| < s$*, for some* $s \geq 1$,
(v) K *is a kernel of derivative order* d.

We refer a kernel satisfying Assumption 4.2 as a derivative kernel of order (s, d).

Assumption 4.2 (Bandwidth) *The bandwidth* $b_N = N^{-\delta}$ *for some* $\delta > 0$.

4.2 Estimation and Inference for $\widehat{\beta}^{\text{pam}}$ and $\widehat{\beta}^{\text{nam}}$

In this section we introduce the estimators for β^{pam} and β^{nam} and present results on the large sample properties of the estimators. We estimate β^{pam} and β^{nam} by substituting nonparametric estimators for the unknown functions $g(w, x)$, $F_W(w)$, and $F_X(x)$:

$$\widehat{\beta}^{\text{pam}} = \frac{1}{N} \sum_{i=1}^{N} \hat{g} \left(\hat{F}_W^{-1}(\hat{F}_X(X_i)), X_i \right), \tag{4.25}$$

and

$$\widehat{\beta}^{\text{nam}} = \frac{1}{N} \sum_{i=1}^{N} \hat{g} \left(\hat{F}_W^{-1}(1 - \hat{F}_X(X_i)), X_i \right). \tag{4.26}$$

It is straightforward to demonstrate consistency for these estimators. The nonparametric estimators $\hat{g}$, $\hat{F}_W$, and $\hat{F}_X$ are uniformly consistent under our assumptions, and consistency of $\hat{\beta}^{\text{pam}}$ follows directly from that. It is more difficult to derive the large sample distributions for these estimators. There are four components to their asymptotic approximations. Here we discuss the decomposition for $\hat{\beta}^{\text{pam}}$. A similar argument holds for $\hat{\beta}^{\text{nam}}$. In both cases the first component corresponds to the estimation error in $g(w, x)$. This component converges at a rate slower than the regular parametric (root-N) rate. This is because we estimate in the first stage a nonparametric regression function with more arguments than we average over in the second stage. As a result $\hat{\beta}^{\text{pam}}$ (and $\hat{\beta}^{\text{nam}}$) is a partial (as opposed to a full) mean in the terminology of Newey (1994). The other three terms converge faster, at the regular root$-N$ rate. There is one term each corresponding to the estimation error in $F_W(w)$ and $F_X(x)$ respectively, and one corresponding to the difference between the average of $g(F_W^{-1}(F_X(X_i)), X_i)$ and its expectation. In describing the large sample properties we include all four of these terms, which leaves a remainder that is $o_p(N^{-1/2})$. In principle one could ignore the three terms of order $O_p(N^{-1/2})$, since they will get dominated by the term describing the uncertainty stemming from estimation of $g(w, x)$, but including the additional terms is likely to lead to more accurate confidence intervals. We provide evidence for this in the simulations in Section 5.

In order to describe the formal properties of the estimator $\hat{\beta}^{\text{pam}}$ it is useful to introduce notation for an intermediate quantity, and some additional functions. Define the average with the true regression function $g(w, x)$ (but still the estimated distribution functions $\hat{F}_W$ and $\hat{F}_X$),

$$\tilde{\beta}^{\text{pam}} = \frac{1}{N} \sum_{i=1}^{N} g \left(\hat{F}_W^{-1} \left(\hat{F}_X(X_i) \right), X_i \right), \tag{4.27}$$

so that we can write $\hat{\beta}^{\text{pam}} - \beta^{\text{pam}} = (\hat{\beta}^{\text{pam}} - \tilde{\beta}^{\text{pam}}) + (\tilde{\beta}^{\text{pam}} - \beta^{\text{pam}})$. Then the first term $\tilde{\beta}^{\text{pam}} - \beta^{\text{pam}} = O_p(N^{-1/2})$, and the second term $\hat{\beta}^{\text{pam}} - \tilde{\beta}^{\text{pam}} = O_p(N^{-1/2}b_N^{-1/2})$. Recall the notation for the derivative of $g(w,x)$ with respect to w,

$$g_W(w,x) = \frac{\partial g}{\partial w}(w,x),$$

and define

$$q^{\text{pam}}(w,x) = \frac{g_W(F_W^{-1}(F_X(x)),x)}{f_W(F_W^{-1}(F_X(x)))} \cdot \left(\mathbf{1}_{F_W(w)\leq F_X(x)} - F_X(x)\right),$$

$$\psi_W^{\text{pam}}(w) = \mathbb{E}\left[q^{\text{pam}}(w,X)\right],$$

$$r^{\text{pam}}(x,z) = \frac{g_W(F_W^{-1}(F_X(z)),z)}{f_W(F_W^{-1}(F_X(z)))} \cdot \left(\mathbf{1}_{x\leq z} - F_X(z)\right),$$

and

$$\psi_X^{\text{pam}}(x) = \mathbb{E}\left[r^{\text{pam}}(x,X)\right].$$

Theorem 4.1 (Large Sample Properties of $\hat{\beta}^{\text{pam}}$)
Suppose Assumptions 2.1, 3.1, 4.1, and 4.2 hold, with $q \geq 2s+1$, $r \geq s+3$, $p \geq 4$, $d \geq s-1$, *and* $1/(2s) < \delta < 1/8$. *Then*

$$\sqrt{N} \cdot \begin{pmatrix} b_N^{1/2}\left(\hat{\beta}^{\text{pam}} - \tilde{\beta}^{\text{pam}}\right) \\ \tilde{\beta}^{\text{pam}} - \beta^{\text{pam}} \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Omega_{11}^{\text{pam}} & 0 \\ 0 & \Omega_{22}^{\text{pam}} \end{pmatrix}\right),$$

where

$$\Omega_{11}^{\text{pam}} = \mathbb{E}\Bigg[\sigma^2\left(F_W^{-1}\left(F_X(X)\right), X_i\right) \cdot \int_{u_1}\left(\int_{u_2} K\left(u_1 + \frac{f_X(X)}{f_W\left(F_W^{-1}\left(F_X(X)\right)\right)} \cdot u_2, u_2\right) du_2\right)^2 du_1 \tag{4.28}$$

$$\cdot f_{W|X}\left(F_W^{-1}\left(F_X(X)\right)|X\right)\Bigg],$$

and

$$\Omega_{22}^{\text{pam}} = \mathbb{E}\left[\left(\psi_W^{\text{pam}}(W) + \psi_X^{\text{pam}}(X) + g(F_W^{-1}(F_X(X)),X) - \beta^{\text{pam}}\right)^2\right].$$

In the expression for the large sample variance, ψ_X^{pam} captures the uncertainty resulting from estimation of $F_X(x)$, and ψ_W^{pam} captures the uncertainty resulting from estimation of $F_W(w)$.

Note that the component of the variance that captures the uncertainty from estimation of $g(w,x)$, Ω_{11}^{pam}, depends on the kernel in a way that involves the distribution of the data. Often when one estimates nonparametric functionals at parametric rates, the dependence on the kernel vanishes asymptotically if one undersmoothes. Here the kernel shows up in the leading term. This is also the case in the discussion of partial means in Newey (1994).

Suppose we wish to construct a 95% confidence interval for β^{pam}. In that case we approximate the variance of $\hat{\beta}^{\text{pam}} - \beta^{\text{pam}}$ by $\hat{\mathbb{V}} = \hat{\Omega}_{11}^{\text{pam}} \cdot N^{-1} \cdot b_N^{-1} + \hat{\Omega}_{22}^{\text{pam}} \cdot N^{-1}$, using suitable plug-in estimators $\hat{\Omega}_{11}^{\text{pam}}$ and $\hat{\Omega}_{22}^{\text{pam}}$, and construct the confidence interval as $(\hat{\beta}^{\text{pam}} - 1.96 \cdot \sqrt{\hat{\mathbb{V}}}, \hat{\beta}^{\text{pam}} + 1.96 \cdot \sqrt{\hat{\mathbb{V}}})$. Although the first term in $\hat{\mathbb{V}}$ will dominate the second term in large samples, in finite samples the second term may still be important. We shall see this in the simulations in Section 5.

Similar results hold for β^{nam}, with some appropriately redefined concepts. Define

$$\tilde{\beta}^{\text{nam}} = \frac{1}{N} \sum_{i=1}^{N} g\left(\hat{F}_W^{-1}\left(1 - \hat{F}_X(X_i)\right), X_i\right), \tag{4.29}$$

$$q^{\text{nam}}(w,x) = \frac{g_W(F_W^{-1}(1-F_X(x)),x)}{f_W(F_W^{-1}(1-F_X(x)))} \cdot \left(\mathbf{1}_{F_W(w)\leq F_X(x)} - F_X(x)\right),$$

$$\psi_W^{\text{nam}}(w) = \mathbb{E}\left[q_{WX}^{\text{nam}}(w,X)\right],$$

$$r^{\text{nam}}(x,z) = \frac{g_W(F_W^{-1}(1-F_X(z)),z)}{f_W(F_W^{-1}(1-F_X(z)))} \cdot \left(\mathbf{1}_{x\leq z} - F_X(z)\right),$$

and

$$\psi_X^{\text{nam}}(x) = \mathbb{E}\left[r_{XZ}^{\text{nam}}(x,X)\right].$$

Theorem 4.2 (Large Sample Properties of $\hat{\beta}^{\text{nam}}$)
Suppose Assumptions 2.1, 3.1, 4.1, and 4.2 hold, with $q \geq 2s+1$, $r \geq s+3$, $p \geq 4$, $d \geq s-1$, and $1/(2s) < \delta < 1/8$. Then

$$\sqrt{N} \cdot \begin{pmatrix} b_N^{1/2}\left(\hat{\beta}^{\text{nam}} - \tilde{\beta}^{\text{nam}}\right) \\ \tilde{\beta}^{\text{nam}} - \beta^{\text{nam}} \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Omega_{11}^{\text{nam}} & 0 \\ 0 & \Omega_{22}^{\text{nam}} \end{pmatrix}\right),$$

where

$$\Omega_{11}^{\text{nam}} = \mathbb{E}\Bigg[\sigma^2\left(F_W^{-1}\left(1-F_X(X)\right),X\right) \cdot \int_{u_1}\left(\int_{u_2} K\left(u_1 + \frac{f_X(X)}{f_W\left(F_W^{-1}\left(1-F_X(X)\right)\right)} \cdot u_2, u_2\right)\right)^2 du_1$$
$$\cdot f_{W|X}\left(F_W^{-1}\left(1-F_X(X)\right)|X\right)\Bigg],$$

and

$$\Omega_{22}^{\text{nam}} = \mathbb{E}\left[\left(\psi_W^{\text{nam}}(W) + \psi_X^{\text{nam}}(X) + g(W,X) - \beta^{\text{nam}}\right)^2\right].$$

4.3 Estimation and Inference for $\beta^{\text{cm}}(\rho,\tau)$

The starting point for estimation of β^{cm} is the representation of $\beta^{\text{cm}}(\rho,0)$ in equation (3.9):

$$\beta^{\text{cm}}(\rho,0) = \int_{w,x} g(w,x) \frac{\phi_c\left(\Phi_c^{-1}(F_W(w)), \Phi_c^{-1}(F_X(x)); \rho\right)}{\phi_c\left(\Phi_c^{-1}(F_W(w))\right)\phi_c\left(\Phi_c^{-1}(F_X(x))\right)} f_W(w) f_X(x) \mathrm{d}w\mathrm{d}x.$$

Note that this expression is an integral over the product of the marginal pdf's of W and X, not the joint. We estimate this by replacing the integrals with sums over the two empirical distribution functions to get analog estimator

$$\widehat{\beta}^{\mathrm{cm}}(\rho,0) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{g}(W_i,X_j)\frac{\phi_c\left(\Phi_c^{-1}(\hat{F}_W(W_i)),\Phi_c^{-1}(\hat{F}_X(X_j));\rho\right)}{\phi_c\left(\Phi_c^{-1}(\hat{F}_W(W_i))\right)\phi_c\left(\Phi_c^{-1}(\hat{F}_X(X_j))\right)}.$$

This estimator would be a standard second order V statistic if we had the true regression function and the true distribution functions. The dependence on the esimated regression function complicates its analysis.

Observe that if $\rho = 0$ (random matching) the ratio of densities on the right hand side is equal to 1, so that

$$\hat{\beta}^{\mathrm{rm}} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{g}(W_i,X_j).$$

For $\tau > 0$, the $\beta^{\mathrm{cm}}(\rho,\tau)$ estimand is a convex combination of average output under the status quo and a correlated matching allocation. The corresponding sample analog is

$$\widehat{\beta}^{\mathrm{cm}}(\rho,\tau) = \tau\cdot\widehat{\beta}^{\mathrm{sq}} + (1-\tau)\cdot\widehat{\beta}^{\mathrm{cm}}(\rho,0),$$

where $\widehat{\beta}^{\mathrm{sq}} = \overline{Y} = \sum_{i=1}^{N} Y_i/N$, the average outcome. This estimator is linear in the nonparametric regression estimator $\hat{g}$ and nonlinear in the empirical CDFs of X and W.

A useful and insightful representation of $\beta^{\mathrm{cm}}(\rho,0)$ is as an average of partial means (c.f., Newey 1994). This representation provides intuition both about the structure of the estimand as well as its large sample properties. Fixing W at $W = w$, but averaging over the distribution of X we get the partial mean:

$$\eta(w) = \mathbb{E}_X\left[g(w,X)\cdot d(w,X)\right], \tag{4.30}$$

where

$$d(w,x) = \frac{\phi_c(\Phi_c^{-1}(F_W(w)),\Phi_c^{-1}(F_X(x));\rho)}{\phi_c(\Phi_c^{-1}(F_W(w)))\phi_c(\Phi_c^{-1}(F_X(x)))}. \tag{4.31}$$

Observe that (4.30) is a weighted averaged of the production function over the distribution of X holding the value of the input to be reallocated W fixed at $W = w$. The weight function $d(w,X)$ depends upon the truncated normal cupola. In particular, the weights give greater emphasis to realizations of $g(w,X)$ that are associated with values of X that will be assigned a value of W close to w as part of the correlated matching reallocation. Thus (4.30) equals the average post-reallocation output for those firms being assigned $W = w$. To give a concrete example (4.30) is the post-reallocation expected achievement of those classrooms that will be assigned a teacher of quality $W = w$.

Equation (4.30) also highlights the value of using the truncated normal copula. Doing so ensures that the denominators of the copula 'weights' in (4.30) are bounded from zero. The copula weights thus play the role similar to fixed trimming weights used by Newey (1994).

If we average these partial means over the marginal distribution of W we get $\beta^{\mathrm{cm}}(\rho,0)$, since

$$\beta^{\mathrm{cm}}(\rho,0) = \mathbb{E}_W\left[\eta(W)\right],$$

yielding average output under the correlated matching reallocation.

¿From the above discussion it is clear that our correlated matching estimator can be viewed as a semiparametric two-step method-of-moments estimator with a moment function of

$$m(Y, W, \beta^{\mathrm{cm}}(\rho,\tau), \eta(W)) = \tau Y + (1-\tau)\,\eta(W) - \beta^{\mathrm{cm}}(\rho,\tau).$$

Our estimator, $\widehat{\beta}^{\mathrm{cm}}(\rho,\tau)$, is the feasible GMM estimator based upon the above moment function after replacing the partial mean ($\eta(w)$ defined in (4.30)) with a consistent estimate. While the above representation is less useful for deriving the asymptotic properties of $\widehat{\beta}^{\mathrm{cm}}(\rho,\tau)$ it does provide some insight as to why we are able to achievement parametric rates of convergence.

To state the large sample properties of the correlated matching estimator we need some additional notation. Define:

$$\begin{aligned}
e_W(w,x) &= \frac{\rho\phi_c(\Phi_c^{-1}(F_W(w)), \Phi_c^{-1}(F_X(x)); \rho)}{(1-\rho^2)\phi_c(\Phi_c^{-1}(F_W(w)))^2\phi_c(\Phi_c^{-1}(\widehat{F}_X(x)))} \times \\
&\quad \left[\Phi_c^{-1}(F_X(x)) - \rho\Phi_c^{-1}(F_W(w))\right], \\
e_X(w,x) &= \frac{\rho\phi_c(\Phi_c^{-1}(F_W(w)), \Phi_c^{-1}(F_X(x)); \rho)}{(1-\rho^2)\phi_c(\Phi_c^{-1}(F_W(w)))\phi_c(\Phi_c^{-1}(\widehat{F}_X(x)))^2} \times \\
&\quad \left[\Phi_c^{-1}(F_W(w)) - \rho\Phi_c^{-1}(F_X(X_k))\right], \\
\omega^{\mathrm{cm}}(w,x) &= \frac{\phi_c\left(\Phi_c^{-1}(F_W(w)), \Phi_c^{-1}(F_X(x)); \rho\right)}{\phi_c\left(\Phi_c^{-1}(F_W(w))\right)\phi_c\left(\Phi_c^{-1}(F_X(x))\right)}, && (4.32)\\
\psi_0^{\mathrm{cm}}(y,w,x) &= && (4.33)\\
&\quad (\mathbb{E}\left[g(W,x)\cdot\omega(W,x)\right] - \beta^{\mathrm{cm}}(\rho,0)) + (\mathbb{E}\left[g(w,X)\cdot\omega(w,X)\right] - \beta^{\mathrm{cm}}(\rho,0)), \\
\psi_g^{\mathrm{cm}}(y,w,x) &= \frac{f_W(w)\cdot f_X(x)}{f_{WX}(w,x)}(y - g(w,x))\omega(w,x), && (4.34)\\
\psi_W^{\mathrm{cm}}(y,w,x) &= \int\int g(s,t)e_W(s,t)\left(\mathbf{1}_{w\le s} - F_W(s)\right) f_W(s)\, f_X(t)\,\mathrm{d}s\mathrm{d}t, && (4.35)
\end{aligned}$$

and

$$\psi_X^{\mathrm{cm}}(y,w,x) = \int\int g(s,t)e_X(s,t)\left(\mathbf{1}_{x\le t} - F_X(t)\right) f_W(s)\, f_X(t)\,\mathrm{d}s\mathrm{d}t. \qquad (4.36)$$

Theorem 4.3 *Suppose Assumptions 2.1, 3.1, 4.1, and 4.2 hold with* $q \geq 2s-1$, $r \geq s+1$, $p \geq 3$, $d \geq s-1$, *and* $(1/2s) < \delta < 1/4$, *then*

$$\widehat{\beta}^{\mathrm{cm}}(\rho,\tau) \xrightarrow{p} \beta^{\mathrm{cm}}(\rho,\tau)$$

and

$$\sqrt{N}(\widehat{\beta}^{\mathrm{cm}}(\rho,\tau) - \beta^{\mathrm{cm}}(\rho,\tau)) \xrightarrow{d} \mathcal{N}(0, \Omega^{\mathrm{cm}}),$$

where

$$\Omega^{\mathrm{cm}} = \mathbb{E}\left[\left(\tau(Y-\beta^{\mathrm{sq}}) + (1-\tau)\,\psi^{\mathrm{cm}}(Y,W,X)\right)^2\right],$$

and

$$\psi^{\mathrm{cm}}(y,w,x) = \psi_0^{\mathrm{cm}}(y,w,x) + \psi_g^{\mathrm{cm}}(y,w,x) + \psi_W^{\mathrm{cm}}(y,w,x) + \psi_X^{\mathrm{cm}}(y,w,x). \qquad (4.37)$$

Note that this estimator is root_N consistent, unlike $\hat{\beta}^{\text{pam}}$ and $\hat{\beta}^{\text{nam}}$.

If there was no estimation error in $\hat{g}(w,x)$, $\hat{F}_W(w)$, and $\hat{F}_X(x)$, the estimator would be root$-N$ consistent with normalized asymptotic variance equal to $[\psi_0^{\text{cm}}(Y_i, W_i, X)i)^2]$. The remaining terms in the influence function, $\psi_W^{\text{cm}}(y,w,x)$, $\psi_X^{\text{cm}}(y,w,x)$, and $\psi_g^{\text{cm}}(y,w,x)$, capture the uncertainty coming from estimation of $F_W(w)$, $F_X(x)$, and $g(w,x)$ respectively.

4.4 Estimation and Inference for β^{lc}

Estimation of β^{lc} proceeds in two steps. First we estimate $g(w,x) = \mathbb{E}[Y|W=w, X=x]$ (and its derivative with respect to w) and $m(w) = \mathbb{E}[X|W=w]$ using kernel methods as in Section 4.1. In the second step we estimate β^{lc} by method-of-moments using the sample analog of the moment condition

$$\mathbb{E}\left[\frac{\partial}{\partial w} g(W,X) \cdot d(W) \cdot (X - m(W)) - \beta^{\text{lc}}\right] = 0.$$

Thus,

$$\hat{\beta}^{\text{lc}} = \frac{1}{N}\sum\nolimits_{i=1}^{N} \frac{\partial}{\partial w}\widehat{g}(W_i, X_i) \cdot d(W_i) \cdot (X_i - \widehat{m}(W_i)). \tag{4.38}$$

Define

$$\begin{aligned}\psi_g^{\text{lc}}(y,w,x) = & -\frac{1}{f_{W,X}(w,x)} \frac{\partial f_{W,X}(w,x)}{\partial W} d(w)\,(y - g(w,x))\,(x - m(w)) \\ & - \frac{\partial m}{\partial w}(w) d(w)\,(y - g(w,x)) \\ & - \frac{\partial d}{\partial w}(w)(y - g(w,x))(x - m(w)).\end{aligned}$$

and

$$\psi_m^{\text{lc}}(y,w,x) = \mathbb{E}\left[\frac{\partial}{\partial w} g(w,X) \middle| W = w\right] \cdot d(w) \cdot (x - m(w)).$$

As in the previous results, the ψ^{lc} are the influence functions, with $\psi_g^{\text{lc}}(y,w,x)$ capturing the uncertainty from estimation of $g(w,x)$, and $\psi_m^{\text{lc}}(y,w,x)$ capturing the uncertainty from estimation of $m(w)$.

The asymptotic properties of $\hat{\beta}^{\text{lc}}$ are summarized by Theorem 4.4.

Theorem 4.4 *Suppose Assumptions 2.1, 3.1, 4.1, and 4.2 hold with* $q \geq 2s+1$, $r \geq s+1$, $p \geq 4$, $d \geq s-1$, *and* $1/(2s) < \delta < 1/12$. *Then*

$$\hat{\beta}^{\text{lc}} \xrightarrow{p} \beta^{\text{lc}},$$

and

$$\sqrt{N}(\hat{\beta}^{\text{lc}} - \beta^{\text{lc}}) \xrightarrow{d} \mathcal{N}(0, \Omega^{\text{lc}}),$$

where

$$\Omega^{\text{lc}} = \mathbb{E}\left[\left(\left(\frac{\partial}{\partial w} g(W,X) \cdot d(W) \cdot (X - m(W)) - \beta^{\text{lc}}\right) + \psi_g^{\text{lc}}(Y,W,X) + \psi_m^{\text{lc}}(Y,W,X)\right)^2\right].$$

5 A Monte Carlo Study

To assess whether the asymptotic properties derived in Section 4 provide useful approximations to finite sample distributions, we carry out a small simulation study. In the interest of brevity we focus on β^{pam} and β^{lc}. We consider the following data generating process. The pair (W_i^*, X_i^*) are drawn from a bivariate normal distribution with both means equal to zero, both variances equal to one, and correlation coefficient equal to ρ. The two covariates W_i and X_i are then constructed as $W_i = 2 \cdot \Phi(W_i^*) - 1$ and $X_i = 2 \cdot \Phi(X_i^*) - 1$, so that both W_i and X_i have a uniform distribution on $[-1, 1]$, with potentially some correlation between them. The outcome is generated as

$$Y_i = W_i + X_i + W_i \cdot X_i + \varepsilon_i, \qquad \varepsilon_i | W_i, X_i \sim \mathcal{N}(0, 0.25).$$

Under this data generating process $\beta^{\text{pam}} = 0.3333$, irrespective of the value of the correlation between the covariates, ρ. The expected outcome under the current allocation is $\mathbb{E}[Y] = 0$ if $\rho = 0$, and $\mathbb{E}[Y] = 0.1212$ if $\rho = 0.5$. We fix the weight function $d(w)$ in the definition of the local complementarity measure at $d(w) = 1 - |w|$. The value of the local reallocation parameter is $\beta^{\text{lc}} = 0.1667$ if $\rho = 0$ and $\beta^{\text{lc}} = 0.1355$ if $\rho = 0.5$.

We estimate β^{pam} using equation (4.26), and β^{lc} using equation (4.38). We use a rectangular kernel on $[-1, 1]$, and local linear regression for estimating $g(w, x)$. The bandwidth for the regression estimation is choosen using cross-validation, after which we divide the bandwidth by two to ensure some undersmoothing. For density estimation we use the Silverman rule of thumb, modified for a uniform kernel. For univariate density estimation this leads to

$$b_N = 1.84 \cdot \sigma \cdot N^{-1/5}.$$

For estimating the bivariate density we use a bivariate uniform kernel, with the bandwidths in each direction equal to

$$b'_N = 1.84 \cdot \sigma \cdot N^{-1/6},$$

where the σ is estimated on the data, and so may differ in the two directions for the bivariate kernel.

We consider four designs, based on two sample sizes, $N = 200$ and $N = 1000$, and two dependence structures, $\rho = 0$ and $\rho = 0.5$. For both designs we calculate the two estimators $\hat{\beta}_{\text{pam}}$ and $\hat{\beta}_{\text{lc}}$, and their variances. In Table 1 we report some summary statistics from the simulations. We report the average and median bias, the standard deviation, the average of the standard errors, the root mean squared error, the median absolute error, and the coverage rates for the nominal 90 and 95% confidence intervals. The estimators appear to work fairly well. Note that the average standard error for $\hat{\beta}^{\text{lc}}$ is large relative to its standard deviation (the ratio is more than six). The reason is that occasionally the estimated standard error is very large. This happens with low probability, so the median standard error is not affected, and the coverage rate is also fine.

The estimators have a complicated structure, with the asymptotic distribution relying on a number of approximations. We further investigate these approximations in Table 2. Define

$$\hat{\beta}_g^{\text{pam}} = \frac{1}{N} \sum_{i=1}^{N} \hat{g}\left(F_W^{-1}\left(F_X(X_i)\right), X_i\right), \tag{5.39}$$

$$\hat{\beta}_W^{\text{pam}} = \frac{1}{N}\sum_{i=1}^{N} g\left(\hat{F}_W^{-1}\left(F_X(X_i)\right), X_i\right), \tag{5.40}$$

$$\hat{\beta}_X^{\text{pam}} = \frac{1}{N}\sum_{i=1}^{N} g\left(F_W^{-1}\left(\hat{F}_X(X_i)\right), X_i\right), \tag{5.41}$$

and

$$\overline{g}^{\text{pam}} = \frac{1}{N}\sum_{i=1}^{N} g\left(F_W^{-1}\left(F_X(X_i)\right), X_i\right). \tag{5.42}$$

Then, as stated formally in Appendix A, Lemma A.15,

$$\hat{\beta}^{\text{pam}} - \beta^{\text{pam}} =$$

$$\left(\hat{\beta}_g^{\text{pam}} - \overline{g}^{\text{pam}}\right) + \left(\hat{\beta}_W^{\text{pam}} - \overline{g}^{\text{pam}}\right) + \left(\hat{\beta}_X^{\text{pam}} - \overline{g}^{\text{pam}}\right) + \left(\overline{g}^{\text{pam}} - \beta^{\text{pam}}\right) + o_p\left(N^{-1/2}\right). \tag{5.43}$$

In Panel A of Table 2, we show the mean and standard deviation of $\hat{\beta}^{\text{pam}} - \beta^{\text{pam}}$, $\hat{\beta}_g^{\text{pam}} - \overline{g}^{\text{pam}}$, $\hat{\beta}_W^{\text{pam}} - \overline{g}^{\text{pam}}$, $\hat{\beta}_X^{\text{pam}} - \overline{g}^{\text{pam}}$, and the remainder term,

$$\text{rem} = \left(\hat{\beta}^{\text{pam}} - \beta^{\text{pam}}\right)$$

$$- \left(\left(\hat{\beta}_g^{\text{pam}} - \overline{g}^{\text{pam}}\right) + \left(\hat{\beta}_W^{\text{pam}} - \overline{g}^{\text{pam}}\right) + \left(\hat{\beta}_X^{\text{pam}} - \overline{g}^{\text{pam}}\right) + \left(\overline{g}^{\text{pam}} - \beta^{\text{pam}}\right)\right).$$

The results in Panel A of Table 2 suggest that the remainder term is indeed small compared to the terms that are taken into account in the asymptotic distribution. Moreover, the relative magnitude of the $O_p(N^{-1/2})$ terms are supportive of the fact that we take into account these terms, not just the leading term which is $N^{-1/2}b_N^{-1/2}$.

In the appendix we also show that

$$N^{1/2}b_N^{1/2} \cdot \left(\hat{\beta}_g^{\text{pam}} - \overline{g}^{\text{pam}}\right) \xrightarrow{d} \mathcal{N}\left(0, \Omega_{11}^{\text{pam}}\right), \tag{5.44}$$

where Ω_{11}^{pam} is defined in (4.28),

$$N^{1/2} \cdot \left(\hat{\beta}_W^{\text{pam}} - \overline{g}^{\text{pam}}\right) \xrightarrow{d} \mathcal{N}\left(0, \mathbb{E}\left[\psi_W^{\text{pam}}(W)^2\right]\right), \tag{5.45}$$

$$N^{1/2} \cdot \left(\hat{\beta}_X^{\text{pam}} - \overline{g}^{\text{pam}}\right) \xrightarrow{d} \mathcal{N}\left(0, \mathbb{E}\left[\psi_X^{\text{pam}}(X)^2\right]\right), \tag{5.46}$$

and

$$N^{1/2} \cdot \left(\overline{g}^{\text{pam}} - \beta^{\text{pam}}\right) \xrightarrow{d} \mathcal{N}\left(0, \mathbb{E}\left[\left(g(F_W^{-1}(F_X(X_i), X_i)) - \beta^{\text{pam}}\right)^2\right]\right). \tag{5.47}$$

To assess the normal approximations we calculate the t-statistics based on these distributions (the point estimates divided by estimates of the standard deviations), and report in Panel B of Table 2 summary statistics for these random variables, which should have approximate normal distributions. The summary statistics we report are averages, standard deviations, and tail frequencies. We find that the actual means, standard deviations, and tail frequencies are close to the nominal ones from the normal distribution.

6 Conclusions

In this paper we introduce a new class of estimands involving reallocation of inputs, and develop statistical methods for analyzing them. We consider a class of problems where a fixed set of inputs is reallocated to a fixed set of units. Whereas a large part of the literature in econometrics has focused on estimating the causal effects of changing inputs for all units, or for a subset of units, here we focus on reallocation rules that take into account resource constraints, by keeping the distribution of the inputs fixed. The effects we focus on depend critically on the degree of complementarity between inputs. We therefore follow a flexible nonparametric approach where the nature of the complementarity is not restricted to a parametric form. We propose estimators for the effects of various reallocation rules, and derive the asymptotic properties of these estimators.

References

Angrist, J. D. and A. B. Krueger. (1999). "Empirical Strategies in Labor Economics," *Handbook of Labor Economics 3A*: 1277 - 1366 (O. Ashenfelter and D. Card, Eds). New York: Elsevier Science.

Athey, S., and G. Imbens (2006), "Identification and Inference in Nonlinear Difference-In-Differences Models," *Econometrica*, 74(2): 431-497.

Athey, S., and S. Stern. (1998). "An Empirical Framework for Testing Theories About Complementarity in Organizational Design", *NBER Working Paper No. 6600.*

Becker, Gary S. and Kevin M. Murphy. (2000). *Social Economics: Market Behavior in a Social Environment.* Cambridge, MA: Harvard University Press.

Bhattacharya, D., (2008), "Inferring Optimal Peer Assignment from Experimental Data,"

forthcoming *Journal of the American Statistical Association*.

Brock, W. and S. Durlauf. (2001). "Interactions-based Models," *Handbook of Econometrics 5*: 3297 - 3380 (J. Heckman & E. Leamer, Eds.). Amsterdam: North-Holland.

Card, D., and A. Krueger. (1992). "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States, " *Journal of Political Economy* 100(1): 1-40.

Dehejia, R. (2005). "Program evaluation as a decision problem," *Journal of Econometrics* 125 (1-2): 141 - 173.

Flores, C. (2005). "Estimation of Dose-Response Functions and Optimal Doses with a Continuous Treatment," *Mimeo.*

Graham, B. (2008). "Identifying Social Interactions Through Conditional Variance Restrictions," *Econometrica* 76(3): 643-660.

Graham, B., G. Imbens, and G. Ridder,(2006a). "Complementarity and the Optimal Allocation of Inputs," *Mimeo.*

Graham, B., G. Imbens, and G. Ridder,(2006b). "Measuring the Average Outcome and Inequality Effects of Segregation in the Presence of Social Spillovers," *Mimeo.*

Glaeser, E., and J. Scheinkman, (2003), "Nonmarket Interactions," *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress* 1: 339 - 369, (M. Dewatripont el al, Eds.). Cambridge: Cambridge University Press.

Härdle, W. and T.M. Stoker. (1989). "Investigating smooth multiple regression by the method of average derivatives," *Journal of the American Statistical Association* 84 (408): 986 - 995.

Heckman, J., R. Lalonde, and J. Smith. (2000). "The Economics and Econometrics of Active Labor Markets Programs," *Handbook of Labor Economics 3A*: 1865-2097 (O. Ashenfelter and D. Card, Eds). New York: Elsevier Science.

Heckman, J., J. Smith, and N. Clements. (1997). "Making The Most Out Of Programme Evaluations and Social Experiments: Accounting For Heterogeneity in Programme Impacts," *Review of Economic Studies* 64 (4): 487 - 535.

Hirano, K. and J. Porter. (2005). "Asymptotics for statistical treatment rules," *Mimeo.*

Imbens, G. (2000). "The Role of the Propensity Score in Estimating Dose-Response Functions," *Biometrika* 87 (3): 706 - 710.

Imbens, G. (2004). "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Survey," *Review of Economics and Statistics* 86 (1): 4 - 30.

Imbens, G., and G. Ridder (2009) "Estimation and Inference for Generalized Full and Partial Means and Derivatives,"unpublished manuscript, dept of economics, Harvard University, http://www.economics.harvard.edu/faculty/imbens/papers_imbens.

Imbens, G., and J. Wooldridge (2009). "New Developments in the Econometrics of Program Evaluation," *Journal of Economic Literature* 47(1): 5 - 86.

Lehman, E. (1998), *Elements of Large-Sample Theory*, Springer-Verlag, New York.

Linton, O., and J. Nielsen. (1995), "A Kernel Method of Estimating Structured Nonparametric Regression Based on Marginal Integration," *Biometrika* 82 (1): 93 - 100.

Manski, C. (1990). "Nonparametric Bounds on Treatment Effects," *American Economic Review* 80(2): 319 - 323.

Manski, C. (1993). "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies* 60(3): 531 - 542.

Manski, C. (2003). *Partial Identification of Probability Distributions*. New York: Springer-Verlag.

Manski, C. (2004). "Statistical Treatment Rules for Heterogenous Populations," *Econometrica* 72(4): 1221 - 1246.

Moffitt, R. (2001), "Policy interventions, low-level equilibria and social interactions,"*Social Dynamics*: 45 - 82 (S. Durlauf and P. Young, Eds.). Cambridge, MA: MIT Press.

Newey, W. (1994). "Kernel Estimation of Partial Means and a General Variance Estimator," *Econometric Theory* 10(2): 233 - 253.

Powell, J., J. Stock and T. Stoker. (1989). "Semiparametric Estimation of Index Coefficients," *Econometrica* 57(6): 1403 - 1430.

Rosenbaum, P., and D. Rubin. (1983). "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika* 70(1): 41 - 55.

Legros, P. and A. Newman. (2004). "Beauty is a beast, frog is a prince: assortative matching with nontransferabilities," Mimeo.

Sacerdote, B. (2001). "Peer effects with random assignment: results for Dartmouth roommates," *Quarterly Journal of Economics* 116(2): 681 - 704.

Table 1: SIMULATION RESULTS FOR $\hat{\beta}^{\text{pam}}$ AND $\hat{\beta}^{\text{lc}}$, 10,000 SIMULATIONS

	$N = 200$				$N = 1000$			
	$\rho = 0.0$		$\rho = 0.5$		$\rho = 0.0$		$\rho = 0.5$	
	$\hat{\beta}^{\text{pam}}$	$\hat{\beta}^{\text{lc}}$	$\hat{\beta}^{\text{pam}}$	$\hat{\beta}^{\text{lc}}$	$\hat{\beta}^{\text{pam}}$	$\hat{\beta}^{\text{lc}}$	$\hat{\beta}^{\text{pam}}$	$\hat{\beta}^{\text{lc}}$
mean bias	-0.009	-0.018	-0.002	-0.016	-0.003	-0.011	-0.000	-0.010
median bias	-0.010	-0.020	-0.003	-0.017	-0.003	-0.011	-0.001	-0.010
s.d	0.093	0.039	0.088	0.043	0.040	0.013	0.039	0.013
ave s.e.	0.085	0.256	0.088	0.578	0.041	0.418	0.044	0.306
median s.e.	0.085	0.051	0.087	0.064	0.040	0.020	0.044	0.039
r.m.s.e.	0.093	0.043	0.088	0.046	0.040	0.017	0.039	0.016
m.a.e.	0.061	0.027	0.060	0.028	0.028	0.013	0.027	0.012
cov rate 90% c.i.	0.871	0.938	0.897	0.959	0.905	0.931	0.935	0.991
cov rate 95% c.i.	0.929	0.968	0.947	0.980	0.953	0.965	0.971	0.997

Table 2: SIMULATION RESULTS: ASSESSING THE ADEQUACY OF THE ASYMPTOTIC APPROXIMATIONS FOR $\hat{\beta}^{\text{pam}}$ $(N = 1000, \rho = 0.0)$

	$\hat{\beta}^{\text{pam}} - \beta^{\text{pam}}$	$\hat{\beta}_g^{\text{pam}} - \overline{g}^{\text{pam}}$	$\hat{\beta}_W^{\text{pam}} - \overline{g}^{\text{pam}}$	$\hat{\beta}_X^{\text{pam}} - \overline{g}^{\text{pam}}$	$\overline{g}^{\text{pam}} - \beta^{\text{pam}}$	remainder
Panel A						
mean	-0.003	-0.002	-0.000	0.001	0.000	-0.001
s.d.	0.040	0.031	0.019	0.019	0.038	0.004
	$\hat{t}_{\text{pam}}$	$\hat{t}_g$	$\hat{t}_W$	$\hat{t}_X$	$\hat{t}_{\overline{g}}$	nominal
Panel B						
mean	-0.074	-0.078	-0.015	0.049	0.002	0.000
s.d.	0.989	1.018	1.005	1.018	1.008	1.000
$\text{pr}(\lvert T\rvert \geq 1.645)$	0.095	0.105	0.102	0.111	0.107	0.100
$\text{pr}(\lvert T\rvert \geq 1.96)$	0.047	0.055	0.053	0.058	0.052	0.050
$\text{pr}(T \geq 1.645)$	0.039	0.043	0.052	0.062	0.053	0.050
$\text{pr}(T \leq -1.645)$	0.056	0.062	0.050	0.049	0.054	0.050
$\text{pr}(T \geq 1.96)$	0.018	0.021	0.026	0.033	0.027	0.025
$\text{pr}(T \leq -1.96)$	0.029	0.034	0.027	0.025	0.025	0.025

Appendix A: Additional Lemmas and Theorems

In this appendix we state a number of additional results that will be used in the proofs of the four Theorems 3.1-4.4. Specifically, Theorem 3.1 uses Lemmas A.1 and A.2. Theorems 4.1 and 4.2 use Lemmas A.3-A.8, A.14, A.15, Theorem A.1, Lemmas A.16-A.18, A.9-A.11. Theorem 4.3 uses Lemmas A.14, Theorem A.1, Lemmas A.9-A.11, Theorem A.3, and Lemmas A.24-A.28. Theorem 4.4 uses Lemma A.13, Theorem A.1, and Lemmas A.19-A.11.

Definition 6.1 (Sobolev Norm) *The norm that we use for functions $g : \mathbb{Z} \subset \mathbb{R}^L \to \mathbb{R}$ that are at least j times continuously differentiable is the Sobolev norm*

$$|g|_j = \sup_{z \in \mathbb{Z}, |\lambda| \leq j} \left| \frac{\partial g^{|\lambda|}}{\partial z^\lambda}(z) \right|.$$

Lemma A.1 *Let $f : \mathbb{X} \mapsto \mathbb{R}$, with $\mathbb{X} = [x_l, x_u]$ a compact subset of $\mathbb{R}$, be a twice continuously differentiable function, and let $g : \mathbb{R} \mapsto \mathbb{R}$ satisfy a Lipschitz condition, $|g(x+y) - g(x)| \leq c \cdot |y|$. Then*

$$\left| f(g(\lambda)) - \left(f(g(0)) + \frac{\partial}{\partial x} f(g(0)) \cdot (g(\lambda) - g(0)) \right) \right| \leq \frac{1}{2} \cdot \sup_{x \in \mathbb{X}} \left| \frac{\partial^2}{\partial x^2} f(x) \right| \cdot c^2 \cdot \lambda^2.$$

Lemma A.2 *Let X be a real-valued random variable with support $\mathbb{X} = [x_l, x_u]$, with density $f_X(x) > 0$ for all $x \in \mathbb{X}$, and let $h : \mathbb{X} \mapsto \mathbb{R}$ be a continous function. Suppose that $\mathbb{E}[|h(X) \cdot X]$ is finite. Then*

$$\operatorname{Cov}(h(X), X) = \mathbb{E}\left[\frac{\partial}{\partial x} h(X) \cdot \gamma(X) \right],$$

where

$$\gamma(x) = \frac{F_X(x) \cdot (1 - F_X(x))}{f_X(x)} \cdot (\mathbb{E}[X|X > x] - \mathbb{E}[X|X \leq x]),$$

and $F_X(x)$ is the cumulative distribution function of X.

For completeness we state a couple of results from Athey and Imbens (2006, AI from hereon).

Lemma A.3 (Lemma A.2 in AI) *Suppose Y is a real-valued, continuously distributed random variable with compact support $\mathbb{Y} = [y_l, y_u]$, with the probability density function $f_Y(y)$ continuous, bounded, and bounded away from zero, on $\mathbb{Y}$. Then, for any $\delta < 1/2$:*

$$\sup_{y \in \mathbb{Y}} N^\delta \cdot |\hat{F}_Y(y) - F_Y(y)| \xrightarrow{p} 0.$$

Lemma A.4 (Lemma A.3 in AI) *Suppose Y is a real-valued, continuously distributed random variable with compact support $\mathbb{Y} = [y_l, y_u]$, with the probability density function $f_Y(y)$ continuous, bounded, and bounded away from zero, on $\mathbb{Y}$. Then, for any $\delta < 1/2$:*

$$\sup_{q \in [0,1]} N^\delta \cdot |\hat{F}_Y^{-1}(q) - F_Y^{-1}(q)| \xrightarrow{p} 0.$$

Lemma A.5 (Lemma A.5 in AI) *Suppose Y is a real-valued random variable with compact support $\mathbb{Y} = [y_l, y_u]$, and suppose that the cumulative distribution function $F_Y(y)$ is twice continuously differentiable on $\mathbb{Y}$, with its first derivative $f_Y(y) = \frac{\partial F_Y}{\partial y}(y)$ bounded away from zero on $\mathbb{Y}$. Then, for $0 < \eta < 3/4$ and $\delta > \max(2\eta - 1, \eta/2)$,*

$$\sup_{y \in \mathbb{Y}, x \leq N^{-\delta}, x+y \in \mathbb{Y}} N^\eta \cdot \left| \hat{F}_Y(y+x) - \hat{F}_Y(y) - (F_Y(y+x) - F_Y(y)) \right| \xrightarrow{p} 0.$$

Lemma A.6 (Lemma A.6 in AI) *Suppose Y is a real-valued random variable with compact support $\mathbb{Y} = [y_l, y_u]$, and suppose that the cumulative distribution function $F_Y(y)$ is twice continuously differentiable on $\mathbb{Y}$, with its first derivative $f_Y(y) = \frac{\partial F_Y}{\partial y}(y)$ bounded away from zero on $\mathbb{Y}$. Then, for all $0 < \eta < 5/7$,*

$$\sup_{q\in[0,1]} N^{\eta} \cdot \left| \hat{F}_Y^{-1}(q) - F_Y^{-1}(q) + \frac{1}{f_Y(F_Y^{-1}(q))} \left(\hat{F}_Y(F_Y^{-1}(q)) - q \right) \right| \xrightarrow{p} 0.$$

Lemma A.7 *Suppose X and Y are real-valued, continuously distributed, random variables with compact support $\mathbb{Y} = [y_l, y_u]$ and $\mathbb{X} = [x_l, x_u]$, with the probability density functions $f_Y(y)$ and $f_X(x)$ continuous, bounded, and bounded away from zero, on $\mathbb{Y}$ and $\mathbb{X}$. Then, for any $\delta < 1/2$:*

$$\sup_{x\in\mathbb{X}} N^{\delta} \cdot \left| \hat{F}_Y^{-1}\left(\hat{F}_X(x)\right) - F_Y^{-1}\left(F_X(x)\right) \right| \xrightarrow{p} 0.$$

Lemma A.8 *Suppose Y is a real-valued random variable with compact support $\mathbb{Y} = [y_l, y_u]$, and the cumulative distribution function $F_Y(y)$ is twice continuously differentiable on $\mathbb{Y}$, with its first derivative $f_Y(y) = \frac{\partial F_Y}{\partial y}(y)$ bounded away from zero on $\mathbb{Y}$. Then, for $0 < \eta < 3/4$ and $\delta > \max(2\eta - 1, \eta/2)$,*

$$\sup_{y\in\mathbb{Y}, x\leq N^{-\delta}, x+y\in\mathbb{Y}} N^{\eta} \cdot \left| \hat{F}_Y(y + x) - \hat{F}_Y(y) - f_Y(y) \cdot x \right| \xrightarrow{p} 0.$$

The next three lemmas are given without proof. Proofs can be found in IR. The first gives a bound on the bias of the NIP estimator.

Lemma A.9 (Bias)
If for $m = 1, 2$ Assumptions 3.1-4.1 hold, and $q \geq 2s - 1$ and $r \geq s - 1$, then

$$\sup_{z\in\mathbb{Z}} \left| \mathbb{E}\left[\hat{h}_{m,\mathrm{nip},s}(z)\right] - h_m(z) \right| = O\left(b^s\right).$$

Note that by matching the order of the kernel and the degree of the polynomial in the NIP estimator we obtain the same reduction in the bias on the full support as on the internal region, i.e. the NIP estimator has a bias that is of the same order as that of the NW estimator on the internal region. The variance is bounded in the following lemma. We only use the following two results for the case with $L = 2$, but for convenience we give the general results.

Lemma A.10 (Variance)
If Assumptions 3.1-4.1 hold and $q \geq s - 1, r \geq s - 1 + L$, then

$$\sup_{z\in\mathbb{Z}} \left| \hat{h}_{m,\mathrm{nip},s}(z) - \mathbb{E}\left[\hat{h}_{m,\mathrm{nip},s}(z)\right] \right| = O_p\left(\left(\frac{\log N}{N b_N^L} \right)^{1/2} \right).$$

This is the same bound as for the NW estimator on the internal set.
The two lemmas imply a uniform rate for the NIP estimator

Lemma A.11 (Uniform Convergence)
If Assumptions 3.1-4.1 hold and $q \geq 2s - 1, r \geq s - 1 + L$, then

$$\sup_{z\in\mathbb{Z}} \left| \hat{h}_{m,\mathrm{nip},s}(z) - h_m(z) \right| = O_p\left(\left(\frac{\log N}{N \cdot b_N^L} \right)^{1/2} + b_N^s \right).$$

Lemma A.12 *If $\hat{h}(z)$ is a nonparametric estimator of $h(z)$ then*

$$\inf_{z\in\mathbb{Z}} |\hat{h}(z)| = \inf_{z\in\mathbb{Z}} |h(z)| + O_p\left(\sup_{z\in\mathbb{Z}} |\hat{h}(z) - h(z)| \right)$$

Therefore if $\sup_{z\in\mathbb{Z}}|\hat{h}(z)-h(z)|=o_p(1)$ and $\inf_{z\in\mathbb{Z}}|h(z)|>0$, then $\inf_{z\in\mathbb{Z}}|\hat{h}(z)|$ converges in probability to a positive number. This lemma is useful if $\hat{h}(z)$ appears in the denominator. In this paper $z=(w,x)$ or $z=w$.

Lemma A.13 *Suppose Assumptions 3.1-4.2 hold. Moreover, suppose that in these assumptions $q\geq 2s-1$, $r\geq s$. Then,*

$$\sup_{w\in\mathbb{W}}|\hat{m}(w)-m(w)|=O_p\left(\left(\frac{\ln(N)}{N\cdot b_N}\right)^{1/2}+b_N^s\right).$$

Lemma A.14 *Suppose Assumptions 3.1-4.2 hold. Moreover, suppose that $q\geq 2s+1$ and $r\geq s+3$, Then, (i)*

$$\sup_{w\in\mathbb{W},x\in\mathbb{X}}|\hat{g}(w,x)-g(w,x)|=O_p\left(\left(\frac{\ln(N)}{N\cdot b_N^2}\right)^{1/2}+b_N^s\right),$$

(ii)

$$\sup_{w\in\mathbb{W},x\in\mathbb{X}}\left|\frac{\partial\hat{g}}{\partial w}(w,x)-\frac{\partial g}{\partial w}(w,x)\right|=O_p\left(\left(\frac{\ln(N)}{N\cdot b_N^4}\right)^{1/2}+b_N^s\right),$$

and iii),

$$\sup_{w\in\mathbb{W},x\in\mathbb{X}}\left|\frac{\partial^2\hat{g}}{\partial w^2}(w,x)-\frac{\partial^2 g}{\partial w^2}(w,x)\right|=O_p\left(\left(\frac{\ln(N)}{N\cdot b_N^6}\right)^{1/2}+b_N^s\right).$$

The next lemma shows that we can separate out the uncertainty in $\hat{\beta}^{\mathrm{pam}}$ into five components: the uncertainty from estimating $g(\cdot)$, the uncertainty from estimating $\hat{F}_W^{-1}(\cdot)$, the uncertainty from estimating $\hat{F}_X(\cdot)$, and the uncertainty from averaging $g(F_W^{-1}(F_X(X_i)),X_i)$ over the sample, and a remainder term that is $o_p\left(N^{-1/2}\right)$. As defined in section 5

$$\hat{\beta}_g^{\mathrm{pam}}=\frac{1}{N}\sum_{i=1}^N\hat{g}\left(F_W^{-1}\left(F_X(X_i)\right),X_i\right),$$

$$\hat{\beta}_W^{\mathrm{pam}}=\frac{1}{N}\sum_{i=1}^N g\left(\hat{F}_W^{-1}\left(F_X(X_i)\right),X_i\right),$$

$$\hat{\beta}_X^{\mathrm{pam}}=\frac{1}{N}\sum_{i=1}^N g\left(F_W^{-1}\left(\hat{F}_X(X_i)\right),X_i\right),$$

and

$$\overline{g}^{\mathrm{pam}}=\frac{1}{N}\sum_{i=1}^N g\left(F_W^{-1}\left(F_X(X_i)\right),X_i\right).$$

Lemma A.15 *Suppose Assumptions 3.1, 4.1, and 4.2 hold with $q\geq 2s+1$, $r\geq s+3$, and $0\leq\delta<1/6$. Then*

$$\hat{\beta}^{\mathrm{pam}}-\beta^{\mathrm{pam}}=$$
$$\left(\hat{\beta}_g^{\mathrm{pam}}-\overline{g}^{\mathrm{pam}}\right)+\left(\hat{\beta}_W^{\mathrm{pam}}-\overline{g}^{\mathrm{pam}}\right)+\left(\hat{\beta}_X^{\mathrm{pam}}-\overline{g}^{\mathrm{pam}}\right)+\left(\overline{g}^{\mathrm{pam}}-\beta^{\mathrm{pam}}\right)+o_p\left(N^{-1/2}\right). \tag{A.1}$$

The next two results are special cases of theorems in Imbens and Ridder (2009). The first one refers to the full mean case, and focuses on the case where we take full means of regression functions and their first derivatives. The second result focuses on partial means of regression functions. The results in Imbens and Ridder (2009) allow for more general dependence on higher order derivatives, even in the partial mean case. Here we also restrict the analysis to the case where the regressors are the pair (W_i,X_i). We also state the conditions that IR invoke.

Let $Z_i=(W_i,X_i)$, with $X_i\in\mathbb{X}\subset\mathbb{R}^{L_X}$, $W_i\in\mathbb{W}\subset\mathbb{R}^{L_W}$, $Z_i\in\mathbb{W}\times\mathbb{X}\subset\mathbb{R}^{L_Z}$, with $L_Z=L_X+L_W$. As before $h(z)=(h_1(z),h_2(z))'$, with $h_1(z)=f_Z(z)$, and $h_2(z)=\mathbb{E}[Y|Z=z]\cdot f_Z(z)$. Let $n:\mathbb{R}^K\mapsto\mathbb{R}$, $t:\mathbb{X}\mapsto\mathbb{W}$, and

$\omega : \mathbb{X} \mapsto \mathbb{R}$, and define $\tilde{Y} = (\tilde{Y}_{i1}\ \tilde{Y}_{i2})'$, with $\tilde{Y}_{i1} = 1$ and $\tilde{Y}_{i2} = Y_i$. We are interested in full means (possibly depending on derivatives) of the regression function,

$$\theta^{\text{fm}} = \mathbb{E}\left[\omega(Z)n\left(h^{[\lambda]}(Z)\right)\right], \tag{A.2}$$

or partial means,

$$\theta^{\text{pm}} = \mathbb{E}\left[\omega(X)n\left(h\left(X, t(X)\right)\right)\right]. \tag{A.3}$$

Note that in the full mean case $\omega : \mathbb{Z} \mapsto \mathbb{R}$, and in the partial mean case $\omega : \mathbb{X} \mapsto \mathbb{R}$: the weight function depends only on the covariates that are being averaged over. In the full mean example $h^{[\lambda]}$ denotes the vector with elements including all derivatives $h^{(\mu)}$ for $\mu \leq \lambda$. The estimators we focus on are

$$\hat{\theta}^{\text{fm}} = \frac{1}{N}\sum_{i=1}^{N} \omega\left(Z_i\right) n\left(\hat{h}^{[\lambda]}_{\text{nip},s}\left(Z_i\right)\right), \qquad \text{and} \quad \hat{\theta}^{\text{pm}} = \frac{1}{N}\sum_{i=1}^{N} \omega(X_i)n\left(\hat{h}_{\text{nip},s}\left(t(X_i), X_i\right)\right).$$

It will also be useful to define the averages over the true regression functions and their derivatives,

$$\overline{\theta}^{\text{fm}} = \frac{1}{N}\sum_{i=1}^{N} \omega(Z_i)n\left(h^{[\lambda]}(Z_i)\right), \qquad \text{and} \quad \overline{\theta}^{\text{pm}} = \frac{1}{N}\sum_{i=1}^{N} \omega(X_i)n\left(h(t(X_i), X_i)\right).$$

Assumption A.1 (Distribution)
(i) $(Y_1, Z_1), (Y_2, Z_2), \ldots,$ *are independent and identically distributed,*
(ii) the support of Z *is* $\mathbb{Z} \subset \mathbb{R}^L$, $\mathbb{Z} = \bigotimes_{m=1}^{L}[z_{ml}, z_{mu}]$, $z_{ll} < z_{ul}$ *for all* $l = 1, \ldots, L$.
(iii) $\sup_{z\in\mathbb{Z}} \mathbb{E}[|Y|^p | Z = z] < \infty$.
(iv) $g(z) = \mathbb{E}[Y|Z = z]$ *is* q *times continuously differentiable on the interior of* $\mathbb{Z}$ *with the* q*-th derivative bounded,*
(v) $f_Z(z)$ *is bounded and bounded away from zero on* $\mathbb{Z}$, *is* q *times continuously differentiable on the interior of* $\mathbb{Z}$ *with the* q*-th derivative bounded.*

Assumption A.2 (Kernel)
(i) $K : \mathbb{R}^L \to \mathbb{R}$, *with* $K(u) = \prod_{l=1}^{L} \mathcal{K}(u_l)$,
(ii) $K(u) = 0$ *for* $u \notin \mathbb{U}$, *with* $\mathbb{U} = [-1, 1]^L$, *and* $\mathbb{U}_1 = [-1, 1]^{L_W}$, *and* $\mathbb{U}_2 = [-1, 1]^{L_X}$,
(iii) K *is* r *times continuously differentiable, with the* r*-th derivative bounded on the interior of* $\mathbb{U}$,
(iv) K *is a kernel of order* s, *so that* $\int_{\mathbb{U}} K(u)du = 1$ *and* $\int_{\mathbb{U}} u^{\lambda}K(u)du = 0$ *for all* λ *such that* $0 < |\lambda| < s$, *for some* $s \geq 1$,
(v) K *is a kernel of derivative order* d.

Assumption A.3 *The bandwidth* $b_N = N^{-\delta}$ *for some* $\delta > 0$.

Assumption A.4 (Smoothness of n and ω)
(i) The function n *is* t *times continuously differentiable with its* t*-th derivative bounded, and*
(ii) the function ω *is* t *times differentiable on* $\mathbb{X}$ *with bounded* t*-th derivative, and* $\frac{\partial^{\mu}\omega}{\partial z^{\mu}}(z)$ *is zero on the boundary of* $\mathbb{Z}$.

Assumption A.5 (Smoothness of t)
The function $t : \mathbb{X} \mapsto \mathbb{W}$ *is twice continuously differentiable on* $\mathbb{X}$ *with its first derivative positive, bounded, and bounded away from zero.*

Theorem A.1 (Generalized Full Mean and Average Derivative, [Theorem 4.2, Imbens and Ridder, 2009])
If Assumptions A.1, A.2, A.3, and A.4 hold with $q \geq |\lambda| + 2s - 1$, $r \geq |\lambda| + s - 1 + L$, $t \geq |\lambda| + s$, $p \geq 3$, $d \geq \max\{\lambda_1, \ldots, \lambda_L\} + s - 1$, *all* $\mu \leq \lambda$, $0 \leq |\mu| \leq |\lambda| - 1$, *and*

$$\frac{1}{2s} < \delta < \min\left\{\frac{2 - \frac{4}{p}}{2L + 4\max\{1, |\lambda|\}}, \frac{1}{2L + 4|\lambda|}\right\}$$

then $\hat{\theta}_{\text{fm}}$ *is asymptotically linear with*

$$\sqrt{N}(\hat{\theta}^{\text{fm}} - \theta^{\text{fm}}) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\omega(Z_i)n(h^{[\lambda]}(Z_i)) - \mathbb{E}\left[\omega(Z_i)n\left(h^{[\lambda]}(Z_i)\right)\right]\right)$$

$$+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\sum_{\kappa\leq\lambda}(-1)^{|\kappa|}\sum_{m=1}^{2}\left(\alpha_{\kappa m}^{(\kappa)}(X_i)\tilde{Y}_{im}-\mathbb{E}[\alpha_{\kappa m}^{(\kappa)}(X)\tilde{Y}_m]\right)\right)+o_p(1).$$

with

$$\alpha_{\kappa 1}^{(\kappa)}(z)=f_X(z)\omega(z)\frac{\partial n}{\partial h_1^{(\kappa)}(z)}(h^{[\lambda]}(z)),\qquad\text{and}\quad \alpha_{\kappa 2}^{(\kappa)}(z)=f_X(z)\omega(z)\frac{\partial n}{\partial h_2^{(\kappa)}(z)}(h^{[\lambda]}(z)),$$

and $\tilde{Y}=(\tilde{Y}_{i1}\ \tilde{Y}_{i2})'$, *with* $\tilde{Y}_{i1}=1$ *and* $\tilde{Y}_{i2}=Y_i$.

The second theorem from IR gives the asymptotic properties of the GPM estimators

Theorem A.2 (Generalized Partial Mean, [Theorem 4.3, Imbens and Ridder, 2009])
If Assumptions A.1, A.2, A.3, A.4, and A.5 hold with $q\geq 2s-1$, $r\geq s-1+L$, $t\geq s$, $p\geq 4$, $d\geq s-1$, *and*

$$\frac{1}{2s}<\delta<\min\left\{\frac{2-\frac{4}{p}}{2L+4},\frac{1}{2L}\right\},$$

then $\hat{\theta}^{\text{pm}}$ *is asymptotically linear with*

$$\sqrt{N}(\hat{\theta}^{\text{pm}}-\theta^{\text{pm}})=\sqrt{N}\cdot\left(\overline{\theta}^{\text{pm}}-\theta^{\text{pm}}\right)$$
$$+\frac{1}{b_N{}^{L_W}\sqrt{N}}\cdot\sum_{i=1}^{N}\sum_{m=1}^{2}\left(\alpha_m(X_i)'\tilde{Y}_{im}\int_{\mathbb{U}_2}K\left(\frac{W_i-t(X_i)}{b_N}+\frac{\partial}{\partial x}t(X_i)\cdot u_2,u_2\right)du_2\right.$$
$$\left.-\mathbb{E}\left[\alpha_m(X)'\tilde{Y}_m\int_{\mathbb{U}_2}K\left(\frac{W-t(X)}{b_N}+\frac{\partial}{\partial x}t(X)\cdot u_2,u_2\right)du_2\right]\right)+o_p(1),$$

with

$$\alpha_1(x)=f_Z(t(x),x)\omega(x)\frac{\partial n}{\partial h_1}(h(t(x),x)),\qquad\text{and}\quad \alpha_2(x)=f_Z(t(x),x)\omega(x)\frac{\partial n}{\partial h_2}(h(t(x),x)).$$

Moreover,

$$\begin{pmatrix}\sqrt{N}\cdot\left(\overline{\theta}^{\text{pm}}-\theta^{\text{pm}}\right)\\ \sqrt{N}b_N^{L_W/2}(\hat{\theta}^{\text{pm}}-\overline{\theta}^{\text{pm}})\end{pmatrix}\xrightarrow{d}\mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}V_1&0\\0&V_2\end{pmatrix}\right),$$

with

$$V_1=\mathbb{E}\left[(\omega(X)n(h(t(X),X))-\theta_{\text{pm}})^2\right],$$

and

$$V_2=\sum_{m=1}^{2}\sum_{m'=1}^{2}\int_{\mathbb{X}}\mu_{mm'}(x,t(x))\alpha_m(x)\alpha_{m'}(x)\int_{\mathbb{U}_2}\left(\int_{\mathbb{U}_1}K\left(u_1,\frac{\partial t}{\partial x}(x)u_1+u_2\right)du_1\right)^2du_2 f_X(x,t(x))dx_1,$$

with $\mu_{mm'}(x)=\mathbb{E}[\tilde{Y}_{im}\tilde{Y}_{im'}|X=x]$ *for* $m,m'=1,2$.

Lemma A.16 *Suppose Assumptions 3.1, 4.1, and 4.2 hold, with* $q\geq 2s-1$, $r\geq s+1$, $p\geq 4$, $d\geq s-1$, *and* $1/(2s)<\delta<1/8$. *Then*

$$\sqrt{N}\left(\hat{\beta}_g^{\text{pam}}-\overline{g}^{\text{pam}}\right)$$
$$=\frac{1}{\sqrt{N}b_N}\sum_{i=1}^{N}(Y_i-g(W_i,X_i))\cdot\int_{u_2}K\left(\frac{W_i-F_W^{-1}(F_X(X_i))}{b_N}+\frac{f_X(X_i)}{f_W\left(F_W^{-1}(F_X(X_i))\right)}\cdot u_2,u_2\right)du_2$$
$$+\frac{1}{\sqrt{N}b_N}\sum_{i=1}^{N}\left\{\left(g(W_i,X_i)-g(F_W^{-1}(F_X(X_i)),X_i)\right)\cdot\int_{u_2}K\left(\frac{W_i-F_W^{-1}(F_X(X_i))}{b_N}+\frac{f_X(X_i)}{f_W\left(F_W^{-1}(F_X(X_i))\right)}\cdot u_2,u_2\right)du_2\right.$$
$$\left.-\mathbb{E}\left[\left(g(W,X)-g(F_W^{-1}(F_X(X)),X)\right)\cdot\int_{u_2}K\left(\frac{W_i-F_W^{-1}(F_X(X))}{b_N}+\frac{f_X(X)}{f_W\left(F_W^{-1}(F_X(X))\right)}\cdot u_2,u_2\right)du_2\right]\right\}$$

$$+o_p(1).$$

and

$$\sqrt{N} b_N^{1/2} \left(\hat{\beta}_g^{\text{pam}} - \overline{g}^{\text{pam}}\right) \xrightarrow{d}$$

$$\mathcal{N}\left(0, \mathbb{E}\left[\sigma^2\left(F_W^{-1}\left(F_X(X)\right), X\right) \cdot \int_{u_1} \left(\int_{u_2} K\left(u_1 + \frac{f_X(X)}{f_W\left(F_W^{-1}\left(F_X(X)\right)\right)} \cdot u_2, u_2\right) du_2\right)^2 du_1 \cdot f_{W|X}\left(F_W^{-1}\left(F_X(X)\right)|X\right)\right]\right).$$

Lemma A.17 *Suppose Assumptions 3.1, 4.1, and 4.2 hold with $q \geq 2$. Then*

$$\hat{\beta}_W^{\text{pam}} - \overline{g}^{\text{pam}} = \frac{1}{N}\sum_{i=1}^N \psi_W^{\text{pam}}(W_i) + o_p\left(N^{-1/2}\right).$$

Lemma A.18 *Suppose Assumptions 3.1, 4.1, and 4.2 hold, with $q \geq 2$. Then*

$$\hat{\beta}_X^{\text{pam}} - \overline{g}^{\text{pam}} = \frac{1}{N}\sum_{i=1}^N \psi_X^{\text{pam}}(X_i) + o_p\left(N^{-1/2}\right).$$

Define

$$\hat{\beta}_g^{\text{lc}} = \frac{1}{N}\sum_{i=1}^N \frac{\partial \hat{g}}{\partial w}(W_i, X_i) \cdot d(W_i) \cdot (X_i - m(W_i)),$$

$$\hat{\beta}_m^{\text{lc}} = \frac{1}{N}\sum_{i=1}^N \frac{\partial g}{\partial w}(W_i, X_i) \cdot d(W_i) \cdot (X_i - \hat{m}(W_i)),$$

and

$$\overline{g}^{\text{lc}} = \frac{1}{N}\sum_{i=1}^N \frac{\partial g}{\partial w}(W_i, X_i) \cdot d(W_i) \cdot (X_i - m(W_i)).$$

Lemma A.19 *Suppose Assumptions 3.1, 4.1, and 4.2 hold. Moreoever, suppose that the estimators for $g(w,x)$ and $m(w)$, $\hat{g}(w,x)$ and $\hat{m}(w)$ respectively, satisfy*

$$\sup_{w\in\mathbb{W}, x\in\mathbb{X}} \left|\frac{\partial \hat{g}}{\partial w}(w,x) - \frac{\partial g}{\partial w}(w,x)\right| = o_p(N^{-\eta}) \qquad \text{and} \qquad \sup_{w\in\mathbb{W}} |\hat{m}(w) - m(w)| = o_p(N^{-\eta}),$$

for some $\eta > 1/4$. Then

$$\hat{\beta}^{\text{lc}} - \beta^{\text{lc}} = \left(\hat{\beta}_g^{\text{lc}} - \overline{g}^{\text{lc}}\right) + \left(\hat{\beta}_m^{\text{lc}} - \overline{g}^{\text{lc}}\right) + \left(\overline{g}^{\text{lc}} - \beta^{\text{lc}}\right) + o_p\left(N^{-1/2}\right). \tag{A.4}$$

Lemma A.20 *Suppose Assumptions 3.1, 4.1, and 4.2 hold, with $q \geq 2s$, $r \geq s$, $p \geq 3$, $d \geq s$, and $1/(2s) < \delta < 1/12$. Then:*

$$\hat{\beta}_g^{\text{lc}} - \overline{g}^{\text{lc}} = \frac{1}{N}\sum_{i=1}^N \psi_g^{\text{lc}}(Y_i, W_i, X_i) + o_p\left(N^{-1/2}\right), \tag{A.5}$$

where

$$\begin{aligned}\psi_g^{\text{lc}}(Y,W,X) = &-\frac{1}{f_{W,X}(W,X)}\frac{\partial f_{W,X}(W,X)}{\partial W}(Y - g(W,X))\, d(W)\,(X - m(W))\\ &-\frac{\partial m(W)}{\partial W} d(W)\,(Y - g(W,X))\\ &+\frac{\partial d}{\partial w}(W)\,(X - m(W))\,(Y - g(W,X)).\end{aligned}$$

Lemma A.21 *Suppose Assumptions 3.1-4.2 hold, with $q \geq 2s-1$, $r \geq s$, $p \geq 3$, $d \geq s-1$, and*

$$\frac{1}{2s} < \delta < \frac{1}{3} - \frac{2}{3p},$$

then

$$\sup_{w \in \mathbb{W}} \left| \frac{1}{\widehat{f}_W(w)} \left(\widehat{f}_W(w) - f_W(w) \right)^2 \right| = o_p\left(N^{-1/2} \right).$$

Lemma A.22 *Let $h(w) = (h_1(w), h_2(w))' = (\mathbb{E}[X | W = w] f_W(w), f_W(w))'$, and suppose Assumptions 3.1-4.2 hold, with $q \geq 2s-1$, $r \geq s$, $p \geq 3$, $d \geq s-1$, and*

$$\frac{1}{2s} < \delta < \frac{1}{3} - \frac{2}{3p},$$

then

$$\sup_{w \in \mathbb{W}} \left| \frac{1}{\widehat{h}_2(w)} \left(\widehat{h}_1(w) - h_1(w) \right) \left(\widehat{h}_2(w) - h_2(w) \right) \right| = o_p\left(N^{-1/2} \right).$$

Lemma A.23 *Suppose Assumptions 3.1-4.2 hold, with $q \geq 2s-1$, $r \geq s$, $p \geq 3$, and*

$$\frac{1}{2s} < \delta < \frac{1}{8}.$$

Then

$$\hat{\beta}_m^{\text{lc}} - \overline{g}^{\text{lc}} = \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}[g_W(W_i, X_i) | W_i] \cdot d(W_i) \cdot (X_i - m(W_i)) + o_p\left(N^{-1/2} \right), \tag{A.6}$$

Before the next theorem we need some additional definitions. We split Z_i into $(Z_{i1}', Z_{i2}')'$, with the dimension of Z_{i1} equal to L_{Z1}, and the dimension of Z_{i2} equal to L_{Z2}, so that $L = L_{Z1} + L_{Z2}$. We are interested in the distribution of

$$V = \sqrt{N} \cdot \left(\frac{1}{N^2} \sum_{j=1}^{N} \sum_{k=1}^{N} n(\hat{h}_{\text{nip},s}(Z_{1j}, Z_{2k})) - \frac{1}{N^2} \sum_{j=1}^{N} \sum_{k=1}^{N} n(h(Z_{1j}, Z_{2k})) \right). \tag{A.7}$$

We show that this is, to first order, equivalent to a single normalized sum.

Theorem A.3 *Suppose that Assumptions A.1-A.4, hold with $q \geq 2s-1$, $r \geq s-1+L$, $1/(2s) < \delta < 1/(2L)$, and $t \geq 2$. Then*

$$V = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left\{ \frac{\partial n}{\partial h}(h(Z_i))' \tilde{Y}_i f_{Z_1}(Z_{1i}) f_{Z_2}(Z_{2i}) - \mathbb{E}_Z \left[\frac{\partial n}{\partial h}(h(Z))' \tilde{Y} f_{Z_1}(Z_{1i}) f_{Z_2}(Z_{2i}) \right] \right\} + o_p(1). \tag{A.8}$$

(To be clear here we index the expectation by the random variable the expectation is taken over, in this case Z.)

Before stating some additional lemmas that will be used for proving Theorem 4.3 we need some additional definitions. Define

$$\overline{g}^{\text{cm}} = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} g(W_i, X_j) \frac{\phi_c\left(\Phi_c^{-1}(F_W(W_i)), \Phi_c^{-1}(F_X(X_j)); \rho \right)}{\phi_c\left(\Phi_c^{-1}(F_W(W_i)) \right) \phi_c\left(\Phi_c^{-1}(F_X(X_j)) \right)}$$

$$\hat{\beta}_g^{\text{cm}} = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \hat{g}(W_i, X_j) \frac{\phi_c\left(\Phi_c^{-1}(F_W(W_i)), \Phi_c^{-1}(F_X(X_j)); \rho \right)}{\phi_c\left(\Phi_c^{-1}(F_W(W_i)) \right) \phi_c\left(\Phi_c^{-1}(F_X(X_j)) \right)}$$

$$\hat{\beta}_W^{\text{cm}} = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} g(W_i, X_j) \frac{\phi_c\left(\Phi_c^{-1}(\hat{F}_W(W_i)), \Phi_c^{-1}(F_X(X_j)); \rho \right)}{\phi_c\left(\Phi_c^{-1}(\hat{F}_W(W_i)) \right) \phi_c\left(\Phi_c^{-1}(F_X(X_j)) \right)}$$

$$\hat{\beta}_X^{\text{cm}} = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} g(W_i, X_j) \frac{\phi_c\left(\Phi_c^{-1}(F_W(W_i)), \Phi_c^{-1}(\hat{F}_X(X_j)); \rho \right)}{\phi_c\left(\Phi_c^{-1}(F_W(W_i)) \right) \phi_c\left(\Phi_c^{-1}(\hat{F}_X(X_j)) \right)}$$

Lemma A.24 *Suppose Assumptions 3.1-4.2 hold with $q \geq 2s+2$, $r \geq s+3$, and $1/(2s) < \delta < 1/4$, then*

$$\begin{aligned}&\widehat{\beta}^{\mathrm{cm}}(\rho,0) - \beta^{\mathrm{cm}}(\rho,0)\\&\quad = \left(\hat{\beta}_g^{\mathrm{cm}} - \overline{g}^{\mathrm{cm}}\right) + \left(\hat{\beta}_W^{\mathrm{cm}} - \overline{g}^{\mathrm{cm}}\right) + \left(\hat{\beta}_X^{\mathrm{cm}} - \overline{g}^{\mathrm{cm}}\right) + \left(\overline{g}^{\mathrm{cm}} - \beta^{\mathrm{cm}}(\rho,0)\right) + o_p\left(N^{-1/2}\right).\end{aligned}$$

Lemma A.25 *Suppose Assumptions 3.1-4.2 hold, then*

$$\hat{\beta}_g^{\mathrm{cm}} - \overline{g}^{\mathrm{cm}} = \frac{1}{N}\sum_{i=1}^{N}\psi_g^{\mathrm{cm}}(Y_i,W_i,X_i) + o_p\left(N^{-1/2}\right),$$

where

$$\psi_g^{\mathrm{cm}}(y,w,x) = \frac{f_W(w)\cdot f_X(x)}{f_{WX}(w,x)}(y-g(w,x))\omega(w,x).$$

Lemma A.26 *Suppose Assumptions 3.1-4.2 hold, then*

$$\hat{\beta}_W^{\mathrm{cm}} - \overline{g}^{\mathrm{cm}} = \frac{1}{N}\sum_{i=1}^{N}\psi_W^{\mathrm{cm}}(Y_i,W_i,X_i) + o_p\left(N^{-1/2}\right),$$

where

$$\psi_W^{\mathrm{cm}}(y,w,x) = \int\int g(s,t)e_W(s,t)\left(1\left(w \leq s\right) - F_W(s)\right)f_W(s)\,f_X(t)\,\mathrm{d}s\mathrm{d}t.$$

Lemma A.27 *Suppose Assumptions 3.1-4.2 hold, then*

$$\hat{\beta}_X^{\mathrm{cm}} - \overline{g}^{\mathrm{cm}} = \frac{1}{N}\sum_{i=1}^{N}\psi_X^{\mathrm{cm}}(Y_i,W_i,X_i) + o_p\left(N^{-1/2}\right).,$$

where

$$\psi_X^{\mathrm{cm}}(y,w,x) = \int\int g(s,t)e_X(s,t)\left(1\left(x \leq t\right) - F_X(t)\right)f_W(s)\,f_X(t)\,\mathrm{d}s\mathrm{d}t.$$

Lemma A.28 *Suppose Assumptions 3.1-4.2 hold, then*

$$\overline{g}^{\mathrm{cm}} - \beta^{\mathrm{cm}}(\rho,0) = \frac{1}{N}\sum_{i=1}^{N}\psi_0^{\mathrm{cm}}(Y_i,W_i,X_i) + o_p\left(N^{-1/2}\right),$$

where

$$\psi_0^{\mathrm{cm}}(w,x) = \left(\mathbb{E}\left[g(W,x)\cdot\omega(W,x)\right] - \beta^{\mathrm{cm}}(\rho,0)\right) + \left(\mathbb{E}\left[g(w,X)\cdot\omega(w,X)\right] - \beta^{\mathrm{cm}}(\rho,0)\right). \tag{A.9}$$

The following theorem is a simplifed version of the V-statistics results in Lehman (1998).

Theorem A.4 (V-STATISTICS) *Suppose $Z_1,\ldots,Z_N$ are independent and identically distributed random vectors with dimension K, with support $\mathbb{Z} \subset \mathbb{R}^K$. Let $\psi : \mathbb{Z}^K \times \mathbb{Z}^K \mapsto \mathbb{R}$ be a real-valued function. Define*

$$\begin{aligned}&\theta = \mathbb{E}\left[\psi(Z_1,Z_2)\right], \qquad \psi_1(z) = \mathbb{E}\left[\psi(z,Z)\right], \qquad \psi_2(z) = \mathbb{E}\left[\psi(Z,z)\right],\\&\sigma^2 = \mathrm{Cov}(\psi(Z_1,Z_2),\psi(Z_1,Z_3)) + \mathrm{Cov}(\psi(Z_2,Z_1),\psi(Z_1,Z_3))\\&\qquad\quad + \mathrm{Cov}(\psi(Z_1,Z_2),\psi(Z_3,Z_1)) + \mathrm{Cov}(\psi(Z_2,Z_1),\psi(Z_3,Z_1)).\end{aligned}$$

and

$$V = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\psi(Z_i,Z_j).$$

Then, if $0 < \sigma^2 < \infty$,

$$V = \frac{1}{N}\sum_{i=1}^{N}\left\{(\psi_1(Z_i) - \theta) + (\psi_2(Z_i) - \theta)\right\} + o_p\left(N^{-1/2}\right),$$

and

$$\sqrt{N}\cdot(V-\theta) \xrightarrow{d} \mathcal{N}\left(0,\sigma^2\right).$$

Appendix B: Proofs of Additional Lemmas and Theorems

In the following proofs c is a generic constant.

Proof of Lemma A.1: Because $f(\cdot)$ is twice continuously differentiable on $\mathbb{X}$, a compact subset of $\mathbb{R}$, it follows that for all $a, b \in \mathbb{X}$, by a Taylor series expansion,

$$f(b) = f(a) + \frac{\partial f}{\partial x}(a) \cdot (b - a) + \frac{1}{2} \cdot \frac{\partial^2 f}{\partial x^2}(c) \cdot (b - a)^2,$$

for some $c \in \mathbb{X}$. Hence

$$\left| f(g(\lambda)) - \left(f(g(0)) + \frac{\partial f}{\partial x}(g(0)) \cdot (g(\lambda) - g(0)) \right) \right| \leq \frac{1}{2} \cdot \sup_{x \in \mathbb{X}} \left| \frac{\partial^2 f}{\partial x^2}(x) \right| \cdot (g(\lambda) - g(0))^2.$$

By the Lipschitz condition on $g(\lambda)$, this is bounded by

$$\frac{1}{2} \cdot \sup_{x \in \mathbb{X}} \left| \frac{\partial^2 f}{\partial x^2}(x) \right| \cdot c^2 \cdot \lambda^2.$$

□

Proof of Lemma A.2: Let $\mu = \mathbb{E}[X]$, and write $h(x) = h(x_l) + \int_{x_l}^{x} \frac{\partial}{\partial x} h(z) \mathrm{d}z$. Then:

$$\begin{aligned}
\mathrm{Cov}(h(X), X) = \mathbb{E}\left[h(X) \cdot (X - \mu)\right] &= \mathbb{E}\left[\left(h(x_l) + \int_{x_l}^{X} \frac{\partial}{\partial x} h(z) \mathrm{d}z \right) \cdot (X - \mu) \right] \\
&= \mathbb{E}\left[\int_{x_l}^{X} \frac{\partial}{\partial x} h(z) \mathrm{d}z \cdot (X - \mu) \right] \\
&= \mathbb{E}\left[\int_{x_l}^{x_u} \mathbf{1}_{X > z} \cdot \frac{\partial}{\partial x} h(z) \mathrm{d}z \cdot (X - \mu) \right] \\
&= \int_{x_l}^{x_u} \frac{\partial}{\partial x} h(z) \cdot \mathbb{E}\left[\mathbf{1}_{X > z} \cdot (X - \mu)\right] \mathrm{d}z \\
&= \int_{x_l}^{x_u} \frac{\partial}{\partial x} h(z) \cdot \mathbb{E}\left[X - \mu | X > z\right] \cdot \Pr(X > z) \mathrm{d}z \\
&= \int_{x_l}^{x_u} \frac{\partial}{\partial x} h(z) \cdot F_X(z) \cdot (1 - F_X(z)) \cdot (\mathbb{E}[X|X > z] - \mathbb{E}[X|X \leq z]) \,\mathrm{d}z \\
&= \int_{x_l}^{x_u} \frac{\partial}{\partial x} h(z) \cdot \frac{F_X(z) \cdot (1 - F_X(z))}{f_X(z)} \cdot (\mathbb{E}[X|X > z] - \mathbb{E}[X|X \leq z])\, f_X(z) \mathrm{d}z \\
&= \mathbb{E}\left[\frac{\partial}{\partial x} h(X) \cdot \gamma(X) \right].
\end{aligned}$$

□

Proof of Lemma A.7: By the triangle inequality

$$\begin{aligned}
&\sup_{x \in \mathbb{X}} N^{\delta} \cdot \left| \hat{F}_Y^{-1}\left(\hat{F}_X(x)\right) - F_Y^{-1}\left(F_X(x)\right) \right| \\
&\quad \leq \sup_{x \in \mathbb{X}} N^{\delta} \cdot \left| \hat{F}_Y^{-1}\left(\hat{F}_X(x)\right) - F_Y^{-1}\left(\hat{F}_X(x)\right) \right| \\
&\qquad + \sup_{x \in \mathbb{X}} N^{\delta} \cdot \left| F_Y^{-1}\left(\hat{F}_X(x)\right) - F_Y^{-1}\left(\hat{F}_X(x)\right) \right| \\
&\quad \leq \sup_{q \in [0,1]} N^{\delta} \cdot \left| \hat{F}_Y^{-1}(q) - F_Y^{-1}(q) \right| \\
&\qquad + \sup_{x \in \mathbb{X}, y \in \mathbb{Y}} N^{\delta} \cdot \frac{1}{f_Y(y)} \left| \hat{F}_X(x) - F_X(x) \right|.
\end{aligned}$$

The first term is $o_p(1)$ by Lemma A.4, and the second by the fact that $f_Y(y)$ is bounded away from zero, in combination with Lemma A.3. □

Proof of Lemma A.8: By the triangle inequality

$$\sup_{y\in\mathbb{Y},x\leq N^{-\delta},x+y\in\mathbb{Y}} N^{\eta}\cdot\left|\hat{F}_Y(y+x)-\hat{F}_Y(y)-f_Y(y)\cdot x\right|$$

$$\leq \sup_{y\in\mathbb{Y},x\leq N^{-\delta},x+y\in\mathbb{Y}} N^{\eta}\cdot\left|\hat{F}_Y(y+x)-(F_Y(y+x)-F_Y(y))\right|$$

$$+ \sup_{y\in\mathbb{Y},x\leq N^{-\delta},x+y\in\mathbb{Y}} N^{\eta}\cdot\left|F_Y(y+x)-F_Y(y)-f_Y(y)\cdot x\right|.$$

The first term on the right-hand side converges to zero in probability by Lemma A.5. To show that the second term converges to zero note that

$$\sup_{y\in\mathbb{Y},x\leq N^{-\delta},x+y\in\mathbb{Y}} N^{\eta}\cdot\left|F_Y(y+x)-F_Y(y)-f_Y(y)\cdot x\right|$$

$$\leq \sup_{y\in\mathbb{Y},x\leq N^{-\delta},x+y\in\mathbb{Y},\lambda\in[0,1]} N^{\eta}\cdot\left|f_Y(y+\lambda x)\cdot x-f_Y(y)\cdot x\right|$$

$$\leq \sup_{y\in\mathbb{Y},z\in\mathbb{Y},x\leq N^{-\delta},x+y\in\mathbb{Y},\lambda\in[0,1]} N^{\eta}\cdot\left|\frac{\partial f_Y}{\partial y}(z)\right|\cdot\lambda x^2$$

$$\leq \sup_{y\in\mathbb{Y},x\leq N^{-\delta}} N^{\eta}x^2\frac{\partial f_Y}{\partial y}(y)\to 0,$$

because $\frac{\partial f_Y}{\partial y}(y)$ is bounded, $x<N^{-\delta}$, and $\delta>\eta/2$. □

Proof of Lemma A.12: By the inequality $|a|\geq|b|-|a-b|$

$$\inf_{z\in\mathbb{Z}}|\hat{h}(z)|\geq\inf_{z\in\mathbb{Z}}|h(z)|-\sup_{z\in\mathbb{Z}}|\hat{h}(z)-h(z)|$$

from which the result follows. □

Proof of Lemma A.13: This follows directly from Theorem 7.1 in IR □

Proof of Lemma A.14: This follows directly from Theorem 7.1 in IR. □

Proof of Lemma A.15: First note that by the assumptions in the Lemma the conditions for Lemma A.14 are satisfied. Moreoever, by the assumption that $0<\delta<1/6$, it follows that $O_p(b^N)=o_p(N^{-\eta})$ for $\eta<\delta\cdot s$, and $O_p(\ln(N)N^{-1}b_N^{-2})=O_p(\ln(N)N^{-1+2\delta})=o_p(1)$, $O_p(\ln(N)N^{-1}b_N^{-4})=O_p(\ln(N)N^{-1+4\delta})=o_p(N^{-\eta})$ for $\eta<1-4\delta$, and $O_p(\ln(N)N^{-1}b_N^{-6})=O_p(\ln(N)N^{-1+6\delta})=o_p(1)$. Hence the results from Lemma A.14 imply

$$\sup_{w\in\mathbb{W},x\in\mathbb{X}}|\hat{g}(w,x)-g(w,x)|=O_p\left(\left(\frac{\ln(N)}{N\cdot b_N^2}\right)^{1/2}+b_N^s\right)=o_p(1), \tag{B.1}$$

$$\sup_{w\in\mathbb{W},x\in\mathbb{X}}\left|\frac{\partial\hat{g}}{\partial w}(w,x)-\frac{\partial g}{\partial w}(w,x)\right|=O_p\left(\left(\frac{\ln(N)}{N\cdot b_N^4}\right)^{1/2}+b_N^s\right)=o_p\left(N^{-\eta}\right), \tag{B.2}$$

for $\eta<\min(1-4\delta,\delta\cdot s)$, and

$$\sup_{w\in\mathbb{W},x\in\mathbb{X}}\left|\frac{\partial^2\hat{g}}{\partial w^2}(w,x)-\frac{\partial^2 g}{\partial w^2}(w,x)\right|=O_p\left(\left(\frac{\ln(N)}{N\cdot b_N^6}\right)^{1/2}+b_N^s\right)=o_p(1). \tag{B.3}$$

Now,

$$\hat{\beta}^{\text{pam}}-\beta^{\text{pam}}=\frac{1}{N}\sum_{i=1}^N\hat{g}\left(\hat{F}_W^{-1}\left(\hat{F}_X(X_i)\right),X_i\right)-\mathbb{E}\left[g\left(F_W^{-1}\left(F_X(X)\right),X\right)\right]$$

$$=\frac{1}{N}\sum_{i=1}^N\hat{g}\left(\hat{F}_W^{-1}\left(\hat{F}_X(X_i)\right),X_i\right)-\frac{1}{N}\sum_{i=1}^N g\left(\hat{F}_W^{-1}\left(\hat{F}_X(X_i)\right),X_i\right) \tag{B.4}$$

$$-\left(\frac{1}{N}\sum_{i=1}^N\hat{g}\left(F_W^{-1}\left(F_X(X_i)\right),X_i\right)-\frac{1}{N}\sum_{i=1}^N g\left(F_W^{-1}\left(F_X(X_i)\right),X_i\right)\right) \tag{B.5}$$

$$+\frac{1}{N}\sum_{i=1}^N\hat{g}\left(F_W^{-1}\left(F_X(X_i)\right),X_i\right)-\frac{1}{N}\sum_{i=1}^N g\left(F_W^{-1}\left(F_X(X_i)\right),X_i\right) \tag{B.6}$$

$$+\frac{1}{N}\sum_{i=1}^{N} g\left(\hat{F}_W^{-1}\left(\hat{F}_X(X_i)\right), X_i\right) - \frac{1}{N}\sum_{i=1}^{N} g\left(F_W^{-1}\left(\hat{F}_X(X_i)\right), X_i\right) \tag{B.7}$$

$$-\left(\frac{1}{N}\sum_{i=1}^{N} g\left(\hat{F}_W^{-1}\left(F_X(X_i)\right), X_i\right) - \frac{1}{N}\sum_{i=1}^{N} g\left(F_W^{-1}\left(F_X(X_i)\right), X_i\right)\right) \tag{B.8}$$

$$+\frac{1}{N}\sum_{i=1}^{N} g\left(\hat{F}_W^{-1}\left(F_X(X_i)\right), X_i\right) - \frac{1}{N}\sum_{i=1}^{N} g\left(F_W^{-1}\left(F_X(X_i)\right), X_i\right) \tag{B.9}$$

$$+\frac{1}{N}\sum_{i=1}^{N} g\left(F_W^{-1}\left(\hat{F}_X(X_i)\right), X_i\right) - \frac{1}{N}\sum_{i=1}^{N} g\left(F_W^{-1}\left(F_X(X_i)\right), X_i\right) \tag{B.10}$$

$$+\frac{1}{N}\sum_{i=1}^{N} g\left(F_W^{-1}\left(F_X(X_i)\right), X_i\right) - \mathbb{E}\left[g\left(F_W^{-1}\left(F_X(X)\right), X\right)\right]. \tag{B.11}$$

Since (B.6) is equal to $\hat{\beta}_{pam,g} - \overline{g}_{pam}$, (B.9) equals $\hat{\beta}_{pam,W} - \overline{g}_{pam}$, (B.10) equals $\hat{\beta}_{pam,X} - \overline{g}_{pam}$, and (B.11) equals $\overline{g}_{pam} - \beta^{\text{pam}}$, we only need to show that the sum of (B.4), (B.5), and that of (B.7), (B.8) are $o_p(N^{-1/2})$. First consider the sum of (B.4) and (B.5) that is equal to

$$\frac{1}{N}\sum_{i=1}^{N} \hat{g}\left(\hat{F}_W^{-1}\left(\hat{F}_X(X_i)\right), X_i\right) - \frac{1}{N}\sum_{i=1}^{N} \hat{g}\left(F_W^{-1}\left(F_X(X_i)\right), X_i\right)$$
$$-\left(\frac{1}{N}\sum_{i=1}^{N} g\left(\hat{F}_W^{-1}\left(\hat{F}_X(X_i)\right), X_i\right) - \frac{1}{N}\sum_{i=1}^{N} g\left(F_W^{-1}\left(F_X(X_i)\right), X_i\right)\right).$$

By a second order Taylor series expansion of $\hat{g}$ and g in $F_W^{-1}(F_X(X_i))$ this is, for some $\tilde{W}_i$ and $\bar{W}_i$, equal to

$$\frac{1}{N}\sum_{i=1}^{N} \frac{\partial \hat{g}}{\partial w}\left(F_W^{-1}\left(F_X(X_i)\right), X_i\right)\left(\hat{F}_W^{-1}\left(\hat{F}_X(X_i)\right) - F_W^{-1}\left(F_X(X_i)\right)\right)$$

$$+\frac{1}{2N}\sum_{i=1}^{N} \frac{\partial^2 \hat{g}}{\partial w^2}\left(\tilde{W}_i, X_i\right)\left(\hat{F}_W^{-1}\left(\hat{F}_X(X_i)\right) - F_W^{-1}\left(F_X(X_i)\right)\right)^2 \tag{B.12}$$

$$-\frac{1}{N}\sum_{i=1}^{N} \frac{\partial g}{\partial w}\left(F_W^{-1}\left(F_X(X_i)\right), X_i\right)\left(\hat{F}_W^{-1}\left(\hat{F}_X(X_i)\right) - F_W^{-1}\left(F_X(X_i)\right)\right)$$

$$-\frac{1}{2N}\sum_{i=1}^{N} \frac{\partial^2 g}{\partial w^2}\left(\bar{W}_i, X_i\right)\cdot\left(\hat{F}_W^{-1}\left(\hat{F}_X(X_i)\right) - F_W^{-1}\left(F_X(X_i)\right)\right)^2. \tag{B.13}$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\partial \hat{g}}{\partial w}\left(F_W^{-1}\left(F_X(X_i)\right), X_i\right) - \frac{\partial g}{\partial w}\left(F_W^{-1}\left(F_X(X_i)\right), X_i\right)\right)$$
$$\times\left(\hat{F}_W^{-1}\left(\hat{F}_X(X_i)\right) - F_W^{-1}\left(F_X(X_i)\right)\right) + o_p\left(N^{-1/2}\right).$$

$$\leq \sup_{x\in\mathbb{X}}\left|\frac{\partial \hat{g}}{\partial w}\left(F_W^{-1}\left(F_X(x)\right), x\right) - \frac{\partial g}{\partial w}\left(F_W^{-1}\left(F_X(x)\right), x\right)\right| \tag{B.14}$$

$$\times \sup_{x\in\mathbb{X}}\left|\left(\hat{F}_W^{-1}\left(\hat{F}_X(x)\right) - F_W^{-1}\left(F_X(x)\right)\right)\right| + o_p\left(N^{-1/2}\right). \tag{B.15}$$

We used the fact that (B.13) is $o_p(N^{-1/2})$ because $\partial^2 g(w,x)/\partial w^2$ is bounded and because $\sup_{x\in\mathbb{X}}(\hat{F}_W^{-1}(\hat{F}_X(x)) - F_W^{-1}(F_X(x)))^2$ is $o_p(N^{-1/2})$ by Lemma A.7. Also (B.12) is $o_p(N^{-1/2})$ by the same argument because the bandwidth choice implies $\sup_{w\in\mathbb{W},x\in\mathbb{X}}|\partial^2\hat{g}(w,x)/\partial w^2 - \partial^2 g(w,x)/\partial w^2| = o_p(1)$ by (B.3), so that

$$\sup_{w\in\mathbb{W},x\in\mathbb{X}}\left|\frac{\partial^2 \hat{g}(w,x)}{\partial w^2}\right| \leq \sup_{w\in\mathbb{W},x\in\mathbb{X}}\left|\frac{\partial^2 g(w,x)}{\partial w^2}\right| + \sup_{w\in\mathbb{W},x\in\mathbb{X}}\left|\frac{\partial^2 \hat{g}(w,x)}{\partial w^2} - \frac{\partial^2 g(w,x)}{\partial w^2}\right| = \sup_{w\in\mathbb{W},x\in\mathbb{X}}\left|\frac{\partial^2 g(w,x)}{\partial w^2}\right| + o_p(1)$$

Finally by Lemma A.7

$$\sup_{x\in\mathbb{X}}\left|\hat{F}_W^{-1}\left(\hat{F}_X(x)\right) - F_W^{-1}\left(F_X(x)\right)\right| = o_p\left(N^{-1/2+\eta}\right)$$

for all $\eta > 0$. By the assumption of the lemma

$$\sup_{x\in\mathbb{X}}\left|\frac{\partial\hat{g}}{\partial w}\left(F_W^{-1}\left(F_X(x)\right),x\right)-\frac{\partial g}{\partial w}\left(F_W^{-1}\left(F_X(x)\right),x\right)\right| = o_p(N^{-\eta})$$

for some $\eta > 0$. We conclude that the sum of (B.4) and (B.5) is $o_p\left(N^{-1/2}\right)$.
Next, consider the sum of (B.7) and (B.8) that is bounded by

$$\sup_{x\in\mathbb{X}}\left|\left[g\left(\hat{F}_W^{-1}\left(\hat{F}_X(x)\right),x\right)-g\left(F_W^{-1}\left(\hat{F}_X(x)\right),x\right)\right]-\left[g\left(\hat{F}_W^{-1}\left(F_X(x)\right),x\right)-g\left(F_W^{-1}\left(F_X(x)\right),x\right)\right]\right|$$

By a second order Taylor series expansion with intermediate values $\tilde{W}(x)$ and $\bar{W}(x)$ and the triangle inequality this is bounded by

$$\sup_{x\in\mathbb{X}}\left|\frac{\partial g}{\partial w}\left(F_W^{-1}\left(\hat{F}_X(x)\right),x\right)\left[\hat{F}_W^{-1}\left(\hat{F}_X(x)\right)-F_W^{-1}\left(\hat{F}_X(x)\right)\right]\right.$$

$$\left.-\frac{\partial g}{\partial w}\left(F_W^{-1}\left(F_X(x)\right),x\right)\left[\hat{F}_W^{-1}\left(F_X(x)\right)-F_W^{-1}\left(F_X(x)\right)\right]\right|+$$

$$\sup_{x\in\mathbb{X}}\frac{1}{2}\left|\frac{\partial^2 g}{\partial w^2}\left(\tilde{W}(x),x\right)\left[\hat{F}_W^{-1}\left(\hat{F}_X(x)\right)-F_W^{-1}\left(\hat{F}_X(x)\right)\right]^2\right|+\frac{1}{2}\sup_{x\in\mathbb{X}}\left|\frac{\partial^2 g}{\partial w^2}\left(\bar{W}(x),x\right)\left[\hat{F}_W^{-1}\left(F_X(x)\right)-F_W^{-1}\left(F_X(x)\right)\right]^2\right|$$

where because the second derivative of $g(w,x)$ is bounded on $\mathbb{W}\times\mathbb{X}$, by Lemma A.4 the expression on the last line is $o_p(N^{-1/2})$. The first term is bounded by

$$\sup_{x\in\mathbb{X}}\left|\left[\frac{\partial g}{\partial w}\left(F_W^{-1}\left(\hat{F}_X(x)\right),x\right)-\frac{\partial g}{\partial w}\left(F_W^{-1}\left(F_X(x)\right),x\right)\right]\left[\hat{F}_W^{-1}\left(\hat{F}_X(x)\right)-F_W^{-1}\left(\hat{F}_X(x)\right)\right]\right|$$

$$+\sup_{x\in\mathbb{X}}\left|\frac{\partial g}{\partial w}\left(F_W^{-1}\left(F_X(x)\right),x\right)\left[\hat{F}_W^{-1}\left(\hat{F}_X(x)\right)-F_W^{-1}\left(\hat{F}_X(x)\right)-\hat{F}_W^{-1}\left(F_X(x)\right)+F_W^{-1}\left(F_X(x)\right)\right]\right|$$

By a first order Taylor series expansion of $\frac{\partial g}{\partial w}\left(F_W^{-1}\left(\hat{F}_X(x)\right),x\right)$ in $F_X(x)$ we have, because the second derivative of $g(w,x)$ is bounded and the density of W is bounded from 0 on its support, that by Lemmas A.4 and A.3, the expression on the first line is $o_p(N^{-1/2})$. The bound on the expression in the second line is proportional to

$$\sup_{x\in\mathbb{X}}\left|\hat{F}_W^{-1}\left(\hat{F}_X(x)\right)-F_W^{-1}\left(\hat{F}_X(x)\right)-\hat{F}_W^{-1}\left(F_X(x)\right)+F_W^{-1}\left(F_X(x)\right)\right|$$

This expression is bounded by

$$\sup_{x\in\mathbb{X}}\left|\frac{1}{f_W\left(F_W^{-1}\left(\hat{F}_X(x)\right)\right)}\left[\hat{F}_W\left(F_W^{-1}\left(\hat{F}_X(x)\right)\right)-\hat{F}_X(x)\right]-\frac{1}{f_W\left(F_W^{-1}\left(F_X(x)\right)\right)}\left[\hat{F}_W\left(F_W^{-1}\left(F_X(x)\right)\right)-F_X(x)\right]\right|$$

$$+\sup_{x\in\mathbb{X}}\left|\hat{F}_W^{-1}\left(\hat{F}_X(x)\right)-F_W^{-1}\left(\hat{F}_X(x)\right)-\frac{1}{f_W\left(F_W^{-1}\left(\hat{F}_X(x)\right)\right)}\left[\hat{F}_W\left(F_W^{-1}\left(\hat{F}_X(x)\right)\right)-\hat{F}_X(x)\right]\right|$$

$$+\sup_{x\in\mathbb{X}}\left|\hat{F}_W^{-1}\left(F_X(x)\right)-F_W^{-1}\left(F_X(x)\right)-\frac{1}{f_W\left(F_W^{-1}\left(F_X(x)\right)\right)}\left[\hat{F}_W\left(F_W^{-1}\left(F_X(x)\right)\right)-F_X(x)\right]\right|$$

By Lemma A.6 the expressions in the last two lines are $o_p(N^{-1/2})$. The expression in the first line is bounded by

$$\sup_{x\in\mathbb{X}}\left|\left[\frac{1}{f_W\left(F_W^{-1}\left(\hat{F}_X(x)\right)\right)}-\frac{1}{f_W\left(F_W^{-1}\left(F_X(x)\right)\right)}\right]\left[\hat{F}_W\left(F_W^{-1}\left(\hat{F}_X(x)\right)\right)-\hat{F}_X(x)\right]\right|$$

$$+\sup_{x\in\mathbb{X}}\left|\frac{1}{f_W\left(F_W^{-1}\left(F_X(x)\right)\right)}\left[\hat{F}_W\left(F_W^{-1}\left(\hat{F}_X(x)\right)\right)-\hat{F}_X(x)-\hat{F}_W\left(F_W^{-1}\left(F_X(x)\right)\right)+F_X(x)\right]\right|$$

The expression in the first line is bounded by

$$\sup_{x\in\mathbb{X}}\left|\frac{1}{f_W\left(F_W^{-1}\left(\hat{F}_X(x)\right)\right)}-\frac{1}{f_W\left(F_W^{-1}\left(F_X(x)\right)\right)}\right|\times\sup_{x\in\mathbb{X}}\left|\hat{F}_W\left(F_W^{-1}\left(\hat{F}_X(x)\right)\right)-\hat{F}_X(x)\right|$$

By a first order Taylor series expansion of $\frac{1}{f_W\left(F_W^{-1}\left(\hat{F}_X(x)\right)\right)}$ in $F_X(x)$, the fact that $f_W(w)$ is bounded from 0 and its derivative bounded on $\mathbb{W}$, and Lemma A.3 the first factor is $o_p(N^{-\delta})$ for all $\delta<1/2$ and by Lemma A.3 the same is true for the second factor, so that the product is $o_p(N^{-1/2})$. Because $f_W(w)$ is bounded from 0 on $\mathbb{W}$, the expression on the second line has a bound that is proportional to

$$\sup_{x\in\mathbb{X}}\left|\hat{F}_W\left(F_W^{-1}\left(\hat{F}_X(x)\right)\right)-\hat{F}_X(x)-\hat{F}_W\left(F_W^{-1}\left(F_X(x)\right)\right)+F_X(x)\right|$$

We rewrite this as

$$\sup_{x\in\mathbb{X}}\left|\hat{F}_W\left(F_W^{-1}\left(\hat{F}_X(x)\right)\right)-\hat{F}_W\left(F_W^{-1}\left(F_X(x)\right)\right)-\left(F_W\left(F_W^{-1}\left(\hat{F}_X(x)\right)-F_W\left(F_W^{-1}\left(F_X(x)\right)\right)\right)\right)\right|\leq$$

$$\sup_{x\in\mathbb{X}}\left|\hat{F}_W\left(F_W^{-1}\left(\hat{F}_X(x)\right)\right)-\hat{F}_W\left(F_W^{-1}\left(F_X(x)\right)\right)-\left(F_W\left(F_W^{-1}\left(\hat{F}_X(x)\right)-F_W\left(F_W^{-1}\left(F_X(x)\right)\right)\right)\right)\right|\times$$
$$1_{\sup_{x\in\mathbb{X}}\left|F_W^{-1}(\hat{F}_X(x))-F_W^{-1}(F_X(x))\right|\leq N^{-\delta}}+4\cdot 1_{\sup_{x\in\mathbb{X}}\left|F_W^{-1}(\hat{F}_X(x))-F_W^{-1}(F_X(x))\right|>N^{-\delta}}$$

By Lemma A.7 and the mean value theorem, the final term is $o_p(1)$ if $1/3<\delta<1/2$. By

$$F_W^{-1}\left(\hat{F}_X(x)\right)=F_W^{-1}\left(F_X(x)\right)+\left[F_W^{-1}\left(\hat{F}_X(x)\right)-F_W^{-1}\left(F_X(x)\right)\right]$$

and defining $\overline{w}=F_W^{-1}\left(F_X(x)\right)$ and $\tilde{w}=F_W^{-1}\left(\hat{F}_X(x)\right)-F_W^{-1}\left(F_X(x)\right)$ we have that the first term on the right hand side is bounded by

$$\sup_{\overline{w}\in\mathbb{W},|\tilde{w}|\leq N^{-\delta},\tilde{w}+\overline{w}\in\mathbb{W}}\left|\hat{F}_W\left(\overline{w}+\tilde{w}\right)-\hat{F}_W\left(\overline{w}\right)-\left(F_W\left(\overline{w}+\tilde{w}\right)-F_W\left(\overline{w}\right)\right)\right|=o_p(N^{-2/3})$$

by Lemma A.5 with $1/3<\delta<1/2,\eta=2/3$, so that we finally conclude that the sum of (B.7) and (B.8) is $o_p(N^{-1/2})$. □

Proof of Lemma A.16: The proof involves checking the conditions for Theorem A.2 from IR (given in Appendix A in the current paper), and simplifying the conclusions from that Theorem to the case at hand.
Define

$$h_1(w,x)=f_{WX}(w,x),\qquad\text{and}\qquad h_2(w,x)=f_{WX}(w,x)\cdot g(w,x),$$
$$n(h)=\frac{h_2}{h_1},$$

so that

$$\omega(x)=1,$$
$$\frac{\partial n}{\partial h_1}(h)=-\frac{h_2}{(h_1)^2}=-\frac{g(\left(F_W^{-1}\left(F_X(x)\right),x\right)}{f_{WX}(\left(F_W^{-1}\left(F_X(x)\right),x\right)},$$
$$\frac{\partial n}{\partial h_2}(h)=\frac{1}{h_1}=\frac{1}{f_{WX}(\left(F_W^{-1}\left(F_X(x)\right),x\right)},$$
$$t(x)=F_W^{-1}\left(F_X(x)\right),\qquad\frac{\partial}{\partial x}t(x)=\frac{f_X(x)}{f_W(F_W^{-1}(F_X(x)))},$$
$$\alpha_1(x)=-g(\left(F_W^{-1}\left(F_X(x)\right),x\right),\qquad\alpha_2(x)=1.$$

With $\tilde{Y}_i=(\tilde{Y}_{i1}\ \tilde{Y}_{i2})'=(1\ Y_i)'$, we have

$$\alpha(x)'\tilde{y}=y-g(\left(F_W^{-1}\left(F_X(x)\right),x\right).$$

Applying the results in Theorem A.2, we have

$$\int_{\mathbb{U}_2}K\left(\frac{W_i-t(X_i)}{b_N}+\frac{\partial t}{\partial x}(X_i)\cdot u_2,u_2\right)du_2=\int_u K\left(u,\frac{W_i-F_W^{-1}(F_X(X_i))}{b_N}+\frac{f_X(X_i)}{f_W(F_W^{-1}(F_X(X_i)))}\cdot u\right)du,$$

Substituting this into the result from Theorem A.2 we get

$$
\begin{aligned}
&\sqrt{N}\left(\hat{\theta}_g^{\text{pam}} - \overline{g}^{\text{pam}}\right)\\
&= \frac{1}{\sqrt{N} b_N} \sum_{i=1}^{N} \left(\left(Y_i - g\left(F_W^{-1}\left(F_X(X_i)\right), X_i\right)\right) \cdot \int_u K\left(\frac{W_i - F_W^{-1}\left(F_X(X_i)\right)}{b_N} + \frac{f_X(X_i)}{f_W\left(F_W^{-1}\left(F_X(X_i)\right)\right)} \cdot u, u \right) du \right.\\
&\qquad \left. - \mathbb{E}\left[\left(Y - g\left(F_W^{-1}\left(F_X(X)\right), X\right)\right) \cdot \int_u K\left(\frac{W - F_W^{-1}\left(F_X(X)\right)}{b_N} + \frac{f_X(X)}{f_W\left(F_W^{-1}\left(F_X(X)\right)\right)} \cdot u, u \right) du \right] \right)\\
&\qquad + o_p(1).
\end{aligned}
$$

Adding and subtracting $g(W_i, X_i)$ in both terms, this is equal to

$$
\begin{aligned}
&\frac{1}{\sqrt{N} b_N} \sum_{i=1}^{N} \left\{ \left(Y_i - g(W_i, X_i)\right) \cdot \int_u K\left(\frac{W_i - F_W^{-1}\left(F_X(X_i)\right)}{b_N} + \frac{f_X(X_i)}{f_W\left(F_W^{-1}\left(F_X(X_i)\right)\right)} \cdot u, u \right) du \right.\\
&\qquad \left. - \mathbb{E}\left[\left(Y - g(W, X)\right) \cdot \int_u K\left(\frac{W - F_W^{-1}\left(F_X(X)\right)}{b_N} + \frac{f_X(X)}{f_W\left(F_W^{-1}\left(F_X(X)\right)\right)} \cdot u, u \right) du \right] \right\}\\
&+ \frac{1}{\sqrt{N} b_N} \sum_{i=1}^{N} \left\{ \left(g(W_i, X_i) - g\left(F_W^{-1}\left(F_X(X_i)\right), X_i\right)\right) \cdot \int_u K\left(\frac{W_i - F_W^{-1}\left(F_X(X_i)\right)}{b_N} + \frac{f_X(X_i)}{f_W\left(F_W^{-1}\left(F_X(X_i)\right)\right)} \cdot u, u \right) du \right.\\
&\qquad \left. - \mathbb{E}\left[\left(g(W, X) - g\left(F_W^{-1}\left(F_X(X)\right), X\right)\right) \cdot \int_u K\left(\frac{W - F_W^{-1}\left(F_X(X)\right)}{b_N} + \frac{f_X(X)}{f_W\left(F_W^{-1}\left(F_X(X)\right)\right)} \cdot u, u \right) du \right] \right\}\\
&\qquad + o_p(1).\\
&= \frac{1}{\sqrt{N} b_N} \sum_{i=1}^{N} \left(Y_i - g(W_i, X_i)\right) \cdot \int_{u_2} K\left(\frac{W_i - F_W^{-1}\left(F_X(X_i)\right)}{b_N} + \frac{f_X(X_i)}{f_W\left(F_W^{-1}\left(F_X(X_i)\right)\right)} \cdot u_2, u_2 \right) du_2\\
&+ \frac{1}{\sqrt{N} b_N} \sum_{i=1}^{N} \left\{ \left(g(W_i, X_i) - g(F_W^{-1}(F_X(X_i)), X_i)\right) \cdot \int_{u_2} K\left(\frac{W_i - F_W^{-1}\left(F_X(X_i)\right)}{b_N} + \frac{f_X(X_i)}{f_W\left(F_W^{-1}\left(F_X(X_i)\right)\right)} \cdot u_2, u_2 \right) du_2 \right.\\
&\quad \left. - \mathbb{E}\left[\left(g(W, X) - g(F_W^{-1}(F_X(X)), X)\right) \cdot \int_{u_2} K\left(\frac{W_i - F_W^{-1}\left(F_X(X)\right)}{b_N} + \frac{f_X(X)}{f_W\left(F_W^{-1}\left(F_X(X)\right)\right)} \cdot u_2, u_2 \right) du_2 \right] \right\}\\
&\qquad + o_p(1).
\end{aligned}
$$

Having checked the conditions for Theorem A.2, the second part of the result in the Lemma follows directly from the second part of the Theorem. □

Proof of Lemma A.17: We prove the result in three parts. First, we show

$$
\begin{aligned}
&\frac{1}{N} \sum_{i=1}^{N} g\left(\hat{F}_W^{-1}\left(F_X(X_i)\right), X_i\right) - \frac{1}{N} \sum_{i=1}^{N} g\left(F_W^{-1}\left(F_X(X_i)\right), X_i\right)\\
&\qquad = \frac{1}{N} \sum_{i=1}^{N} g_W\left(F_W^{-1}\left(F_X(X_i)\right), X_i\right) \cdot \left(\hat{F}_W^{-1}\left(F_X(X_i)\right) - F_W^{-1}\left(F_X(X_i)\right)\right) + o_p\left(N^{-1/2}\right)
\end{aligned} \tag{B.16}
$$

Second, we will prove that

$$
\begin{aligned}
&\frac{1}{N} \sum_{i=1}^{N} g_W\left(F_W^{-1}\left(F_X(X_i)\right), X_i\right) \cdot \left(\hat{F}_W^{-1}\left(F_X(X_i)\right) - F_W^{-1}\left(F_X(X_i)\right)\right)\\
&= \frac{1}{N} \sum_{i=1}^{N} \frac{g_W\left(F_W^{-1}\left(F_X(X_i)\right), X_i\right)}{f_W\left(F_W^{-1}\left(F_X(X_i)\right)\right)} \cdot \left(\hat{F}_W\left(F_W^{-1}\left(F_X(X_i)\right)\right) - F_X(X_i)\right) + o_p\left(N^{-1/2}\right).
\end{aligned} \tag{B.17}
$$

Third, we will show that

$$
\frac{1}{N} \sum_{i=1}^{N} \frac{g_W\left(F_W^{-1}\left(F_X(X_i)\right), X_i\right)}{f_W\left(F_W^{-1}\left(F_X(X_i)\right)\right)} \cdot \left(\hat{F}_W\left(F_W^{-1}\left(F_X(X_i)\right)\right) - F_X(X_i)\right)
$$

$$= \frac{1}{N}\sum_{i=1}^{N}\psi_W^{\text{pam}}(W_i) + o_p\left(N^{-1/2}\right). \tag{B.18}$$

Together these three claims, (B.16)-(B.18), imply the result in the Lemma.
First we prove (B.16).

$$\Bigg|\frac{1}{N}\sum_{i=1}^{N} g\left(\hat{F}_W^{-1}\left(F_X\left(X_i\right)\right), X_i\right) - \frac{1}{N}\sum_{i=1}^{N} g\left(F_W^{-1}\left(F_X\left(X_i\right)\right), X_i\right)$$

$$-\frac{1}{N}\sum_{i=1}^{N} g_W\left(F_W^{-1}\left(F_X\left(X_i\right)\right), X_i\right)\cdot\left(\hat{F}_W^{-1}\left(F_X\left(X_i\right)\right) - F_W^{-1}\left(F_X\left(X_i\right)\right)\right)\Bigg|$$

$$\leq \sup_{x\in\mathbb{X}}\Big|g\left(\hat{F}_W^{-1}\left(F_X\left(x\right)\right), x\right) - g\left(F_W^{-1}\left(F_X\left(x\right)\right), x\right)$$

$$-g_W\left(F_W^{-1}\left(F_X\left(x\right)\right), x\right)\cdot\left(\hat{F}_W^{-1}\left(F_X\left(x\right)\right) - F_W^{-1}\left(F_X\left(x\right)\right)\right)\Big|$$

$$\leq \frac{1}{2}\cdot\sup_{w\in\mathbb{W},x\in\mathbb{X}}\left|\frac{\partial^2}{\partial w^2}g(w,x)\right|\cdot\sup_{q\in[0,1]}\left|\hat{F}_W^{-1}\left(q\right) - F_W^{-1}\left(q\right)\right|^2.$$

By Lemma A.3 it follows that for all $\delta < 1/2$, $\sup_{q\in[0,1]} N^{\delta}\cdot\left|\hat{F}_W^{-1}\left(q\right) - F_W^{-1}\left(q\right)\right| = o_p(1)$. In combination with the fact that $\frac{\partial^2 g}{\partial w^2}(w,x)$ is bounded this implies that

$$\sup_{w\in\mathbb{W},x\in\mathbb{X}}\left|\frac{\partial^2 g}{\partial w^2}(w,x)\right|\cdot\sup_{q\in[0,1]}\left|\hat{F}_W^{-1}\left(q\right) - F_W^{-1}\left(q\right)\right|^2 = o_p\left(N^{-1/2}\right).$$

This finishes the proof of (B.16).
Next, we prove (B.17).

$$\Bigg|\frac{1}{N}\sum_{i=1}^{N} g_W\left(F_W^{-1}\left(F_X\left(X_i\right)\right), X_i\right)\cdot\left(\hat{F}_W^{-1}\left(F_X\left(X_i\right)\right) - F_W^{-1}\left(F_X\left(X_i\right)\right)\right)$$

$$+\frac{1}{N}\sum_{i=1}^{N}\frac{g_W\left(F_W^{-1}\left(F_X\left(X_i\right)\right), X_i\right)}{f_W\left(F_W^{-1}\left(F_X\left(X_i\right)\right)\right)}\cdot\left(\hat{F}_W\left(F_W^{-1}\left(F_X\left(X_i\right)\right)\right) - F_X\left(X_i\right)\right)\Bigg|$$

$$\leq \sup_{w\in\mathbb{W},x\in\mathbb{X},q\in[0,1]}\left|g_W\left(w,x\right)\cdot\left(\hat{F}_W^{-1}\left(q\right) - F_W^{-1}\left(q\right)\right) + \frac{g_W\left(w,x\right)}{f_W\left(F_W^{-1}(q)\right)}\cdot\left(\hat{F}_W\left(F_W^{-1}\left(q\right)\right) - q\right)\right|$$

$$\leq \sup_{w\in\mathbb{W},x\in\mathbb{X}}\left|g_W\left(w,x\right)\right|\cdot\sup_{q\in[0,1]}\left|\left(\hat{F}_W^{-1}\left(q\right) - F_W^{-1}\left(q\right)\right) + \frac{1}{f_W\left(F_W^{-1}(q)\right)}\cdot\left(\hat{F}_W\left(F_W^{-1}\left(q\right)\right) - q\right)\right|,$$

so that Lemma A.6 implies that (B.17) holds.
Finally, let us prove (B.18).

$$\frac{1}{N}\sum_{i=1}^{N}\frac{g_W\left(F_W^{-1}\left(F_X\left(X_i\right)\right), X_i\right)}{f_W\left(F_W^{-1}\left(F_X\left(X_i\right)\right)\right)}\cdot\left(\hat{F}_W\left(F_W^{-1}\left(F_X\left(X_i\right)\right)\right) - F_X\left(X_i\right)\right)$$

$$= \frac{1}{N}\sum_{i=1}^{N}\frac{g_W\left(F_W^{-1}\left(F_X\left(X_i\right)\right), X_i\right)}{f_W\left(F_W^{-1}\left(F_X\left(X_i\right)\right)\right)}\cdot\left(\frac{1}{N}\sum_{j=1}^{N}1_{W_j\leq F_W^{-1}\left(F_X\left(X_i\right)\right)} - F_X\left(X_i\right)\right)$$

$$= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{g_W\left(F_W^{-1}\left(F_X\left(X_i\right)\right), X_i\right)}{f_W\left(F_W^{-1}\left(F_X\left(X_i\right)\right)\right)}\cdot\left(1_{F_W\left(W_j\right)\leq F_X\left(X_i\right)} - F_X\left(X_i\right)\right).$$

This is a two-sample V-statistic. The projection is the sample average of the sum of the expectation over W_j if we fix $X_i = x$ (this expectation is zero), and the expectation over X_i if we fix $W_j = w$, which gives $\psi_W^{\text{pam}}(w)$. Thus,

$$\frac{1}{N}\sum_{i=1}^{N}\frac{g_W\left(F_W^{-1}\left(F_X\left(X_i\right)\right), X_i\right)}{f_W\left(F_W^{-1}\left(F_X\left(X_i\right)\right)\right)}\cdot\left(\hat{F}_W\left(F_W^{-1}\left(F_X\left(X_i\right)\right)\right) - F_X\left(X_i\right)\right) = \frac{1}{N}\sum_{i=1}^{N}\psi_W^{\text{pam}}(W_i) + o_p\left(N^{-1/2}\right),$$

which is the claim in (B.18). $\square$

Proof of Lemma A.18: We prove this result in two steps. First we prove

$$\Bigg| \frac{1}{N}\sum_{i=1}^{N} g\left(F_w^{-1}\left(\hat{F}_X\left(X_i\right)\right), X_i\right) - \frac{1}{N}\sum_{i=1}^{N} g\left(F_w^{-1}\left(F_X\left(X_i\right)\right), X_i\right)$$

$$-\frac{g_W\left(F_W^{-1}\left(F_X\left(X_i\right)\right), X_i\right)}{f_w\left(F_W^{-1}\left(F_X\left(X_i\right)\right)\right)} \cdot \left(\hat{F}_X(X_i) - F_X(X_i)\right)\Bigg| = o_p\left(N^{-1/2}\right). \quad \text{(B.19)}$$

Second, we prove

$$\frac{1}{N}\sum_{i=1}^{N} \frac{g_W\left(F_W^{-1}\left(F_X\left(X_i\right)\right), X_i\right)}{f_W\left(F_W^{-1}\left(F_X\left(X_i\right)\right)\right)} \cdot \left(\hat{F}_X(X_i) - F_X(X_i)\right) = \frac{1}{N}\sum_{i=1}^{N} \psi_X^{\text{pam}}(X_i) + o_p\left(N^{-1/2}\right). \quad \text{(B.20)}$$

Together these two results imply the claim in Lemma A.18.

First we prove (B.19). By a second order Taylor series expansion, using the fact that $g(w,x)$ is at least twice continuously differentiable,

$$\Bigg| \frac{1}{N}\sum_{i=1}^{N} g\left(F_w^{-1}\left(\hat{F}_X\left(X_i\right)\right), X_i\right) - \frac{1}{N}\sum_{i=1}^{N} g\left(F_w^{-1}\left(F_X\left(X_i\right)\right), X_i\right)$$

$$- \frac{g_W\left(F_W^{-1}\left(F_X\left(X_i\right)\right), X_i\right)}{f_W\left(F_W^{-1}\left(F_X\left(X_i\right)\right)\right)} \cdot \left(\hat{F}_X(X_i) - F_X(X_i)\right)\Bigg|$$

$$\leq \sup_{x\in\mathbb{X}} \left| g\left(F_W^{-1}\left(\hat{F}_X\left(x\right)\right), x\right) - g\left(F_W^{-1}\left(F_X\left(x\right)\right), x\right) - \frac{g_W\left(F_W^{-1}\left(F_X\left(x\right)\right), x\right)}{f_W\left(F_W^{-1}\left(F_X\left(x\right)\right)\right)} \cdot \left(\hat{F}_X(x) - F_X(x)\right)\right|$$

$$\leq \frac{1}{2} \sup_{w\in\mathbb{W}, x\in\mathbb{X}} \left| \frac{\frac{\partial^2 g}{\partial w^2}(w,x)}{f_W(w)} - \frac{g_W(w,x)\cdot\frac{\partial f}{\partial w}(w)}{(f_W(w))^2}\right| \sup_{x\in\mathbb{X}} \left|\hat{F}_X(x) - F_X(x)\right|^2 = o_p\left(N^{-1/2}\right),$$

by Lemma A.3. This finishes the proof of (B.19).

Second we prove (B.20).

$$\frac{1}{N}\sum_{i=1}^{N} \frac{g_W\left(F_W^{-1}\left(F_X\left(X_i\right)\right), X_i\right)}{f_W\left(F_W^{-1}\left(F_X\left(X_i\right)\right)\right)} \cdot \left(\hat{F}_X(X_i) - F_X(X_i)\right)$$

$$= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} \frac{g_W\left(F_W^{-1}\left(F_X\left(X_i\right)\right), X_i\right)}{f_W\left(F_W^{-1}\left(F_X\left(X_i\right)\right)\right)} \cdot \left(\mathbf{1}_{X_j\leq X_i} - F_X(X_i)\right)$$

This is a one-sample V-statistic. To obtain the projection we first fix $X_i = x$ and take the expectation over X_j. This gives 0 for all x. Second, we fix $X_j = x$ and take the expectation over X_i. This gives $\psi_X^{\text{pam}}(x)$ defined above. This finishes the proof of (B.20), and thus completes the proof of Lemma A.18. $\square$

Proof of Lemma A.19: Adding and subtracting terms we have

$$\hat{\beta}^{\text{lc}} - \beta^{\text{lc}}$$

$$= \frac{1}{N}\sum_{i=1}^{N} \frac{\partial \hat{g}}{\partial w}(W_i, X_i)\cdot d(W_i)\cdot (X_i - \hat{m}(W_i)) - \frac{1}{N}\sum_{i=1}^{N} \frac{\partial g}{\partial w}(W_i, X_i)\cdot d(W_i)\cdot (X_i - \hat{m}(W_i)) \quad \text{(B.21)}$$

$$-\left(\frac{1}{N}\sum_{i=1}^{N} \frac{\partial \hat{g}}{\partial w}(W_i, X_i)\cdot d(W_i)\cdot (X_i - m(W_i)) - \frac{1}{N}\sum_{i=1}^{N} \frac{\partial g}{\partial w}(W_i, X_i)\cdot d(W_i)\cdot (X_i - m(W_i))\right) \quad \text{(B.22)}$$

$$+\frac{1}{N}\sum_{i=1}^{N} \frac{\partial \hat{g}}{\partial w}(W_i, X_i)\cdot d(W_i)\cdot (X_i - m(W_i)) - \frac{1}{N}\sum_{i=1}^{N} \frac{\partial g}{\partial w}(W_i, X_i)\cdot d(W_i)\cdot (X_i - m(W_i)) \quad \text{(B.23)}$$

$$+\frac{1}{N}\sum_{i=1}^{N} \frac{\partial g}{\partial w}(W_i, X_i)\cdot d(W_i)\cdot (X_i - \hat{m}(W_i)) - \frac{1}{N}\sum_{i=1}^{N} \frac{\partial g}{\partial w}(W_i, X_i)\cdot d(W_i)\cdot (X_i - m(W_i)) \quad \text{(B.24)}$$

$$+\frac{1}{N}\sum_{i=1}^{N} \frac{\partial g}{\partial w}(W_i, X_i)\cdot d(W_i)\cdot (X_i - m(W_i)) - \beta^{\text{lc}}. \quad \text{(B.25)}$$

Because (B.23) is equal to $\beta_g^{\text{lc}} - \overline{g}^{\text{lc}}$, (B.24) is equal to $\beta_m^{\text{lc}} - \overline{g}^{\text{lc}}$, and (B.25) is equal to $\overline{g}^{\text{lc}} - \beta^{\text{lc}}$, it follows that it is sufficient for the proof of Lemma A.19 to show that the sum of (B.21) and (B.22) is $o_p(N^{-1/2})$. We can write the sum of (B.21) and (B.22) as

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\partial \hat{g}}{\partial w}(W_i,X_i)\cdot d(W_i)\cdot(X_i-\hat{m}(W_i)) - \frac{1}{N}\sum_{i=1}^{N}\frac{\partial g}{\partial w}(W_i,X_i)\cdot d(W_i)\cdot(X_i-\hat{m}(W_i))$$

$$-\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\partial \hat{g}}{\partial w}(W_i,X_i)\cdot d(W_i)\cdot(X_i-m(W_i)) - \frac{1}{N}\sum_{i=1}^{N}\frac{\partial g}{\partial w}(W_i,X_i)\cdot d(W_i)\cdot(X_i-m(W_i))\right)$$

$$=\frac{1}{N}\sum_{i=1}^{N} d(W_i)\cdot\left(\frac{\partial \hat{g}}{\partial w}(W_i,X_i)-\frac{\partial g}{\partial w}(W_i,X_i)\right)\cdot(m(W_i)-\hat{m}(W_i))$$

$$\leq \sup_{w\in\mathbb{W}}|d(w)|\cdot\sup_{w\in\mathbb{W},x\in\mathbb{X}}\left|\frac{\partial \hat{g}}{\partial w}(w,x)-\frac{\partial g}{\partial w}(x,x)\right|\cdot\sup_{w\in\mathbb{W}}|\hat{m}(w)-m(w)| = C\cdot o_p\left(N^{-\eta}\right)\cdot o_p\left(N^{-\eta}\right),$$

for some $\eta > 1/4$, and so this expression is $o_p(N^{-1/2})$. □

Proof of Lemma A.20: The proof consists of checking the conditions for Theorem A.1, and specializing the result in Theorem A.1 to the case in the Lemma.

We apply Theorem A.1 with $z = (z_1\ z_2)' = (w\ x)'$, $Z_i = (W_i\ X_i)'$, $\omega(z) = d(z_1)\cdot(z_2 - m(z_1)) = d(w)\cdot(x - m(w))$ (so that $\omega(z)$ goes smoothly to zero on the boundary of $\mathbb{Z}$), $L = 2$, and $\lambda = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Then $\{\kappa : \kappa \leq \lambda\} = \{\kappa_0, \kappa_1\} = \left\{\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}\right\}$, and

$$h^{[\lambda]}(w,x) = \begin{pmatrix} h_1^{(\kappa_0)}(w,x) \\ h_2^{(\kappa_0)}(w,x) \\ h_1^{(\kappa_1)}(w,x) \\ h_2^{(\kappa_1)}(w,x) \end{pmatrix},$$

with

$$h_1^{(\kappa_0)}(w,x) = f_{WX}(w,x)$$

$$h_2^{(\kappa_0)}(w,x) = f_{WX}(w,x)\cdot g(w,x)$$

$$h_1^{(\kappa_1)}(w,x) = \frac{\partial}{\partial w}f_{WX}(w,x)$$

$$h_2^{(\kappa_1)}(w,x) = g(w,x)\cdot\frac{\partial}{\partial w}f_{WX}(w,x) + f_{WX}(w,,x)\cdot\frac{\partial}{\partial w}g(w,x).$$

The functional of interest is

$$n\left(h^{[\lambda]}\right) = \frac{\partial}{\partial w}g(\cdot) = \frac{h_2^{(\kappa_1)}}{h_1^{(\kappa_0)}} - \frac{h_2^{(\kappa_0)}\cdot h_1^{(\kappa_1)}}{\left(h_1^{(\kappa_0)}\right)^2}$$

The derivatives of this functional are

$$\frac{\partial}{\partial h_1^{(\kappa_0)}}n\left(h^{[\lambda]}\right) = -\frac{h_2^{(\kappa_1)}}{\left(h_1^{(\kappa_0)}\right)^2} + 2\frac{h_2^{(\kappa_0)}\cdot h_1^{(\kappa_1)}}{\left(h_1^{(\kappa_0)}\right)^3}$$

$$= -\frac{f_{WX}(w,x)\cdot\frac{\partial}{\partial w}g(w,x) + g(w,x)\cdot\frac{\partial}{\partial w}f_{WX}(w,x)}{\left(f_{WX}(w,x)\right)^2} + 2\frac{g(w,x)\cdot f_{WX}(w,x)\cdot\frac{\partial}{\partial w}f_{WX}(w,x)}{\left(f_{WX}(w,x)\right)^3}$$

$$= -\frac{\frac{\partial}{\partial w}g(w,x)}{f_{WX}(w,x)} + \frac{g(w,x)\cdot\frac{\partial}{\partial w}f_{WX}(w,x)}{\left(f_{WX}(w,x)\right)^2}$$

$$\frac{\partial}{\partial h_2^{(\kappa_0)}}n\left(h^{[\lambda]}\right) = -\frac{h_1^{(\kappa_1)}}{\left(h_1^{(\kappa_0)}\right)^2} = -\frac{\frac{\partial}{\partial w}f_{WX}(w,x)}{\left(f_{WX}(w,x)\right)^2}$$

$$\frac{\partial}{\partial h_1^{(\kappa_1)}} n\left(h^{[\lambda]}\right) = -\frac{h_2^{(\kappa_0)}}{\left(h_1^{(\kappa_0)}\right)^2} = -\frac{g(w,x)\cdot f_{WX}(w,x)}{(f_{WX}(w,x))^2} = -\frac{g(w,x)}{f_{WX}(w,x)}$$

$$\frac{\partial}{\partial h_2^{(\kappa_1)}} n\left(h^{[\lambda]}\right) = \frac{1}{h_1^{(\kappa_0)}} = \frac{1}{f_{WX}(w,x)}.$$

$$\begin{aligned}\alpha_{\kappa_0,1}(w,x) &= d(w)\cdot(x-m(w))\cdot f_W(w,x)\cdot\left(-\frac{\frac{\partial}{\partial w}g(w,x)}{f_{WX}(w,x)} + \frac{g(w,x)\cdot\frac{\partial}{\partial w}f_{WX}(w,x)}{(f_{WX}(w,x))^2}\right)\\ &= d(w)\cdot(x-m(w))\cdot\left(-\frac{\partial}{\partial w}g(w,x) + \frac{g(w,x)\cdot\frac{\partial}{\partial w}f_{WX}(w,x)}{f_{WX}(w,x)}\right)\end{aligned}$$

$$\alpha_{\kappa_0,2}(w,x) = d(w)\cdot(x-m(w))\cdot f_{WX}(w,x)\cdot\left(-\frac{\frac{\partial}{\partial w}f_{WX}(w,x)}{(f_{WX}(w,x))^2}\right) = -d(w)\cdot(x-m(w))\cdot\frac{\frac{\partial}{\partial w}f_{WX}(w,x)}{f_{WX}(w,x)}$$

$$\alpha_{\kappa_1,1}(w,x) = d(w)\cdot(x-m(w))\cdot f_{WX}(w,x)\cdot\left(-\frac{g(w,x)}{f_{WX}(w,x)}\right) = -d(w)\cdot(x-m(w))\cdot g(w,x)$$

$$\alpha_{\kappa_1,2}(w,x) = d(w)\cdot(x-m(w))\cdot f_{WX}(w,x)\cdot\frac{1}{f_{WX}(w,x)} = d(w)\cdot(x-m(w))$$

$$(-1)^{|\kappa_0|}\alpha_{\kappa_0,1}^{(\kappa_0)}(w,x) = \alpha_{\kappa_0,1}(w,x) = d(w)\cdot(x-m(w))\cdot\left(-\frac{\partial}{\partial w}g(w,x) + \frac{g(w,x)\cdot\frac{\partial}{\partial w}f_{WX}(w,x)}{f_{WX}(w,x)}\right)$$

$$(-1)^{|\kappa_0|}\alpha_{\kappa_0,2}^{(\kappa_0)}(w,x) = \alpha_{\kappa_0,2}(w,x) = -d(w)\cdot(x-m(w))\cdot\frac{\frac{\partial}{\partial w}f_{WX}(w,x)}{f_{WX}(w,x)}$$

$$\begin{aligned}(-1)^{|\kappa_1|}\alpha_{\kappa_1,1}^{(\kappa_1)}(w,x) &= \frac{\partial}{\partial w}\Big(d(w)\cdot(x-m(w))\cdot g(w,x)\Big)\\ &= d(w)\cdot(x-m(w))\cdot\frac{\partial}{\partial w}g(w,x) + g(w,x)\cdot(x-m(w))\cdot\frac{\partial}{\partial w}d(w) - g(w,x)\cdot d(w)\cdot\frac{\partial}{\partial w}m(w)\end{aligned}$$

$$(-1)^{|\kappa_1|}\alpha_{\kappa_1,2}^{(\kappa_1)}(w,x) = -\frac{\partial}{\partial w}\left(d(w)\cdot(x-m(w))\right) = -(x-m(w))\cdot\frac{\partial}{\partial w}d(w) + d(w)\cdot\frac{\partial}{\partial w}m(w)$$

Then

$$\begin{aligned}&\sum_{\kappa\le\lambda}(-1)^{|\kappa|}\sum_{m=1}^{2}\alpha_{\kappa m}^{(\kappa)}(w,x)\tilde{y}_{im}\\ &\quad= (-1)^{|\kappa_0|}\alpha_{\kappa_0,1}^{(\kappa_0)}(w,x) + Y_i\cdot(-1)^{|\kappa_0|}\alpha_{\kappa_0,2}^{(\kappa_0)}(w,x) + (-1)^{|\kappa_1|}\alpha_{\kappa_1,1}^{(\kappa_1)}(w,x) + Y_i\cdot(-1)^{|\kappa_1|}\alpha_{\kappa_1,2}^{(\kappa_1)}(w,x)\\ &\quad= d(w)\cdot(x-m(w))\cdot\left(-\frac{\partial}{\partial w}g(w,x) + \frac{g(w,x)\cdot\frac{\partial}{\partial w}f_{WX}(w,x)}{f_{WX}(w,x)}\right)\\ &\qquad -Y_i\cdot d(w)\cdot(x-m(w))\cdot\frac{\frac{\partial}{\partial w}f_{WX}(w,x)}{f_{WX}(w,x)}\\ &\qquad +d(w)\cdot(x-m(w))\cdot\frac{\partial}{\partial w}g(w,x) + g(w,x)\cdot(x-m(w))\cdot\frac{\partial}{\partial w}d(w) - g(w,x)\cdot d(w)\cdot\frac{\partial}{\partial w}m(w)\end{aligned}$$

$$+Y_i \cdot \left(-(x-m(w)) \cdot \frac{\partial}{\partial w} d(w) + d(w) \cdot \frac{\partial}{\partial w} m(w) \right)$$

$$= -(Y - g(W,X)) \cdot \left(\frac{\frac{\partial}{\partial w} f_{WX}(W,X)}{f_{WX}(W,X)} \cdot d(w) \cdot (X - m(W)) + (X - m(W)) \cdot \frac{\partial}{\partial w} d(W) - d(W) \cdot \frac{\partial}{\partial w} m(W) \right).$$

Since

$$\mathbb{E}\left[-(y - g(w,x)) \cdot \left(\frac{\frac{\partial}{\partial w} f_{WX}(w,x)}{f_{WX}(w,x)} \cdot d(w) \cdot (x - m(w)) + (x - m(w)) \cdot \frac{\partial}{\partial w} d(w) - d(w) \cdot \frac{\partial}{\partial w} m(w) \right) \right] = 0,$$

it follows that

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\sum_{\kappa \le \lambda} (-1)^{|\kappa|} \sum_{m=1}^{2} \mathbb{E}\left[\alpha_{\kappa m}^{(\kappa)}(W_i, X_i) \tilde{Y}_{im} \right] \right) = 0,$$

and therefore

$$\sqrt{N}(\hat{\beta}^{\mathrm{lc}} - \beta^{\mathrm{lc}}) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\sum_{\kappa \le \lambda} (-1)^{|\kappa|} \sum_{m=1}^{2} \alpha_{\kappa m}^{(\kappa)}(W_i, X_i) \tilde{Y}_{im} \right) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \psi^{\mathrm{lc}}(Y_i, W_i, X_i).$$

where

$$\psi^{\mathrm{lc}}(y,w,x) = -(y - g(w,x)) \cdot \left(\frac{\frac{\partial}{\partial w} f_{WX}(w,x)}{f_{WX}(w,x)} \cdot d(w) \cdot (x - m(w)) + (x - m(w)) \cdot \frac{\partial}{\partial w} d(w) - d(w) \cdot \frac{\partial}{\partial w} m(w) \right).$$

□.

Proof of Lemma A.21: We start with the inequality

$$\sup_{w \in \mathbb{W}} \left| \frac{1}{\widehat{f}_W(w)} \left(\widehat{f}_W(w) - f_W(w) \right)^2 \right| \le \frac{\left(\sup_{w \in \mathbb{W}} \left| \widehat{f}_W(w) - f_W(w) \right| \right)^2}{\inf_{w \in W} \left| \widehat{f}_W(w) \right|}.$$

Under the stated restriction on δ the bandwidth sequence satisfies

$$\frac{N^{1/4}}{\sqrt{\ln(N)}} b_N^{1/2} \to \infty, \qquad N^{1/4} b_N^s \to 0,$$

which, by Lemma A.11, implies

$$\left(\sup_{w \in \mathbb{W}} \left| \widehat{f}_W(w) - f_W(w) \right| \right)^2 = o_p\left(N^{-1/2} \right).$$

Now observe that the the denominator is bounded away from zero since, by the TI, we have $\left|\widehat{f}_W(w)\right| + |f_W(w)| \ge \left|\widehat{f}_W(w) - f_W(w)\right|$ and therefore $\inf_{w \in W} \left|\widehat{f}_W(w)\right| \ge \sup_{w \in \mathbb{W}} \left|\widehat{f}_W(w) - f_W(w)\right| - \inf_{w \in W} |f_W(w)| \ge \inf_{w \in W} |f_W(w)| - \sup_{w \in \mathbb{W}} \left|\widehat{f}_W(w) - f_W(w)\right|$. By Assumption 3.1 $\inf_{w \in W} |f_W(w)|$ is bounded away from zero, with the result then following. □

Proof of Lemma A.22: We start with the inequality

$$\sup_{w \in \mathbb{W}} \left| \frac{1}{\widehat{h}_2(w)} \left(\widehat{h}_1(w) - h_1(w) \right) \left(\widehat{h}_2(w) - h_2(w) \right) \right| \le \frac{\sup_{w \in \mathbb{W}} \left| \widehat{h}_1(w) - h_1(w) \right| \times \sup_{w \in \mathbb{W}} \left| \widehat{h}_2(w) - h_2(w) \right|}{\inf_{w \in W} \left| \widehat{h}_2(W_i) \right|}.$$

The remainder of the proof is along the lines of that to Lemma A.21. □

Proof of Lemma A.23: Let $h(w) = (h_1(w), h_2(w))' = (m(w) \cdot f_W(w), f_W(w))'$, then

$$\hat{\beta}_{\text{lc},m} = \frac{1}{N}\sum_{i=1}^{N} g_W(W_i, X_i) \cdot d(W_i) \cdot \left(X_i - \frac{\widehat{h}_{1,nip}(W_i)}{\widehat{h}_{2,nip}(W_i)} \right)$$

$$= \frac{1}{N}\sum_{i=1}^{N} g_W(W_i, X_i) \cdot d(W_i) \cdot (X_i - m(W_i)) \tag{B.26}$$

$$- \frac{1}{N}\sum_{i=1}^{N} g_W(W_i, X_i) \cdot d(W_i) \cdot \left(\frac{\widehat{h}_{1,nip}(W_i)}{\widehat{h}_{2,nip}(W_i)} - \frac{h_1(W_i)}{h_2(W_i)} \right). \tag{B.27}$$

Expanding the ratio[4] in (B.27) yields

$$\frac{1}{N}\sum_{i=1}^{N} g_W(W_i, X_i) \cdot d(W_i) \cdot \left(\frac{h_1(W_i)}{h_2(W_i)} - \frac{\widehat{h}_{1,nip}(W_i)}{\widehat{h}_{2,nip}(W_i)} \right)$$

$$= -\frac{1}{N}\sum_{i=1}^{N} g_W(W_i, X_i) \cdot d(W_i) \cdot \frac{1}{f_W(W_i)} \left(\widehat{h}_{1,nip}(W_i) - m(W_i)\widehat{h}_{2,nip}(W_i) \right) \tag{B.28}$$

$$- \frac{1}{N}\sum_{i=1}^{N} g_W(W_i, X_i) \cdot d(W_i) \cdot \frac{h_1(W_i)}{h_2(W_i)^2 \,\widehat{h}_{2,nip}(W_i)} \left(\widehat{h}_{2,nip}(W_i) - h_2(W_i) \right)^2 \tag{B.29}$$

$$- \frac{1}{N}\sum_{i=1}^{N} g_W(W_i, X_i) \cdot d(W_i) \cdot \frac{h_1(W_i)}{h_2(W_i)\,\widehat{h}_{2,nip}(W_i)} \left(\widehat{h}_{1,nip}(W_i) - h_1(W_i) \right) \left(\widehat{h}_{2,nip}(W_i) - h_2(W_i) \right). \tag{B.30}$$

First consider (B.29). By Lemma A.12,

$$\left| \frac{1}{N}\sum_{i=1}^{N} g_W(W_i, X_i) \cdot d(W_i) \cdot \frac{h_1(W_i)}{h_2(W_i)^2 \,\widehat{h}_{2,nip}(W_i)} \left(\widehat{h}_{2,nip}(W_i) - h_2(W_i) \right)^2 \right|$$

$$\leq \sup_{w \in \mathbb{W}, x \in \mathbb{X}} \left| \frac{1}{f_W(w)} g_W(w, x) \cdot d(w) \cdot m(w) \right| \sup_{w \in \mathbb{W}} \left| \frac{1}{\widehat{f}_W(w)} \left(\widehat{f}_W(w) - f_W(w) \right)^2 \right|$$

$$= o_p\left(N^{-1/2}\right)$$

if the NIP estimator is uniformly $o_p(N^{-1/4})$ which holds if $\frac{1}{4s} < \delta < \frac{1}{8}$. An analogous application of Lemma A.12 can be used to show that (B.30) is $o_p\left(N^{-1/2}\right)$ under the same condition.
Now consider (B.28) that we express as the sum of a variance and a bias term

$$-\frac{1}{N}\sum_{i=1}^{N} g_W(W_i, X_i) \cdot d(W_i) \cdot \frac{1}{f_W(W_i)} \left(\widehat{h}_{1,nip}(W_i) - \mathbb{E}\left[\widehat{h}_{1,nip}(W_i) \right] - m(W_i) \left(\widehat{h}_{2,nip}(W_i) - \mathbb{E}\left[\widehat{h}_{2,nip}(W_i) \right] \right) \right) +$$

$$\frac{1}{N}\sum_{i=1}^{N} g_W(W_i, X_i) \cdot d(W_i) \cdot \frac{1}{f_W(W_i)} \left(h_1(W_i) - \mathbb{E}\left[\widehat{h}_{1,nip}(W_i) \right] - m(W_i) \left(h_2(W_i) - \mathbb{E}\left[\widehat{h}_{2,nip}(W_i) \right] \right) \right)$$

The bias term is $O_p(N^{-1/2})$ if $\delta > \frac{1}{2s}$. After substitution of the NIP estimator the variance term is

$$-\frac{1}{N}\sum_{i=1}^{N} g_W(W_i, X_i) \cdot d(W_i) \cdot \frac{1}{f_W(W_i)} \sum_{j=0}^{s-1} \sum_{|\mu|=j} \frac{1}{\mu!}$$

$$\left(\hat{h}^{(\mu)}_{1,NW} - \mathbb{E}\left[\hat{h}^{(\mu)}_{1,NW}(r_b(W_i)) \right] - m(W_i) \left(\hat{h}^{(\mu)}_{2,NW}(r_b(W_i)) - \mathbb{E}\left[\hat{h}^{(\mu)}_{2,NW}(r_b(W_i)) \right] \right) (r_b(W_i))(W_i - r_b(W_i))^{\mu} \right)$$

---

[4] The ratio expansion is of the form

$$\frac{\widehat{a}}{\widehat{b}} - \frac{a}{b} = \frac{1}{b}\left(\widehat{a} - \frac{a}{b}\widehat{b} \right) + \frac{a}{b^2 \widehat{b}}(\widehat{b} - b)^2 - \frac{a}{b\widehat{b}}(\widehat{a} - a)(\widehat{b} - b).$$

We consider separately

$$-\frac{1}{N}\sum_{i=1}^{N} g_W\left(W_i,X_i\right)\cdot d(W_i)\cdot\frac{1}{f_W\left(W_i\right)}\sum_{j=0}^{s-1}\sum_{|\mu|=j}\frac{1}{\mu!}\left(\hat{h}^{(\mu)}_{1,NW}(r_b(W_i))-\mathbb{E}\left[\hat{h}^{(\mu)}_{1,NW}(r_b(W_i))\right]\right)(W_i-r_b(W_i))^{\mu} \quad \text{(B.31)}$$

and

$$\frac{1}{N}\sum_{i=1}^{N} g_W\left(W_i,X_i\right)\cdot d(W_i)\cdot\frac{m(W_i)}{f_W\left(W_i\right)}\sum_{j=0}^{s-1}\sum_{|\mu|=j}\frac{1}{\mu!}\left(\hat{h}^{(\mu)}_{2,NW}(r_b(W_i))-\mathbb{E}\left[\hat{h}^{(\mu)}_{2,NW}(r_b(W_i))\right]\right)(W_i-r_b(W_i))^{\mu} \quad \text{(B.32)}$$

We show that (B.31) is asymptotically equivalent to an average. The same method shows that (B.32) is also asymptotically equivalent to an average, but we omit the details. The expression (B.31) is a linear combination of terms

$$D_\mu=-\frac{1}{N}\sum_{i=1}^{N} g_W\left(W_i,X_i\right)\cdot d(W_i)\cdot\frac{1}{f_W\left(W_i\right)}\left(\hat{h}^{(\mu)}_{1,NW}(r_b(W_i))-\mathbb{E}\left[\hat{h}^{(\mu)}_{1,NW}(r_b(W_i))\right]\right)(W_i-r_b(W_i))^{\mu}=$$

$$-\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} a_{N,\mu}(W_i,X_i,X_j,W_j)$$

with

$$a_{N,\mu}(W_i,X_i,X_j,W_j)=\frac{g_W\left(W_i,X_i\right)d(W_i)}{f_W\left(W_i\right)}\left(\frac{1}{b_N^{1+|\mu|}}X_jK^{(\mu)}\left(\frac{W_j-r_{b_N}(W_i)}{b_N}\right)-\right.$$

$$\left.\mathbb{E}\left[\frac{1}{b_N^{1+|\mu|}}XK^{(\mu)}\left(\frac{W-r_{b_N}(W_i)}{b_N}\right)\right]\right)(W_i-r_{b_N}(W_i))^{\mu}$$

Therefore D_μ is a V-statistic with a kernel that depends on N so that the usual projection theorem does not apply directly. Instead we derive the projection directly. First we bound the second moments of $a_{N,\mu}(W_i,X_i,X_j,W_j)$. For $j\neq i$ we have

$$\mathbb{E}\left[a_{N,\mu}(W_i,X_i,X_j,W_j)^2\right]\leq C\frac{\sup_{w\in\mathbb{W}}|w-r_b(w)|^{2|\mu|}}{b_N^{2|\mu|+2}}\mathbb{E}\left[X_j^2K^{(\mu)}\left(\frac{W_j-r_{b_N}(W_i)}{b_N}\right)^2\right]\leq$$

$$\frac{C}{b_N^2}\mathbb{E}\left[K^{(\mu)}\left(\frac{W_j-r_{b_N}(W_i)}{b_N}\right)^2\right]$$

because the conditional variance of X given W is bounded. Because given $W_i=\tilde{w}$

$$\mathbb{E}\left[K^{(\mu)}\left(\frac{W_j-r_{b_N}(W_i)}{b_N}\right)^2|W_i=\tilde{w}\right]=\int_{\mathbb{W}}K^{(\mu)}\left(\frac{w-r_{b_N}(\tilde{w})}{b_N}\right)^2f_W(w)\mathrm{d}w$$

we have by a change of variables to $t=(w-r_{b_N}(\tilde{w}))/b_N$ with Jacobian b_N and the boundedness of $K^{(\mu)}(t)$ and $f_W(w)$ that this integral is bounded by Cb_N and we conclude

$$\mathbb{E}\left[a_{N,\mu}(W_i,X_i,X_j,W_j)^2\right]=O(b_N^{-1})$$

For $j=i$ we have

$$\mathbb{E}\left[a_{N,\mu}(W_i,X_i,X_i,W_i)^2\right]=\frac{1}{b_N^{2+2|\mu|}}\mathbb{E}\left[\frac{g_W\left(W_i,X_i\right)^2d(W_i)^2}{f_W\left(W_i\right)^2}X_i^2K^{(\mu)}\left(\frac{W_i-r_{b_N}(W_i)}{b_N}\right)^2(W_i-r_{b_N}(W_i))^{2\mu}\right]+$$

$$\frac{1}{b_N^{2+2|\mu|}}\mathbb{E}\left[\frac{g_W\left(W_i,X_i\right)^2d(W_i)^2}{f_W\left(W_i\right)^2}\mathbb{E}\left[XK^{(\mu)}\left(\frac{W-r_{b_N}(W_i)}{b_N}\right)\right]^2(W_i-r_{b_N}(W_i))^{2\mu}\right]-$$

$$\frac{2}{b_N^{2+2|\mu|}}\mathbb{E}\left[\frac{g_W\left(W_i,X_i\right)^2d(W_i)^2}{f_W\left(W_i\right)^2}X_iK^{(\mu)}\left(\frac{W_i-r_{b_N}(W_i)}{b_N}\right)\mathbb{E}\left[XK^{(\mu)}\left(\frac{W-r_{b_N}(W_i)}{b_N}\right)\right](W_i-r_{b_N}(W_i))^{2\mu}\right]$$

The first term on the right hand side is bounded by

$$\frac{C}{b_N^2}\mathbb{E}\left[K^{(\mu)}\left(\frac{W_i - r_{b_N}(W_i)}{b_N}\right)^2\right] = \frac{C}{b_N^2}\int_{\mathbb{W}_{b_N}^I} K^{(\mu)}\left(\frac{w - r_{b_N}(w)}{b_N}\right)^2 f_W(w)\mathrm{d}w +$$

$$\frac{C}{b_N^2}\int_{\mathbb{W}\setminus\mathbb{W}_{b_N}^I} K^{(\mu)}\left(\frac{w - r_{b_N}(w)}{b_N}\right)^2 f_W(w)\mathrm{d}w$$

with $\mathbb{W}_{b_N}^I$ the internal set of the support. Because the argument of $K^{(\mu)}$ is 0 on the interior set, the first integral is obviously $O(b_N^{-2})$. The second integral is

$$\frac{C}{b_N^2}\int_{w_l}^{w_l+b_N} K^{(\mu)}\left(\frac{w - w_l}{b_N} - 1\right)^2 f_W(w)\mathrm{d}w + \frac{C}{b_N^2}\int_{w_u-b_N}^{w_u} K^{(\mu)}\left(\frac{w - w_u}{b_N} + 1\right)^2 f_W(w)\mathrm{d}w$$

Because the kernel has support $[-1,1]$ and its derivatives op to order μ are bounded so that

$$K^{(\mu)}\left(\frac{w - w_l}{b_N} - 1\right)^2 \leq C1_{w_l\leq w\leq w_l+2b_N} \qquad K^{(\mu)}\left(\frac{w - w_u}{b_N} + 1\right)^2 \leq C1_{w_u-2b_N\leq w\leq w_u}$$

so that the second integral by the boundedness of f_W is $O(b_N^{-1})$. The second term on the right hand side is bounded by

$$\frac{C}{b_N^2}\int_{\mathbb{W}}\left(\int_{\mathbb{W}} K^{(\mu)}\left(\frac{w - r_{b_N}(\tilde{w})}{b_N}\right) f_W(w)\mathrm{d}w\right)^2 f_W(\tilde{w})\mathrm{d}\tilde{w} \leq \frac{C}{b_N^2}\int_{\mathbb{W}}\int_{\mathbb{W}} K^{(\mu)}\left(\frac{w - r_{b_N}(\tilde{w})}{b_N}\right)^2 f_W(w)f_W(\tilde{w})\mathrm{d}w\mathrm{d}\tilde{w}$$

This integral is $O(b_N^{-1})$ by a change of variables with Jacobian b_N in the inner integral. The third term on the right hand side is bounded by

$$\frac{C}{b_N^2}\left|\int_{\mathbb{W}} K^{(\mu)}\left(\frac{\tilde{w} - r_{b_N}(\tilde{w})}{b_N}\right)\int_{\mathbb{W}} K^{(\mu)}\left(\frac{w - r_{b_N}(\tilde{w})}{b_N}\right) f_W(w)\mathrm{d}w f_W(\tilde{w})\mathrm{d}\tilde{w}\right| = O(b_N^{-1})$$

by a change of variables in the inner integral. We conclude

$$\mathbb{E}\left[a_{N,\mu}(W_i, X_i, X_i, W_i)^2\right] = O(b_N^{-2})$$

The next step is to express D_μ as an average. Define

$$c_{N,\mu}(X_j, W_j) = \frac{1}{b_N^{1+|\mu|}}\cdot$$

$$\int_{\mathbb{X}}\int_{\mathbb{W}} \frac{g_W(w,x)\,d(x)}{f_W(w)}\left(X_j K^{(\mu)}\left(\frac{W_j - r_{b_N}(w)}{b_N}\right) - \mathbb{E}\left[XK^{(\mu)}\left(\frac{W - r_{b_N}(w)}{b_N}\right)\right]\right)(w - r_{b_N}w))^\mu f_{WX}(w,x)\mathrm{d}w\mathrm{d}x$$

and

$$E_\mu = -\frac{1}{N}\sum_{j=1}^N c_{N,\mu}(X_j, W_j)$$

Then

$$D_\mu - E_\mu = \frac{N(N-1)}{N^2}(D_{\mu,1} - E_\mu) + \left(\frac{N(N-1)}{N^2} - 1\right)E_\mu + D_{\mu,2}$$

with

$$D_{\mu,1} = -\frac{1}{N(N-1)}\sum_{i\neq j=1}^N a_{N,\mu}(W_i, X_i, X_j, W_j) \qquad D_{\mu,2} = -\frac{1}{N^2}\sum_{i=1}^N a_{N,\mu}(W_i, X_i, X_i, W_i)$$

Now

$$D_{\mu,1} - E_\mu = -\frac{1}{N(N-1)}\sum_{i\neq j=1}^N (a_{N,\mu}(W_i, X_i, X_j, W_j) - c_{N,\mu}(X_j, W_j))$$

with

$$\mathbb{E}[(a_{N,\mu}(W_i,X_i,X_j,W_j)-c_{N,\mu}(X_j,W_j))(a_{N,\mu}(W_{i'},X_{i'},X_{j'},W_{j'})-c_{N,\mu}(X_{j'},W_{j'}))]=0$$

if (i) $i\neq i'$, $j\neq j'$, (ii) $i=i'$, $j\neq j'$ (iii) $i\neq i'$, $j=j'$, because

$$\begin{aligned}
\mathbb{E}[a_{N,\mu}(W_i,X_i,X_j,W_j)|W_i,X_i] &= 0\\
\mathbb{E}[a_{N,\mu}(W_i,X_i,X_j,W_j)] &= 0\\
\mathbb{E}[a_{N,\mu}(W_i,X_i,X_j,W_j)|X_j,W_j] &= c_{N,\mu}(W_j,X_j)\\
\mathbb{E}[c_{N,\mu}(W_j,X_j)] &= 0
\end{aligned}$$

Therefore

$$\mathbb{E}[(D_\mu-E_\mu)^2]=$$

$$\frac{1}{N^2(N-1)^2}\sum_{i\neq j}\sum_{i'\neq j'}\mathbb{E}[(a_{N,\mu}(W_i,X_i,X_j,W_j)-c_{N,\mu}(X_j,W_j))(a_{N,\mu}(W_{i'},X_{i'},X_{j'},W_{j'})-c_{N,\mu}(X_{j'},W_{j'}))]=$$

$$\frac{1}{N^2(N-1)^2}\sum_{i\neq j}\mathbb{E}[(a_{N,\mu}(W_i,X_i,X_j,W_j)-c_{N,\mu}(X_j,W_j))^2]=$$

Because

$$\mathbb{E}[(a_{N,\mu}(W_i,X_i,X_j,W_j)-c_{N,\mu}(X_j,W_j))^2]=\mathbb{E}[(a_{N,\mu}(W_i,X_i,X_j,W_j)^2]-\mathbb{E}[c_{N,\mu}(X_j,W_j))^2]\leq$$

$$\mathbb{E}[(a_{N,\mu}(W_i,X_i,X_j,W_j)^2]=O(b_N^{-1})$$

we have

$$\mathbb{E}[(D_\mu-E_\mu)^2]=O(N^{-2}b_N^{-1})$$

so that

$$\frac{N(N-1)}{N^2}(D_{\mu,1}-E_\mu)=O_p(N^{-1}b_N^{-1/2})$$

Also

$$\mathbb{E}[c_{N,\mu}(X_j,W_j)^2]\leq$$

$$\frac{1}{b_N^{2+2|\mu|}}\mathbb{E}\left[\left(\int_{\mathbb{X}}\int_{\mathbb{W}}\frac{g_W\left(w,x\right)d(x)}{f_W\left(w\right)}X_jK^{(\mu)}\left(\frac{W_j-r_{b_N}(w)}{b_N}\right)(w-r_{b_N}w))^{\mu}f_{WX}(w,x)\mathrm{d}w\mathrm{d}x\right)^2\right]\leq$$

$$\frac{1}{b_N^{2+2|\mu|}}\mathbb{E}\left[\int_{\mathbb{X}}\int_{\mathbb{W}}\frac{g_W\left(w,x\right)^2d(x)^2}{f_W\left(w\right)^2}X_j^2K^{(\mu)}\left(\frac{W_j-r_{b_N}(w)}{b_N}\right)^2(w-r_{b_N}w))^{2\mu}f_{WX}(w,x)\mathrm{d}w\mathrm{d}x\right]\leq$$

$$\frac{C}{b_N^2}\int_{\mathbb{W}}\int_{\mathbb{X}}\int_{\mathbb{W}}K^{(\mu)}\left(\frac{\tilde{w}-r_{b_N}(w)}{b_N}\right)^2f_{WX}(w,x)\mathrm{d}w\mathrm{d}xf_W(\tilde{w})\mathrm{d}\tilde{w}=O(b_N^{-1})$$

by a change of variables in the outer integral, so that

$$\left(\frac{N(N-1)}{N^2}-1\right)E_\mu=O_p(N^{-1}b_N^{-1/2})$$

Finally

$$\mathbb{E}[|D_{\mu,2}|]\leq\frac{1}{N}\mathbb{E}[|a_{N,\mu}(W_i,X_i,X_i,W_i)|]\leq\frac{1}{N}\sqrt{\mathbb{E}[a_{N,\mu}(W_i,X_i,X_i,W_i)^2]}=O(N^{-1}b_N^{-1})$$

so that

$$D_{\mu,2}=O_p(N^{-1}b_N^{-1})$$

Therefore if $\delta<1/2$ then

$$D_\mu=E_\mu+o_p(N^{-1/2})$$

Under the same condition (B.32) is a linear combination of terms

$$F_\mu = \frac{1}{N}\sum_{i=1}^{N} \frac{g_W\left(W_i, X_i\right) d(W_i) m(W_i)}{f_W\left(W_i\right)} \left(\hat{h}^{(\mu)}_{2,NW}(r_b(W_i)) - \mathbb{E}\left[\hat{h}^{(\mu)}_{2,NW}(r_b(W_i))\right]\right)(W_i - r_b(W_i))^\mu =$$

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} e_{N,\mu}(W_i, X_i, W_j)$$

with

$$e_{N,\mu}(W_i, X_i, W_j) = \frac{g_W\left(W_i, X_i\right) d(W_i) m(W_i)}{f_W\left(W_i\right)} \left(\frac{1}{b_N^{1+|\mu|}} K^{(\mu)}\left(\frac{W_j - r_{b_N}(W_i)}{b_N}\right) - \right.$$

$$\left.\mathbb{E}\left[\frac{1}{b_N^{1+|\mu|}} K^{(\mu)}\left(\frac{W - r_{b_N}(W_i)}{b_N}\right)\right]\right)(W_i - r_{b_N}(W_i))^\mu$$

such that

$$F_\mu = G_\mu + o_p(N^{-1/2})$$

with

$$G_\mu = \frac{1}{N}\sum_{i=1}^{N} f_{N,\mu}(W_j)$$

and

$$f_{N,\mu}(W_j) = \frac{1}{b_N^{1+|\mu|}} \cdot$$

$$\int_{\mathbb{X}}\int_{\mathbb{W}} \frac{g_W\left(w,x\right) d(x) m(w)}{f_W\left(w\right)} \left(K^{(\mu)}\left(\frac{W_j - r_{b_N}(w)}{b_N}\right) - \mathbb{E}\left[K^{(\mu)}\left(\frac{W - r_{b_N}(w)}{b_N}\right)\right]\right)(w - r_{b_N} w))^\mu f_{WX}(w,x) \mathrm{d}w\mathrm{d}x$$

The final step is to show that

$$E_0 = -\frac{1}{N}\sum_{j=1}^{N} c_{N,\mu}(X_j, W_j) = -\frac{1}{N}\sum_{j=1}^{N}(\zeta_j - \mathbb{E}[\zeta_j]) + o_p(N^{-1/2})$$

with

$$\zeta_j = X_j \mathbb{E}[g_W(W_j, X) d(X) | W_j]$$

where the expectation is over the conditional distribution of X given W, and

$$G_0 = \frac{1}{N}\sum_{j=1}^{N} c_{N,\mu}(X_j, W_j) = \frac{1}{N}\sum_{j=1}^{N}(\xi_j - \mathbb{E}[\xi_j]) + o_p(N^{-1/2})$$

with

$$\xi_j = m(W_j)\mathbb{E}[g_W(W_j, X) d(X) | W_j]$$

and

$$E_\mu = o_p(N^{-1/2}) \qquad\qquad G_\mu = o_p(N^{-1/2})$$

for $|\mu| \geq 1$. We only consider E_0 and E_μ. The proof for G_0 and G_μ is analogous. Define

$$\psi_{N,\mu,j} = \frac{1}{b_N^{1+|\mu|}} \int_{\mathbb{X}}\int_{\mathbb{W}} \frac{g_W\left(w,x\right) d(x)}{f_W\left(w\right)} X_j K^{(\mu)}\left(\frac{W_j - r_{b_N}(w)}{b_N}\right)(w - r_{b_N} w))^\mu f_{WX}(w,x) \mathrm{d}w\mathrm{d}x$$

so that $c_{N,\mu}(X_j, W_j) = \psi_{N,\mu,j} - \mathbb{E}[\psi_{N,\mu,j}]$. Now

$$\psi_{N,0,j} = \psi_{N,0,j,0} + \psi_{N,0,j,1}$$

with

$$\psi_{N,0,i,0} = \frac{1}{b_N}\int_{\mathbb{X}}\int_{w_l+b_N}^{w_u-b_N} \frac{g_W\left(w,x\right)d(x)}{f_W\left(w\right)}X_jK\left(\frac{W_j-w}{b_N}\right)f_{WX}(w,x)\mathrm{d}w\mathrm{d}x$$

and

$$\psi_{N,0,j,1} = \frac{1}{b_N}\int_{\mathbb{X}}\int_{w_l}^{w_l+b_N} \frac{g_W\left(w,x\right)d(x)}{f_W\left(w\right)}X_jK\left(\frac{W_j-w_l}{b_N}\right)f_{WX}(w,x)\mathrm{d}w\mathrm{d}x+$$

$$\frac{1}{b_N}\int_{\mathbb{X}}\int_{w_u-b_n}^{w_u} \frac{g_W\left(w,x\right)d(x)}{f_W\left(w\right)}X_jK^{(\mu)}\left(\frac{W_j-w_u}{b_N}\right)f_{WX}(w,x)\mathrm{d}w\mathrm{d}x$$

so that

$$E_0 = -\frac{1}{N}\sum_{j=1}^{N}(\psi_{N,0,j,0}-\mathbb{E}[\psi_{N,0,j,0}])-\frac{1}{N}\sum_{j=1}^{N}(\psi_{N,0,j,1}-\mathbb{E}[\psi_{N,0,j,1}])$$

Obviously

$$\mathbb{E}\left[\left(-\frac{1}{N}\sum_{i=1}^{N}(\psi_{N,0,j,0}-\mathbb{E}[\psi_{N,0,j,0}])+\frac{1}{N}\sum_{j=1}^{N}(\zeta_j-\mathbb{E}[\zeta_j])\right)^2\right] \leq \frac{1}{N}\mathbb{E}[(\psi_{N,0,j,0}-\zeta_j)^2]$$

By a change of variables to $t=(W_j-w)/b_N$ with Jacobian b_N

$$\psi_{N,0,j,0} = \int_{\mathbb{X}}\int_{-1}^{1} 1_{1+\frac{W_j-w_u}{b_N}\leq t\leq -1+\frac{W_j-w_l}{b_N}}\frac{g_W\left(W_j-b_Nt,x\right)d(x)}{f_W\left(W_j-b_Nt\right)}X_jK\left(t\right)f_{WX}(W_j-b_Nt,x)\mathrm{d}t\mathrm{d}x$$

so that

$$|\psi_{N,0,j,0}-\zeta_i| \leq \int_{\mathbb{X}}\int_{-1}^{1} 1_{1+\frac{W_j-w_u}{b_N}\leq t\leq -1+\frac{W_j-w_l}{b_N}}$$

$$\left|\frac{g_W\left(W_j-b_Nt,x\right)d(x)}{f_W\left(W_j-b_Nt\right)}f_{WX}(W_j-b_Nt,x)-\frac{g_W\left(W_j,x\right)d(x)}{f_W\left(W_j\right)}f_{WX}(W_j,x)\right||X_j||K\left(t\right)|\mathrm{d}t\mathrm{d}x+$$

$$\int_{\mathbb{X}}\left|\frac{g_W\left(W_j,x\right)d(x)}{f_W\left(W_j\right)}f_{WX}(W_j,x)\right|\mathrm{d}x|X_j|\int_{-1}^{1}\left|1_{1+\frac{W_j-w_u}{b_N}\leq t\leq -1+\frac{W_j-w_l}{b_N}}-1\right||K\left(t\right)|\mathrm{d}t+$$

By the mean value theorem the first term on the right hand side is $b_N|X_j|p(W_j)$ with $p(W_j)$ a (generic) bounded function of W_j. The second term on the right hand side is $|X_j|p(W_j)(1-\Pr(w_l+2b_N\leq W_j\leq w_u-2b_N))$. Therefore

$$|\psi_{N,0,j,0}-\zeta_i| \leq |X_j|p(W_j)(b_N+(1-\Pr(w_l+2b_N\leq W_j\leq w_u-2b_N)))$$

so that

$$\mathbb{E}[(\psi_{N,0,j,0}-\zeta_j)^2]=O(b_N)$$

and

$$\left|-\frac{1}{N}\sum_{i=1}^{N}(\psi_{N,0,j,0}-\mathbb{E}[\psi_{N,0,j,0}])+\frac{1}{N}\sum_{j=1}^{N}(\zeta_j-\mathbb{E}[\zeta_j])\right|=o_p(N^{-1/2})$$

if $\delta<\frac{1}{2}$. For $\psi_{N,0,j,1}$ we consider the first term on the right hand side

$$\left|K\left(\frac{W_j-w_l}{b_N}\right)X_j\frac{1}{b_N}\int_{w_l}^{w_l+b_N}\int_{\mathbb{X}}\frac{g_W\left(w,x\right)d(x)}{f_W\left(w\right)}f_{WX}(w,x)\mathrm{d}x\mathrm{d}w\right| \leq C|X_j|1_{w_l\leq W_j\leq w_l+b_N}$$

For the other term on the right hand side we get a similar bound and we conclude

$$\mathbb{E}[\psi_{N,0,j,1}^2]=O(b_N)$$

so that if $\delta < \frac{1}{2}$

$$\left|\frac{1}{N}\sum_{j=1}^{N}(\psi_{N,0,j,1} - \mathbb{E}[\psi_{N,0,j,1}])\right| = o_p(N^{-1/2})$$

Finally if $\mu \geq 1$

$$\psi_{N,\mu,j} = \frac{1}{b_N^{1+|\mu|}}\int_{\mathbb{X}}\int_{w_l}^{w_l+b_N} \frac{g_W(w,x)\,d(x)}{f_W(w)} X_j K^{(\mu)}\left(\frac{W_j - w_l}{b_N}\right)(w-w_l)^{\mu} f_{WX}(w,x)\mathrm{d}w\mathrm{d}x +$$

$$\frac{1}{b_N^{1+|\mu|}}\int_{\mathbb{X}}\int_{w_u-b_N}^{w_u} \frac{g_W(w,x)\,d(x)}{f_W(w)} X_j K^{(\mu)}\left(\frac{W_j - w_u}{b_N}\right)(w-w_u)^{\mu} f_{WX}(w,x)\mathrm{d}w\mathrm{d}x$$

The first term on the right hand side is bounded by

$$\left|K^{(\mu)}\left(\frac{W_j - w_l}{b_N}\right)\right| |X_j| \frac{1}{b_N}\int_{\mathbb{X}}\int_{w_l}^{w_l+b_N}\left|\frac{g_W(w,x)\,d(x)}{f_W(w)} f_{WX}(w,x)\right|\mathrm{d}w\mathrm{d}x \leq C|X_j| 1_{w_l \leq W_j \leq w_l + b_N}$$

so that

$$\mathbb{E}[\psi_{N,\mu,j}^2] = O(b_N)$$

and therefore

$$E_\mu = -\sum_{j=1}^{N}(\psi_{N,\mu,j} - \mathbb{E}[\psi_{N,\mu,j}]) = o_p(N^{-1/2})$$

if $\delta < \frac{1}{2}$. □

Proof of Theorem A.3: Because the class of 'doubly averaged' estimators has not been considered previously, we provide a somewhat detailed proof. The proof consists of four steps. In the first we approximate the estimator by a linear function of the kernel estimator $\hat{h}_{\text{nip},s}$ (linearization). Formally, we show that

$$V = \frac{1}{N\sqrt{N}}\sum_{j=1}^{N}\sum_{k=1}^{N}\frac{\partial n}{\partial h'}(h(Z_{1j},Z_{2k}))\left(\hat{h}_{\text{nip},s}(Z_{1j},Z_{2k}) - h(Z_{1j},Z_{2k})\right) + O_p\left(\sqrt{N}\left|\hat{h}_{\text{nip},s} - h\right|^2\right). \quad \text{(B.33)}$$

By the assumptions, and Lemma A.11, the remainder term is $o_p(1)$.

In the second step we express the difference between the linearized estimator and the estimand as the sum of a bias term (that is asymptotically negligible) and a variance term (bias-variance decomposition). The bias term will be shown to satisfy

$$\left|\frac{1}{N\sqrt{N}}\sum_{j=1}^{N}\sum_{k=1}^{N}\frac{\partial n}{\partial h'}(h(Z_{1j},Z_{2k}))(\mathbb{E}[\hat{h}_{\text{nip},s}(Z_{1j},Z_{2k})] - h(Z_{1j},Z_{2k}))\right| = O\left(\sqrt{N}b_N^p\right). \quad \text{(B.34)}$$

By the assumption on the bandwidth rate, the remainder term is $o(1)$. Note that by $\mathbb{E}[\hat{h}(Z_{i1},Z_{i2}]$ we mean the expectation of $\hat{h}(z_1,z_2)$, evaluated at $z_1 = Z_{1i}$ and $z_2 = Z_{2j}$: the expectation is taken over the estimator of the function $h(\cdot)$.

The second step leaves us with

$$V = W + O_p\left(\sqrt{N}\left|\hat{h}_{\text{nip},s} - h\right|^2\right) + O(\sqrt{N}b_N^p),$$

where

$$W = \frac{1}{N\sqrt{N}}\sum_{j=1}^{N}\sum_{k=1}^{N}\frac{\partial n}{\partial h'}(h(Z_{1j},Z_{2k}))\left(\hat{h}(Z_{1j},Z_{2k}) - \mathbb{E}\left[\hat{h}_{\text{nip},s}(Z_{1j},Z_{2k})\right]\right). \quad \text{(B.35)}$$

Define

$$\nu(z_1,z_2) = \frac{\partial n}{\partial h'}(h(z_1,z_2)),$$

and

$$a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k})$$
$$= \nu(Z_{1j}, Z_{2k})' \cdot \left(\frac{1}{b_N^{L+|\mu|}} \tilde{Y}_i K^{(\mu)} \left(\frac{Z_i - r_{b_N}(Z_{1j}, Z_{2k})}{b_N} \right) \right.$$
$$\left. - \mathbb{E}_{\tilde{Y}Z} \left[\frac{1}{b_N^{L+|\mu|}} \tilde{Y} K^{(\mu)} \left(\frac{Z - r_{b_N}(Z_{1j}, Z_{2k})}{b_N} \right) \right] \right) \cdot \left(\begin{pmatrix} Z_{1j} \\ Z_{2k} \end{pmatrix} - r_{b_N}(Z_{1j}, Z_{2k}) \right)^{\mu}$$

so that

$$W = \sum_{\mu:|\mu| \leq s-1} \frac{1}{\mu!} W_\mu, \tag{B.36}$$

where

$$W_\mu = \frac{1}{N^2\sqrt{N}} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k})$$

Define

$$c_{N,\mu}(\tilde{Y}_i, Z_i) = \frac{1}{b_N^{L+|\mu|}} \int_{\mathcal{Z}_2} \int_{\mathcal{Z}_1} \nu(z_1, z_2)' \left(\tilde{Y}_i K^{(\mu)} \left(\frac{Z_i - r_{b_N}(z_1, z_2)}{b_N} \right) - \mathbb{E}_{\tilde{Y}Z} \left[\tilde{Y} K^{(\mu)} \left(\frac{Z - r_{b_N}(z_1, z_2)}{b_N} \right) \right] \right)$$
$$\times \left((z_1' \; z_2')' - r_{b_N}(z_1, z_2) \right)^{\mu} f_{Z_1}(z_1) f_{Z_2}(z_2) \mathrm{d}z_1 \mathrm{d}z_2$$

and

$$U_\mu = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} c_{N,\mu}(\tilde{Y}_i, Z_i),$$

or, equivalently

$$U_\mu =$$
$$\sqrt{N} \int_{\mathcal{Z}_2} \int_{\mathcal{Z}_1} \nu(z_1, z_2)' (\hat{h}_{NW}^{(\mu)}(r_{b_N}(z_1, z_2)) - \mathbb{E}[\hat{h}_{NW}^{(\mu)}(r_{b_N}(z_1, z_2))])((z_1' \; z_2')' - r_{b_N}(z_1, z_2))^{\mu} f_{Z_1}(z_1) f_{Z_2}(z_2) \mathrm{d}z_1 \mathrm{d}z_2.$$

In the third step we show that

$$W_\mu = U_\mu + O_p\left(N^{-1/2} b_N^{-L}\right) \tag{B.37}$$

In the fourth step we show that

$$U_0 = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left\{ \frac{\partial n}{\partial h}(h(Z_i))' \tilde{Y}_i f_{Z_1}(Z_{1i}) f_{Z_2}(Z_{2i}) - \mathbb{E}_Z \left[\frac{\partial n}{\partial h}(h(Z))' \tilde{Y} f_{Z_1}(Z_{1i}) f_{Z_2}(Z_{2i}) \right] \right\} + o_p(1). \tag{B.38}$$

which gives us the representation in the Theorem.
In the fifth and final step we show that we can ignore U_μ for μ such that $|\mu| \geq 1$, because for such μ,

$$U_\mu = O_p(b_N). \tag{B.39}$$

Proving these statements implies the result in the Theorem.
Now we turn to proving each of the statements (B.33), (B.34), (B.37), (B.38), and (B.39).

Step 1: Linearization In the first step of the proof we prove equality (B.33). First define

$$d(z_1, z_2) \equiv n(\hat{h}_{\mathrm{nip},s}(z_1, z_2)) - n(h(z_1, z_2)) - \frac{\partial n}{\partial h'}(h(z_1, z_2))(\hat{h}_{\mathrm{nip},s}(z_1, z_2) - h(z_1, z_2)).$$

By a second order Taylor series expansion of $n(\hat{h}_{\mathrm{nip},s}(Z_{1i}, Z_{2j}))$ around $h(Z_{1i}, Z_{2j})$ we have,

$$|d(z_1, z_2)| = \frac{1}{2} \left| (\hat{h}_{\mathrm{nip},s}(z_1, z_2) - h(z_1, z_2))' \frac{\partial^2 n}{\partial h \partial h'}(\overline{h}(z_1, z_2))(\hat{h}_{\mathrm{nip},s}(z_1, z_2) - h(z_1, z_2)) \right|$$

$$\leq \sup_{z} \left| \frac{\partial^2 n}{\partial h \partial h'}(h(z)) \right| \cdot \left| \hat{h}_{\text{nip},s}(z_1, z_2) - h(z_1, z_2) \right|^2$$

$$\leq C \cdot \left| \hat{h}_{\text{nip},s} - h \right|^2,$$

with $\overline{h}(z_1, z_2)$ intermediate between $\hat{h}_{\text{nip},s}(z_1, z_2)$, and $h(z_1, z_2)$ so that

$$\left| \frac{1}{N^2} \sum_{j=1}^{N} \sum_{k=1}^{N} d(Z_{1j}, Z_{2k}) \right| \leq C \left| \hat{h}_{\text{nip},s} - h \right|^2 = O_p\left(\left| \hat{h}_{\text{nip},s} - h \right|^2 \right).$$

Hence

$$\frac{1}{N\sqrt{N}} \sum_{j=1}^{N} \sum_{k=1}^{N} \left[n\left(\hat{h}_{\text{nip},s}(Z_{1j}, Z_{2k}) \right) - n\left(h(Z_{1j}, Z_{2k}) \right) \right]$$

$$= \frac{1}{N\sqrt{N}} \sum_{j=1}^{N} \sum_{k=1}^{N} \frac{\partial n}{\partial h'}(h(Z_{1j}, Z_{2k})) \left(\hat{h}_{\text{nip},s}(Z_{1j}, Z_{2k}) - h(Z_{1j}, Z_{2k}) \right)$$

$$+ \frac{1}{N\sqrt{N}} \sum_{j=1}^{N} \sum_{k=1}^{N} d(Z_{1j}, Z_{2k})$$

$$= \frac{1}{N\sqrt{N}} \sum_{j=1}^{N} \sum_{k=1}^{N} \frac{\partial n}{\partial h'}(h(Z_{1j}, Z_{2k})) \left(\hat{h}_{\text{nip},s}(Z_{1j}, Z_{2k}) - h(Z_{1j}, Z_{2k}) \right) + O_p\left(\sqrt{N} \left| \hat{h}_{\text{nip},s} - h \right|^2 \right)$$

so that the linearization remainder has the same stochastic order as $\sqrt{N} \left| \hat{h}_{\text{nip},s} - h_0 \right|^2$.

Step 2: Bias-variance decomposition In the second step of the proof we proof equation (B.34). Define

$$E = \frac{1}{N\sqrt{N}} \sum_{j=1}^{N} \sum_{k=1}^{N} \frac{\partial n}{\partial h'}(h(Z_{1j}, Z_{2k})) \left(\hat{h}_{\text{nip},s}(Z_{1j}, Z_{2k}) - h(Z_{1j}, Z_{2k}) \right).$$

so that

$$V = E + O_p\left(\sqrt{N} \left| \hat{h}_{\text{nip},s} - h \right|^2 \right).$$

We decompose E into a bias and variance part, $E = EU_{\text{bias}} + W$, where

$$E_{\text{bias}} = \frac{1}{N\sqrt{N}} \sum_{j=1}^{N} \sum_{k=1}^{N} \frac{\partial n}{\partial h'}(h(Z_{1j}, Z_{2k})) \left(\mathbb{E}\left[\hat{h}_{\text{nip},s}(Z_{1j}, Z_{2k}) \right] - h(Z_{1j}, Z_{2k}) \right),$$

and W is defined in (B.35). The bias part is bounded by

$$\left| \frac{1}{N\sqrt{N}} \sum_{j=1}^{N} \sum_{k=1}^{N} \frac{\partial n}{\partial h'}(h(Z_{1j}, Z_{2k})) \left(\mathbb{E}\left[\hat{h}_{\text{nip},s}(Z_{1j}, Z_{2k}) \right] - h(Z_{1j}, Z_{2k}) \right) \right|$$

$$\leq \sup_{z \in \mathbb{Z}} \left| \frac{\partial n}{\partial h}(h(z)) \right| \sqrt{N} \left| \mathbb{E}[\hat{h}_{\text{nip},s}] - h \right| = O(\sqrt{N} b_N^p),$$

due to smoothness of the function and Lemma A.9.

Step 3: Projection In the third step of the proof we prove equation (B.37), $W_\mu = U_\mu + O_p(N^{-1/2} b_N^{-L})$. This is the most complicated step. First, note that

$$W_\mu = \frac{1}{N^2\sqrt{N}} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k}),$$

is a third order V-statistic with kernel (that depends on N) $a_{N,\mu}$. We show that this V-statistic is asymptotically equivalent to a projection that is a single sum. Because the kernel depends on N we cannot use a standard result.

The projection of W_μ is

$$U_\mu = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} c_{N,\mu}(\tilde{Y}_i, Z_i),$$

with

$$c_{N,\mu}(\tilde{Y}_i, Z_i) = \frac{1}{b_N^{L+|\mu|}} \int_{\mathbb{Z}_2} \int_{\mathbb{Z}_1} \nu(z_1,z_2)' \left(\tilde{Y}_i K^{(\mu)} \left(\frac{Z_i - r_{b_N}(z_1,z_2)}{b_N} \right) - \mathbb{E}_{\tilde{Y}Z} \left[\tilde{Y} K^{(\mu)} \left(\frac{Z - r_{b_N}(z_1,z_2)}{b_N} \right) \right] \right) \times ((z_1' \; z_2')' - r_{b_N}(z_1,z_2))^\mu f_{Z_1}(z_1) f_{Z_2}(z_2) \mathrm{d}z_1 \mathrm{d}z_2.$$

The projection remainder is

$$W_\mu - U_\mu = \frac{N(N-1)(N-2)}{N^3}(W_{\mu,1} - U_\mu) + \left(\frac{N(N-1)(N-2)}{N^3} - 1 \right) U_\mu + W_{\mu,2} + W_{\mu,3} + W_{\mu,4} + W_{\mu,5} \quad \text{(B.40)}$$

with

$$W_{\mu,1} \equiv \frac{\sqrt{N}}{N(N-1)(N-2)} \sum_{i \neq j \neq k} a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k})$$

$$W_{\mu,2} \equiv \frac{\sqrt{N}}{N^3} \sum_{i=j \neq k} a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1i}, Z_{2k})$$

$$W_{\mu,3} \equiv \frac{\sqrt{N}}{N^3} \sum_{i=k \neq j} a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2i})$$

$$W_{\mu,4} \equiv \frac{\sqrt{N}}{N^3} \sum_{i \neq j = k} a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2j})$$

$$W_{\mu,5} \equiv \frac{\sqrt{N}}{N^3} \sum_{i=j=k} a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1i}, Z_{2k})$$

We prove that the projection remainder $W_\mu - U_\mu = O_p(N^{-1/2} b_N^{-L})$ by proving the following six equalities:

$$W_{\mu,1} - U_\mu = O_p\left(N^{-1} b_N^{-L/2}\right), \quad \text{(B.41)}$$

$$\left(\frac{N(N-1)(N-2)}{N^3} - 1 \right) U_\mu = O_p\left(N^{-1} b_N^{-L/2}\right), \quad \text{(B.42)}$$

$$W_{\mu,2} = O_p\left(N^{-1/2} b_N^{-L+L_2/2}\right), \quad \text{(B.43)}$$

$$W_{\mu,3} = O_p\left(N^{-1/2} b_N^{-L+L_1/2}\right), \quad \text{(B.44)}$$

$$W_{\mu,4} = O_p\left(N^{-3} b_N^{-L/2}\right), \quad \text{(B.45)}$$

$$W_{\mu,5} = O_p\left(N^{-1/2} b_N^{-L}\right). \quad \text{(B.46)}$$

In order to prove these results, we establish bounds on the second moment of $a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k})$. This will be relatively straightforward if $i \neq j$ and $i \neq k$. The derivation of the bound is more involved if $i = j$ and/or $i = k$. We could simplify the proof by omitting these observations and redefining the estimator by restricting the averaging to observations with $i \neq j$ and $i \neq k$. This would amount to redefining the kernel estimator in (A.7) by omitting observations $i = j$ and $i = k$ in $\hat{h}_{\mathrm{nip},s}$. We will keep these observations and derive bounds on all second moments. We derive the following bounds, considering four separate cases (note that the bounds do not depend on μ)

$$\mathbb{E}[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k})^2] = O(b_N^{-L}) \qquad j \neq i, \text{ and } k \neq i, \quad \text{(B.47)}$$

$$\mathbb{E}[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1i}, Z_{2i})^2] = O(b_N^{-2L}) \qquad i = j = k, \quad \text{(B.48)}$$

$$\mathbb{E}[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1i}, Z_{2k})^2] = O(b_N^{-2L+L_2}) \qquad k \neq i = j, \quad \text{(B.49)}$$

$$\mathbb{E}[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k})^2] = O(b_N^{-2L+L_1}) \qquad j \neq i = k. \quad \text{(B.50)}$$

Step 3A: Equation (B.47) First if $j \neq i$ and $k \neq i$

$$\begin{aligned}
&\mathbb{E}[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k})^2] \\
&\leq \frac{1}{b_N^{2L+2|\mu|}} \mathbb{E}\left[K^{(\mu)}\left(\frac{Z_i - r_{b_N}(Z_{1j}, Z_{2k})}{b_N}\right)^2 \nu(Z_{1j}, Z_{2k})'\tilde{Y}_i\tilde{Y}_i'\nu(Z_{1j}, Z_{2k})((Z_{1j}'\ Z_{2k}')' - r_{b_N}(Z_{1j}, Z_{2k}))^{2\mu}\right] \\
&\leq \frac{\sup_{z\in\mathbb{Z}} |z - r_{b_N}(z)|^{2\mu}}{b_N^{2L+2|\mu|}} \mathbb{E}\left[K^{(\mu)}\left(\frac{Z_i - r_{b_N}(Z_{1j}, Z_{2k})}{b_N}\right)^2 \nu(Z_{1j}, Z_{2k})'\tilde{Y}_i\tilde{Y}_i'\nu(Z_{1j}, Z_{2k})\right] \\
&\leq \frac{1}{b_N^{2L}} \mathbb{E}\left[K^{(\mu)}\left(\frac{Z_i - r_{b_N}(Z_{1j}, Z_{2k})}{b_N}\right)^2 \nu(Z_{1j}, Z_{2k})'\tilde{Y}_i\tilde{Y}_i'\nu(Z_{1j}, Z_{2k})\right]
\end{aligned}$$

Now by the Cauchy-Schwartz inequality

$$\begin{aligned}
&\mathbb{E}\left[K^{(\mu)}\left(\frac{Z_i - r_{b_N}(Z_{1j}, Z_{2k})}{b_N}\right)^2 \nu(Z_{1j}, Z_{2k})'\tilde{Y}_i\tilde{Y}_i'\nu(Z_{1j}, Z_{2k})\right] \\
&= \mathbb{E}\left[K^{(\mu)}\left(\frac{Z_i - r_{b_N}(Z_{1j}, Z_{2k})}{b_N}\right)^2 (\nu(Z_{1j}, Z_{2k})'\,\tilde{Y}_i)^2\right] \\
&\leq \mathbb{E}\left[K^{(\mu)}\left(\frac{Z_i - r_{b_N}(Z_{1j}, Z_{2k})}{b_N}\right)^2 |\nu(Z_{1j}, Z_{2k}|^2 |\tilde{Y}_i|^2\right] \\
&= \mathbb{E}\left[K^{(\mu)}\left(\frac{Z_i - r_{b_N}(Z_{1j}, Z_{2k})}{b_N}\right)^2 |\nu(Z_{1j}, Z_{2k}|^2 \mathbb{E}\left[|\tilde{Y}_i|^2 | Z_i\right]\right]
\end{aligned}$$

By Assumption 3.1 $\mathbb{E}\left[|\tilde{Y}|^2 | Z = z\right]$ an ν are bounded on $\mathbb{Z}$ so that this is bounded by (condition on Z_{1j} and Z_{2k})

$$C\mathbb{E}\left[K^{(\mu)}\left(\frac{Z_i - r_{b_N}(Z_1, Z_2)}{b_N}\right)^2\right] = C\int_{\mathbb{Z}} K^{(\mu)}\left(\frac{z - r_{b_N}(Z_1, Z_2)}{b_N}\right)^2 f_Z(z)\mathrm{d}z$$

and by a change of variables to $t = (z - r_{b_N}(Z_1, Z_2))/b_N$ with Jacobian b_N^L we obtain

$$Cb_N^L \int_{\{t|t=(z-r_{b_N}(Z_1,Z_2))/b_N, z\in\mathbb{Z}\}} K^{(\mu)}(t)^2 f_Z(b_N t + r_{b_N}(Z_1, Z_2))\mathrm{d}t \leq C_1 b_N^L \int_{\mathcal{U}} K^{(\mu)}(t)^2 \mathrm{d}t \leq C_2 b_N^L$$

by Assumptions 3.1 and 4.1. We conclude

$$\mathbb{E}[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k})^2] = O(b_N^{-L}) \tag{B.51}$$

The same proof and the same bound holds if $j \neq k \neq i$ or $j = k \neq i$.

Step 3B: Equation (B.48) Next, we consider $\mathbb{E}[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1i}, Z_{2i})^2]$ where we note that $\mathbb{E}[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1i}, Z_{2i})] \neq 0$. Because

$$\begin{aligned}
&a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1i}, Z_{2i}) \\
&= \frac{1}{b_N^{L+|\mu|}}\left[\nu(Z_i)'\tilde{Y}_i K^{(\mu)}\left(\frac{Z_i - r_{b_N}(Z_i)}{b_N}\right)(Z_i - r_{b_N}(Z_i))^\mu - \mathbb{E}_Z\left(\nu(Z_i)'^{(\mu)}\left(\frac{Z - r_{b_N}(Z_i)}{b_N}\right)(Z_i - r_{b_N}(Z_i))^\mu\right)\right]
\end{aligned}$$

we have

$$\mathbb{E}[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1i}, Z_{2i})^2] \tag{B.52}$$

$$= \frac{1}{b_N^{2L+2|\mu|}} \mathbb{E}\left[K^{(\mu)}\left(\frac{Z_i - r_{b_N}(Z_i)}{b_N}\right)^2 (\nu(Z_i)'\tilde{Y}_i)^2 (Z_i - r_{b_N}(Z_i))^{2\mu}\right] \tag{B.53}$$

$$-\frac{2}{b_N^{2L+2|\mu|}} \mathbb{E}_{Z_i}\left[\left\{\nu(Z_i)'g(Z_i)K^{(\mu)}\left(\frac{Z_i - r_{b_N}(Z_i)}{b_N}\right)\right.\right. \tag{B.54}$$

$$\left.\left. -\ \mathbb{E}_Z\left[\nu(Z)g(Z)K^{(\mu)}\left(\frac{Z - r_{b_N}(Z_i)}{b_N}\right)\right](Z_i - r_{b_N}(Z_i))^{2\mu}\right\}\right]$$

$$+\frac{1}{b_N^{2L+2|\mu|}}\mathbb{E}_{Z_i}\left[\left(\nu(Z_i)'\mathbb{E}_Z\left(g(Z)K^{(\mu)}\left(\frac{Z-r_{b_N}(Z_i)}{b_N}\right)\right)\right)^2(Z_i-r_{b_N}(Z_i))^{2\mu}\right] \tag{B.55}$$

By Assumption 3.1 and smoothness (B.53) is bounded by

$$\frac{C}{b_N^{2L}}\int_{\mathbb{Z}}K^{(\mu)}\left(\frac{z-r_{b_N}(z)}{b_N}\right)^2 f_Z(z)\mathrm{d}z=\frac{C}{b_N^{2L}}\int_{\mathbb{Z}^I_{b_N}}K^{(\mu)}\left(\frac{z-r_{b_N}(z)}{b_N}\right)^2 f_Z(z)\mathrm{d}z$$
$$+\frac{C}{b_N^{2L}}\int_{\mathbb{Z}\setminus\mathbb{Z}^I_{b_N}}K^{(\mu)}\left(\frac{z-r_{b_N}(z)}{b_N}\right)^2 f_Z(z)\mathrm{d}z$$

If $r_b(z)$ is the projection on the internal set, then $z-r_b(z)=0$ if z is in the internal set. Therefore

$$\frac{C}{b_N^{2L}}\int_{\mathbb{Z}^I_{b_N}}K^{(\mu)}\left(\frac{z-r_{b_N}(z)}{b_N}\right)^2 f_Z(z)\mathrm{d}z\leq\frac{CK^{(\mu)}(0)^2}{b_N^{2L}}$$

Next we consider the second integral. If $s\in\mathbb{Z}^B_{b_N}\equiv\mathbb{Z}\setminus\mathbb{Z}^I_{b_N}$, then at least one component of z is in the boundary region. We can subdivide $\mathbb{Z}^B_{b_N}$ into disjoint subsets $\mathbb{Z}^B_{b_N,p}, p=1,\ldots,2^L-1$ and in each such subset $L_p\geq 1$ components of z are within b_N from the boundary. We further partition $\mathbb{Z}^B_{b_N,p}$ into disjoint sets $\mathbb{Z}^B_{b_N,p,r}, r=1,\ldots 2^{L_p}$ with $0\leq K_r\leq L_p$ components with $z_{ll}\leq Z_l\leq z_{ll}+b_N$ and the remaining L_p-K_r components with $z_{ul}-b_N\leq Z_l\leq z_{ul}$. Without loss of generality we assume that the first K_r components of z are near the lower bound, the next L_p-K_r are near the upper bound and the rest is in the internal region, so that

$$\frac{C}{b_N^{2L}}\int_{\mathbb{Z}^B_{b_N,p,r}}K^{(\mu)}\left(\frac{z-r_{b_N}(z)}{b_N}\right)^2 f_Z(z)\mathrm{d}z$$
$$=\int_{z_{l1}}^{z_{l1}+b_N}\cdots\int_{z_{l,K_r}}^{z_{l,K_r}+b_N}\int_{z_{u,K_r+1}-b_N}^{z_{u,K_r+1}}\cdots\int_{z_{u,L_p}-b_N}^{z_{u,L_p}}\int_{z_{l,L_p+1}+b_N}^{z_{u,L_p+1}-b_N}\cdots\int_{z_{lL}+b_N}^{z_{uL}-b_N}\prod_{l=1}^{K_r}\mathcal{K}_l^{(\mu_l)}\left(\frac{Z_l-z_{ll}}{b_N}-1\right)^2$$
$$\times\prod_{l=K_r+1}^{L_p}\mathcal{K}_l^{(\mu_l)}\left(\frac{Z_l-z_{ul}}{b_N}+1\right)^2\prod_{l=L_p+1}^{L}\mathcal{K}_l^{(\mu_l)}(0)^2 f_Z(z)\mathrm{d}z$$

Because the support of the kernel is $[-1,1]$ and by Assumption 4.1 $\mathcal{K}_l^{(\mu_l)}$ is bounded on this support we have

$$\mathcal{K}_l^{(\mu_l)}\left(\frac{Z_l-z_{ll}}{b_N}-1\right)\leq C\cdot 1\left(z_{ll}\leq z_l\leq z_{ll}+2b_N\right)\qquad\mathcal{K}_l^{(\mu_l)}\left(\frac{z_l-z_{ul}}{b_N}+1\right)\leq C\cdot 1\left(z_{ul}-2b_N\leq z_l\leq z_{ul}\right)$$

and substitution gives the upper bound

$$\frac{C_1}{b_N^{2L}}\int_{z_{l1}}^{z_{l1}+b_N}\cdots\int_{z_{lK_r}}^{z_{lK_r}+b_N}\int_{z_{u,K_r+1}-b_N}^{z_{u,K_r+1}}\cdots\int_{z_{u,L_p}-b_N}^{z_{u,L_p}}\prod_{l=1}^{K_r}1\left(z_{ll}\leq z_l\leq z_{ll}+2b_N\right)$$
$$\times\prod_{l=K_r+1}^{L_p}1\left(z_{ul}-2b_N\leq z_l\leq z_{ul}\right)f_Z(z_1,\ldots z_{L_p})\mathrm{d}z_1\ldots\mathrm{d}z_{L_p}\leq\frac{C_2}{b_N^{2L-L_p}}$$

Because $L_p\geq 1$ the integral over the boundary region is $O(b_N^{-2L+1})$. Combining the results we have that

$$\mathbb{E}[a_{N,\mu}(\tilde{Y}_i,Z_i,Z_{1i},Z_{2i})^2]=O(b_N^{-2L}) \tag{B.56}$$

which is larger than the bound in (B.51) and could be a reason to omit the terms $i=j=k$ (and redefine the kernel estimator).

Step 3C: Equation (B.49) Third, we consider $\mathbb{E}[a_{N,\mu}(\tilde{Y}_i,Z_i,Z_{1i},Z_{2k})^2]$. Again we have $\mathbb{E}[a_{N,\mu}(\tilde{Y}_i,Z_i,Z_{1i},Z_{2k})]\neq 0$. We have

$$\mathbb{E}[a_{N,\mu}(\tilde{Y}_i,Z_i,Z_{1i},Z_{2k})^2]$$
$$=\frac{1}{b_N^{2L+2|\mu|}}\mathbb{E}\left[K^{(\mu)}\left(\frac{Z_i-r_{b_N}(Z_{1i},Z_{2k})}{b_N}\right)^2\nu(Z_{1i},Z_{2k})'\tilde{Y}_i\tilde{Y}_i'\nu(Z_{1i},Z_{2k})((Z_{1i}'\ Z_{2k}')'-r_{b_N}(Z_{1i},Z_{2k}))^{2\mu}\right]$$

$$\tag{B.57}$$

$$-\frac{2}{b_N^{2L+2|\mu|}}\mathbb{E}_{Z_iZ_{2k}}\left[\nu(Z_{1i},Z_{2k})'g(Z_{1i},Z_{2k})K^{(\mu)}\left(\frac{Z_i-r_{b_N}(Z_{1i},Z_{2k})}{b_N}\right)\right. \tag{B.58}$$
$$\left.\times\mathbb{E}_Z\left(g(Z)'^{(\mu)}\left(\frac{Z-r_{b_N}(Z_{1i},Z_{2k})}{b_N}\right)\right)((Z'_{1i}\;Z'_{2k})'-r_{b_N}(Z_{1i},Z_{2k}))^{2\mu}\right]$$

$$+\frac{1}{b_N^{2L+2|\mu|}}\mathbb{E}_{Z_iZ_{2k}}\left[\left(\nu(Z_{1i},Z_{2k})'\mathbb{E}_Z\left(g(Z)K^{(\mu)}\left(\frac{Z-r_{b_N}(Z_{1i},Z_{2k})}{b_N}\right)\right)\right)^2\times((Z'_{1i}\;Z'_{2k})'-r_{b_N}(Z_{1i},Z_{2k}))^{2\mu}\right] \tag{B.59}$$

By Assumptions 3.1 and smoothness (B.57) is bounded by

$$\frac{C}{b_N^{2L}}\int_{\mathbb{Z}_2}\int_{\mathbb{Z}}K^{(\mu)}\left(\frac{z-r_{b_N}(z_1,\tilde{z}_2)}{b_N}\right)^2 f_Z(z)f_{Z_2}(\tilde{z}_2)\mathrm{d}z\mathrm{d}\tilde{z}_2$$
$$=\frac{C}{b_N^{2L}}\int_{\mathbb{Z}_2}\int_{\mathbb{Z}^I_{b_N,1}}\int_{\mathbb{Z}_2}K^{(\mu)}\left(\frac{z-r_{b_N}(z_1,\tilde{z}_2)}{b_N}\right)^2 f_Z(z_1,z_2)f_{Z_2}(\tilde{z}_2)\mathrm{d}z_2\mathrm{d}z_1\mathrm{d}\tilde{z}_2$$
$$+\frac{C}{b_N^{2L}}\int_{\mathbb{Z}_2}\int_{\mathbb{Z}_1\setminus\mathbb{Z}^I_{b_N,1}}\int_{\mathbb{Z}_2}K^{(\mu)}\left(\frac{z-r_{b_N}(z_1,\tilde{z}_2)}{b_N}\right)^2 f_Z(z_1,z_2)f_{Z_2}(\tilde{z}_2)\mathrm{d}z_2\mathrm{d}z_1\mathrm{d}\tilde{z}_2$$

Because $z_1-r_b(z_1,\tilde{z}_2)=0$ if $z_1\in\mathbb{Z}^I_{b_N,1}$, the first term on the right hand side is equal to

$$\frac{CK_1^{(\mu_1)}(0)}{b_N^{2L}}\int_{\mathbb{Z}_2}\int_{\mathbb{Z}^I_{b_N,1}}\int_{\mathbb{Z}_2}\mathcal{K}^{(\mu_2)}\left(\frac{z_2-r_{b_N}(z_1,\tilde{z}_2)}{b_N}\right)^2 f_Z(z_1,z_2)f_{Z_2}(\tilde{z}_2)\mathrm{d}z_2\mathrm{d}z_1\mathrm{d}\tilde{z}_2=O(b_N^{-2L+L_2}),$$

(where $\mathcal{K}(u)$ is the univariate kernel), by a change of variables to $t_2=(z_2-r_{b_N}(z_1,\tilde{z}_2)/b_N$ with Jacobian $b_N^{L_2}$. For the second integral we partition $\mathbb{Z}^B_{1,b_N}\equiv\mathbb{Z}_1\setminus\mathbb{Z}^I_{b_N,1}$ into sets $\mathbb{Z}^B_{1,b_N,p},p=1,\dots,2^{L_1}-1$ in which $1\le L_{1p}\le L_1$ components of z_1 are in the boundary region. Each $\mathbb{Z}^B_{1,b_N,p}$ is partitioned further into sets $\mathbb{Z}^B_{1,b_N,p,r},r=1,\dots,2^{L_{1p}}$ in which $0\le K_{1r}\le L_{1r}$ components of z_1 are near the lower, $L_{1r}-K_{1r}$ are near the upper boundary, and the remaining L_1-L_{1p} components are in the internal set. Hence, if we without loss of generality assume that the first K_r components of z_1 are near the lower boundary, the next $L_{1p}-K_{1r}$ are near the upper boundary, and the remaining components are in the internal set,

$$\frac{C}{b_N^{2L}}\int_{\mathbb{Z}_2}\int_{\mathbb{Z}_1\setminus\mathbb{Z}^I_{b_N,1}}\int_{\mathbb{Z}_2}K^{(\mu)}\left(\frac{z-r_{b_N}(z_1,\tilde{z}_2)}{b_N}\right)^2 f_Z(z_1,z_2)f_{Z_2}(\tilde{z}_2)\mathrm{d}z_2\mathrm{d}z_1\mathrm{d}\tilde{z}_2$$
$$=\frac{C}{b_N^{2L}}\int_{\mathbb{Z}_2}\int_{z_{l,11}}^{z_{l,l1}+b_N}\cdots\int_{z_{l1,K_{1r}}}^{z_{l1,K_{1r}}+b_N}\int_{z_{u1,K_{1r}+1}-b_N}^{z_{u1,K_{1r}+1}}$$
$$\cdots\int_{z_{u1,L_{1p}}-b_N}^{z_{u1,L_{1p}}}\int_{z_{l1,L_{1p}+1}+b_N}^{z_{u1,L_{1p}+1}-b_N}\cdots\int_{z_{l1,L_1}+b_N}^{z_{u1,L_1}-b_N}\int_{\mathbb{Z}_2}\prod_{l=1}^{K_{1r}}\mathcal{K}_l^{(\mu_l)}\left(\frac{z_{1l}-z_{l1l}}{b_N}-1\right)^2$$
$$\times\prod_{l=K_{1r}+1}^{L_{1p}}\mathcal{K}_l^{(\mu_l)}\left(\frac{z_{1l}-z_{u1l}}{b_N}+1\right)^2\prod_{l=L_{1p}+1}^{L_1}\mathcal{K}_l^{(\mu_l)}(0)^2\,K_2^{(\mu_2)}\left(\frac{z_2-r_{b_N}(z_1,\tilde{z}_2)}{b_N}\right)^2 f_Z(z_1,z_2)f_{Z_2}(\tilde{z}_2)\mathrm{d}z_2\mathrm{d}z_1\mathrm{d}\tilde{z}_2$$

After a change of variables to $t_2=(z_2-r_{b_N}(z_1,\tilde{z}_2)/b_N$ with Jacobian $b_N^{L_2}$ we have by analogous argument as above that this term is $O(b_N^{-2L+L_2+L_{1p}})$. Because $L_{1p}\ge1$ we have by combining the results

$$\mathbb{E}[a_{N,\mu}(\tilde{Y}_i,Z_i,Z_{1i},Z_{2k})^2]=O(b_N^{-2L+L_2}) \tag{B.60}$$

Step 3D: Equation (B.50)

An analogous argument gives

$$\mathbb{E}[a_{N,\mu}(\tilde{Y}_i,Z_i,Z_{1j},Z_{2i})^2]=O(b_N^{-2L+L_1}) \tag{B.61}$$

This finishes the derivation of the bounds on the second moments of the kernel of the V-statistic.
Now we turn to the proofs of equalities (B.41)-(B.46).

Step 3E: Equation (B.41) For the first term

$$W_{\mu,1} - U_\mu = \frac{\sqrt{N}}{N(N-1)(N-2)} \sum_{i \neq j \neq k} (a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k}) - c_{N,\mu}(\tilde{Y}_i, Z_i)) \tag{B.62}$$

so that

$$\begin{aligned} &\mathbb{E}[(W_{\mu,1} - U_\mu)^2] \\ &= \frac{N}{N^2(N-1)^2(N-2)^2} \sum_{i \neq j \neq k} \sum_{i' \neq j' \neq k'} \mathbb{E}[(a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k}) - c_{N,\mu}(\tilde{Y}_i, Z_i)) \\ &\times (a_{N,\mu}(\tilde{Y}_{i'}, Z_{i'}, Z_{1j'}, Z_{2k'}) - c_{N,\mu}(\tilde{Y}_{i'}, Z_{i'}))] \end{aligned}$$

This expression can be simplified using

$$\mathbb{E}[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k})] = 0 \tag{B.63}$$

$$\mathbb{E}[c_{N,\mu}(\tilde{Y}_i, Z_i)] = 0 \tag{B.64}$$

$$\mathbb{E}[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k})|\tilde{Y}_i, Z_i] = c_{N,\mu}(\tilde{Y}_i, Z_i) \tag{B.65}$$

$$\mathbb{E}[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k}) a_{N,\mu}(V_{i'}, Z_{i'}, Z_{1j'}, Z_{2k})|Z_{2k}] = 0 \tag{B.66}$$

$$\mathbb{E}[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k}) a_{N,\mu}(V_{i'}, Z_{i'}, Z_{1j}, Z_{2k'})|Z_{1j}] = 0 \tag{B.67}$$

$$\mathbb{E}[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k}) a_{N,\mu}(V_{i'}, Z_{i'}, Z_{1j}, Z_{2k'})|Z_{1j}, Z_{2k}] = 0 \tag{B.68}$$

Therefore

$$\mathbb{E}[(a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k}) - c_{N,\mu}(\tilde{Y}_i, Z_i))(a_{N,\mu}(V_{i'}, Z_{i'}, Z_{1j'}, Z_{2k'}) - c_{N,\mu}(V_{i'}, Z_{i'}))] = 0$$

if $i \neq i', j \neq j', k \neq k'$ by (B.63) and (B.64), if $i = i', j \neq j', k \neq k'$ by (B.65), if $i \neq i', j \neq j', k = k'$ by (B.66), and if $i \neq i', j = j', k \neq k'$ by (B.67), and if $i \neq i', j = j', k = k'$ by (B.68). Using this we obtain

$$\begin{aligned} &\mathbb{E}[(W_{\mu,1} - U_\mu)^2] \\ &= \frac{N}{N^2(N-1)^2(N-2)^2} \sum_{i \neq j \neq k} \mathbb{E}[(a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k}) - c_{N,\mu}(\tilde{Y}_i, Z_i))^2] \\ &+ \frac{N}{N^2(N-1)^2(N-2)^2} \sum_{i \neq j \neq k \neq k'} \mathbb{E}[(a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k}) - c_{N,\mu}(\tilde{Y}_i, Z_i))(a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k'}) - c_{N,\mu}(\tilde{Y}_i, Z_i))] \\ &+ \frac{N}{N^2(N-1)^2(N-2)^2} \sum_{i \neq k \neq j \neq j'} \mathbb{E}[(a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k}) - c_{N,\mu}(\tilde{Y}_i, Z_i))(a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j'}, Z_{2k}) - c_{N,\mu}(\tilde{Y}_i, Z_i))] \end{aligned} \tag{B.69}$$

Because $\mathbb{E}[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k})|\tilde{Y}_i, Z_i] = c_{N,\mu}(\tilde{Y}_i, Z_i)$ we have $\mathbb{E}[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k}) c_{N,\mu}(\tilde{Y}_i, Z_i)] = \mathbb{E}[c_{N,\mu}(\tilde{Y}_i, Z_i)^2]$ so that by the bounds on the second moment of $a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k})$, given in (B.47)-(B.50),

$$\begin{aligned} \mathbb{E}[(a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k}) - c_{N,\mu}(\tilde{Y}_i, Z_i))^2] &= \mathbb{E}[(a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k})^2] - \mathbb{E}[c_{N,\mu}(\tilde{Y}_i, Z_i)^2] \\ &\leq \mathbb{E}[(a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k})^2] = O(b_N^{-L}). \end{aligned}$$

Further (note that $\mathbb{E}[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k}) a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k'})] = \mathbb{E}[(\mathbb{E}_{Z_2}[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_2)])^2] \geq 0$)

$$\begin{aligned} &\mathbb{E}[(a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k}) - c_{N,\mu}(\tilde{Y}_i, Z_i))(a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k'}) - c_{N,\mu}(\tilde{Y}_i, Z_i))] \\ &= \mathbb{E}[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k}) a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k'})] - \mathbb{E}[c_{N,\mu}(\tilde{Y}_i, Z_i)^2] \\ &\leq \mathbb{E}[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k}) a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k'})] \end{aligned}$$

and

$$
\begin{aligned}
&\mathbb{E}[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k}) a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k'})] \\
&= \frac{1}{b_N^{2|\mu|+2L}} \mathbb{E}\left[\nu(Z_{1j}, Z_{2k})' \tilde{Y}_i \tilde{Y}_i' \nu(Z_{1j}, Z_{2k'}) K^{(\mu)}\left(\frac{Z_i - r_{b_N}(Z_{1j}, Z_{2k})}{b_N}\right) K^{(\mu)}\left(\frac{Z_i - r_{b_N}(Z_{1j}, Z_{2k'})}{b_N}\right)\right. \\
&\left.\times((Z_{1j}' \ Z_{2k}')' - r_{b_N}(Z_{1j}, Z_{2k}))^{\mu}((Z_{1j}' \ Z_{2k'}')' - r_{b_N}(Z_{1j}, Z_{2k'}))^{\mu}\right] \\
&- \frac{1}{b_N^{2|\mu|+2L}} \mathbb{E}\left[\nu(Z_{1j}, Z_{2k})' \mathbb{E}_{\tilde{Y}Z}\left[\tilde{Y} K^{(\mu)}\left(\frac{S - r_{b_N}(Z_{1j}, Z_{2k})}{b_N}\right)\right] \mathbb{E}_{\tilde{Y}Z}\left[\tilde{Y} V'^{(\mu)}\left(\frac{Z - r_{b_N}(Z_{1j}, Z_{2k'})}{b_N}\right)\right] \nu(Z_{1j}, Z_{2k'})\right. \\
&\left.\times((Z_{1j}' \ Z_{2k}')' - r_{b_N}(Z_{1j}, Z_{2k}))^{\mu}((Z_{1j}' \ Z_{2k'}')' - r_{b_N}(Z_{1j}, Z_{2k'}))^{\mu}\right] \\
&\leq \frac{1}{b_N^{2|\mu|+2L}} \mathbb{E}\left[\nu(Z_{1j}, Z_{2k})' \tilde{Y}_i \tilde{Y}_i' \nu(Z_{1j}, Z_{2k'}) K^{(\mu)}\left(\frac{Z_i - r_{b_N}(Z_{1j}, Z_{2k})}{b_N}\right) K^{(\mu)}\left(\frac{Z_i - r_{b_N}(Z_{1j}, Z_{2k'})}{b_N}\right)\right. \\
&\left.\times((Z_{1j}' \ Z_{2k}')' - r_{b_N}(Z_{1j}, Z_{2k}))^{\mu}((Z_{1j}' \ Z_{2k'}')' - r_{b_N}(Z_{1j}, Z_{2k'}))^{\mu}\right]
\end{aligned}
$$

because both expectations are nonnegative. By Assumptions 3.1 and smoothness this is bounded by

$$
\frac{C_1}{b_N^{2L}} \mathbb{E}\left[K^{(\mu)}\left(\frac{Z_i - r_{b_N}(Z_{1j}, Z_{2k})}{b_N}\right) K^{(\mu)}\left(\frac{Z_i - r_{b_N}(Z_{1j}, Z_{2k'})}{b_N}\right)\right] \leq C_2 b_N^{-L}
$$

by a change of variables to $t = (Z_i - r_{b_N}(Z_{1j}, Z_{2k}))/b_N$ with Jacobian b_N^L and Assumption 4.1. By interchanging the roles of j and k we obtain a bound of the same order for the third term on the right hand side of (B.69). Combining these results we find

$$
\mathbb{E}[(W_{\mu,1} - U_{\mu})^2] = O(N^{-2} b_N^{-L}) + O(N^{-1} b_N^{-L}) = O(N^{-1} b_N^{-L}) \tag{B.70}
$$

so that by the Markov inequality the first term in the projection remainder (B.40) is $O_p(N^{-1/2} b_N^{-L/2})$.

Step 3F: Equation (B.42) For the second term of the projection remainder (B.40) we have by the Cauchy-Schwartz inequality

$$
\begin{aligned}
&\mathbb{E}[c_{N,\mu}(\tilde{Y}_i, Z_i)^2] \\
&\leq \frac{1}{b_N^{2L+2|\mu|}} \mathbb{E}\left[\left(\int_{\mathbb{Z}_2} \int_{\mathbb{Z}_1} \nu(z_1, z_2)' \tilde{Y}_i K^{(\mu)}\left(\frac{Z_i - r_{b_N}(z_1, z_2)}{b_N}\right) ((z_1' \ z_2')' - r_{b_N}(z_1, z_2))^{\mu} f_{Z_1}(z_1) f_{Z_2}(z_2) \mathrm{d}z_1 \mathrm{d}z_2\right)^2\right] \\
&\leq \frac{1}{b_N^{2L+2|\mu|}} \mathbb{E}\left[|\tilde{Y}_i|^2 \left(\int_{\mathbb{Z}_2} \int_{\mathbb{Z}_1} |\nu(z_1, z_2)| \left|K^{(\mu)}\left(\frac{Z_i - r_{b_N}(z_1, z_2)}{b_N}\right)\right| \left|((z_1' \ z_2')' - r_{b_N}(z_1, z_2)^{\mu}\right| f_{Z_1}(z_1) f_{Z_2}(z_2) \mathrm{d}z_1 \mathrm{d}z_2\right)^2\right] \\
&\leq \frac{C_1}{b_N^{2L}} \mathbb{E}_{Z_i}\left[\left(\int_{\mathbb{Z}_2} \int_{\mathbb{Z}_1} \left|K^{(\mu)}\left(\frac{Z_i - r_{b_N}(z_1, z_2)}{b_N}\right)\right| f_{Z_1}(z_1) f_{Z_2}(z_2) \mathrm{d}z_1 \mathrm{d}z_2\right)^2 \mathbb{E}(|\tilde{Y}_i|^2 | Z_i)\right] \\
&\leq \frac{C_2}{b_N^{2L}} \int_{\mathbb{Z}} \left(\int_{\mathbb{Z}_2} \int_{\mathbb{Z}_1} \left|K^{(\mu)}\left(\frac{\tilde{z} - r_{b_N}(z_1, z_2)}{b_N}\right)\right| f_{Z_1}(z_1) f_{Z_2}(z_2) \mathrm{d}z_1 \mathrm{d}z_2\right)^2 f_Z(\tilde{z}) \mathrm{d}\tilde{z}
\end{aligned}
$$

by Assumptions 3.1 and smoothness. By a change of variables $t = (\tilde{z} - r_{b_N}(z_1, z_2))/b_N$ with Jacobian b_N^L we conclude that

$$
\mathbb{E}[c_{N,\mu}(\tilde{Y}_i, Z_i)^2] = O(b_N^{-L}) \tag{B.71}
$$

Therefore

$$
\mathbb{E}[U_{\mu}^2] = \mathbb{E}[c_{N,\mu}(\tilde{Y}_i, Z_i)^2] = O(b_N^{-L})
$$

so that the second term of the projection remainder is $O_p(N^{-1} b_N^{-L/2})$.

Step 3F: Equations (B.43)-(B.46) The other terms of the projection remainder can be bounded using (B.47)-(B.50). For the third term (note $\mathbb{E}[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1i}, Z_{2k})] \neq 0$) by (B.47)-(B.50)

$$\begin{aligned}\mathbb{E}[|W_{\mu,2}|] &\leq \frac{\sqrt{N}(N-1)}{N^2}\mathbb{E}[|a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1i}, Z_{2k})|] \\ &\leq \frac{\sqrt{N}(N-1)}{N^2}\sqrt{\mathbb{E}[|a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1i}, Z_{2k})|^2]} \\ &= O\left(N^{-\frac{1}{2}}b_N^{-L+\frac{L_2}{2}}\right)\end{aligned}$$

so that that term is $O_p\left(N^{-\frac{1}{2}}b_N^{-L+\frac{L_2}{2}}\right)$. In the same way by (B.47)-(B.50) the fourth term of the remainder is $O_p\left(N^{-\frac{1}{2}}b_N^{-L+\frac{L_1}{2}}\right)$. For the fifth term (note $a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2j}) = 0$)

$$\mathbb{E}[W_{\mu,4}^2] = \frac{N^2(N-1)}{N^6}\mathbb{E}[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2j})^2] = O\left(N^{-3}b_N^{-L}\right)$$

so that that term is $O_p\left(N^{-3/2}b_N^{-L/2}\right)$. Finally, the sixth term (note $\mathbb{E}[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1i}, Z_{2i})] \neq 0$) is by a similar argument as for the third term and by (B.47)-(B.50), $O_p\left(N^{-1/2}b_N^{-L}\right)$. This is the largest term in the projection remainder.

This finishes the proof of

$$W_\mu = U_\mu + O_p\left(N^{-\frac{1}{2}}b_N^{-L}\right) \tag{B.72}$$

Note again that the remainder is smaller if we redefine the kernel estimators. In that case the sixth term of the projection remainder is 0.

Step 4: Asymptotic distribution The fourth step in the proof is the derivation of the asymptotically normal distribution of the projection U_μ. In particular, we show that U_0 is asymptotically normal and we obtain the variance of that distribution. We show that U_μ/b_N also converges to a normal distribution for $|\mu| \geq 1$ so that $U_\mu = O_p(b_N)$ if $|\mu| \geq 1$. Because W in (B.36) is a linear combination of the W_μ that are asymptotically equivalent to the U_μ if a rate condition is met, W is asymptotically equivalent to U_0 under that rate condition. Define

$$\psi_{N,\mu,i} \equiv \frac{1}{b_N^{L+|\mu|}}\int_{\mathbb{Z}_2}\int_{\mathbb{Z}_1}\nu(z_1,z_2)'\tilde{Y}_iK^{(\mu)}\left(\frac{Z_i - r_{b_N}(z_1,z_2)}{b_N}\right)((z_1' \; z_2')' - r_{b_N}(z_1,z_2))^\mu f_{Z_1}(z_1)f_{Z_2}(z_2)\mathrm{d}z_1\mathrm{d}z_2$$

so that

$$c_{N,\mu}(\tilde{Y}_i, Z_i) = \psi_{N,\mu,i} - \mathbb{E}[\psi_{N,\mu,i}].$$

We have

$$\psi_{N,0,i} \equiv \frac{1}{b_N^L}\int_{\mathbb{Z}_2}\int_{\mathbb{Z}_1}\nu(z_1,z_2)'\tilde{Y}_iK\left(\frac{Z_i - r_{b_N}(z_1,z_2)}{b_N}\right)f_{Z_1}(z_1)f_{Z_2}(z_2)\mathrm{d}z_1\mathrm{d}z_2$$

The integration region $\mathbb{Z}_1 \times \mathbb{Z}_2$ can be partitioned into a set where all components of z_1 and z_2 are in the internal region, $\mathbb{Z}_{1,b_N}^I \times \mathbb{Z}_{2,b_N}^I$, and its complement, $\mathbb{Z}_1 \times \mathbb{Z}_2 \setminus \mathbb{Z}_{1,b_N}^I \times \mathbb{Z}_{2,b_N}^I$. We define

$$\psi_{N,0,i,0} \equiv \frac{1}{b_N^L}\int_{\mathbb{Z}_{2,b_N}^I}\int_{\mathbb{Z}_{1,b_N}^I}\nu(z_1,z_2)'\tilde{Y}_iK_1\left(\frac{Z_{1i} - z_1}{b_N}\right)K_2\left(\frac{Z_{2i} - z_2}{b_N}\right)f_{Z_1}(z_1)f_{Z_2}(z_2)\mathrm{d}z_1\mathrm{d}z_2$$

and

$$U_{0,0} \equiv \frac{1}{\sqrt{N}}\sum_{I=1}^{N}(\psi_{N,0,i,0} - \mathbb{E}[\psi_{N,0,i,0}])$$

We apply the Liapounov central limit theorem for triangular arrays that requires

$$\frac{N^2\left(\mathbb{E}\left[|\psi_{N,0,i,0}-\mathbb{E}[\psi_{N,0,i,0}]|^3\right]\right)^2}{N^3\mathrm{Var}(\psi_{N,0,i,0})}\to 0$$

and a sufficient condition is that $\mathbb{E}\left[|\psi_{N,0,i,0}|^m\right]<\infty$ for $m=1,2,3$. By a change of variables to $t_1=(Z_{1i}-z_1)/b_N$ and $t_2=(Z_{2i}-z_2)/b_N$ with Jacobians $b_N^{L_1}$ and $b_N^{L_2}$, respectively

$$\begin{aligned}
&|\psi_{N,0,i,0}|^m\\
&=\Bigg|\int_{\mathcal{U}_1}\int_{\mathcal{U}_2}\prod_{l=1}^{L_1}1\left(1+\frac{Z_{1li}-z_{u1l}}{b_N}\le t_{1l}\le -1+\frac{Z_{1li}-z_{l1l}}{b_N}\right)\prod_{l=1}^{L_2}1\left(1+\frac{Z_{2li}-z_{u2l}}{b_N}\le t_{2l}\le -1+\frac{Z_{2li}-z_{l2l}}{b_N}\right)\\
&\times\nu\left(Z_{1i}-b_Nt_1,Z_{2i}-b_Nt_2\right)'\tilde{Y}_iK_1(t_1)K_2(t_2)f_{Z_1}(Z_{1i}-b_Nt_1)f_{Z_2}(Z_{2i}-b_Nt_2)\mathrm{d}t_1\mathrm{d}t_2\Bigg|^m\\
&\le\Bigg(\int_{\mathcal{U}_1}\int_{\mathcal{U}_2}\prod_{l=1}^{L_1}1\left(1+\frac{Z_{1li}-z_{u1l}}{b_N}\le t_{1l}\le -1+\frac{Z_{1li}-z_{l1l}}{b_N}\right)\prod_{l=1}^{L_2}1\left(1+\frac{Z_{2li}-z_{u2l}}{b_N}\le t_{2l}\le -1+\frac{Z_{2li}-z_{l2l}}{b_N}\right)\\
&\times|\nu\left(Z_{1i}-b_Nt_1,Z_{2i}-b_Nt_2\right)|\cdot|\tilde{Y}_i|\cdot|K_1(t_1)|\cdot|K_2(t_2)|f_{Z_1}(Z_{1i}-b_Nt_1)f_{Z_2}(Z_{2i}-b_Nt_2)\mathrm{d}t_1\mathrm{d}t_2\Bigg)^m
\end{aligned}$$

by the Cauchy-Schwartz inequality . Because $\max\left\{-1,1+\frac{Z_{jli}-z_{ujl}}{b_N}\right\}\le t_{jl}\le\min\left\{1,-1+\frac{Z_{jli}-z_{ljl}}{b_N}\right\}, j=1,2$ if and only if $z_{ljl}+b_N\le Z_{jli}-b_Nt_{jl}\le z_{ujl}-b_N$, we obtain by Assumptions 3.1, 4.1 and smoothness

$$|\psi_{N,0,i,0}|^m\le C|\tilde{Y}_i|^m \tag{B.73}$$

and $\mathbb{E}[|\tilde{Y}|^3]$ is finite by Assumption 3.1. Therefore the condition of the Liapounov theorem holds. The above expressions also show that for almost all Z_{1i},Z_{2i}

$$\psi_{N,0,i,0}\to\nu\left(Z_{1i},Z_{2i}\right)'\tilde{Y}_if_{Z_1}(Z_{1i})f_{Z_2}(Z_{2i})$$

and by (B.73) $\mathbb{E}[\psi^m_{N,0,i,0}]$ converges to the corresponding expectation by dominated convergence. The conclusion is that $U_{0,0}$ has the same asymptotic distribution as

$$\frac{1}{\sqrt{N}}\sum_{i=1}^N\left\{\nu\left(Z_{1i},Z_{2i}\right)'\tilde{Y}_if_{Z_1}(Z_{1i})f_{Z_2}(Z_{2i})-\mathbb{E}[\nu\left(Z_1,Z_2\right)'\tilde{Y}f_{Z_1}(Z_1)f_{Z_2}(Z_2)]\right\} \tag{B.74}$$

We still have to derive the stochastic order of

$$U_{0,1}\equiv\frac{1}{\sqrt{N}}\sum_{I=1}^N(\psi_{N,0,i,1}-\mathbb{E}[\psi_{N,0,i,1}])$$

with the integration region in $\psi_{N,0,i,1}$, i.e. $\mathbb{Z}_1\times\mathbb{Z}_2\setminus\mathbb{Z}^I_{1,b_N}\times\mathbb{Z}^I_{2,b_N}$, such that at least one component of z_1 or z_2 is in the boundary region. We partition $\mathbb{Z}_1\times\mathbb{Z}_2\setminus\mathbb{Z}^I_{1,b_N}\times\mathbb{Z}^I_{2,b_N}$ into subsets $\mathbb{Z}^B_{1,b_N,p_1}\times\mathbb{Z}^B_{1,b_N,p_2}, p_1=1,\dots,2^{L_1}, p_2=1,\dots,2^{L_2},\min\{p_1,p_2\}\ge 1$ and in each such set $0\le L_{1p_1}\le L_1, 0\le L_{2p_2}\le L_2,\min\{L_{1p_1}L_{1p_1}\}\ge 1$ components of z_1 and z_2 are near the boundary. We take without loss of generality $\mathbb{Z}^B_{1,b_N,1}=\mathbb{Z}^I_{1,b_N}$ and $\mathbb{Z}^B_{2,b_N,1}=\mathbb{Z}^I_{2,b_N}$ so that we exclude the set with $p_1=p_2=1$ because in that set all components are in the internal region. For $j=1,2$ each $\mathbb{Z}^B_{j,b_N,p_j}$ is partitioned further into sets $\mathbb{Z}^B_{j,b_N,p_j,r_j}, r_j=1,\dots,2^{L_{1p_j}}$ in which $0\le K_{jr_j}\le L_{jr_j}$ components of z_j are near the lower, $L_{jr_j}-K_{1r_j}$ are near the upper boundary, and the remaining $L_j-L_{jp_j}$ components are in the internal set. Without loss of generality we assume that the first K_{jr_j} components of z_j are near the lower boundary, the next $L_{jp_j}-K_{jr_j}$ are near the upper boundary, and the remaining components are in the internal set, $j = 1, 2$. Therefore

$$
\begin{aligned}
|\psi_{N,0,i,1}|^m &= \left| \frac{1}{b_N^L} \int_{\mathbb{Z}_2 \times \mathbb{Z}_1 \setminus \mathbb{Z}^I_{1,b_N} \times \mathbb{Z}^I_{2,b_N}} \nu(z_1, z_2)' \tilde{Y}_i K\left(\frac{Z_i - r_{b_N}(z_1, z_2)}{b_N}\right) f_{Z_1}(z_1) f_{Z_2}(z_2) \mathrm{d}z_1 \mathrm{d}z_2 \right|^m \qquad \text{(B.75)} \\
&\leq \left(\frac{1}{b_N^L} \int_{\mathbb{Z}_2 \times \mathbb{Z}_1 \setminus \mathbb{Z}^I_{1,b_N} \times \mathbb{Z}^I_{2,b_N}} |\nu(z_1, z_2)| |\tilde{Y}_i| \left| K\left(\frac{Z_i - r_{b_N}(z_1, z_2)}{b_N}\right) \right| f_{Z_1}(z_1) f_{Z_2}(z_2) \mathrm{d}z_1 \mathrm{d}z_2 \right)^m \\
&\leq \sum_{p_1} \sum_{p_2} \sum_{r_1} \sum_{r_2} \left(\frac{1}{b_N^L} \int_{z_{l21}}^{z_{l21}+b_N} \cdots \int_{z_{l2K_{2r_2}}}^{z_{l2K_{2r_2}}+b_N} \int_{z_{u2,K_{2r_2}+1}-b_N}^{z_{u2,K_{2r_2}+1}} \cdots \int_{z_{u2,L_{2p_2}}-b_N}^{z_{u2,L_{2p_2}}} \int_{z_{l2,L_{2p_2}+1}+b_N}^{z_{u2,L_{2p_2}+1}-b_N} \cdots \int_{z_{l2,L_2}+b_N}^{z_{u2,L_2}-b_N} \right. \\
&\int_{z_{l,l1}}^{z_{l,l1}+b_N} \cdots \int_{z_{l,1K_{1r_1}}}^{z_{l,1K_{1r_1}}+b_N} \int_{z_{u1,K_{1r_1}+1}-b_N}^{z_{u1,K_{1r_1}+1}} \cdots \int_{z_{u1,L_{1p_1}}-b_N}^{z_{u1,L_{1p_1}}} \int_{z_{u1,L_{1p_1}+1}+b_N}^{z_{u1,L_{1p_1}+1}-b_N} \cdots \int_{z_{l1,L_1}+b_N}^{z_{u1,L_1}-b_N} |\nu(z_1, z_2)| |\tilde{Y}_i| \\
&\prod_{l=1}^{K_{1r_1}} \left| \mathcal{K}_{1l}\left(\frac{Z_{1li} - z_{l1l}}{b_N} - 1\right) \right| \prod_{l=K_{1r_1}+1}^{L_{1p_1}} \left| \mathcal{K}_{1l}\left(\frac{Z_{1li} - z_{l1l}}{b_N} + 1\right) \right| \prod_{l=L_{1p_1}+1}^{L_1} \left| \mathcal{K}_{1l}\left(\frac{Z_{1li} - z_{1l}}{b_N}\right) \right| \\
&\prod_{l=1}^{K_{2r_1}} \left| \mathcal{K}_{2l}\left(\frac{Z_{2li} - z_{l2l}}{b_N} - 1\right) \right| \prod_{l=K_{2r_2}+1}^{L_{2p_2}} \left| \mathcal{K}_{2l}\left(\frac{Z_{2li} - z_{u2l}}{b_N} + 1\right) \right| \\
&\left. \prod_{l=L_{2p_2}+1}^{L_2} \left| \mathcal{K}_{2l}\left(\frac{Z_{2li} - z_{2l}}{b_N}\right) \right| f_{Z_1}(z_1) f_{Z_2}(z_2) \mathrm{d}z_1 \mathrm{d}z_2 \right)^m
\end{aligned}
$$

By a change of variables to $t_{1l} = (Z_{1li} - z_{1l})/b_N, l = L_{1p_1} + 1, L_1$ and $t_{2l} = (Z_{2li} - z_{2l})/b_N, l = L_{2p_2} + 1, L_2$ with Jacobian $b_N^{L - L_{1p_1} - L_{2p_2}}$ we have

$$
\begin{aligned}
|\psi_{N,0,i,1}|^m &\leq \sum_{p_1} \sum_{p_2} \sum_{r_1} \sum_{r_2} \left(\frac{1}{b_N^{m(L_{1p_1} + L_{2p_2})}} \int_{z_{l,21}}^{z_{l,21}+b_N} \cdots \int_{z_{l,2K_{2r_2}}}^{z_{l,2K_{2r_2}}+b_N} \int_{z_{u2,K_{2r_2}+1}-b_N}^{z_{u2,K_{2r_2}+1}} \cdots \int_{z_{u2,L_{2p_2}}-b_N}^{z_{u2,L_{2p_2}}} \int_{-1}^{1} \right. \\
&\cdots \int_{-1}^{1} \int_{z_{l,l1}}^{z_{l,l1}+b_N} \cdots \int_{z_{l,1K_{1r_1}}}^{z_{l,1K_{1r_1}}+b_N} \int_{z_{u1,K_{1r_1}+1}-b_N}^{z_{u1,K_{1r_1}+1}} \cdots \int_{z_{u1,L_{1p_1}}-b_N}^{z_{u1,L_{1p_1}}} \int_{-1}^{1} \\
&\cdots \int_{-1}^{1} \prod_{j=1}^{2} \prod_{l=L_{jp_j}+1}^{L_j} 1\left(\frac{Z_{jli} - z_{ujl}}{b_N} + 1 \leq t_{jl} \leq \frac{Z_{jli} - z_{ljl}}{b_N} - 1\right) \\
&|\nu(z_{11}, \ldots, z_{1L_{1p_1}}, Z_{1,L_{1p_1}+1,i} - b_N t_{1L_{1p_1}+1}, \ldots, Z_{1,L_1,i} \\
&- b_N t_{1L_1}, z_{21}, \ldots, z_{2L_{2p_2}}, Z_{2,L_{2p_2}+1,i} - b_N t_{2L_{2p_2}+1}, \ldots, Z_{2,L_2,i} - b_N t_{2L_2})| |\tilde{Y}_i| \\
&\prod_{l=1}^{K_{1r_1}} \left| \mathcal{K}_{1l}\left(\frac{Z_{1li} - z_{l1l}}{b_N} - 1\right) \right| \prod_{l=K_{1r_1}+1}^{L_{1p_1}} \left| \mathcal{K}_{1l}\left(\frac{Z_{1li} - z_{l1l}}{b_N} + 1\right) \right| \\
&\prod_{l=1}^{K_{2r_1}} \left| \mathcal{K}_{2l}\left(\frac{Z_{2li} - z_{l2l}}{b_N} - 1\right) \right| \prod_{l=K_{2r_2}+1}^{L_{2p_2}} \left| \mathcal{K}_{2l}\left(\frac{Z_{2li} - z_{u2l}}{b_N} + 1\right) \right| \\
&\prod_{l=L_{1p_1}+1}^{L_1} |\mathcal{K}_{1l}(t_{1l})| \prod_{l=L_{2p_2}+1}^{L_2} |\mathcal{K}_{2l}(t_{2l})| f_{Z_1}(z_{11}, \ldots, z_{1L_{1p_1}}, Z_{1,L_{1p_1}+1,i} - b_N t_{1L_{1p_1}+1}, \ldots, Z_{1,L_1,i} - b_N t_{1L_1}) \\
&f_{Z_2}(z_{21}, \ldots, z_{2L_{2p_2}}, Z_{2,L_{2p_2}+1,i} - b_N t_{2L_{2p_2}+1}, \ldots, Z_{2,L_2,i} - b_N t_{2L_2}) \mathrm{d}z_{11} \cdots \mathrm{d}z_{1L_{1p_1}} \mathrm{d}t_{1,L_{1p_1}+1} \cdots \mathrm{d}t_{1L_1} \\
&\left. \mathrm{d}z_{21} \cdots \mathrm{d}z_{2L_{2p_2}} \mathrm{d}t_{2,L_{2p_2}+1} \cdots \mathrm{d}t_{2L_2} \right)^m
\end{aligned}
$$

In this integral the function ν takes only values in the support $\mathbb{Z}$ and this function and the kernel functions are bounded by smoothness and Assumption 4.1 so that

$$
\begin{aligned}
&|\psi_{N,0,i,1}|^m \\
&\leq C|\tilde{Y}_i|^m \sum_{p_1}\sum_{p_2}\sum_{r_1}\sum_{r_2} \frac{1}{b_N^{m(L_{1p_1}+L_{2p_2})}} \\
&\times \left(\prod_{l=1}^{K_{1r_1}} \left| \mathcal{K}_{1l}\left(\frac{Z_{1li}-z_{l,11}}{b_N}-1\right)\right| \prod_{l=K_{1r_1}+1}^{L_{1p_1}} \left|\mathcal{K}_{1l}\left(\frac{Z_{1li}-z_{l1l}}{b_N}+1\right)\right| \right. \\
&\prod_{l=1}^{K_{2r_1}} \left| \mathcal{K}_{2l}\left(\frac{Z_{2li}-z_{l2l}}{b_N}-1\right)\right| \prod_{l=K_{2r_2}+1}^{L_{2p_2}} \left|\mathcal{K}_{2l}\left(\frac{Z_{2li}-z_{u2l}}{b_N}+1\right)\right| \\
&\int_{z_{l,21}}^{z_{l,21}+b_N} \cdots \int_{z_{l,2K_{2r_2}}}^{z_{l,2K_{2r_2}}+b_N} \int_{z_{u2,K_{2r_2}+1}-b_N}^{z_{u2,K_{2r_2}+1}} \cdots \int_{z_{u2,L_{2p_2}}-b_N}^{z_{u2,L_{2p_2}}} \int_{-1}^{1} \\
&\cdots \int_{-1}^{1} \int_{z_{l,l1}}^{z_{l,l1}+b_N} \cdots \int_{z_{l,1K_{1r_1}}}^{z_{l,1K_{1r_1}}+b_N} \int_{z_{u1,K_{1r_1}+1}-b_N}^{z_{u1,K_{1r_1}+1}} \cdots \int_{z_{u1,L_{1p_1}}-b_N}^{z_{u1,L_{1p_1}}} \int_{-1}^{1} \cdots \int_{-1}^{1} \\
&f_{Z_1}(z_{11},\ldots,z_{1L_{1p_1}}, Z_{1,L_{1p_1}+1,i}-b_N t_{1L_{1p_1}+1},\ldots,Z_{1,L_1,i}-b_N t_{1L_1}) \\
&\times f_{Z_2}(z_{21},\ldots,z_{2L_{2p_2}}, Z_{2,L_{2p_2}+1,i}-b_N t_{2L_{2p_2}+1},\ldots,Z_{2,L_2,i}-b_N t_{2L_2}) \\
&\left. \mathrm{d}z_{11}\cdots \mathrm{d}z_{1L_{1p_1}} \mathrm{d}t_{1,L_{1p_1}+1}\cdots \mathrm{d}t_{1L_1} \mathrm{d}z_{21}\cdots \mathrm{d}z_{2L_{2p_2}} \mathrm{d}t_{2,L_{2p_2}+1}\cdots \mathrm{d}t_{2L_2}\right)^m
\end{aligned}
$$

Because the density is bounded, the integral is bounded by $Cb_N^{L_{1p_1}+L_{2p_2}}$. Moreover because the kernel has support $[-1,1]^L$ and is bounded on that support we have that

$$
\begin{aligned}
&\prod_{l=1}^{K_{1r_1}} \left| \mathcal{K}_{1l}\left(\frac{Z_{1li}-z_{l1l}}{b_N}-1\right)\right| \prod_{l=K_{1r_1}+1}^{L_{1p_1}} \left|\mathcal{K}_{1l}\left(\frac{Z_{1li}-z_{l1l}}{b_N}+1\right)\right| \\
&\times \prod_{l=1}^{K_{2r_1}} \left| \mathcal{K}_{2l}\left(\frac{Z_{2li}-z_{l2l}}{b_N}-1\right)\right| \prod_{l=K_{2r_2}+1}^{L_{2p_2}} \left|\mathcal{K}_{2l}\left(\frac{Z_{2li}-z_{u2l}}{b_N}+1\right)\right| \\
&\leq C \prod_{l=1}^{K_{1r_1}} 1\left(z_{l1l} \leq Z_{1li} \leq z_{l1l}+2b_N\right) \prod_{l=K_{1r_1}+1}^{L_{1p_1}} 1\left(z_{l1l}-2b_N \leq Z_{1li} \leq z_{l1l}\right) \\
&\qquad \times \prod_{l=1}^{K_{2r_1}} 1\left(z_{l2l} \leq Z_{2li} \leq z_{l2l}+2b_N\right) \prod_{l=K_{2r_2}+1}^{L_{2p_2}} 1\left(z_{u2l}-2b_N \leq Z_{2li} \leq z_{u2l}\right)
\end{aligned}
$$

Therefore

$$
\begin{aligned}
|\psi_{N,0,i,1}|^m \leq C|\tilde{Y}_i|^m \sum_{p_1}\sum_{p_2}\sum_{r_1}\sum_{r_2} \prod_{l=1}^{K_{1r_1}} 1\left(z_{l1l} \leq Z_{1li} \leq z_{l1l}+2b_N\right) \prod_{l=K_{1r_1}+1}^{L_{1p_1}} 1\left(z_{l1l}-2b_N \leq Z_{1li} \leq z_{u1l}\right) \\
\times \prod_{l=1}^{K_{2r_1}} 1\left(z_{l2l} \leq Z_{2li} \leq z_{l2l}+2b_N\right) \prod_{l=K_{2r_2}+1}^{L_{2p_2}} 1\left(z_{u2l}-2b_N \leq Z_{2li} \leq z_{u2l}\right)
\end{aligned}
$$

and because $\mathbb{E}[|\tilde{Y}|^3|Z=z]$ is bounded on $\mathbb{Z}$ and the density of Z is bounded, we have because $L_{1p_1}+L_{2p_2} \geq 1$ for $m=1,2,3$

$$\mathbb{E}\left[|\psi_{N,0,i,1}|^m\right] = O(b_N)$$

By the Liapounov central limit theorem U_{01}/b_N converges in distribution and hence

$$U_{01} = O_p(b_N) \tag{B.76}$$

Step 5: Ignoring Higher Order Terms The final step is to show that U_μ is asymptotically negligible if $|\mu| \geq 1$. Note that if $|\mu| \geq 1$, then the integrand in $\psi_{N,\mu,i}$ is 0 if z_1 and z_2 are both in the internal region. Hence we can take the integration region such that at least one component of either z_1 or z_2 is in the boundary region

$$
\begin{aligned}
&|\psi_{N,\mu,i}|^m \\
&= \left| \frac{1}{b_N^{L+|\mu|}} \int_{\mathbb{Z}_2\times\mathbb{Z}_1\setminus\mathbb{Z}^I_{1,b_N}\times\mathbb{Z}^I_{2,b_N}} \nu(z_1,z_2)'\tilde{Y}_i K^{(\mu)}\left(\frac{Z_i - r_{b_N}(z_1,z_2)}{b_N}\right)((z_1'\ z_2')' - r_{b_N}(z_1,z_2))^\mu f_{Z_1}(z_1)f_{Z_2}(z_2)\mathrm{d}z_1\mathrm{d}z_2 \right|^m \\
&\leq \Bigg(\frac{1}{b_N^{L+|\mu|}} \int_{\mathbb{Z}_2\times\mathbb{Z}_1\setminus\mathbb{Z}^I_{1,b_N}\times\mathbb{Z}^I_{2,b_N}} |\nu(z_1,z_2)||\tilde{Y}_i| \left| K^{(\mu)}\left(\frac{Z_i - r_{b_N}(z_1,z_2)}{b_N}\right)\right| \\
&|(z_1'\ z_2')' - r_{b_N}(z_1,z_2)|^{|\mu|} f_{Z_1}(z_1)f_{Z_2}(z_2)\mathrm{d}z_1\mathrm{d}z_2 \Bigg)^m \\
&\leq \left(\frac{1}{b_N^{L}} \int_{\mathbb{Z}_2\times\mathbb{Z}_1\setminus\mathbb{Z}^I_{1,b_N}\times\mathbb{Z}^I_{2,b_N}} |\nu(z_1,z_2)||\tilde{Y}_i| \left| K^{(\mu)}\left(\frac{Z_i - r_{b_N}(z_1,z_2)}{b_N}\right)\right| f_{Z_1}(z_1)f_{Z_2}(z_2)\mathrm{d}z_1\mathrm{d}z_2 \right)^m
\end{aligned}
$$

We obtained a bound on the right hand side in (B.75). Therefore by the Liapounov central limit theorem $\frac{U_\mu}{b_N}$ converges in distribution so that if $|\mu| \geq 1$

$$U_\mu = O_p(b_N) \tag{B.77}$$

By (B.33) (linearization), (B.34) (bias), (B.72) (projection), (B.76) (boundary remainder), and (B.77) (NIP remainder) we have that

$$
\begin{aligned}
\sqrt{N}(\hat{\theta} - \theta) = {} & \frac{1}{N\sqrt{N}} \sum_{j=1}^N \sum_{k=1}^N (n(h_0(Z_{1j}, Z_{2k})) - \theta) \qquad \text{(B.78)} \\
& + \frac{1}{\sqrt{N}} \sum_{i=1}^N \left\{ \frac{\partial n}{\partial h}(h_0(Z_i))'\tilde{Y}_i f_{Z_1}(Z_{1i}) f_{Z_2}(Z_{2i}) - \mathbb{E}_{\tilde{Y}Z}\left[\frac{\partial n}{\partial h}(h_0(S))'\tilde{Y} f_{Z_1}(Z_1)f_{Z_2}(Z_2)\right]\right\} \\
& + O_p\left(\sqrt{N}\left|\hat{h}_{\mathrm{nip},s} - h_0\right|^2\right) + O(\sqrt{N}b_N^p) + O_p(N^{-1}b_N^{-L}) + O_p(b_N)
\end{aligned}
$$

The first term on the right hand side is a V statistic that is asymptotically equivalent to

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \left\{ (\mathbb{E}[n(h_0(Z_{1i}, Z_2)) - \theta) + \mathbb{E}[n(h_0(Z_1, Z_{2i})) - \theta) \right\}.$$

□

Proof of Lemma A.24: Using Lemma A.14, the assumptions imply that

$$\sup_{w\in\mathbb{W},x\in\mathbb{X}} |\hat{g}(w,x) - g(w,x)| = O_p\left(\left(\frac{\ln(N)}{N\cdot b_N^2}\right)^{1/2} + b_N^s\right) = o_p\left(N^{-\eta}\right), \tag{B.79}$$

For $1/4 < \delta < 1/4s$ we can find an $\eta > 1/4$ such that this holds. Using the definitions preceding the statement of the Lemma, we have, by adding and subtracting terms,

$$
\begin{aligned}
\widehat{\beta}^{\mathrm{cm}}(\rho,0) - \beta^{\mathrm{cm}}(\rho,0) = {} & (\widehat{\beta}^{\mathrm{cm}}(\rho,0) - \hat{\beta}_g^{\mathrm{cm}}) \qquad \text{(B.80)} \\
& - (\hat{\beta}_W^{\mathrm{cm}} - \overline{g}^{\mathrm{cm}}) \qquad \text{(B.81)} \\
& - (\hat{\beta}_X^{\mathrm{cm}} - \overline{g}^{\mathrm{cm}}) \qquad \text{(B.82)} \\
& + \left(\hat{\beta}_g^{\mathrm{cm}} - \overline{g}^{\mathrm{cm}}\right) + \left(\hat{\beta}_W^{\mathrm{cm}} - \overline{g}^{\mathrm{cm}}\right) + \left(\hat{\beta}_X^{\mathrm{cm}} - \overline{g}^{\mathrm{cm}}\right) + (\overline{g}^{\mathrm{cm}} - \beta^{\mathrm{cm}}(\rho,0)).
\end{aligned}
$$

The result then follows if we can show that the sum of (B.80), (B.81) and (B.82) is $o_p\left(N^{-1/2}\right)$. Define

$$\hat{\omega}^{\mathrm{cm}}(w,x) = \frac{\phi_c\left(\Phi_c^{-1}(\hat{F}_W(w)), \Phi_c^{-1}(\hat{F}_X(x)); \rho\right)}{\phi_c\left(\Phi_c^{-1}(\hat{F}_W(w))\right)\phi_c\left(\Phi_c^{-1}(\hat{F}_X(x))\right)},$$

$$\hat{\omega}_W^{\text{cm}}(w,x) = \frac{\phi_c\left(\Phi_c^{-1}(\hat{F}_W(w)), \Phi_c^{-1}(F_X(x)); \rho\right)}{\phi_c\left(\Phi_c^{-1}(\hat{F}_W(w))\right)\phi_c\left(\Phi_c^{-1}(F_X(x))\right)},$$

and

$$\hat{\omega}_X^{\text{cm}}(w,x) = \frac{\phi_c\left(\Phi_c^{-1}(F_W(w)), \Phi_c^{-1}(\hat{F}_X(x)); \rho\right)}{\phi_c\left(\Phi_c^{-1}(F_W(w))\right)\phi_c\left(\Phi_c^{-1}(\hat{F}_X(x))\right)},$$

Then, using the definition of $\omega^{\text{cm}}(w,x)$ given in (4.32), we can write the sum of these three components as

$$\begin{aligned}
&(\widehat{\beta}^{\text{cm}}(\rho,0) - \hat{\beta}_g^{\text{cm}}) - (\hat{\beta}_W^{\text{cm}} - \overline{g}^{\text{cm}}) - (\hat{\beta}_X^{\text{cm}} - \overline{g}^{\text{cm}}) \\
&\quad = \frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N \hat{g}(W_i,X_j)\left[\widehat{\omega}^{\text{cm}}(W_i,X_j) - \omega^{\text{cm}}(W_i,X_j)\right] \\
&\qquad - \frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N g(W_i,X_j)\left[\widehat{\omega}_W(W_i,X_j) - \omega^{\text{cm}}(W_i,X_j)\right] \\
&\qquad - \frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N g(W_i,X_j)\left[\widehat{\omega}_X(W_i,X_j) - \omega^{\text{cm}}(W_i,X_j)\right] \\
&\quad = \frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N \hat{g}(W_i,X_j)\left[\widehat{\omega}^{\text{cm}}(W_i,X_j) - \omega^{\text{cm}}(W_i,X_j)\right] \\
&\quad - \frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N g(W_i,X_j)\left[\widehat{\omega}^{\text{cm}}(W_i,X_j) - \omega^{\text{cm}}(W_i,X_j)\right] + \frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N g(W_i,X_j)\left[\widehat{\omega}^{\text{cm}}(W_i,X_j) - \widehat{\omega}_W(W_i,X_j)\right] \\
&\qquad - \frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N g(W_i,X_j)\left[\widehat{\omega}_X(W_i,X_j) - \omega^{\text{cm}}(W_i,X_j)\right] \\
&\quad = \frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N \left[\hat{g}(W_i,X_j) - g(W_i,X_j)\right]\left[\widehat{\omega}^{\text{cm}}(W_i,X_j) - \omega^{\text{cm}}(W_i,X_j)\right] && \text{(B.83)} \\
&\qquad + \frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N g(W_i,X_j)\left[\widehat{\omega}^{\text{cm}}(W_i,X_j) - \widehat{\omega}_W(W_i,X_j) - \widehat{\omega}_X(W_i,X_j) + \omega^{\text{cm}}(W_i,X_j)\right] && \text{(B.84)}
\end{aligned}$$

It remains to be shown that both (B.83) and (B.84) are $o_p\left(N^{-1/2}\right)$.
Now define

$$k(z_1,z_2) = \frac{\phi_c\left(\Phi_c^{-1}(z_1), \Phi_c^{-1}(z_2); \rho\right)}{\phi_c\left(\Phi_c^{-1}(z_1)\right)\cdot\phi_c\left(\Phi_c^{-1}(z_2)\right)} \qquad \text{so that} \quad \widehat{\omega}^{\text{cm}}(w,x) = k(\hat{F}_W(w), \hat{F}_X(x)). \tag{B.85}$$

By a second order Taylor expansion we have

$$\begin{aligned}
\widehat{\omega}^{\text{cm}}(w,x) - \omega^{\text{cm}}(w,x) &= \frac{\partial k}{\partial z_1}(F_W(w),F_X(x))(\hat{F}_W(w) - F_W(w)) + \frac{\partial k}{\partial z_2}(F_W(w),F_X(x))(\hat{F}_X(x) - F_X(x)) \\
&\quad + \frac{1}{2}\frac{\partial^2 k}{\partial z_1^2}(\overline{F}_W(w),\overline{F}_X(x))(\hat{F}_W(w) - F_W(w))^2 + \frac{1}{2}\frac{\partial^2 k}{\partial z_2^2}(\overline{F}_W(w),\overline{F}_X(x))(\hat{F}_X(x) - F_X(x))^2 \\
&+ \frac{1}{2}\frac{\partial^2 k}{\partial z_1 \partial z_2}(\overline{F}_W(w),\overline{F}_X(x))(\hat{F}_W(w) - F_W(w))(\hat{F}_X(x) - F_X(x))
\end{aligned}$$

with $\overline{F}_W(w)$ and $\overline{F}_X(x)$ intermediate values. By Lemma A.3 it follows that for any $0 < \delta < 1/2$, $\sup_x |\hat{F}_X(x) - F_X(x)| = o_p(N^{-\delta})$, and $\sup_w |\hat{F}_W(w) - F_W(w)| = o_p(N^{-\delta})$. In combination with the fact that $|\partial^2 k/\partial z_1^2|$, $|\partial^2 k/\partial z_2^2|$, and $|\partial^2 k/\partial z_1 \partial z_2|$ are bounded, this implies that

$$\begin{aligned}
&\widehat{\omega}^{\text{cm}}(w,x) - \omega^{\text{cm}}(w,x) && \text{(B.86)} \\
&\quad = \frac{\partial k}{\partial z_1}(F_W(w),F_X(x))(\hat{F}_W(w) - F_W(w)) + \frac{\partial k}{\partial z_2}(F_W(w),F_X(x))(\hat{F}_X(x) - F_X(x)) + o_p\left(N^{-1/2}\right).
\end{aligned}$$

The same argument implies that

$$\widehat{\omega}_W(w,x) - \omega^{\rm cm}(w,x) = \frac{\partial k}{\partial z_1}(F_W(w), F_X(x))(\hat{F}_W(w) - F_W(w)) + o_p\left(N^{-1/2}\right),$$

and

$$\widehat{\omega}_X(w,x) - \omega^{\rm cm}(w,x) = \frac{\partial k}{\partial z_2}(F_W(w), F_X(x))(\hat{F}_X(x) - F_X(x)) + o_p\left(N^{-1/2}\right).$$

Substituting in these results, it follows that (B.84) is $o_p\left(N^{-1/2}\right)$.
Equation (B.86) also implies, by Lemma A.3, that

$$\widehat{\omega}^{\rm cm}(w,x) - \omega^{\rm cm}(w,x) = o_p\left(N^{-1/4}\right).$$

In combination with (B.79), this implies that (B.83) is also $o_p\left(N^{-1/2}\right)$. $\square$

Proof of Lemma A.25: The proof of this Lemma make use of an application of Theorem A.3. Using the notation of that Theorem we have $Z_1 = W, Z_2 = X$, $\tilde{Y} = (Y, 1)'$,

$$h\left(w,x\right) = \begin{pmatrix} g(w,x) \cdot f_{WX}(w,x) \\ f_{WX}(w,x) \end{pmatrix}$$

and

$$n(h(w,x)) = \frac{h_1\left(w,x\right)}{h_2\left(w,x\right)}\omega^{\rm cm}\left(w,x\right) = g(w,x)\cdot\omega^{\rm cm}(w,x).$$

In terms of this notation we can write this in the form of Theorem A.3:

$$\hat{\beta}_g^{\rm cm} - \overline{g}^{\rm cm} = \frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N n\left(\hat{h}(W_i, X_j)\right) - \frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N n\left(h(W_i, X_j)\right).$$

We also have

$$\frac{\partial n}{\partial h}(h(w,x)) = \begin{pmatrix} \frac{1}{h_2(w,x)} \\ -\frac{h_1(w,x)}{h_2(w,x)^2} \end{pmatrix}\omega^{\rm cm}\left(w,x\right) = \begin{pmatrix} \frac{1}{f_{WX}(w,x)} \\ -\frac{g(w,x)}{f_{WX}(w,x)} \end{pmatrix}\omega^{\rm cm}\left(w,x\right),$$

and hence

$$\frac{\partial n}{\partial h}(h(w,x))'\tilde{y}f_W(w)f_X(x)) = \frac{f_W(w)f_X(x)}{f_{WX}(w,x)}\cdot(y - g(w,x))\cdot\omega^{\rm cm}\left(w,x\right),$$

which is mean zero. Therefore, by the result of Theorem A.3, we have

$$\hat{\beta}_g^{\rm cm} - \beta^{\rm cm}(\rho,0) = \frac{1}{N}\sum_{i=1}^N\frac{\partial n}{\partial h}(h(W_i,X_i))'\tilde{Y}_i f_W(W_i)f_X(X_i)) - \mathbb{E}\left[\frac{\partial n}{\partial h}(h(W,X))'\tilde{y}Y f_W(W)f_X(X))\right] + o_p\left(N^{-1/2}\right)$$

$$= \frac{1}{N}\sum_{i=1}^N\frac{f_W(W_i)f_X(X_i)}{f_{WX}(W_i,X_i)}\cdot(Y_i - g(W_i,X_i))\cdot\omega^{\rm cm}\left(W_i,X_i\right) - \mathbb{E}\left[\frac{f_W(W)f_X(X)}{f_{WX}(W,X)}\cdot(Y - g(W,X))\cdot\omega^{\rm cm}\left(W,X\right)\right] + o_p\left(N^{-1/2}\right)$$

$$= \frac{1}{N}\sum_{i=1}^N\frac{f_W(W_i)f_X(X_i)}{f_{WX}(W_i,X_i)}\cdot(Y_i - g(W_i,X_i))\cdot\omega^{\rm cm}\left(W_i,X_i\right) + o_p\left(N^{-1/2}\right)$$

$$= \frac{1}{N}\sum_{i=1}^N\psi_g^{\rm cm}(Y_i,W_i,X_i) + o_p(N^{-1/2}) + o_p\left(N^{-1/2}\right).$$

$\square$

Proof of Lemma A.26: Using the definition of $k(z_1, z_2)$ in (B.85) and the Taylor expansion in the proof of Lemma A.24 we have

$$\hat{\beta}_W^{\rm cm} - \overline{g}^{\rm cm} = \frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N g(W_i,X_j)\frac{\partial k}{\partial z_1}(F_W(W_i), F_X(X_j))(\hat{F}_W(W_i) - F_W(W_i))$$

$$+\frac{1}{2}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} g(W_i,X_j)\frac{\partial^2 k}{\partial z_1^2}(\overline{F}_W(W_i),\overline{F}_X(X_j))(\hat{F}_W(W_i)-F_W(W_i))^2.$$

By Lemma A.3 $\sup_w |\hat{F}_W(w)-F_W(w)| = o_p(N^{-\delta})$ for all $\delta < 1/2$, and using the fact that the second derivatives of $k(z_1,z_2)$ are bounded, this implies

$$\hat{\beta}_W^{\text{cm}} - \overline{g}^{\text{cm}} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} g(W_i,X_j)\frac{\partial k}{\partial z_1}(F_W(W_i),F_X(X_i))(\hat{F}_W(W_i)-F_W(W_i)) + o_p\left(N^{-1/2}\right).$$

Inspection of the definition of $e_W(w,x)$ shows that $e_W(w,x) = \frac{\partial k}{\partial s_1}(F_W(w),F_X(x))$ and therefore

$$\hat{\beta}_W^{\text{cm}} - \overline{g}^{\text{cm}} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} g(W_i,X_j)e_W(W_i,X_j)(\hat{F}_W(W_i)-F_W(W_i)) + o_p\left(N^{-1/2}\right)$$

$$= \frac{1}{N^3}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} g(W_i,X_j)e_W(W_i,X_j)(1(W_k \le W_i) - F_W(W_i)) + o_p\left(N^{-1/2}\right).$$

This is, up to the $o_p(N^{-1/2})$ term, a third order V-statistic,

$$\hat{\beta}_W^{\text{cm}} - \overline{g}^{\text{cm}} = V + o_p\left(N^{-1/2}\right)$$

where

$$V = \frac{1}{N^3}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\psi(W_i,X_i,W_j,X_j,W_k,X_k),$$

with

$$\psi(w_1,x_1,w_2,x_2,w_3,x_3) = g(w_1,x_2)e_W(w_1,x_2)(1(w_3 \le w_1) - F_W(w_1).$$

Define

$$\psi_1(w,x) = \mathbb{E}\left[\psi(w,x,W_2,X_2,W_3,X_3)\right], \qquad \psi_2(w,x) = \mathbb{E}\left[\psi(W_1,X_1,w,x,W_3,X_3)\right],$$

$$\psi_3(w,x) = \mathbb{E}\left[\psi(W_1,X_1,W_2,X_2,w,x)\right], \qquad \text{and} \quad \theta = \mathbb{E}[\psi(W_1,X_1,W_2,X_2,W_3,X_3)].$$

Using V-statistic theory, this V-statistic can be approximated as

$$V = \frac{1}{N}\sum_{i=1}^{N}\{(\psi_1(W_i,X_i)-\theta) + (\psi_2(W_i,X_i)-\theta) + (\psi_3(W_i,X_i)-\theta)\} + o_p\left(N^{-1/2}\right).$$

Note that $\mathbb{E}[\psi(w_1,x_1,w_2,x_2,W,X)] = 0$. Hence $\theta = 0$, $\psi_1(w,x) = 0$, and $\psi_2(w,x) = 0$. Thus,

$$V = \frac{1}{N}\sum_{i=1}^{N}\psi_3(W_i,X_i) + o_p\left(N^{-1/2}\right).$$

$$= \frac{1}{N}\sum_{k=1}^{N}\int\int g(s,t)e_W(s,t)(1(W_i \le s) - F_W(s))f_W(s)f_X(t)\,\mathrm{d}s\mathrm{d}t + o_p\left(N^{-1/2}\right),$$

$$= \frac{1}{N}\sum_{i=1}^{N}\psi_W^{\text{cm}}(Y_i,W_i,X_i) + o_p\left(N^{-1/2}\right),$$

as required. □

Proof of Lemma A.27: The proof is entirely analogous to that of Lemma A.26 and therefore omitted. □

Proof of Lemma A.28: Define

$$\psi(w,x) = g(w,x)\cdot\omega^{\text{cm}}(w,x),$$

$$\psi_1(w) = \mathbb{E}\left[\psi(w,X)\right] = \mathbb{E}\left[g(w,X)\cdot\omega^{\text{cm}}(w,X)\right],$$

and

$$\psi_2(x) = \mathbb{E}\left[\psi(W,x)\right] = \mathbb{E}\left[g(W,x)\cdot\omega^{\text{cm}}(W,x)\right].$$

Then, by the V-Statistic Projection Theorem, given as Theorem A.4 in Appendix A, it follows that

$$\overline{g}^{\text{cm}} - \beta^{\text{cm}}(\rho,0) = \frac{1}{N}\sum_{i=1}^{N}\left\{(\psi_1(W_i) - \beta^{\text{cm}}(\rho,0)) + (\psi_2(X_i) - \beta^{\text{cm}}(\rho,0))\right\} + o_p\left(N^{-1/2}\right)$$

$$= \frac{1}{N}\sum_{i=1}^{N}\psi_0^{\text{cm}}(Y_i,W_i,X_i) + o_p\left(N^{-1/2}\right).$$

□

Proof of Theorem A.4: Define $\phi(z_1,z_2) = (\psi(z_1,z_2) + \psi(z_2,z_1))/2$. Then $V = \sum_{i=1}^{N}\sum_{j=1}^{N}\phi(Z_i,Z_j)/N^2$ is a V-statistic with a symmetric kernel. In the notation of Lehman (1998),

$$\sigma_1^2 = \text{Cov}\left(\phi(Z_i,Z_j),\phi(Z_i,Z_k)\right),$$

for i,j,k distinct, which simplifies to $\sigma_1^2 = \sigma^2/4$. Therefore, by Theorems 6.1.2 (with $a = 2$) and 6.2.1 in Lehman (1998), the result follows. □.

Appendix C: Proofs of Theorems in Text

Proof of Theorem 3.1 Define

$$V_{\lambda,i} = \lambda \cdot X_i \cdot d(W_i) + W_i,$$

$$h(\lambda, a) = \mathrm{pr}(V_\lambda \le a) = F_{V_\lambda}(a), \qquad \text{and} \quad k(w, x, \lambda) = h(\lambda, \lambda \cdot x \cdot d(w, x) + w).$$

First we focus on

$$\beta^{\mathrm{lr,v}}(\lambda) = \mathbb{E}\left[g(F_W^{-1}(F_{V_\lambda}(V_{\lambda,i})), X)\right] = \mathbb{E}\left[g(F_W^{-1}(k(W_i, X_i, \lambda)), X_i)\right].$$

We then prove four results. First, we show that for small λ, $\beta^{\mathrm{lr,v}}(\lambda)$ and $\beta^{\mathrm{lr}}(\lambda)$ are close, or

$$\beta^{\mathrm{lr,v}}(\lambda) = \beta^{\mathrm{lr}}(\lambda) + o(\lambda). \tag{C.1}$$

Second, we show that

$$\beta^{\mathrm{lr,v}}(\lambda) = \mathbb{E}[g(W, X)] + \mathbb{E}\left[\frac{\partial g}{\partial w}(W_i, X_i) \frac{1}{f_W(W_i)} \left(k(W_i, X_i, \lambda) - k(W_i, X_i, 0)\right)\right] + o(\lambda). \tag{C.2}$$

Next we show that $\beta^{\mathrm{lr,v}}(\lambda)$ has the two representations in Theorem 3.1. In particular, the third part of the proof shows that $\beta^{\mathrm{lc,v}} = \frac{\partial \beta^{\mathrm{lr,v}}}{\partial \lambda}(0)$ satisfies

$$\beta^{\mathrm{lc,v}} = \mathbb{E}\left[\frac{\partial g}{\partial w}(W_i, X_i) \cdot \left(X_i \cdot d(W_i, X_i) - \mathbb{E}\left[X_i \cdot d(W_i, X_i) \middle| W_i\right]\right)\right]. \tag{C.3}$$

Fourth, we show that $\beta^{\mathrm{lc,v}}$ satisfies

$$\beta^{\mathrm{lc,v}} = \mathbb{E}\left[\delta(W_i, X_i) \cdot \frac{\partial^2 g}{\partial w \partial x}(W_i, X_i)\right]. \tag{C.4}$$

We start with the proof of (C.1). Define

$$u(w, x, \lambda) = \lambda \cdot x \cdot d(w, x)^{1-|\lambda|} + \sqrt{1 - \lambda^2} \cdot w, \qquad \text{and} \quad u(w, x, \lambda) = \lambda \cdot x \cdot d(w, x) + w.$$

Then

$$\sup_{w \in \mathbb{W}, x \in \mathbb{X}} |u(w, x, \lambda) - v(w, x, \lambda)| = O\left(\lambda^2\right).$$

Define also

$$h_U(\lambda, a) = \mathrm{pr}(U_\lambda \le a), \qquad \text{and} \quad k_U(w, x, \lambda) = h_U(\lambda, u(w, x, \lambda)).$$

Then

$$\sup_a |h_U(\lambda, a) - h(\lambda, a)| = O\left(\lambda^2\right),$$

and

$$\sup_{w \in \mathbb{W}, x \in \mathbb{X}} |k_U(w, x, \lambda) - k(w, x, \lambda)| = O\left(\lambda^2\right).$$

Combined with the smoothness assumptions, this implies that

$$\beta^{\mathrm{lr,v}}(\lambda) - \beta^{\mathrm{lr}}(\lambda) = \mathbb{E}\left[g(F_W^{-1}(k(W_i, X_i, \lambda)), X_i)\right] - \mathbb{E}\left[g(F_W^{-1}(k_U(W_i, X_i, \lambda)), X_i)\right] = O(\lambda^2).$$

This finishes the proof of (C.1).
Next, we prove (C.2). Let c_1 and c_2 satisfy

$$\sup_{x, w, \gamma, \lambda} |k(w, x, \lambda + \gamma) - k(w, x, \lambda)| \le c_1 \cdot \gamma,$$

and

$$\sup_{w\in\mathbb{W},x\in\mathbb{X}}\left|\frac{\partial^2}{\partial w^2}g(w,x)\right|\le c_2,$$

respectively. Then, applying Lemma A.1 with $f(a)=g(F_W^{-1}(a),x)$ and $h(\lambda)=k(w,x,\lambda)$, we obtain

$$\left|g(F_W^{-1}(k(w,x,\lambda),x)-\left(g(F_W^{-1}(k(w,x,0),x)+\frac{\frac{\partial}{\partial w}g(F_W^{-1}(k(w,x,0),x)}{f_W(F_W^{-1}(k(w,x,0)))}\left(k(w,x,\lambda)-k(w,x,0)\right)\right)\right|$$
$$\le c_2c_1^2\lambda^2=o(\lambda).$$

Since the bound does not depend on x and w, we can average over W and X and it follows that

$$\left|\mathbb{E}\left[g(F_W^{-1}(k(W,X,\lambda),X)\right]-\mathbb{E}\left[g(W,X)\right]-\mathbb{E}\left[\frac{\frac{\partial}{\partial w}g(W,X)}{f_W(W)}\left(k(W,X,\lambda)-W\right)\right]\right|,$$

where we also use the fact that $k(w,x,0)=F_W(w)$. This finishes the proof of (C.2).
Now we prove (C.3). By definition,

$$\begin{aligned}
h(\lambda,a)&=\Pr\left(V_{\lambda,i}<a\right)=\Pr\left(V_{\lambda,i}<a,W_i<w_m\right)+\Pr\left(V_{\lambda,i}<a,W_i\ge w_m\right)\\
&=\Pr\left(\lambda\cdot X_i\cdot d(W_i,X_i)+W_i\le a,W_i<w_m\right)\\
&\quad+\Pr\left(\lambda\cdot X_i\cdot d(W_i,X_i)+W_i\le a,W_i\ge w_m\right).\\
&=\Pr\left(\lambda\cdot X_i\cdot (W_i-w_l)+W_i\le a,W_i<w_m\right)\\
&\quad+\Pr\left(\lambda\cdot X_i\cdot (w_u-W_i)+W_i\le a,W_i\ge w_m\right)\\
&=\Pr\left(W_i\le\min\left(w_m,\frac{a+\lambda\cdot X_i\cdot w_l}{1+\lambda\cdot X_i}\right)\right)\\
&\quad+\Pr\left(w_m\le W_i\le\frac{a-\lambda\cdot X_i\cdot w_u}{1-\lambda\cdot X_i}\right).
\end{aligned}$$

For λ sufficiently close to zero, we can write this as

$$\begin{aligned}
h(\lambda,a)&=1_{a>w_m}\cdot\Pr(W_i\le w_m)+1_{a\le w_m}\cdot\Pr\left(W_i\le\frac{a+\lambda\cdot X_i\cdot w_l}{1+\lambda\cdot X_i}\right)\\
&\quad+1_{a>w_m}\cdot\Pr\left(w_m<W_i\le\frac{a-\lambda\cdot X_i\cdot w_u}{1-\lambda\cdot X_i}\right)\\
&=1_{a\le w_m}\cdot\Pr\left(W_i\le\frac{a+\lambda\cdot X_i\cdot w_l}{1+\lambda\cdot X_i}\right)\\
&\quad+1_{a>w_m}\cdot\Pr\left(W_i\le\frac{a-\lambda\cdot X_i\cdot w_u}{1-\lambda\cdot X_i}\right)\\
&=1_{a\le w_m}\cdot\mathbb{E}\left[\Pr\left(W_i\le\frac{a+\lambda\cdot X_i\cdot w_l}{1+\lambda\cdot X_i}\middle|X_i\right)\right]\\
&\quad+1_{a>w_m}\cdot\mathbb{E}\left[\Pr\left(W_i\le\frac{a-\lambda\cdot X_i\cdot w_u}{1-\lambda\cdot X_i}\middle|X_i\right)\right]\\
&=1_{a\le w_m}\cdot\int F_{W|X}\left(\frac{a+\lambda\cdot z\cdot w_l}{1+\lambda\cdot z}\middle|z\right)f_X(z)\mathrm{d}z\\
&\quad+1_{a>w_m}\cdot\int F_{W|X}\left(\frac{a-\lambda\cdot z\cdot w_u}{1-\lambda\cdot z}\middle|z\right)f_X(z)\mathrm{d}z
\end{aligned}$$

Substituting $a=\lambda\cdot x\cdot d(w,x)+w$, we get

$$\begin{aligned}
k(w,x,\lambda)&=1_{\lambda\cdot x\cdot d(w,x)+w\le w_m}\cdot\int F_{W|X}\left(\frac{\lambda\cdot x\cdot d(w,x)+w+\lambda\cdot z\cdot w_l}{1+\lambda\cdot z}\middle|z\right)f_X(z)\mathrm{d}z\\
&\quad+1_{\lambda\cdot x\cdot d(w,x)+w>w_m}\cdot\int F_{W|X}\left(\frac{\lambda\cdot x\cdot d(w,x)+w-\lambda\cdot z\cdot w_u}{1-\lambda\cdot z}\middle|z\right)f_X(z)\mathrm{d}z
\end{aligned}$$

$$= 1_{\lambda\cdot x\cdot(w-w_l)+w\le w_m} 1_{w\le w_m} \cdot \int F_{W|X}\left(\frac{\lambda\cdot x\cdot(w-w_m)+w+\lambda\cdot z\cdot w_l}{1+\lambda\cdot z} \middle| z \right) f_X(z)\mathrm{d}z$$

$$+1_{\lambda\cdot x\cdot(w_u-w)+w\le w_m} 1_{w> w_m} \cdot \int F_{W|X}\left(\frac{\lambda\cdot x\cdot(w_u-w)+w+\lambda\cdot z\cdot w_l}{1+\lambda\cdot z} \middle| z \right) f_X(z)\mathrm{d}z$$

$$+1_{\lambda\cdot x\cdot(w-w_l)+w> w_m} \cdot 1_{w\le w_m} \int F_{W|X}\left(\frac{\lambda\cdot x\cdot(w-w_m)+w-\lambda\cdot z\cdot w_u}{1-\lambda\cdot z} \middle| z \right) f_X(z)\mathrm{d}z$$

$$+1_{\lambda\cdot x\cdot(w_u-w)+w> w_m} \cdot 1_{w> w_m} \int F_{W|X}\left(\frac{\lambda\cdot x\cdot(w_u-w)+w-\lambda\cdot z\cdot w_u}{1-\lambda\cdot z} \middle| z \right) f_X(z)\mathrm{d}z$$

$$= 1_{w\le w_m(1+\lambda x w_l/w_m)/(1+\lambda x)} \cdot \int F_{W|X}\left(\frac{\lambda\cdot x\cdot(w-w_m)+w+\lambda\cdot z\cdot w_l}{1+\lambda\cdot z} \middle| z \right) f_X(z)\mathrm{d}z$$

$$+0\cdot \int F_{W|X}\left(\frac{\lambda\cdot x\cdot(w_u-w)+w+\lambda\cdot z\cdot w_l}{1+\lambda\cdot z} \middle| z \right) f_X(z)\mathrm{d}z$$

$$+1_{w_m(1+\lambda x w_l/w_m)/(1+\lambda x)\le w\le w_m} \cdot \int F_{W|X}\left(\frac{\lambda\cdot x\cdot(w-w_m)+w-\lambda\cdot z\cdot w_u}{1-\lambda\cdot z} \middle| z \right) f_X(z)\mathrm{d}z$$

$$+1_{w\ge w_m} \cdot \int F_{W|X}\left(\frac{\lambda\cdot x\cdot(w_u-w)+w-\lambda\cdot z\cdot w_u}{1-\lambda\cdot z} \middle| z \right) f_X(z)\mathrm{d}z.$$

The last equality uses the following four facts: *(i)*, $\lambda\cdot x\cdot(w-w_l)+w\le w_m$ implies $w\le w_m(1+\lambda x w_l/w_m)/(1+\lambda x)\le w_m$, *(ii)* $\lambda\cdot x\cdot(w_u-w)+w\le w_m$ implies $w\le w_m(1-\lambda x w_u/w_m)/(1-\lambda x)<w_m$, *(iii)*, $\lambda\cdot x\cdot(w-w_l)+w>w_m$ implies $w\ge w_m(1+\lambda x w_l/w_m)/(1+\lambda x)$, and *(iv)* $\lambda\cdot x\cdot(w_u-w)+w>w_m$ implies $w\ge w_m(1-\lambda x w_u/w_m)/(1-\lambda x)$. Now we will look at

$$\mathbb{E}\left[\frac{\partial g}{\partial w}(W_i,X_i)\frac{1}{f_W(W_i)}k(W_i,X_i,\lambda)\right]$$

$$= \int_{x_l}^{x_u}\int_{w_l}^{w_u}\frac{\partial g}{\partial w}(w,x)\frac{1}{f_W(w)}k(w,x,\lambda)f_{W,X}(w,x)\mathrm{d}w\mathrm{d}x.$$

Substituting the three terms of $k(w,x,\lambda)$ in here we get

$$\mathbb{E}\left[\frac{\partial g}{\partial w}(W_i,X_i)\frac{1}{f_W(W_i)}k(W_i,X_i,\lambda)\right]$$

$$= \int_{x_l}^{x_u}\int_{w_l}^{w_m\frac{1+\lambda x w_l/w_m}{1+\lambda x}} \frac{\frac{\partial g}{\partial w}(w,x)}{f_W(w)}\int F_{W|X}\left(\frac{\lambda\cdot x\cdot(w-w_l)+w+\lambda\cdot z\cdot w_l}{1+\lambda\cdot z} \middle| z \right) f_X(z)\mathrm{d}z f_{W,X}(w,x)\mathrm{d}w\mathrm{d}x \tag{C.5}$$

$$+ \int_{x_l}^{x_u}\int_{w_m\frac{(1+\lambda x w_l/w_m}{1+\lambda x}}^{w_m} \frac{\frac{\partial g}{\partial w}(w,x)}{f_W(w)}\int F_{W|X}\left(\frac{\lambda\cdot x\cdot(w-w_l)+w-\lambda\cdot z\cdot w_u}{1-\lambda\cdot z} \middle| z \right) f_X(z)\mathrm{d}z f_{W,X}(w,x)\mathrm{d}w\mathrm{d}x \tag{C.6}$$

$$+ \int_{x_l}^{x_u}\int_{w_m}^{w_u} \frac{\frac{\partial g}{\partial w}(w,x)}{f_W(w)}\int F_{W|X}\left(\frac{\lambda\cdot x\cdot(w_u-w)+w-\lambda\cdot z\cdot w_u}{1-\lambda\cdot z} \middle| z \right) f_X(z)\mathrm{d}z f_{W,X}(w,x)\mathrm{d}w\mathrm{d}x \tag{C.7}$$

Next, we take the derivative with respect to λ for each of these three terms, and evaluate that derivative at $\lambda=0$. For the first term, (C.5) this derivative consists of two terms, one corresponding to the derivative with respect to the λ in the bounds of the integral, and one corresponding to the derivative with respect to λ in the integrand. For the second term we only have the term corresponding to the derivative with respect to the λ in the bounds of the integral since the other term vanishes when we evaluate it at $\lambda=0$. The third term, (C.6) only has λ in the integrand. So,

$$\frac{\partial}{\partial\lambda}\,\mathbb{E}\left[\frac{\partial g}{\partial w}(W_i,X_i)\frac{1}{f_W(W_i)}k(W_i,X_i,\lambda)\right]\bigg|_{\lambda=0}$$

$$= (w_l-w_m)\cdot\mathbb{E}\left[\frac{\partial}{\partial w}g(w_m,X_i)\middle| W_i=w_m\right]$$

$$+\int_{x_l}^{x_u}\int_{w_l}^{w_m}\frac{\partial}{\partial w}g(w,x)\frac{1}{f_W(w)}\int_{x_l}^{x_u} f_{W|X}(w|z)\left(x\cdot(w-w_l)+z\cdot w_l-z\cdot w\right)f_X(z)\mathrm{d}z f_{W,X}(w,x)\mathrm{d}w\mathrm{d}x$$

$$-(w_l-w_m)\cdot\mathbb{E}\left[\left.\frac{\partial}{\partial w}g(w_m,X_i)\right| W_i=w_m\right]$$

$$+\int_{x_l}^{x_u}\int_{w_l}^{w_m}\frac{\partial}{\partial w}g(w,x)\frac{1}{f_W(w)}\int_{x_l}^{x_u} f_{W|X}(w|z)\left(x\cdot(w_u-w)+z\cdot w-z\cdot w_u\right)f_X(z)\mathrm{d}z f_{W,X}(w,x)\mathrm{d}w\mathrm{d}x$$

$$=\int_{x_l}^{x_u}\int_{w_l}^{w_m}\frac{\partial}{\partial w}g(w,x)\int_{x_l}^{x_u} f_{X|W}(z|w)\left(x\cdot d(w,x)-z\cdot d(w,z)\right)\mathrm{d}z f_{W,X}(w,x)\mathrm{d}w\mathrm{d}x$$

$$+\int_{x_l}^{x_u}\int_{w_l}^{w_m}\frac{\partial}{\partial w}g(w,x)\int_{x_l}^{x_u} f_{X|W}(z|w)\left(x\cdot d(w,x))-z\cdot d(w,z)\right)\mathrm{d}z f_{W,X}(w,x)\mathrm{d}w\mathrm{d}x$$

$$=\int_{x_l}^{x_u}\int_{w_l}^{w_u}\frac{\partial}{\partial w}g(w,x)\int_{x_l}^{x_u} f_{X|W}(z|w)\left(x\cdot d(w,x)-z\cdot d(w,z)\right)\mathrm{d}z f_{W,X}(w,x)\mathrm{d}w\mathrm{d}x$$

$$=\int_{x_l}^{x_u}\int_{w_l}^{w_u}\frac{\partial}{\partial w}g(w,x)\left(\left(X_i\cdot d(w,X_i)-\mathbb{E}\left[X_i\cdot d(w,X_i))\right| W_i=w\right]\right)f_{W,X}(w,x)\mathrm{d}w\mathrm{d}x$$

$$=\mathbb{E}\left[\frac{\partial g}{\partial w}\left(W_i,X_i\right)\cdot\left(X_i\cdot d(W_i,X_i)-\mathbb{E}\left[X_i\cdot d(W_i,X_i)\right| W_i\right]\right)\right]=\beta^{\mathrm{lc,v}}.$$

This finishes the proof of (C.3).

Finally, we show (C.4), by showing the equality of

$$\beta^{\mathrm{lc,v}}=\mathbb{E}\left[\frac{\partial g}{\partial w}\left(W_i,X_i\right)\cdot\left(X_i\cdot d(W_i)-\mathbb{E}\left[X_i\cdot d(W_i)\right| W_i\right]\right)\right], \tag{C.8}$$

and

$$\mathbb{E}\left[\delta\left(W_i,X_i\right)\cdot\frac{\partial^2 g}{\partial w\partial x}\left(W_i,X_i\right)\right]. \tag{C.9}$$

Define

$$b(w)=\mathbb{E}\left[\left.\frac{\partial g}{\partial w}\left(w,X_i\right)\cdot\left(X_i\cdot d(w)-\mathbb{E}\left[X_i\cdot d(w)\right| W_i=w\right]\right)\right| W_i=w\right]$$

$$=\mathbb{E}\left[\left.\frac{\partial g}{\partial w}\left(w,X_i\right)\cdot d(w)\cdot\left(X_i-\mathbb{E}\left[X_i\right| W_i=w\right]\right)\right| W_i=w\right],$$

so that $\beta^{\mathrm{lc,v}}=\mathbb{E}[b(W)]$. Apply Lemma A.2, with $h(x)=\frac{\partial g}{\partial w}\left(w,x\right)\cdot d(w)$, to get

$$b(w)=\mathbb{E}\left[\frac{\partial^2}{\partial w\partial x}g(w,X)\cdot\delta(w,X)\right]$$

with

$$\delta(w,x)=d(w)\cdot\frac{F_{X|W}(x|w)\cdot(1-F_{X|W}(x|w))}{f_{X|W}(x|w)}\cdot\left(\mathbb{E}\left[X|X>x,W=w\right]-\mathbb{E}\left[X|X\leq x,W=w\right]\right).$$

Thus

$$\beta^{\mathrm{lc,v}}=\mathbb{E}[b(W)]=\mathbb{E}\left[\frac{\partial^2}{\partial w\partial x}g(W,X)\cdot\delta(W,X)\right].$$

□

Proof of Theorem 4.1: We apply Lemmas A.15-A.18. The assumptions in the theorem imply that the conditions for those lemmas are satisfied. □

Proof of Theorem 4.2: The proof is essentially the same as that for Theorem 4.1 and is omitted. □

Proof of Theorem 4.3: We apply Lemma's A.24-A.28 to get an asymptotic linear representation for $\hat\beta^{\mathrm{cm}}(\rho,\tau)$. The assumptions in the Theorem imply that the conditions for the applications of these lemmas are satisfied. Therefore, by Lemma A.24, we have

$$\hat\beta^{\mathrm{cm}}(\rho,0)=\beta^{\mathrm{cm}}(\rho,0)+\left(\hat\beta_g^{\mathrm{cm}}-\overline{g}^{\mathrm{cm}}\right)+\left(\hat\beta_W^{\mathrm{cm}}-\overline{g}^{\mathrm{cm}}\right)+\left(\hat\beta_X^{\mathrm{cm}}-\overline{g}^{\mathrm{cm}}\right)+\left(\overline{g}^{\mathrm{cm}}-\beta^{\mathrm{cm}}(\rho,0)\right)+o_p\left(N^{-1/2}\right).$$

By Lemmas A.25-A.28, this is equal to

$$\beta^{\mathrm{cm}}(\rho,0)+\frac{1}{N}\sum_{i=1}^{N}\left\{\psi_g^{\mathrm{cm}}(Y_i,W_i,X_i)+\psi_W^{\mathrm{cm}}(Y_i,W_i,X_i)+\psi_X^{\mathrm{cm}}(Y_i,W_i,X_i)+\psi_0^{\mathrm{cm}}(Y_i,W_i,X_i)\right\}+o_p\left(N^{-1/2}\right)$$
$$=\beta^{\mathrm{cm}}(\rho,0)+\frac{1}{N}\sum_{i=1}^{N}\psi(Y_i,W_i,X_i)+o_p\left(N^{-1/2}\right),$$

with $\psi_g^{\mathrm{cm}}(y,w,x)$ given in (4.34), $\psi_W^{\mathrm{cm}}(y,w,x)$ given in (4.35), $\psi_X^{\mathrm{cm}}(y,w,x)$ given in (4.36), $\psi_0^{\mathrm{cm}}(y,w,x)$ given in (4.33), and $\psi(y,w,x)$ given in (4.37). Then we have an asymptotic linear representation for $\hat\beta^{\mathrm{cm}}(\rho,\tau)$:

$$\begin{aligned}\hat\beta^{\mathrm{cm}}(\rho,\tau)&=\tau\cdot\overline{Y}+(1-\tau)\cdot\hat\beta^{\mathrm{cm}}(\rho,0)\\&=\beta^{\mathrm{cm}}(\rho,\tau)+\tau\cdot(\overline{Y}-\beta^{\mathrm{cm}}(\rho,1))+(1-\tau)\cdot\left(\hat\beta^{\mathrm{cm}}(\rho,0)-\beta^{\mathrm{cm}}(\rho,0)\right)\\&=\beta^{\mathrm{cm}}(\rho,\tau)+\tau\cdot(\overline{Y}-\beta^{\mathrm{cm}}(\rho,1))+(1-\tau)\cdot\frac{1}{N}\sum_{i=1}^{N}\psi(Y_i,W_i,X_i).\end{aligned}$$

Since by a law of large numbers $\overline{Y}\to\beta^{\mathrm{cm}}(\rho,1)$, and $\sum_i\psi(Y_i,W_i,X_i)/N\to\mathbb{E}[\psi(Y_i,W_i,X_i)]=0$, it follows that $\hat\beta^{\mathrm{cm}}(\rho,\tau)\to\beta^{\mathrm{cm}}(\rho,\tau)$. By a central limit theorem the second part of the Theorem follows. □

Proof of Theorem 4.4: The proof uses Lemmas A.13, A.14, A.19, A.20, and A.23.
By the conditions on q, r, s, and δ, Lemma A.13 implies that for some $\eta>1/4$

$$\sup_{w\in\mathbb{W}}|\hat m(w)-m(w)|=o_p\left(N^{-\eta}\right).$$

Moreover, by the same conditions, Lemma A.14 implies that for some $\eta>1/4$,

$$\sup_{w\in\mathbb{W},x\in\mathbb{X}}\left|\frac{\partial\hat g}{\partial w}(w,x)-\frac{\partial g}{\partial w}(w,x)\right|=o_p\left(N^{-\eta}\right).$$

Then, the conditions for Lemma A.19 are satisfied, so we can write

$$\begin{aligned}&\sqrt{N}\left(\hat\beta^{\mathrm{lc}}-\beta^{\mathrm{lc}}\right)\\&=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\partial g}{\partial w}(W_i,X_i)\cdot d(W_i)\cdot(X-m(W_i))-\sqrt{N}\cdot\beta^{\mathrm{lc}}\\&\quad+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\partial\hat g}{\partial w}(W_i,X_i)\cdot d(W_i)\cdot(X-m(W_i))-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\partial g}{\partial w}(W_i,X_i)\cdot d(W_i)\cdot(X-m(W_i))\\&\quad+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\partial g}{\partial w}(W_i,X_i)\cdot d(W_i)\cdot(X-\hat m(W_i))-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\partial g}{\partial w}(W_i,X_i)\cdot d(W_i)\cdot(X-m(W_i))+o_p(1).\end{aligned}$$

By Lemma A.20,

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\partial}{\partial w}\widehat{g}\left(W_i,X_i\right)\cdot d(W_i)\cdot\left(X_i-m\left(W_i\right)\right)-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\partial}{\partial w}g\left(W_i,X_i\right)\cdot d(W_i)\cdot\left(X_i-m\left(W_i\right)\right)$$
$$=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\psi_g^{\mathrm{lc}}(Y_i,W_i,X_i)+o_p(1),$$

where

$$\begin{aligned}\psi_g^{\mathrm{lc}}\left(y,w,x\right)=&-\frac{1}{f_{W,X}\left(w,x\right)}\frac{\partial f_{W,X}\left(w,x\right)}{\partial W}\left(y-g\left(w,x\right)\right)d(w)\left(x-m\left(w\right)\right)\\&-\frac{\partial m\left(w\right)}{\partial W}d(w)\left(y-g\left(w,x\right)\right)\\&+\frac{\partial}{\partial w}d(w)\left(x-m(w)\right)\left(y-g(w,x)\right).\end{aligned}$$

By Lemma A.23,

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\partial}{\partial w}g\left(W_i,X_i\right)\cdot d(W_i)\cdot\widehat{m}\left(W_i\right)-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\partial}{\partial w}g\left(W_i,X_i\right)\cdot d(W_i)\cdot m\left(W_i\right)$$

$$=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\psi_m^{\mathrm{lc}}(Y_i,W_i,X_i)+o_p(1),$$

where

$$\psi_m^{\mathrm{lc}}(y,w,x)=\mathbb{E}\left[\left.\frac{\partial g(w,X_i)}{\partial W}\right|W_i=w\right]\cdot d(w)\cdot(x-m(w)).$$

Combining these results implies that

$$\sqrt{N}\left(\hat{\beta}^{\mathrm{lc}}-\beta^{\mathrm{lc}}\right)=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\psi^{\mathrm{lc}}(Y_i,W_i,X_i)+o_p(1),$$

with

$$\psi^{\mathrm{lc}}(y,w,x)=\left(\frac{\partial g(w,x)}{\partial w}\cdot d(w)\cdot(x-m(w))-\beta^{\mathrm{lc}}\right)+\psi_g^{\mathrm{lc}}(y,w,x)+\psi_m^{\mathrm{lc}}(y,w,x).$$

Using a law of large numbers then implies the first result in the theorem, and using a central limit theorem implies the second result in the Theorem. □

Notation: (page number indicates where it was first introduced)

(Y_i, W_i, X_i, V_i) observed variables for unit i, $i = 1, \ldots, N$. Y_i, W_i, X_i are scalars, V_i is vector. (page 3)

$k(w, x, v, \varepsilon)$ is production function (page 3)

$g(w, x, v)$ is average production function (conditional expectation of Y given (W, X, V)). (page 4, equation 2.2)

$\sigma^2(w, x, v)$ is conditional variance of Y given (W, X, V)). (page 4, equation 2.3)

$g_W(w, x, v)$ is derivative of average production function. (page 4, equation 2.4)

$h_{W|X,V}(w|x, v)$ is potential conditional distribution of W given X and V (page 6)

$f_{W|X,V}(w|x, v)$ is conditional distribution of W given X and V (page 6)

β_h^{are} is output given new allocation indexed by h. (page 6)

$F_{W|V}(w|v)$ denotes conditional distribution function of W given V. (page 6)

β^{pam} is positive assortive matching output (page 6, equation (3.6))

$\beta^{\text{pam}-\text{pop}}$ is alternative positive assortive matching output (page 7, equation (3.7))

β^{nam} is negative assortive matching output (page 7, equation (3.8))

β^{sq} is *status quo* output (page 8)

β^{rm} is random matching output (page 8)

$\phi(x_1, x_2, \rho)$ bivariate normal density with correlation ρ, (page 8)

$\Phi(x_1, x_2, \rho)$ bivariate normal distribution with correlation ρ, (page 8)

$\phi_c(x_1, x_2, \rho)$ truncated bivariate normal density with correlation ρ, (page 8)

$\Phi_c(x_1, x_2, \rho)$ truncated bivariate normal distribution with correlation ρ, (page 9)

$H_{W,X}(w, mx)$ joint distribution function from truncated bivariate normal cupola (page 9).

$h_{W,X}(w, mx)$ joint density function from truncated bivariate normal cupola (page 9).

$\beta^{\text{cm}}(\rho, \tau)$ correlated matching estimand (page 8, and page 9, equation (3.10))

$d(w)$ weight function in local complementarity measure (page 10)

U_λ combination of W and X for local allocation (page 10)

$\beta^{\text{lr}}(\lambda)$ path of local reallocations page 10, equation (3.11))

β^{lc} local reallocation measure (page 11, equation (3.12)

$\mathbb{W}$ support of W (page 11, assumption 3.1)

$\mathbb{X}$ support of X (page 11, assumption 3.1)

$\delta(w,x)$ weight function in local complementarity measure in representation as weighted average of cross derivative (page 11)

q is the number of derivatives of g and f_{WX} (page 11).

$\hat{F}_W(w)$ estimate of cumulative distribution function for W (page 12)

$m(w)$ (page 12, equation 4.15)

$h_1(w,x) = f_{W,X}(w,x)$ notation for density and product of density and regression function (page 12)

$h_2(w,x) = g(w,x) \cdot F_{W,X}(w,x)$ notation for density and product of density and regression function (page 12, equation (4.16))

$\tilde{Y} = (\tilde{Y}_{i1}, \tilde{Y}_{i2})$ with $\tilde{Y}_{i1} = 1$, $\tilde{Y}_{i2} = Y_i$ (page 12)

$\hat{h}_{\text{nw},m}(w,x)$ nadaraya-watson kernel estimator for $h_m(w,x)$ (page 12, equation (4.17))

$z = (w,x)'$ and $Z = (W,X)'$ compact notation for pair of covariates (page 13)

L dimension of Z (is equal to 2 (page 13)

λ vector of nonnegative integers of dimension $L = 2$ (page 13)

$z^\lambda = \prod_{l=1}^{L} z_l^{\lambda_l}$ (page 13)

$g^{(\lambda)}(z) = \frac{\partial g^{|\lambda|}}{\partial z^\lambda}(z)$ (page 13)

$\mathbb{Z}_b^I$ internal region (page 13, equation (4.18)

$\mathbb{Z}_b^B$ boundary region (page 13, equation (4.19)

$t(z; g, r, p)$ taylor series expansion evaluated at z, equation (4.20)

$r_b(z)$ projection on internal region (page 14, equation 4.21)

$\hat{h}_{m,\text{nip},s}(z)$ NIP estimator for $h_m(z)$ (page 14, equation 4.22)

$\hat{g}_{\text{nip},s}(w,x)$ NIP estimator for $g(w,x)$ (page 14, equation 4.23)

$\widehat{\frac{\partial g_{\text{nip},s}}{\partial w}}(w,x)$ NIP estimator for derivative of $g(w,x)$ 4.24

$\hat{g}(w,x) = \hat{g}_{\text{nip},s}(w,x)$ NIP estimator for $g(w,x)$ short hand for NIP estimator (page 14)

derivative order of kernel is defined in definition 4.1 on page 14

$K(\cdot)$ bivariate kernel (page 15)

$\mathcal{K}(\cdot)$ univariate kernel (page 15)

$\mathbb{U}$ support of bivariate kernel (page 15).

r is number of derivatives of kernel $K(u)$ (page 15)

s is order of kernel $K(u)$, and order of NIP kernel estimator (page 14, 15)

d is derivative order of kernel $K(u)$ (page 15)

$b_N = N^{-\delta}$ is bandwidth (page 15)

$\hat{\beta}^{\text{pam}}$ estimator for β^{pam} (page 15, equation 4.25)

$\hat{\beta}^{\text{nam}}$ estimator for β^{nam} (page 15, equation 4.26)

$\tilde{\beta}^{\text{pam}}$ estimator for β^{pam} given known $g(\cdot)$ (page 15, equation 4.27)

$q^{\text{pam}}(w, x)$ (page 16)

$\psi_W^{\text{pam}}(w)$ (page 16)

$r^{\text{pam}}(x, z)$ (page 16)

$\psi_X^{\text{pam}}(x)$ (page 16)

Ω_{11}^{pam} (page 16)

Ω_{22}^{pam} (page 16)

$\tilde{\beta}^{\text{nam}}$ (page 16, equation 4.29)

$q^{\text{nam}}(w, x)$ (page 17)

$\psi_W^{\text{nam}}(w)$ (page 17)

$r_{XZ}^{\text{nam}}(x, z)$ (page 17)

$\psi_X^{\text{nam}}(x)$ (page 17)

Ω_{11}^{nam} (page 17)

Ω_{22}^{nam} (page 17)

$\hat{\beta}^{\text{cm}}(\rho, \tau)$ (page 17)

$\hat{\beta}^{\text{sq}}$ (page 17)

$\eta(w)$ (page 18, equation 4.30)

$d(w, x)$ (page 18, equation 4.31)

$m(Y, W, \beta^{\text{cm}}(\rho, \tau), \eta(W))$ moment function (page 18)

$e_W(w, x)$ (page 18)

$e_X(w, x)$ (page 18)

$\omega^{\text{cm}}(w, x)$ (page 18)

$\psi_0^{\mathrm{cm}}(y, w, x)$ (page 19)

$\psi_g^{\mathrm{cm}}(y, w, x)$ (page 19)

$\psi_W^{\mathrm{cm}}(y, w, x)$ (page 19)

$\psi_X^{\mathrm{cm}}(y, w, x)$ (page 19)

Ω^{cm} (page 19)

$\psi^{\mathrm{cm}}(y, w, x)$ (page 19)

Ω^{lc} (page 20)

$\psi_g^{\mathrm{lc}}(y, w, x)$ (page 20)

$\psi_m^{\mathrm{lc}}(y, w, x)$ (page 20)

$\hat{\beta}_g^{\mathrm{pam}}$ (page 21, equation 5.39)

$\hat{\beta}_W^{\mathrm{pam}}$ (page 21, equation 5.40)

$\hat{\beta}_X^{\mathrm{pam}}$ (page 21, equation 5.41)

$\overline{g}^{\mathrm{pam}}$ (page 21, equation 5.42)

$Z = (W, X)'$ (page 29)

$\omega(Z)$ and $\omega(X)$ (page 29)

$n(h^{[\lambda]})$ and $n(h)$ (page 29)

$t(x)$ (page 29)

θ^{fm} (page 29, A.2)

$\theta^{\pm}$ (page 29, A.3)

$\hat{\theta}^{\mathrm{fm}}$ (page 29)

$\hat{\theta}^{\mathrm{pm}}$ (page 29)

$\overline{\theta}^{\mathrm{fm}}$ (page 29)

$\overline{\theta}^{\mathrm{pm}}$ (page 29)

$\alpha_{\kappa 1}^{(\kappa)}(z)$ (page 30)

$\alpha_m(z)$ (page 30)

V_1, V_2 (page 30)

$\hat{\beta}_g^{\mathrm{lc}}$ (page 31)

$\hat{\beta}_m^{\mathrm{lc}}$ (page 31)