THE PARTIAL MONOTONICITY PARAMETER: A GENERALIZATION OF REGRESSION MONOTONICITY*

SCOTT KOSTYSHAK AND YE LUO

ABSTRACT. We define the partial monotonicity parameter (PMP) as the proportion of the population for which a small increase in an explanatory variable is associated with an increase in the outcome variable. The PMP is a novel approach useful in three classes of applications: (1) Even though monotonicity may be the most common predicted relationship among variables in economics, it is rarely tested in practice because even a small violation causes it to be rejected. The PMP fills this gap between theory and practice by allowing for estimation of an interpretable parameter that includes standard monotonicity as a special case. (2) The PMP generalizes from binary categorizations to continuums some classical properties, such as a good being normal, and relations, such as inputs being complements. (3) In the presence of heterogeneous effects, inference on the PMP provides answers to policy-relevant questions, including whether an increase in a variable would benefit the majority of the population. We provide results for parametric and non-parametric inference for the PMP, as well as results for joint inference with average-effect parameters.

KEY WORDS. Regression Monotonicity, Heterogeneous Effects, Treatment Effects, Program Evaluation, Semi-Parametric Regression, Non-Linear Model

1. INTRODUCTION

This paper proposes a new parameter, the partial monotonicity parameter (PMP), which answers the question "For what proportion of the population is a small increase in x associated with an increase in y?" This question is important in many applications, including program evaluation, tests of model predictions, heterogeneous treatment effects, and any empirical analysis in which it is of interest to identify the proportion of the population that would benefit from, or be harmed by, an increase in x. The PMP also generalizes from binary categorizations to continuums some classical properties in microeconomics, such as a good being normal, and relations, such as inputs being complements.

Let (X, Y) be a random vector drawn from a population of interest, and for the sake of terminology suppose a larger value of Y is always "better." Let the relationship between X and Y be modeled as

$$Y = f(X) + \epsilon, \tag{1.1}$$

Date: March 3, 2023.

where $E(\epsilon) = 0$, X has support $\mathcal{X}$, and for simplicity a causal interpretation of f holds.[1] When $\mathcal{X}$ is one-dimensional, we define the PMP as $\delta^+ := \mathrm{P}[f'(X) > 0]$, the proportion of the population for which a small increase in X is associated with an increase in Y.

This simple framework can be generalized in various ways when $\mathcal{X}$ is multi-dimensional. For the explanatory variable of interest x_j and corresponding partial derivative $f_{x_j} := \frac{\partial f}{\partial x_j}$, we consider the parameter $\delta^+_{x_j} := \mathrm{P}\left[f_{x_j}(X) > 0\right]$, which allows for multiple variables to drive heterogeneity in the sign of f_{x_j}. We also consider parameters of higher-order partial derivatives of f. In Example 1 below, we consider a bivariate production function f with capital (k) and labor (l) as inputs and define multiple partial monotonicity parameters of interest: $\delta^+_{ii} := \mathrm{P}[f_{ii}(K,L) > 0]$ for $i \in \{k,l\}$ to capture partial increasing returns in individual inputs; $\delta^+_{kl} := \mathrm{P}[f_{kl}(K,L) > 0]$ to capture partial complementarity of inputs; and $\delta^+_{k;l} := \mathrm{P}[f_k(K,L) > 0, f_l(K,L) > 0]$ to capture partial monotonicity in a multivariate sense. Extending these definitions to an arbitrary number of inputs is straightforward.

A separate generalization of the PMP is to consider joint probabilities of the signs of derivatives of different orders. For example, the production function in some industries is predicted to exhibit decreasing positive marginal returns, which corresponds to the function being both monotone and concave. The extent to which this prediction holds could be measured by the parameters $\delta_{i;ii^-} := \mathrm{P}[f_i(K,L) > 0, f_{ii}(K,L) < 0]$ for $i \in \{k,l\}$. Similarly, on a more aggregate level, most growth theories predict monotone concave convergence dynamics.

An important feature of these generalizations of PMP is that the parameter space is always the one-dimensional unit interval, $[0,1]$. Specifically, the dimension of the parameter space of the non-parametric PMP estimator we propose does not increase with the dimension of $\mathcal{X}$ or the number of events included in the joint probability. Few non-parametric estimators in econometrics have this convenient property, which makes reporting and interpreting results relatively simple and concise.

Many theories in economics predict monotone relationships between variables. However, monotonicity is rarely tested by empiricists because it is viewed as too strict in practice: A theory that is mostly supported by the data could be rejected because of a minor violation of monotonicity arising from selection, measurement error, competing effects, or a violation of an assumption by a small subgroup. In such cases, one way is to quantify the degree of monotonicity by performing inference on the proportion of data points in the population such that a monotonicity constraint holds in a local sense. The PMP fulfills this role and thus

[1] This assumption could either be satisfied by the nature of the vector (X, ϵ), such as assuming $E(\epsilon|X) = 0$, or relaxed in an extended model by employing instrumental variables, panel data techniques, and other econometric methods. As long as an estimator of f can be interpreted causally, the same holds for the estimator of the PMP. If the required assumptions cannot be satisfied for a causal interpretation of f, the PMP still has a predictive interpretation.

reconciles the asymmetry between a prediction of monotonicity and the common practice of testing a weaker prediction, such as the average effect being positive.

In the presence of heterogeneous effects, it has long been recognized that estimators of average effects, such as individual ordinary least squares (OLS) coefficients, are sensitive to a minority's strong positive effect dominating a majority's weak negative effect. Although empiricists commonly add polynomial and interaction terms to model heterogeneous effects, we are not aware of any paper that performs *inference* on a non-binary generalization of monotonicity. Further, the polynomial terms are often added in an ad hoc manner.[2] Reporting results of the combined fit of the polynomial terms in an interpretable way essentially requires a graph of the derivative of the fit and confidence band, as well as some representation of the marginal density of X (e.g., a rug plot) for the reader to understand whether a violation of monotonicity in the fit corresponds to a considerable portion of the population. If heterogeneity is driven by other variables (i.e., interaction terms), then f_x is multi-dimensional, which makes it even more difficult to interpret the visualization of the fit and relevant joint marginal distribution.

In the context of parametric estimation, we make the contribution of providing results for inference on the PMP using polynomials and interaction terms to capture heterogeneous effects. The PMP naturally integrates over the joint marginal density of the variable of interest and interaction variables, and the result can be represented in a single table cell as opposed to a graph. We additionally provide results for semi-parametric inference. Because of the non-increasing dimensionality of the PMP, semi-parametric estimation offers a practical alternative that is just as easy to report as individual OLS coefficients and is root-n consistent.

By abstracting away from magnitude, inference on the PMP captures information that complements average effect estimators. Similar to the common practice of reporting the mean and median of a variable, average effects and the PMP together provide a more complete picture of the regression function being studied than average effects alone. We are the first to provide joint inference results for parametric and non-parametric frameworks to test the simple hypothesis that an increase would have a positive average effect *and* would benefit the majority of the population.

Relative to current tests of monotonicity, inference on the PMP allows an empiricist to estimate a parameter rather than perform a test based on a statistic that is difficult to interpret directly and is useful only to generate a p-value. The output from current tests of monotonicity

[2] A common approach is to add terms as long as they are significant. If an insignificant coefficient on a polynomial term has a wide confidence interval, that could provide important information for inference, e.g., for the confidence interval of the PMP. However, global polynomial terms of high degree often have high standard errors in practice. These ad hoc approaches thus lead to disadvantages compared with semi-parametric methods that avoid global polynomials.

is analogous to the hypothetical situation in which OLS regression would report only the p-value corresponding to the null hypothesis that a coefficient is zero, without reporting the actual estimated coefficient itself.

A test of standard monotonicity based on the PMP can be performed by testing the null hypothesis $H_0 \colon \delta^+ = 1$. Performing inference on the PMP has an advantage over classical tests of monotonicity even when the researcher desires a test of standard monotonicity. Although a failure to reject monotonicity by any test could be due to low power, the PMP is unique: Because it is a parameter that is directly interpretable, the confidence interval indicates whether power is likely low. If a confidence interval for δ^+ is concentrated around 1, then the failure to reject H_0 is not likely due to low power. In contrast, in general for classical tests of monotonicity, it would be necessary to perform power analysis to assess whether a failure to reject is likely due to low power.

The PMP generalizes some classical properties in economics from binary categorizations to continuums, which allows researchers to perform inference on the extent to which the properties hold. For example, whether demand for a good increases with additional income depends on the baseline level of income as well as other consumer characteristics. Partial monotonicity allows for inference on the extent to which demand increases with additional income, taking into account the joint marginal distribution of baseline income and consumer characteristics, and thus extending the binary notion of a good being normal to an index from zero to one. Similarly, in the multivariate production function of Example 1 below, we explore partial monotonicity of the cross-derivative in order to perform inference on the extent to which two inputs are complementary. Underlying these examples are potentially complex relationships affected by multiple variables, and the PMP can assess the extent to which violations of monotonicity manifest from the heterogeneous relationships.

We provide results for inference on the PMP that can be applied by techniques as simple as adding polynomial and interaction terms to a regression. The results apply to OLS, quantile regression, probit and logit, and many other common parametric estimation techniques. We also derive results for non-parametric estimation of the PMP based on local polynomial estimation. Implementation of PMP estimation has appealing computational properties and does not require, for example, quadratic programming for projections. We offer practical advice to empiricists, as well as a user-friendly R package that implements both parametric and non-parametric PMP estimation.

Literature review. The paper most closely related to ours is Chernozhukov, Fernández-Val & Luo (2018), which proposes uniform estimation of the "sorted effect curve," which completely represents the range of heterogeneous effects. Chernozhukov et al. (2018) allows us to construct inference of one-dimensional functions defined on level sets under reasonable smoothness assumptions. For a variable of interest, Z, when the function $g := \frac{\partial f(X)}{\partial Z}$ is one-dimensional, we are able to formulate an inference procedure based on Chernozhukov et al. (2018). However, their method does not apply directly to multi-dimensional problems and requires the functional central limit theorem of the original function. In this paper, we relax this assumption in the analysis of the PMP with application to local-linear estimation. In addition, the PMP is a scalar that summarizes heterogeneity of the sign of the partial effects in order to preserve a connection to monotonicity and provide results that can be presented in a table cell, rather than providing a complete representation of the partial effects.

There is a rich literature on the inference of monotonicity or monotonized functions. See, for example, the two recent papers Chetverikov (2019) and Fang & Seo (2021) which both provide uniformly valid and asymptotically non-conservative tests of monotonicity. See also Du, Parmeter & Racine (2013) and Du, Parmeter & Racine (2021), which propose to constrain the weights in kernel regression of the data points to achieve the desired restriction. Finally, see the recent survey on shape restrictions by Chetverikov, Santos & Shaikh (2018).

Assuming a model similar to (1.1), Powell, Stock & Stoker (1989) and Härdle & Stoker (1989) propose to estimate the average derivative of the regression function, $E\left[f'(X)\right]$. This parameter captures similar information to OLS, and does so in a more flexible way (f can be modeled semi-parametrically). However, like OLS, this average-effect parameter is sensitive to magnitude and thus misses information on the average *sign* of the derivative of f, as discussed in the context of individual OLS coefficients above.

Outline. The rest of this paper is organized as follows. In the next section we present examples of how the PMP framework is useful in several settings. Section 3 discusses the asymptotic theory for parametric PMP estimation. Section 4 extends the theory to non-parametric estimation via local polynomial regression, and Section 5 has results on further extensions. Section 6 provides results from Monte Carlo simulations to explore the finite-sample properties of PMP estimators. Section 7 concludes.

2. Examples

We first introduce some notation that allows for higher-order partial derivatives as well as joint probabilities. Let X be a random vector drawn from a population of interest and have support $\mathcal{X} \subset \mathbb{R}^k$. For $f : \mathcal{X} \mapsto \mathbb{R}$ and $j_1, j_2, \ldots, j_p \in \{1, 2, \ldots, k\}$, we use $f_{x_{j_1} x_{j_2} \cdots x_{j_k}} := \frac{\partial^p f}{\partial x_{j_1} \partial x_{j_2} \cdots \partial x_{j_p}}$. Similarly, we define $\delta^+_{x_{j_1} x_{j_2} \cdots x_{j_p}} := \mathrm{P}\left[f_{x_{j_1} x_{j_2} \cdots x_{j_p}}(X_1, X_2, \ldots, X_k) > 0\right]$.

Example 1 (Production functions)**.** It is often important to understand the extent to which a production function exhibits certain properties that are implied by belonging to classes of shapes. For example, convexity implies increasing returns to scale. These properties affect the presence of monopolies, inform policies on subsidies and industry protection, and can explain structural differences across industries. Consider a bivariate production function f with capital (K) and labor (L) as inputs, modeled as $Y = f(K, L) + \epsilon$. The inputs K and L are defined as being "q-complements" if $f_{kl}(k, l) > 0$ for all pairs $(k, l) \in \mathcal{X}$. We apply the concept of partial monotonicity to the cross-derivative of f in order to generalize this property to accommodate production functions in which the inequality holds for some pairs $(k, l) \in \mathcal{X}$ but not necessarily all: $\delta^+_{kl} \coloneqq \mathrm{P}\left[f_{kl}(K, L) > 0\right]$. For example, suppose each worker can operate up to, but no more than, two machines simultaneously. In this case, we might expect f_{kl} to be strictly positive over the lower half-space $\{(k, l) \in \mathbb{R}^2 : l < 2k\}$ but not over the upper half-space counterpart.

On a more basic level, we might seek to check that the production function is increasing as the levels of inputs increase. We thus consider the following parameter: $\delta^+_{k;l} \coloneqq \mathrm{P}\left[f_k(K, L) > 0, f_l(K, L) > 0\right]$. Although in theory monotonicity is a fundamental property of production functions, in an empirical setting $\hat{f}$ might be biased in a way that monotonicity is violated over part of $\mathcal{X}$. For example, productivity or product quality could be unobserved and correlated with base levels of capital and labor. Inference on the PMP can be employed to check whether the data generating process is *close* to satisfying the fundamental property of monotonicity.

The above examples extend trivially to higher dimensions: For a production function $f : \mathbb{R}^n \to \mathbb{R}$ with n inputs and $i, j \in \{1, \ldots, n\}$, we define $\delta^+_{ij} \coloneqq \mathrm{P}\left[f_{ij}(X) > 0\right]$, $\delta^+_{1:n} \coloneqq \mathrm{P}\left[f_1(X) > 0, \ldots, f_n(X) > 0\right]$.

Example 2 (The policy maker's null)**.** Policy makers might be interested in testing whether an increase in a policy-influenced variable x would have a positive average effect *and* would be beneficial for the majority of the population. Let $\beta \coloneqq E\left[f'(X)\right]$ be the average effect. We can then specify the policy maker's null hypothesis in terms of δ^+ and β as follows:

$$H_0\colon \;\; \beta \leq 0 \quad \text{or} \quad \delta^+ \leq 0.5$$

$$\text{vs.}$$

$$H_A\colon \;\; \beta > 0 \quad \text{and} \quad \delta^+ > 0.5.$$

δ^+ captures information on the average *sign* of the derivative, while β captures information on the average *magnitude*. Together, they provide a rich summary of the distribution of $f'(X)$ that is useful for practitioners. We provide results for joint inference on the vector (δ^+, β) in Section 5.1.

Example 3 (Life-cycle predictions)**.** Some theories in economics make predictions about the shape of features of the joint distribution of variables over the life cycle, such as the conditional mean function or conditional variance function.[3] Suppose a theory predicts that a function

[3] In this paper we focus on estimation of the conditional mean function, but we mention the conditional variance function here to emphasize that the framework of partial monotonicity has many potential applications.

$f : \mathbb{R} \mapsto \mathbb{R}^+$ defined by $f(\text{age}) = E(y|\text{age})$ or by $f(\text{age}) = V(y|\text{age})$ belongs to the class $\mathcal{F}_1$ of monotone functions, the class $\mathcal{F}_{2^-}$ of concave functions, or the class $\mathcal{F}_1 \cap \mathcal{F}_{2^-}$ of monotone concave functions. For example, Storesletten, Telmer & Yaron (2004) theoretically and empirically examines the monotonicity and concavity of consumption inequality over age in the United States; and Van Landeghem (2012) explores convexity of life satisfaction over age in the United States. We can embed the testing of these classes of prediction into the PMP framework by considering partial monotonicity and partial convexity of f. A notable feature of this example is that since the prediction is about the life cycle, the strength of the prediction is most appropriately assessed with the *unweighted* shape of f'. That is, we define $\delta^{U+}_{(1)} := \mathrm{P}_U \left[f'(\text{age}) > 0\right]$, $\delta^{U+}_{(2^-)} := \mathrm{P}_U \left[f''(\text{age}) < 0\right]$, and $\delta^{U+}_{(1);(2^-)} := \mathrm{P}_U \left[f'(\text{age}) > 0, f''(\text{age}) < 0\right]$, where the probability is taken with respect to the uniform distribution $U \sim \left[\underline{\text{age}}, \overline{\text{age}}\right]$, where $\left[\underline{\text{age}}, \overline{\text{age}}\right]$ is the interval of ages over which the model makes a prediction. If instead we used the $\delta^+ \equiv \delta^{age+}$ counterparts, the parameters would implicitly integrate over the marginal distribution of age; and if the population has more young people than old people, the parameters would assess the model's prediction with a higher weight on the predicted sign for young people. In Section 5.2, we provide results for inference on the counterfactually-weighted uniform PMP, δ^{U+}, and more generally for δ^{Z+} where Z has any distribution that satisfies certain regularity conditions.

Example 4 (Heterogeneous treatment effects)**.** We are often interested in heterogeneous treatment effects in which the heterogeneity is potentially driven by more than one variable. We can consider inference on the PMP when we are interested in heterogeneity of the *sign* of the derivative. For example, whether a certain drug or vaccine benefits people could depend on their age, weight, blood pressure, diet, and medical history. Suppose that an outcome of interest, Y, is modeled as $Y = f(X_1, \ldots, X_k) + \epsilon$, where x_j is the treatment of interest for $1 \leq j \leq k$, and f is estimated either parametrically (in which case the heterogeneous effects are captured by interaction terms) or non-parametrically. Displaying the full results of estimation of $f_{x_j}(x_1, \ldots, x_k)$ and corresponding confidence band would be complex, even in the case of a parametric specification. Further, in order to take into account how much of the population is affected, we would need to present additional inference on the joint distribution of X. The PMP simplifies reporting results by considering inference on the scalar $\delta^+_{x_j} := P\left[f_{x_j}(X_1, \ldots, X_k) > 0\right]$.[4]

Example 5 (Slutsky restrictions)**.** Consider a sample of individuals purchasing G goods. Let $q_i \in \mathbb{R}^G$ denote the vector of quantities purchased by consumer i facing prices $p_i \in \mathbb{R}^G$ with income $y_i \in \mathbb{R}$. Let demand be modeled as $Q_i = x(P_i, Y_i) + \epsilon_i$, where $x\colon \mathbb{R}^G \times \mathbb{R} \to \mathbb{R}^G$ and $\epsilon_i \in R^G$ is an independent error term with mean 0. For consumer behavior to be consistent

[4]For simplicity, we focus on the case in which the partial derivative exists, but the PMP can be estimated when the treatment is binary as well with the parameter $\delta^+_{x_j} := P\left[f(X_1, \ldots, X_{j-1}, 1, X_{j+1}, \ldots, X_k) - f(X_1, \ldots, X_{j-1}, 0, X_{j+1}, \ldots, X_k) > 0\right]$.

with utility maximization, x must satisfy the Slutsky restrictions, which are conditions on the $\mathbb{R}^G \times \mathbb{R}^G$ matrix

$$S(p, y) = \frac{\partial x(p, y)}{\partial p^T} + \frac{\partial x(p, y)}{\partial y} x(p, y)^T.$$

S represents the derivative of the Hicksian demand with respect to price. The Slutsky restrictions require that $S(p, y)$ be symmetric and negative semi-definite (NSD) for every $(p, y) \in \mathbb{R}^{G+1}$. Intuitively, S is the substitution effect of a price increase, and consumer demand theory suggests it should be negative regardless of a good's price or a consumer's income.

Several papers have provided tests of the Slutsky restrictions.[5] The present paper suggests an alternative approach to the previous strict tests of the Slutsky restrictions by proposing inference for the parameter $\mathrm{P}\left[S(P, Y) \text{ is NSD}\right]$. Even if $S(p, y)$ is symmetric for all (p, y), it may be that the estimator $\hat{S}(p, y)$ is almost surely not symmetric for any pair in the sample, even for arbitrarily large samples. However, the restriction of symmetry can be accommodated by considering an upper-bound estimator for the parameter $\mathrm{P}\left[S(P, Y) \text{ is NSD and symmetric}\right]$[6]

This PMP approach is applicable to similar frameworks as well. For example, recently Bhattacharya (2021) presented sufficient and necessary conditions for binary choice probabilities to be consistent with utility maximization. For simplicity, suppose demand is smooth and let $z(P_i, Y_i)$ denote the fraction of a population which, if they possessed income Y_i and faced prices P_i, would purchase a good. For z to be consistent with utility maximization, the following two inequalities must hold for all (p, y): $z_p(p, y) \leq 0$, $z_p(p, y) + z_y(p, y) \leq 0$. To embed these conditions in a PMP framework, we consider the following parameter:

$$\delta^+ := \mathrm{P}\left[z_p(P, Y) \leq 0, \;\; z_p(P, Y) + z_y(P, Y) \leq 0\right].$$

3. Parametric Estimation

Define the domain of X as $\mathcal{X}$. Denote $p := dim(X)$. Define $g = \frac{\partial f(X)}{\partial Z}$. This definition reduces to $f'(X)$ when $Z = X$ and $p = 1$. In general, Z can be a subvector of X, or any other variables that are transformations of X. For any $y \in \mathbb{R}$ and real-valued function h, denote $\mathcal{M}_h(y) := \{x | h(x) = y\}$ as the contour set of h at level y. For any $y \in \mathbb{R}$, we call y a critical value of h if there exists $x \in \mathcal{M}_h(y)$ such that $\nabla h(x) = 0$, where ∇h is the gradient of h. Otherwise, we call y a regular value of h. Appendix A has definitions for $\mathcal{C}^r$, manifolds, connected branches, and volume of a manifold, which are used in this section.

---

[5]See Banks, Blundell & Lewbel (1997), Blundell, Browning & Crawford (2003), Dette, Hoderlein & Neumeyer (2016), Epstein & Yatchew (1985), Gallant & Souza (1991), Härdle, Hildenbrand & Jerison (1991), Hoderlein (2011), Lewbel (1995), Fang & Seo (2021).

[6]The PMP framework is applicable even for parameters defined by boundary conditions, rather than inequalities. Although the naive plug-in estimator based on $\hat{f}$ is unable to pick up whether an equality holds exactly, $\hat{V}[\hat{f}]$ can be employed to construct interval estimators that asymptotically detect any *violation* of the equality.

We consider a general setting where a preliminary estimator $\hat{g}$ of g exists. To be more specific, we assume that there exists $a_n \to \infty$ such that $a_n(\hat{g}(x) - g(x)) \to_d \mathbb{G}(x)$, where $\mathbb{G}(x)$ is a Gaussian process with tight paths.

Example 6 (Parametric formulation)**.** One way to obtain such $\hat{g}(x)$ is when $f(x)$ is parametrized as $f(x, \beta)$. The plugged in estimator $g(x, \hat{\beta})$ given $\hat{\beta}$ as an estimator of β can, in general, approximate $g(x, \beta)$ in a Gaussian process.

Specifically, assume that $y_i = f(x_i, \beta) + \epsilon_i$ holds. In practice, f is often chosen as polynomials of x_i, e.g., quadratic.

We consider the least squares estimation $\hat{\beta} = argmax_{\beta \in \Theta} \sum_{i=1}^{n} (y_i - f(x_i, \beta))^2$, where Θ is a compact set that contains the true parameter β_0.

Under mild regularity conditions of f, one can conclude that $\hat{\beta}$ converges to β_0. In addition, we usually have $\sqrt{n}(\hat{\beta} - \beta_0) \to_d N(0, \Omega)$.

If g is also $\mathcal{C}^1$ on $\mathcal{X}$, then $\hat{g}(X) = \frac{\partial f(x, \hat{\beta})}{\partial Z}$ converges to $g(X) = \frac{\partial f(x, \beta_0)}{\partial Z}$. In addition, by the Delta-method,

$\sqrt{n}(\hat{g}(x) - g(x)) \to_d N(0, G(x)'\Omega G(x))$, where $G(x) = \frac{\partial g(x, \beta)}{\partial \beta}|_{\beta = \beta_0}$.

As shown in the previous section, the PMP is defined as $\delta^+ := \mathrm{P}(g(X) > 0)$. The estimator is $\hat{\delta}^+ := \mathrm{P}(\hat{g}(x) > 0)$. To show the asymptotic property of $\hat{\delta}^+$, we use techniques from Chernozhukov et al. (2018).

Suppose $g(x) = (g_1(x), \ldots, g_p(x))$ is differentiable in x, and denote $\nabla g_i(x) := \frac{\partial g_i}{\partial x}$, $i = 1, 2, \ldots, p$. For now, assume that all x are continuous. This setting can be easily generalized to the case where discrete variables are allowed.

Assume that the domain of x, denoted as $\mathcal{X}$ is bounded and open. We also assume that $y = 0$ is a regular value of g.

Similar to Chernozhukov et al. (2018), we require the following assumption for all $g(\cdot) = g_i(\cdot)$, $i = 1, 2, \ldots, p$.

Assumption 1. *(1) The part of the domain of the PE function $x \mapsto \Delta(x)$ of interest, $\mathcal{X}$, is open and its closure $\overline{\mathcal{X}}$ is compact. The distribution μ is absolutely continuous with respect to the Lebesgue measure with density μ'. There exists an open set $B(\mathcal{X})$ containing $\overline{\mathcal{X}}$ such that $x \mapsto \Delta(x)$ is $\mathcal{C}^1$ on $B(\mathcal{X})$, and $x \mapsto \mu'(x)$ is continuous on $B(\mathcal{X})$ and is zero outside the domain of interest, i.e., $\mu'(x) = 0$ for any $x \in B(\mathcal{X}) \setminus \mathcal{X}$.*

(2) For any regular value y of g on $\overline{\mathcal{X}}$, we assume that the closure of $\mathcal{M}_g(\delta)$ has a finite number of connected branches.

(3) $\widehat{g}$, the estimator of g, belongs to $\mathcal{F}$ with probability approaching 1 and obeys a functional central limit theorem, namely,

$$a_n(\widehat{g} - g) \rightsquigarrow G_\infty \text{ in } \ell^\infty(B(\mathcal{X})),$$

where a_n is a sequence such that $a_n \to \infty$ as $n \to \infty$, and $x \mapsto G_\infty(x)$ is a tight process that has almost surely uniformly continuous sample paths on $B(\mathcal{X})$.

(4) Define $\mathbb{H}$ as the set of all bounded linear operators H on $\mathcal{G}$ of the form

$$g \mapsto H(g)$$

that are uniformly continuous on $g \in \mathcal{G}$ under the $L^2(\mu)$ norm. The function $x \mapsto \widehat{\mu}(x)$ is a distribution over $B(\mathcal{X})$ obeying in $\mathbb{H}$,

$$b_n(\widehat{\mu} - \mu) \rightsquigarrow H_\infty, \tag{3.1}$$

where $g \mapsto H_\infty(g)$ is a.s. an element of $\mathbb{H}$ (i.e. it has almost surely uniformly continuous sample paths on $\mathcal{G}$ with respect to the $L^2(\mu)$ metric) and b_n is a sequence such that $b_n \to \infty$ as $n \to \infty$.

Let $r_n := a_n \wedge b_n$, the slowest of the rates of convergence of $\widehat{\Delta}$ and $\widehat{\mu}$. Assume $r_n/a_n \to s_\Delta \in [0,1]$ and $r_n/b_n \to s_\mu \in [0,1]$, where $s_g = 0$ when $b_n = o(a_n)$ and $s_\mu = 0$ when $a_n = o(b_n)$. For example, $s_\mu = 0$ if μ is treated as known.

Define $\hat{\delta}_j^+ := \frac{1}{n}\sum_{i=1}^n 1(\hat{g}_j(x_i) > 0)$ as the PMP in j^{th} covariate x_j, $j = 1, 2, \ldots, p$.

Below we present the asymptotic property of the PMP when $\delta_1^+, \ldots, \delta_p^+$ are in the interior.

Theorem 1. *Suppose Assumption 1 holds. Assume that $y = 0$ is a regular value for all $g_1, \ldots, g_p$, and $\delta_1^+, \ldots, \delta_p^+$ are real fixed numbers in $(0, 1)$. Assume that G and H are jointly Gaussian. Then, the PMP estimator satisfies that:*

(1) $r_n((\hat{\delta}_1^+, \ldots, \hat{\delta}_p^+) - (\delta_1^+, \ldots, \delta_p^+)) \rightsquigarrow N(0, \Omega_1)$, where Ω_1 is a semi-positive definite matrix defined in (A.4).

(2) Assume that the manifolds $\mathcal{M}_{g_j}(0)$, $j = 1, 2, \ldots, n$, intersect each other with 0 measure in volume in $d_x - 1$ dimensional space, where volume of manifold.

Then, $r_n(\hat{\delta}^+ - \delta^+) \rightsquigarrow Z_\infty$, where Z_∞ is a Gaussian distribution defined in (A.5).

As shown in Chernozhukov et al. (2018), bootstrap gives consistent inference of δ^+. Below we state a corollary that allows us to perform inference stated in Theorem 1.

Define $\hat{\delta}^{*+}$ as the PMP computed via bootstrap data $x_1^B, \ldots, x_n^B$. More specifically, we assume that the estimator using the bootstrap data satisfies: $a_n(\hat{g}^* - \hat{g}) \rightsquigarrow G(x)$, and $b_n(\hat{\mu}^* - \hat{\mu}) \rightsquigarrow H_\infty$. $\hat{\delta}^{*+}$ is defined as

$$\frac{1}{n}\sum_{i=1}^n 1(\hat{g}^*(x_i^B) > 0).$$

Define $\hat{\delta}_j^{*+} := \sum_{i=1}^{n} 1(\hat{g}_j^*(x_i^B) > 0)$ as the PMP in j^{th} covariate x_j. Below we present the asymptotic property of the PMP.

Corollary 1. *Suppose all assumptions in Theorem 1 hold. Then, the bootstrapped PMP satisfies:*

(1) $r_n((\hat{\delta}_1^{*+}, \dots, \hat{\delta}_p^{*+}) - (\hat{\delta}_1^{+}, \dots, \hat{\delta}_p^{+})) \rightsquigarrow N(0, \Omega_1)$.

(2) $r_n(\hat{\delta}^{*+} - \hat{\delta}^{+}) \rightsquigarrow Z_\infty$.

4. Non-Parametric methods: PMP via local polynomial regression

When parametric form of the target function is unavailable, researchers can consider to use non-parametric methods instead. To address our problem, we propose to use locally linear to estimate partial derivatives of the function $f(X)$ as:

$$\hat{\alpha}(x), \hat{\beta}(x) := argmin_{\alpha,\beta} \sum_{i=1}^{n} K_h(x_i - x)(y_i - \theta^K P^K(\frac{x - x_i}{h}))^2, \tag{4.1}$$

where

$$K_h(u) := \frac{1}{h} K(\frac{u}{h}), \tag{4.2}$$

$K(\cdot)$ is a kernel function, e.g., Epanechnikov kernel, and $P^K(\cdot)$ is a multi-variable polynomial up to the K^{th} order. We assume that K is fixed such that $K \geq p$.

There are a few reasons for us to consider local-linear estimator, rather than the other non-parametric kernel based methods. First, as we are interested in make inference on the PMP, which is a function that depends on the partial derivatives of the function $f(\cdot)$, local polynomial provides a feasible and convenient method to estimate $\frac{\partial f}{\partial X}$. Second, local polynomial is robust to boundaries, which is good for our analysis. Third, since we are not investigating derivatives with order larger than one, local polynomial estimator is enough for our purpose. Based on the same spirit, similar methods such as locally quadratic, can be used if convexity measurement similar to PMP is to be considered.

We consider a non-parametric mean model

$$y_i = f(x_i) + \epsilon_i, \tag{4.3}$$

$i = 1, 2, \dots, n$. Assume that (x_i, ϵ_i) are i.i.d. across i, and $\mathbb{E}[\epsilon_i|X] = 0, Var(\epsilon_i|X) = \sigma^2$.

We adopt a set of common assumptions in Kernel function $K(\cdot)$ as follows:

Assumption 2 (Kernel)**.** *Assume that the kernel function satisfies the following conditions:*

(1) The support of $K(\cdot)$ *is contained in* $\{u : \|u\| \leq 1\}$. $\int_{\|u\|\leq 1} K(u)du = 1$.

(2) $K(u) = K(-u)$.

(3) $K(\cdot)$ is a bounded and continuous kernel function on $\mathbb{R}^p$. The class of functions $\mathcal{K} := \{k(ht+x) : h > 0, x \in \mathcal{R}^d\}$ is VC type with envelop $||k||_\infty$.

Remark 1. *Statements (1) and (2) are common for kernel functions. Statement (3) is required to utilize results in Chernozhukov, Chetverikov & Kato (2014). Such condition holds for well-known kernels such as the Epanechnikov kernel.*

For each $x \in \mathcal{X}$, we can define

$$W_h(x) := \left(K(\frac{x_1 - x}{h}), \ldots, K(\frac{x_n - x}{h})\right)^\intercal \tag{4.4}$$

as the weight function of x. For any random variables $z_1, \ldots, z_n$, we define

$$\mathbb{E}_{n,h}[z_i] = \frac{1}{nh^p} \sum_{i=1}^{n} z_i W_{h,i}(x) \tag{4.5}$$

as the weighted average of $z_1, \ldots, z_n$ given weights vector $W_h(x)$, where $W_{h,i}(x)$ is the i^{th} component of $W_h(x)$.

Denote

$$Q_{n,h} := \mathbb{E}_{n,h}[P^K(\frac{x - x_i}{h})(P^K(\frac{x - x_i}{h}))^\intercal]. \tag{4.6}$$

Therefore, (4.1) implies that

$$(\hat{\alpha}(x), \hat{\beta}(x), \hat{\beta}_H(x))^\intercal = Q_{n,h}^{-1}\mathbb{E}_{n,h}[P^K(\frac{x - x_i}{h})y_i],$$

where $\hat{\beta}_H(x)$ contains all the coefficients in $P^K(\cdot)$ with order > 1.

We consider the non-parametric PMP defined as:

$$\hat{\delta}_j^+ := \int_{x \in \mathcal{X}} 1(\hat{\beta}_j(x) > 0) d\widetilde{\mu}(x), \tag{4.7}$$

where $\widetilde{\mu}(x)$ is a probability measure of x on $\mathcal{X}$, $j = 1, 2, \ldots, p$.

Below we state a set of assumptions for $f(x) := \mathbb{E}[y|x]$, density function $\mu(x)$, and residual ϵ.

Assumption 3. *Assume that the function $f(x)$ satisfies the following conditions:*

(1) For any $x \in \mathcal{X}$, $f(x)$ is twice continuously differentiable.

(2) For any $j = 1, 2, \ldots, p$, the j^{th} derivative of $f(x)$, denoted as $\nabla f_j(x)$, 0 is a regular value of $\nabla f_j(x)$.

(3) For any $s \in \{1, 2, \ldots, K+1\}$, we have: $\sup_{x \in \mathcal{X}} ||\frac{\partial^s f(x)}{\partial^{s_1} x_1 \partial^{s_2} x_2 \ldots \partial^{s_p} x_p}|| \leq M$ for some generic positive constant $M > 0$, where $s_1, \ldots, s_p$ can be any arbitrary sequence of non-negative integers such that $s_1 + \cdots + s_p = s$.

(4) Density $\mu(x)$ of x is continuous and bounded from below by a constant $c_\mu > 0$ for any $x \in \mathcal{X}$.

(5) $\mathbb{E}[\epsilon_i^q | X] < \infty$ for some $q \geq 4$.

(6) $\mu(x)$ is twice differentiable over all $x \in \mathcal{X}$.

Remark 2. *Assumptions 2 and 3 allow us to establish strong approximation results of $Q_{n,h} - \mathbb{E}[Q_{n,h}]$ and $\mathbb{E}_{n,h}[P^K(\frac{x-x_i}{h})y_i] - \mathbb{E}[\mathbb{E}_{n,h}[P^K(\frac{x-x_i}{h})y_i]]$. Such approximation is crucial in order to approximate $\sqrt{nh^p}(\hat{\beta}(x) - \beta(x))$ as a tight local-empirical process. Consequently, the integration by partition technique in Chernozhukov et al. (2018) can be used and therefore, CLT is established in the following Theorem 2.*

Theorem 2 (Inference of Non-parametric PMP)**.** *Assume that Assumptions 1–3 hold. Assume that $\delta_1^+, \ldots, \delta_p^+$ are fixed real numbers in $(0,1)$. Suppose $K \geq p$. Assume that h satisfies the growth conditions that $n^{1-\frac{2}{q}} h^p \to \infty$, $nh^{2K+2} \to 0$, $nh^{2p} \to \infty$. If $\widetilde{\mu}(x) = \mu(x)$, then the PMP defined in (4.7) satisfies:*

$$\sqrt{n}((\hat{\delta}_1^+, \ldots, \hat{\delta}_k^+) - (\delta_1^+, \ldots, \delta_k^+)) \rightsquigarrow N(0, diag(\sigma_1^2, \ldots, \sigma_p^2)) \tag{4.8}$$

where σ_j^2 is defined in (A.46), $j = 1, 2, \ldots, p$.

Remark 3. *Theorem 2 establishes the asymptotic distribution of $\sqrt{n}((\hat{\delta}_1^+, \ldots, \hat{\delta}_k^+) - (\delta_1^+, \ldots, \delta_k^+))$, which allows us to perform inference on $(\delta_1^+, \ldots, \delta_k^+)$. Though the derivatives are estimated locally, the PMP aggregates the information from the local derivatives and thus attains the $\sqrt{n}$ convergence rate.*

5. Extensions

The main results presented in the previous section can be extended to other PMP measures for different interests. Subsection 5.1 considers joint inference on PMP and the average treatment effect. Subsection 5.2 investigates PMP under a different measure than the measure of X.

5.1. Joint distribution with average effect.

Let $\hat{\beta} \coloneqq \mathbb{E}_n[\hat{g}(x_i)]$. We motivated the interest in the joint distribution of $\hat{\beta}$ and $\hat{\delta}^+$ in Example 2. We now provide the result.

Corollary 2. *Assume that all assumptions in Theorem 1 hold. Then, $r_n(\hat{\beta} - \beta)$ and $r_n(\hat{\delta}^+ - \delta^+)$ is jointly normal with covariance matrix defined in (A.50).*

5.2. Counterfactually-weighted PMP.

Define $\hat{\delta}^{Z+} \coloneqq \int_{x \in \mathcal{X}, z \in \mathcal{Z}} 1(\hat{g}(x) > 0) \widetilde{\mu}(z, x) dx dz$, where $\mathcal{Z}$ is the support of z, $\widetilde{\mu}(z, x)$ is the joint distribution of Z, X, and $\hat{g}$ is an estimator of g. One special case that we consider is that $x = z$ and $\widetilde{\mu}(z, x)$ reduces to $\widetilde{\mu}(z)$ where $\widetilde{\mu}$ may not be the same as μ. In such a case, we have the following corollary.

Corollary 3. *Assume that all assumptions in Theorem 1 hold. In addition, we assume that Assumption 1 holds for the joint support of X and Z, i.e., $\mathcal{X} \cup \mathcal{Z}$, and $\tilde{z}$ is known. Then,* $a_n(\hat{\delta}^{Z+} - \delta^{Z+}) \to N(0, \Omega_Z)$*, where Ω_Z is defined in* (A.53).

Remark 4. *Corollary 3 covers only the case where $\mathcal{Z} \subset \mathcal{X}$. When $\mathcal{Z}$ is not a subset of $\mathcal{X}$, generalizibility of estimator $\hat{g}$ can be problematic, and hence leads to bad estimation of δ^{Z+}.*

6. Simulations

Parametric example. Recall Example 1. We explore estimation of a "partial complements parameter" (PCP) by considering the following parametric specification:

$$f(k,l) = \beta_0 + \beta_1 l + \beta_2 l^2 + \beta_3 k + \beta_4 k^2 + \beta_5 lk + \beta_6 kl^2 + \beta_7 lk^2 + \beta_8 l^2 k^2$$

A similar specification was used in Chen, Chernozhukov, Fernández-Val, Kostyshak & Luo (2021) and allows for production functions that satisfy a variety of properties.[7] The Hessian matrix evaluated at l and k is

$$H(l,k) = \begin{bmatrix} 2\beta_2 + 2\beta_6 k + 2\beta_8 k^2 & \beta_5 + 2\beta_6 l + 2\beta_7 k + 4\beta_8 lk \\ \beta_5 + 2\beta_6 l + 2\beta_7 k + 4\beta_8 lk & 2\beta_4 + 2\beta_7 l + 2\beta_8 l^2 \end{bmatrix}.$$

The PCP thus equals the following:

$$\begin{aligned} \delta_{lk}^{+} &= \mathrm{P}\left[f_{lk}\left(L, K\right) > 0\right] \\ &= \mathrm{P}\left[\beta_5 + 2\beta_6 L + 2\beta_7 K + 4\beta_8 LK > 0\right]. \end{aligned}$$

We simulate data by drawing K and L independently from the uniform $[0,1]$ distribution, draw ϵ from $N(0,1)$, and construct y as $y = f(k,l) + \epsilon$. For the simulations we set $\beta_j = 0$ for $j = 0, 1, \ldots, 7$, and vary β_8 to vary the PCP. We show results for $\beta_8 \in \{-1.25, -1.5, -3.5\}$. We set the number of bootstrap replications to n (the sample size).

Table 1 shows results for 5 different types of bootstraps: the basic bootstrap, the percentile method, the normal approximation, the adjusted bootstrap percentile (BCa) method, and as a conservative approach the union of the percentile interval and the BCa interval. Where relevant, we intersect the interval with the parameter space, $[0,1]$, which does not affect coverage but reduces length.

For all three values of β_8, the simulations show that for large n the confidence intervals have good coverage and the length of the confidence intervals decreases. Because of finite-sample bias, the confidence interval has poor properties in some cases for small n.

[7] In Chen et al. (2021) a tensor product of third-degree global polynomials was used. Here, we use a tensor product of second-degree polynomials to keep the algebra (e.g., analytical derivation of PMP), concise.

Table 1. Monte Carlo Results for PCP

		n				
PCP (β_8)		100	200	500	1,000	2,500
1 (-1.25)	length	0.485	0.446	0.402	0.356	0.279
	covers	0.043	0.059	0.205	0.459	0.828
	mse	0.146	0.106	0.060	0.034	0.014
0.96 (-1.5)	length	0.487	0.449	0.412	0.374	0.305
	covers	0.096	0.148	0.339	0.603	0.866
	mse	0.134	0.100	0.058	0.034	0.014
0.55 (-3.5)	length	0.483	0.438	0.391	0.354	0.298
	covers	0.966	0.940	0.903	0.878	0.874
	mse	0.026	0.026	0.025	0.021	0.013

7. Conclusion

PMP generalizes monotonicity, providing researchers with a flexible way to test the empirical accuracy of several classes of predictions and to assess effect profiles, with both reduced-form and structural estimation methods. Our results accommodate estimators based on simple methods—such as OLS with polynomial and interaction terms—to non-parametric estimators such as local-linear regression. Semi-parametric methods are an attractive alternative because the dimension of the parameter space of the PMP is the same as in parametric estimation and does not increase as the dimension of the support increases. Extending the PMP to a multivariate context is simple which allows for practical applications where heterogeneous effects are driven by potentially several variables. Results for inference are provided and easy to apply. The PMP framework can be generalized even further in future research, to allow for multiple treatments and multiple outcome variables.

By adding PMP estimation to their toolbox, researchers can complement statistics that are sensitive to magnitude. The simple interpretation of PMP makes it easy to assess power (e.g., for a test of standard monotonicity), to report results concisely, and to convey results to a general audience.

Appendix A. Proofs

We first introduce some additional notation. For any positive integer r, $\mathcal{C}^r$ is the class of all differentiable functions whose derivative is in $\mathcal{C}^{r-1}$, where $\mathcal{C}^0$ is the class of all continuous functions.

We will require the following definitions.

Definition 1 (Manifold)**.** *Let d_k, d_x and r be positive integers such that $d_x \geq d_k$. Suppose that $\mathcal{M}$ is a subspace of $\mathbb{R}^{d_x}$ that satisfies the following property: for each point $m \in \mathcal{M}$, there is a set $\mathcal{V}$ containing m that is open in $\mathcal{M}$, a set $\mathcal{K}$ that is open in $\mathbb{R}^{d_k}$, and a continuous map*

$\alpha_m : \mathcal{K} \to \mathcal{V}$ carrying $\mathcal{K}$ onto $\mathcal{V}$ in a one-to-one fashion, such that: (1) α_m is of class $\mathcal{C}^r$ on $\mathcal{K}$, (2) $\alpha_m^{-1} : \mathcal{V} \to \mathcal{K}$ is continuous, and (3) the Jacobian matrix of α_m, $D\alpha_m(k)$, has rank d_k for each $k \in \mathcal{K}$. Then $\mathcal{M}$ is called a d_k-manifold without boundary in $\mathbb{R}^{d_x}$ of class $\mathcal{C}^r$. The map α_m is called a coordinate patch on $\mathcal{M}$ about m. A set of coordinate patches that covers $\mathcal{M}$ is called an atlas.

Definition 2 (Connected Branch)**.** *For any subset $\mathcal{M}$ of a topological space, if any two points m_1 and m_2 cannot be connected via path in $\mathcal{M}$, then we say that m_1 and m_2 are not connected. Otherwise, we say that m_1 and m_2 are connected. We say that $\mathcal{V} \subseteq \mathcal{M}$ is a connected branch of $\mathcal{M}$ if all points of $\mathcal{V}$ are connected to each other and do not connect to any points in $\mathcal{M} \setminus \mathcal{V}$.*

Definition 3 (Volume)**.** *For a $d_x \times d_k$ matrix $A = (x_1, x_2, \ldots, x_{d_k})$ with $x_i \in \mathbb{R}^{d_x}$, $1 \leq i \leq d_k \leq d_x$, let $\mathrm{Vol}(A) = \sqrt{\det(A'A)}$, which is the volume of the parallelepiped $P(A)$ with edges given by the columns of A, $P(A) = \{c_1 x_1 + \cdots + c_{d_k} x_{d_k} : 0 \leq c_i \leq 1, i = 1, \ldots, d_k\}$.*

The volume measures the amount of mass in $\mathbb{R}^{d_k}$ of a d_k-dimensional parallelepiped in $\mathbb{R}^{d_x}$, $d_k \leq d_x$. This concept is essential for integration on manifolds, which we will discuss shortly. First we recall the concept of integration on parameterized manifolds:

Definition 4 (Integration on a parametrized manifold)**.** *Let $\mathcal{K}$ be open in $\mathbb{R}^{d_k}$, and let $\alpha : \mathcal{K} \to \mathbb{R}^{d_x}$ be of class $\mathcal{C}^r$ on $\mathcal{K}$, $r \geq 1$. The set $\mathcal{M} = \alpha(\mathcal{K})$ together with the map α constitute a parametrized d_k-manifold in $\mathbb{R}^{d_x}$ of class $\mathcal{C}^r$. Let g be a real-valued continuous function defined at each point of $\mathcal{M}$. The integral of g over $\mathcal{M}$ with respect to volume is defined by*

$$\int_{\mathcal{M}} g(m) d\mathrm{Vol} := \int_{\mathcal{K}} (g \circ \alpha)(k) \mathrm{Vol}(D\alpha(k)) dk, \tag{A.1}$$

provided that the right side integral exists. Here $D\alpha(k)$ is the Jacobian matrix of the mapping $k \mapsto \alpha(k)$, and $\mathrm{Vol}(D\alpha(k))$ is the volume of matrix $D\alpha(k)$ as defined in Definition 3.

The above definition coincides with the usual interpretation of integration. The integral can be extended to manifolds that do not admit a global parametrization α using the notion of partition of unity. This partition is a set of smooth local functions defined in a neighborhood of the manifold. The following Lemma shows the existence of the partition of unity and is proven in Lemma 25.2 in Munkres (1991).

A.1. **Proof of Theorem 1.**

Proof of Theorem 1. (1) Given the conditions stated in Theorem 1, Theorem 4.1 of Chernozhukov et al. (2018) shows that for each $j = 1, 2, \ldots, p$,

$$\hat{\delta}_j^+ := \sum_{i=1}^n 1(\hat{g}_j(x_i) > 0) \rightsquigarrow s_g T_{j,\infty}(0) + s_\mu H_\infty(g_{j,0}), \tag{A.2}$$

where $T_{j,\infty}(0) := \int_{\mathcal{M}_{g_j}(0)} \frac{G_{j,\infty}(x)\mu'(x)}{\|\partial g_j(x)\|} d\mathrm{Vol}$, $G_{j,\infty}$ is the j^{th} component of G_∞, and $g_{j,0} := 1(g_j(x) > 0)$.

Since we can apply (A.2) to $j = 1, 2, \ldots, p$, therefore, the convergence law stated in (A.2) holds for the entire vector of g. More specifically,

$r_n((\hat{\delta}_1^+, \ldots, \hat{\delta}_p^+) - (\delta_1^+, \ldots, \delta_p^+)) \rightsquigarrow (Z_1, \ldots, Z_p)$, where

$$Z_j := s_g T_{j,\infty}(0) + s_\mu H_\infty(g_{j,0}). \tag{A.3}$$

Since G and H are jointly Gaussian, define

$$\Omega_1 := \mathrm{Var}(Z_1, \ldots, Z_p), \tag{A.4}$$

therefore, $r_n((\hat{\delta}_1^+, \ldots, \hat{\delta}_p^+) - (\delta_1^+, \ldots, \delta_p^+)) \rightsquigarrow N(0, \Omega_1)$.

(2) We can define $g_{min}(x) := \min_{1 \le j \le k} g_j(x)$. Therefore, $1(g(x) > 0) = 1(g_{min}(x) > 0)$. The benefit of doing so is that the value of $g_{min}(x)$ is one-dimensional.

By assumption, $\mathcal{M}_{g_j}(0)$, $j = 1, 2, \ldots, p$, do not intersect each other with positive $d_x - 1$ dimensional volume. Hence, $g_{min}(x)$ is differentiable almost everywhere around $\mathcal{M}_{g_{min}}(0)$, and $\mathcal{M}_{g_{min}}(0)$ can be divided into a finite collection of $p-1$ dimensional manifold.

Denote $\mathcal{M} := \cup_{j=1}^p \mathcal{M}^*_{g_j}(0)$, where $\mathcal{M}^*_{g_j}(0)$ is a submanifold of $\mathcal{M}_{g_j}(0)$, and $\mathcal{M}^*_{g_j}(0)$ are disjoint of each other.

Therefore, $r_n(\hat{\delta}^+ - \delta^+) \rightsquigarrow \sum_{j=1}^p s_g \int_{\mathcal{M}^*_{g_j}(0)} \frac{G_{j,\infty}(x)\mu'(x)}{\|\partial g_j(x)\|} d\mathrm{Vol} + s_\mu H_\infty(g_{min,0})$, where $g_{min,0} := 1(g_{min}(x) > 0)$. Since G_∞ and H_∞ are jointly Gaussian, therefore, $r_n(\hat{\delta}^+ - \delta^+)$ converges to a Gaussian distribution

$$Z_\infty := \sum_{j=1}^p s_g \int_{\mathcal{M}^*_{g_j}(0)} \frac{G_{j,\infty}(x)\mu'(x)}{\|\partial g_j(x)\|} d\mathrm{Vol} + s_\mu H_\infty(g_{min,0}). \tag{A.5}$$

□

Proof of Corollary 1. Similar to the proof of Theorem 1, the proof of this corollary follows Theorem 4.3 of Chernozhukov et al. (2018). Therefore, we abbreviate the results. □

A.2. **Proof of Theorem 2.** Define

$$Q_h(x) := \frac{1}{h^p} \mathbb{E}\left[K\left(\frac{x_i - x}{h}\right) P^K(\frac{x - x_i}{h})(P^K(\frac{x - x_i}{h}))^\intercal \right].$$

It is easy to know that

$$Q_h(x) \to Q(x) := \mu(x) Q_K, \tag{A.6}$$

where $Q_K := \int_{||x_i - x|| \le 1} K(x_i - x) P^K(\frac{x - x_i}{h})(P^K(\frac{x - x_i}{h}))^\intercal$ is a fixed positive definite matrix given $\mu(x) > 0$.

More specifically, by assumption, we have:

$$\|Q_h(x) - Q(x)\| \leq C_Q h^2 \tag{A.7}$$

for some generic constant $C_Q > 0$ given that $\mu(x)$ is twice-continuously differentiable.

Before we prove Theorem 2, we need the following lemma.

Lemma 1 (Strong approximation of $Q_{n,h}(x)$)**.** *Assume that Assumptions 2 and 3 hold. Assume that $h = Cn^{-\gamma}$ for some fixed $C, \gamma > 0$, and $1 - \gamma p > 0$. Assume that $\mathcal{X}$ is compact. We have:*

$$\sqrt{nh^p} \sup_{x \in \mathcal{X}} ||Q_{n,h} - Q_h|| = O_P\left((nh^p)^{-\frac{1}{6}} \ln n\right). \tag{A.8}$$

Proof. Since $x_i \in \mathcal{X}$ is bounded, $i = 1, 2, \ldots, n$, by Chernozhukov et al. (2014), $W_n := \sqrt{nh^p}(Q_{n,h} - Q_h)$ is can be approximated by a tight local empirical process. The conclusion of the lemma follows directly from Proposition 3.1 of Chernozhukov et al. (2014). □

Proof of Theorem 2. For any $j \in \{1, 2, \ldots, p\}$, by definition, we know that

$$\delta_j^+ = \int_{x \in \mathcal{X}} 1(\nabla f_j(x) > 0) d\mu(x).$$

By Assumption 3, for any x' that is close enough to x and any $f(x) = f_j(x)$, we have:

$$||f(x') - f(x) - \nabla f(x)(x' - x) - P_1^K(x' - x)|| \leq C||x' - x||^{K+1}, \tag{A.9}$$

where $P_1^K(x' - x)$ is a polynomial of $(x' - x)$ with degree ranging from 2 to K^{th} order with all coefficients $\leq M$.

Denote $r_1(x', x) := f(x') - f(x) - \nabla f(x)(x' - x) - P_1^K(x' - x)$. So we have that $|r_1(x)| \leq C\|x' - x\|^{K+1}$.

(4.1) implies that:

$$(\hat{\alpha}(x), \hat{\beta}(x), \hat{\beta}_H(x))^\intercal = Q_{n,h}^{-1} \mathbb{E}_{n,h}[P^K(\frac{x - x_i}{h})(f(x) + \nabla f(x) \frac{(x_i - x)}{h} + P_1^K(x_i - x) + r_1(x_i, x) + \epsilon_i)] \tag{A.10}$$

$$= (f(x), \nabla f(x), f_H(x))^\intercal + Q_{n,h}^{-1} \mathbb{E}_{n,h}[P^K(\frac{x_i - x}{h})(r_1(x_i, x) + \epsilon_i)], \tag{A.11}$$

where $f_H(x)$ is some function of x bounded by M.

Define $S_n(x) := \sqrt{nh^p} \mathbb{E}_{n,h}[P^K(\frac{x_i - x}{h})\epsilon_i]$. It is easy to see that $\mathbb{E}[S_n(x)] = 0$ for all $x \in \mathcal{X}$ since $\mathbb{E}[\epsilon_i | x_i] = 0$. Define $W_n(x) = \sup_{x \in \mathcal{X}} |S_n(x)|$.

Under our assumptions, by Proposition 3.2 of Chernozhukov et. al (2017), $S_n(x)$ can be well approximated by a local Gaussian process. More specifically, there exists a tight Gaussian process $B_n(x), x \in \mathcal{X}$ such that $\widetilde{W}_n(x) := \sup_{x \in \mathcal{X}} B_n(x)$ and

$$|\widetilde{W}_n(x) - W_n(x)| = O_P\left((nh^p)^{-\frac{1}{6}} \log n + (nh^p)^{-\frac{1}{4}} \log^{\frac{5}{4}} n + (n^{1-2/q} h^p)^{-\frac{1}{2}} \log^{\frac{3}{2}} n\right). \tag{A.12}$$

Choose h such that $n^{1-2/q}h^p \log^{-3} n \to \infty$ and $nh^p \log^{-6} n \to \infty$, then

$$|\widetilde{W}_n(x) - W_n(x)| = O_P\left(\max\{(nh^p)^{-\frac{1}{6}} \log n, (n^{1-2/q}h^p)^{-\frac{1}{2}} \log^{\frac{3}{2}} n\}\right) = o_p(1). \tag{A.13}$$

That said, $W_n(x) = O_p(1)$, since $B_n(x)$ is a tight Gaussian process.

Denote $L_n(x) := Q_{n,h}^{-1}\mathbb{E}_{n,h}[(1, \frac{x_i-x}{h})^\intercal r_1(x_i, x)]$.

Since the kernel $K(\cdot)$ is r^{th} order, it is easy to know that

$$\|\mathbb{E}[\mathbb{E}_{n,h}[P^K(\frac{x_i - x}{h})r_1(x_i - x)]]\| \tag{A.14}$$
$$\leq M\sqrt{p+1}h^{K+1}. \tag{A.15}$$

for some constant $C > 0$.

Define $W_n^L(x) := \sup_{x\in\mathcal{X}} \|\sqrt{nh^p}(\mathbb{E}_{n,h}[P^K(\frac{x_i-x}{h})r_1(x_i - x)] - \mathbb{E}[\mathbb{E}_{n,h}[P^K(\frac{x_i-x}{h})r_1(x_i - x)]])\|$.

Similar to W_n, we have that $W_n^L(x) = O_p(1)$.

Therefore, $\sup_{x\in\mathcal{X}} |\hat{\beta}(x) - \beta(x)| \leq O_p(\sqrt{p+1}h^{K+1}) + \|Q_{n,h}\|^{-1}\frac{1}{\sqrt{nh^p}} \sup_{x\in\mathcal{X}} |W_n(x)| = O_P(h^{K+1} + \frac{1}{\sqrt{nh^p}})$.

Denote $t_n := \frac{1}{\sqrt{nh^p}}$, define:

$$G_n(x) := S_n(x) + t_n^{-1}L_n(x).$$

We have that $\sup_{x\in\mathcal{X}} \|G_n(x)\| \leq \sup_{x\in\mathcal{X}} \|S_n(x)\| + \sup_{x\in\mathcal{X}} \|\sqrt{nh^p}L_n(x)\| = O_p(1)$.

Denote $\Delta_j(x) := \nabla f_j(x)$, $j = 1, 2, \ldots, p$.

Given that $G_n(x) = O_p(1)$, Lemma E.4 of Chernozhukov et al. (2018) shows that, for continuous function $\widetilde{\mu}(x)$, $t_n \to 0$ and $G_n(x)$ as a bounded element,

$$\frac{\int_{\mathcal{X}}[1(\Delta_j(x) + t_nG_n(x) > 0) - 1(\Delta_j(x) > 0)]\widetilde{\mu}(x)d_x}{t_n} = -\int_{\mathcal{M}_{\Delta_j}} \frac{\widetilde{\mu}(x)G_n(x)d\mathrm{Vol}}{||\nabla\Delta_j(x)||} + o(1). \tag{A.16}$$

Moreover, when $g_j(x)$ is twice differentiable, i.e., $\Delta_j(x)$ is differentiable, then we have:

$$\frac{\int_{\mathcal{X}}[1(\Delta_j(x) + t_nG_n(x) > 0) - 1(\Delta_j(x) > 0)]\widetilde{\mu}(x)d_x}{t_n} = -\int_{\mathcal{M}_{\Delta_j}} \frac{\widetilde{\mu}(x)G_n(x)d\mathrm{Vol}}{||\nabla\Delta_j(x)||} + O(t_n). \tag{A.17}$$

Denote $G_n = (G_{n,0}, G_{n,1}, \ldots, G_{n,p})^\intercal$ where $G_{n,j}$ is the component of G_n that corresponds to $\hat{\beta}_j$, $j = 1, 2, \ldots, p$. Similarly, we call $S_{n,j}$, $L_{n,j}$ as the component in S_n, L_n that correspond to $\hat{\beta}_j$, respectively.

Given that $\widetilde{\mu} = \mu$, by (A.17), with probability going to one, we have:

$$\hat{\delta}_j^+ - \delta_j^+ \tag{A.18}$$

$$= t_n \frac{\int_{x \in \mathcal{X}} [1(\Delta_j(x) + t_n G_{n,j}(x) > 0) - 1(\Delta_j(x) > 0)]\mu(x)dx}{t_n} \tag{A.19}$$

$$= -t_n \int_{\mathcal{M}_{\Delta_j}} \frac{\mu(x) G_{n,j}(x) d\text{Vol}}{||\nabla \Delta_j(x)||} + O(t_n^2) \tag{A.20}$$

$$= -t_n \int_{\mathcal{M}_{\Delta_j}} \frac{\mu(x) S_{n,j}(x) d\text{Vol}}{||\nabla \Delta_j(x)||} + L_{n,0,j} + O(t_n^2), \tag{A.21}$$

where

$$L_{n,0} = -t_n \int_{\mathcal{M}_{\Delta_j}} \frac{\mu(x)\sqrt{nh^p} L_n(x) d\text{Vol}}{||\nabla \Delta_j(x)||} \tag{A.22}$$

$$= - \int_{\mathcal{M}_{\Delta_j}} \frac{\mu(x) Q_{n,h}^{-1} \mathbb{E}_{n,h}[P^K(\frac{x_i - x}{h}) r_1(x_i, x)] d\text{Vol}}{||\nabla \Delta_j(x)||} \tag{A.23}$$

$$= -\mathbb{E}_{n,h} \int_{\mathcal{M}_{\Delta_j}} \frac{\mu(x) P^K(\frac{x_i - x}{h}) r_1(x_i, x) d\text{Vol}}{||\nabla \Delta_j(x)||} \tag{A.24}$$

We analyze $S_{n,0}$ first in (A.21).

Notice that in (A.21), denote the leading term for $j = 1, 2, \ldots, p$ in (A.21) as

$$S_{n,0,j} := -t_n \int_{\mathcal{M}_{\Delta_j}} \frac{\mu(x) S_{n,j}(x) d\text{Vol}}{||\nabla \Delta_j(x)||}.$$

Denote $Q_j^{-1}(x), Q_{n,h,j}^{-1}(x)$ as the $j+1^{th}$ row of $Q^{-1}(x), Q_{n,h}^{-1}(x)$ respectively, $j = 1, 2, \ldots, p$. We can rewrite $S_{n,0,j}$ as:

$$S_{n,0,j} = -t_n \int_{\mathcal{M}_{\Delta_j}} \frac{\mu(x) S_{n,j}(x) d\text{Vol}}{||\nabla \Delta_j(x)||} \tag{A.25}$$

$$= - \int_{\mathcal{M}_{\Delta_j}} \frac{\mu(x) Q_{n,h,j}^{-1}(x) \mathbb{E}_{n,h}[P^K(\frac{x_i - x}{h}) \epsilon_i] d\text{Vol}}{||\nabla \Delta_j(x)||} \tag{A.26}$$

$$= \frac{1}{n} \sum_{i=1}^{n} (W_{n,0,j}(i)\epsilon_i + W_{n,1,j}(i)\epsilon_i), \tag{A.27}$$

where

$$W_{n,0,j}(i) := - \int_{\mathcal{M}_{\Delta_j}} \frac{h^{-p} \mu(x) Q_j(x)^{-1} K((x_i - x)/h) P^K(\frac{x_i - x}{h})] d\text{Vol}}{||\nabla \Delta_j(x)||} \tag{A.28}$$

is an i.i.d. random variable that depends on x_i only and it is uncorrelated with ϵ_i, and

$$W_{n,1,j}(i) := - \int_{\mathcal{M}_{\Delta_j}} \frac{h^{-p} \mu(x) (Q_{n,h,j}(x)^{-1} - Q_j^{-1}(x)) P^K(\frac{x_i - x}{h}) d\text{Vol}}{||\nabla \Delta_j(x)||} \tag{A.29}$$

is a residual term that is also uncorrelated with ϵ_i.

It is easy to know that by law of iterated expectations, we have:

$$\mathbb{E}[W_{n,l,j}(i)\epsilon_i] = \mathbb{E}[\mathbb{E}[W_{n,l,j}(i)\mathbb{E}[\epsilon_i|X]] = 0,$$

for $l = 0, 1$.

By Lemma 1,

$$\sup_{x\in\mathcal{X}} ||Q^{-1}_{n,h}(x) - Q^{-1}_h(x)|| \precsim_p (nh^p)^{-\frac{2}{3}} \log n.$$

Therefore,

$$\sup_{x\in\mathcal{X}} \|Q^{-1}_{n,h,j}(x) - Q^{-1}_j(x)\| \leq \sup_{x\in\mathcal{X}} \|Q^{-1}_{n,h,j}(x) - Q^{-1}_{h,j}(x)\| + \|Q^{-1}_{h,j}(x) - Q^{-1}_j(x)\| \tag{A.30}$$

$$\precsim_p (nh^p)^{-\frac{2}{3}} \log n + h^2. \tag{A.31}$$

That said, when $h = Cn^{-\gamma}$ such that $1 - p\gamma > 0$ and $\gamma > 0$, $\sup_{x\in\mathcal{X}} \|Q^{-1}_{n,h,j} - Q^{-1}_j(x)\| = o_p(1)$. Therefore, $\frac{1}{n}\sum_{i=1}^n W_{n,1,j}(i)\epsilon_i$ must be dominated by $\frac{1}{n}\sum_{i=1}^n W_{n,0,j}(i)\epsilon_i$, since the second moment of $\frac{1}{n}\sum_{i=1}^n W_{n,1,j}(i)\epsilon_i$ is of higher order of the second moment of $\frac{1}{n}\sum_{i=1}^n W_{n,0,j}(i)\epsilon_i$.

Now, let's focus on $W_{n,0,j}(i)$.

Recall that

$$W_{n,0,j}(i) = -\int_{\mathcal{M}_{\Delta_j}} \frac{h^{-p}K((x_i - x)/h)Q^{-1}_{K,1,j}P^K(\frac{x_i-x}{h})d\text{Vol}}{||\nabla\Delta_j(x)||}, \tag{A.32}$$

where $Q^{-1}_{K,1,j}$ is the $j+1^{th}$ row of Q^{-1}_K that correspond to β_j.

Define $\mathcal{M}_{\Delta_j}(h)$ as

$$\mathcal{M}_{\Delta_j}(h) := \{x \in \mathcal{X} | d_H(x, \mathcal{M}_{\Delta_j}) \leq h\}.$$

Notice that $W_{n,0,j} = 0$ if $x_i \notin \mathcal{M}_{\Delta_j}(h)$.

Denote $B_h(x)$ as the ball centered at x with radius h.

Therefore,

$$\mathbb{E}[W_{n,0,j}(i)^2] = h^{-2p}\mathbb{E}[\left(\int_{B_h(x_i)\cap\mathcal{M}_{\Delta_j}} \frac{K(\frac{x_i-x}{h})Q^{-1}_{K,1,j}P^K(\frac{x_i-x}{h})d\text{Vol}}{||\nabla\Delta_j(x)||}\right)^2] \tag{A.33}$$

$$= h^{-2p}\int_{x\in\mathcal{M}_{\Delta_j}(h)} \mu(x_i)\left(\int_{B_h(x_i)\cap\mathcal{M}_{\Delta_j}} \frac{K(\frac{x_i-x}{h})Q^{-1}_{K,1,j}P^K(\frac{x_i-x}{h})d\text{Vol}}{||\nabla\Delta_j(x)||}\right)^2 dx_i. \tag{A.34}$$

Denote $K_{p-1}(z) := \int_{x'=x+w, w^\intercal(x'-x)=0, \|w\|=z} K(x' - x)dx'$ as the integration of kernel $K(\cdot)$ on a $p-1$ dimensional hyperplane with distance to 0 equals to $z \geq 0$, for fixed w with $\|w\| = z$. It is easy to know that $\int_{z=0}^1 K_{p-1}(z)dz = \frac{1}{2}$ by definition of kernel $K(\cdot)$.

For any $x_i \in \mathcal{M}_{\Delta_j}(h)$, we can define the projection of x_i on $\mathcal{M}_{\Delta_j}$ as x_i^0. Locally, $\mu(x_i) = \mu(x_i^0) + O(h)$ by differentiability of $\mu(\cdot)$. Define the tangent space of $\mathcal{M}_{\Delta_j}$ at x_i^0 as $T(x_i^0)$. Then,

$$\int_{B_h(x_i)\cap\mathcal{M}_{\Delta_j}} \frac{K(\frac{x_i-x}{h})Q_{K,1,j}^{-1}P^K(\frac{x_i-x}{h})d\text{Vol}}{||\nabla\Delta_j(x)||} \tag{A.35}$$

$$= \int_{B_h(x_i)\cap\mathcal{M}_{\Delta_j}} \frac{K(\frac{x_i-x}{h})Q_{K,1,j}^{-1}P^K(\frac{x_i-x_i^0}{h})d\text{Vol}}{||\nabla\Delta_j(x)||} \tag{A.36}$$

$$+ \int_{B_h(x_i)\cap\mathcal{M}_{\Delta_j}} \frac{K(\frac{x_i-x}{h})Q_{K,1,j}^{-1}P^K(\frac{x_i-x_i^0}{h})d\text{Vol}}{||\nabla\Delta_j(x)||} \tag{A.37}$$

In (A.36), we know that

$$\begin{aligned}&\int_{B_h(x_i)\cap\mathcal{M}_{\Delta_j}} \frac{K(\frac{x_i-x}{h})Q_{K,1,j}^{-1}P^K(\frac{x_i-x_i^0}{h})d\text{Vol}}{||\nabla\Delta_j(x)||}\\ &= h^p(Q_{K,1,j}^{-1})^\intercal\nabla\Delta_j(x_i^0)\left(K_{p-1}(\|x_i-x_i^0\|/h)\frac{(x_i-x_i^0)/h}{||\nabla\Delta_j(x_i^0)\|^2}+O(h)\right).\end{aligned} \tag{A.38}$$

And in (A.37), since the kernel is symmetric, we have:

$$\int_{B_h(x_i)\cap\mathcal{M}_{\Delta_j}} \frac{K(\frac{x_i-x}{h})Q_{K,1,j}^{-1}P^K(\frac{x_i-x_i^0}{h})d\text{Vol}}{||\nabla\Delta_j(x)||} = O(h^p). \tag{A.39}$$

Plugging in (A.38) and (A.39) back to (A.34), we have:

$$h^{-2p}\int_{x\in\mathcal{M}_{\Delta_j}(h)}\mu(x_i)\left(\int_{B_h(x_i)\cap\mathcal{M}_{\Delta_j}} \frac{K(\frac{x_i-x}{h})Q_{K,1,j}^{-1}P^K(\frac{x_i-x_i^0}{h})d\text{Vol}}{||\nabla\Delta_j(x)||}\right)^2 dx_i \tag{A.40}$$

$$= h^{-2p}2\int_{z:=\|x_i-x_i^0\|/h\in[0,1],x_i^0\in\mathcal{M}_{\Delta_j}}(1+o(1))\mu(x_i^0)h^{2p}(1+o(1))\frac{K_{p-1}(z)^2z^2((Q_{K,1,j}^{-1})^\intercal\nabla\Delta_j(x_i^0))^2}{||\nabla\Delta_j(x_i^0)||^4}d\text{Vol} \tag{A.41}$$

$$\to K_{2,p-1}\int_{x_i^0\in\mathcal{M}_{\Delta_j}}\frac{\mu(x_i^0)((Q_{K,1,j}^{-1})^\intercal\nabla\Delta_j(x_i^0))^2d\text{Vol}}{||\nabla\Delta_j(x_i^0)||^4}, \tag{A.42}$$

where $K_{2,p-1} := 2\int_{z\in[0,1]}K_{p-1}(z)^2z^2$ is a generic constant that only depends on $K(\cdot)$.

Hence, by the Central Limit Theorem,

$$\sqrt{n}(\frac{1}{n}\sum_{i=1}^n W_{n,0,j}(i)\epsilon_i) \to_d N(0, K_{2,p-1}\int_{x_i^0\in\mathcal{M}_{\Delta_j}}\frac{\mu(x_i^0)((Q_{K,1,j}^{-1})^\intercal\nabla\Delta_j(x_i^0))^2d\text{Vol}}{||\nabla\Delta_j(x_i^0)||^4}\sigma^2), \tag{A.43}$$

where $Var(\epsilon_i) = \sigma^2$. That said, by (A.21), we have:

$$\hat{\delta}_j^+ - \delta_j^+ = O_p(\frac{1}{\sqrt{n}} + h^{K+1} + \frac{1}{nh^p}). \tag{A.44}$$

For the asymptotic distribution of $\delta_j^+ - \delta_j$, if: (1) $\sqrt{n}h^{K+1} \to 0$ (2) $\frac{1}{\sqrt{n}h^p} \to 0$, then,

$$\sqrt{n}(\hat{\delta}_j^+ - \delta_j^+) \to_d N(0, \sigma_j^2), \tag{A.45}$$

where

$$\sigma_j^2 := K_{2,p-1} \int_{x_i^0 \in \mathcal{M}_{\Delta_j}} \frac{\mu(x_i^0)((Q_{K,1,j}^{-1})^\intercal \nabla \Delta_j(x_i^0))^2 d\text{Vol}}{||\nabla \Delta_j(x_i^0)||^4} \sigma^2. \tag{A.46}$$

Assuming that $\mathcal{M}_{\Delta_j}$, $j = 1, 2, \ldots, p$, do not intersect each other with positive volume. Then, as $h \to 0$, the correlation between $W_{n,0,j}$ and $W_{n,0,j'}$ for $j \neq j'$ must go to 0 as

$$\mathbb{P}(x_i \in \mathcal{M}_{\Delta_j}(h) \cap \mathcal{M}_{\Delta_{j'}}(h)) = o(h).$$

Therefore, $\sqrt{n}(\hat{\delta}_j^+ - \delta_j^+)$ is uncorrelated with each other as $h \to 0$, for any pairs $j = j_1, j_2$. That said,

$$\sqrt{n}(\hat{\delta}^+ - \delta^+) \to_d N(0, diag(\sigma_1^2, \ldots, \sigma_p^2)). \tag{A.47}$$

□

A.3. **Proofs of Section 5.**

Proof of Corollary 2. We know that

$$\hat{\beta} - \beta = \int_x \hat{g}(x)\hat{\mu}(x)dx - \int_x g(x)\mu(x)dx \tag{A.48}$$

$$= \int_x (\hat{g}(x) - g(x))\mu(x)dx + \int_x g(x)(\hat{\mu}(x) - \mu(x))dx + \int_x (\hat{g}(x) - g(x))(\hat{\mu}(x) - \mu(x))dx. \tag{A.49}$$

By Assumption 1, the third term in (A.49) is negligible compared to the first two terms, and we have: $r_n(\hat{\beta} - \beta) \to_d s_g \int_x G_\infty(x)\mu(x)dx + s_\mu H_\infty(g(\cdot))$. □

Therefore, define

$$\Omega_{Joint} := Var(Z_1, \ldots, Z_p, s_g \int_x G_\infty(x)\mu(x)dx + s_\mu H_\infty(g(\cdot))), \tag{A.50}$$

where Z_j is defined in (A.3), we have:

$$r_n(\hat{\delta}^+ - \delta^+, \hat{\beta} - \beta) \to_d N(0, \Omega_{Joint}). \tag{A.51}$$

Proof of Corollary 3. We can apply similar techniques of Theorem 1 to this corollary, by replacing μ with $\widetilde{\mu}$. Notice that in this case, $\hat{\mu} = \mu = \widetilde{\mu}$. Therefore,

$$a_n(\hat{\delta}^{Z+} - \delta^{Z+}) \to_d (W_1, \ldots, W_p), \tag{A.52}$$

where $W_j := \int_{x \in \mathcal{M}_{g_j}} \frac{G_{j,\infty}(x)\widetilde{\mu}(x) d\text{Vol}}{\|\nabla g_j(x)\|}$, $j = 1, 2, \ldots, p$. Define

$$\Omega_Z := \text{Var}(W_1, \ldots, W_p), \tag{A.53}$$

then, $a_n(\hat{\delta}^{Z+} - \delta^{Z+}) \to_d N(0, \Omega_Z)$. □

References

Banks, J., Blundell, R. & Lewbel, A. (1997), 'Quadratic engel curves and consumer demand', *Review of Economics and statistics* **79**(4), 527–539.

Bhattacharya, D. (2021), 'The empirical content of binary choice models', *Econometrica* **89**(1), 457–474.

Blundell, R. W., Browning, M. & Crawford, I. A. (2003), 'Nonparametric engel curves and revealed preference', *Econometrica* **71**(1), 205–240.

Chen, X., Chernozhukov, V., Fernández-Val, I., Kostyshak, S. & Luo, Y. (2021), 'Shape-enforcing operators for point and interval estimators'.

Chernozhukov, V., Chetverikov, D. & Kato, K. (2014), 'Gaussian approximation of suprema of empirical processes', *The Annals of Statistics* **42**(4), 1564–1597.

Chernozhukov, V., Fernández-Val, I. & Luo, Y. (2018), 'The sorted effects method: Discovering heterogeneous effects beyond their averages', *Econometrica* **86**(6), 1911–1938.

Chetverikov, D. (2019), 'Testing regression monotonicity in econometric models', *Econometric Theory* **35**(4), 729–776.

Chetverikov, D., Santos, A. & Shaikh, A. M. (2018), 'The econometrics of shape restrictions', *Annual Review of Economics* **10**, 31–63.

Dette, H., Hoderlein, S. & Neumeyer, N. (2016), 'Testing multivariate economic restrictions using quantiles: The example of slutsky negative semidefiniteness', *Journal of Econometrics* **191**(1), 129–144.

Du, P., Parmeter, C. & Racine, J. (2013), 'Nonparametric kernel regression with multiple predictors and multiple shape constraints', *Statistica Sinica* pp. 1347–1371.

Du, P., Parmeter, C. & Racine, J. (2021), 'Shape constrained kernel pdf and pmf estimation', *Available at SSRN 3812484* .

Epstein, L. G. & Yatchew, A. J. (1985), 'Non-parametric hypothesis testing procedures and applications to demand analysis', *Journal of Econometrics* **30**(1), 149–169.

Fang, Z. & Seo, J. (2021), 'A projection framework for testing shape restrictions that form convex cones', *arXiv preprint arXiv:1910.07689* .

Gallant, A. & Souza, G. (1991), 'On the asymptotic normality of fourier flexible form estimates', *Journal of Econometrics* **50**(3), 329–353.

Härdle, W., Hildenbrand, W. & Jerison, M. (1991), 'Empirical evidence on the law of demand', *Econometrica* **59**(6), 1525–1549.

Härdle, W. & Stoker, T. M. (1989), 'Investigating smooth multiple regression by the method of average derivatives', *Journal of the American Statistical Association* **84**(408), 986–995.

Hoderlein, S. (2011), 'How many consumers are rational?', *Journal of Econometrics* **164**(2), 294–309.

Lewbel, A. (1995), 'Consistent nonparametric hypothesis tests with an application to slutsky symmetry', *Journal of Econometrics* **67**(2), 379–401.

Munkres, J. R. (1991), *Analysis on manifolds*, Westview Press.

Powell, J. L., Stock, J. H. & Stoker, T. M. (1989), 'Semiparametric estimation of index coefficients', *Econometrica* **57**(6), 1403–1430.

Storesletten, K., Telmer, C. I. & Yaron, A. (2004), 'Consumption and risk sharing over the life cycle', *Journal of Monetary Economics* **51**(3), 609–633.

Van Landeghem, B. (2012), 'A test for the convexity of human well-being over the life cycle: Longitudinal evidence from a 20-year panel', *Journal of Economic Behavior & Organization* **81**(2), 571–582.