A Test for the Practical Relevance of Endogeneity in Semiparametric Distribution Regression

Master's Thesis

Jonathan Goedeke

University of Cologne

Faculty of Management, Economics and Social Sciences

Specialization Empirical Methods and Data Analysis

Supervisor: Prof. Dr. Dominik Wied

2026

Abstract

We develop a test for the *practical* relevance of endogeneity in semiparametric distribution regression (DR). Classical tests ask whether the endogenous regressor is *exactly* exogenous, a question that is uninformative in large samples because even negligible deviations are detected with probability approaching one. We instead test whether correcting for endogeneity changes the estimated conditional distribution by more than an economically meaningful threshold $\Delta > 0$. The object of interest is the L^2-distance $d(x, y_2) = \|F^{\mathrm{N}}(\cdot|x, y_2) - F^{\mathrm{V}}(\cdot|x, y_2)\|_2$ between the ordinary DR estimator and the control-function IV-DR estimator of Wied (2024). The relevant null hypothesis is $H_0 \colon d(x, y_2) \leq \Delta$ against $H_1 \colon d(x, y_2) > \Delta$.

We derive the joint influence-function representation of the naive and IV-DR estimators—both estimated from the same sample—state a high-level sequential asymptotic linearity condition that yields the uniform weak convergence of the sequential process of their difference, and adapt the self-normalization device of Dette et al. (2025) to this one-sample setting. The resulting test statistic is asymptotically pivotal: its limiting distribution under the boundary $d = \Delta$ is that of a ratio of Brownian-motion functionals, whose quantiles can be obtained by simulation once and for all. The test is consistent and detects local alternatives at rate $n^{-1/2}$.

Contents

1 Introduction

A central question in applied microeconometrics is whether a regressor of interest is endogenous, and if so, whether this endogeneity materially distorts inference. The standard approach is to run a Hausman-type test of exact exogeneity and proceed with IV estimation only if the null is rejected. This strategy has a well-known weakness: in large samples, even economically negligible deviations from exogeneity are detected with probability approaching one (Berkson, 1938; Hodges and Lehmann, 1954). The econometrician is then forced to adopt an IV estimator that may be far less precise than the ordinary least squares alternative, at a cost that outweighs the bias removed.

The appropriate question is not *whether* endogeneity is exactly zero but *whether it is large enough to matter practically.* This perspective is familiar from the equivalence-testing and minimum-detectable-effect literatures (Wellek, 2010; Lakens, 2017), but it has received little attention in the context of distributional models.

Distribution regression. Distributional regression (DR), introduced as a formal inference framework by Foresi and Peracchi (1995) and further developed by Choi and Hall (2000) and Chernozhukov et al. (2013), models the entire conditional distribution function of an outcome Y given covariates X as

$$F_{Y|X}(y|x) = \Lambda(x'\beta(y)) \,,$$

where Λ is a known link function and $\beta(y)$ varies with the threshold y. Each binary regression $\mathbb{1}\{Y_i \leq y\} = \Lambda(X_i'\beta(y)) + U_i(y)$ is estimated separately for each y, yielding a flexible nonparametric description of the full conditional distribution. DR handles mass points and non-smooth distributions naturally, unlike quantile regression, and avoids bandwidth choices, unlike kernel methods. Rothe and Wied (2013) establish a specification test for the DR class; Chernozhukov et al. (2013) provide comprehensive large-sample theory.

Endogeneity in distribution regression. When the regressor Y_2 is endogenous—correlated with the latent error $U(y)$—the standard DR estimator is inconsistent for the structural conditional distribution $F^{\mathrm{V}}_{Y|X,Y_2}(y|x,y_2)$. Using instruments Z and the control-function approach of Rivers and Vuong (1988), Wied (2024) proposes a three-step estimator $\widehat{F}^{\mathrm{V}}$ and establishes its uniform weak convergence to a Gaussian process. The ordinary DR estimator $\widehat{F}^{\mathrm{N}}$ consistently estimates a *different* object, $F^{\mathrm{N}}_{Y|X,Y_2}(y|x,y_2) = \Lambda(x'_{\text{full}}\beta^{\mathrm{N}}(y))$, which conflates structural and endogeneity effects.

The relevant-endogeneity hypothesis. We propose to test whether the L^2-distance between these two objects exceeds a threshold $\Delta > 0$:

$$H_0 \colon \left\| F^{\mathrm{N}}(\cdot|x, y_2) - F^{\mathrm{V}}(\cdot|x, y_2) \right\|_{2,I} \leq \Delta, \tag{1}$$

$$H_1 \colon \left\| F^{\mathrm{N}}(\cdot|x, y_2) - F^{\mathrm{V}}(\cdot|x, y_2) \right\|_{2,I} > \Delta, \tag{2}$$

where $\|g\|_{2,I}^2 = \int_I g(y)^2 \, dy$ and $I \subset \mathbb{R}$ is the outcome interval of interest. Rejecting H_0 means that correcting for endogeneity changes the estimated conditional distribution by more than Δ over I; failing to reject means endogeneity is practically irrelevant at that covariate profile, regardless of statistical significance.

Contributions. This paper makes three main contributions. First, we define and analyze the endogeneity-distance functional $d(x, y_2) = \|F^{\mathrm{N}}(\cdot|x, y_2) - F^{\mathrm{V}}(\cdot|x, y_2)\|_{2,I}$ and relate it to structural parameters of the model. Second, we derive the joint influence-function representation of the pair $(\widehat{F}^{\mathrm{N}}, \widehat{F}^{\mathrm{V}})$ estimated from the *same* data, a step that requires combining the weak convergence of the naive DR estimator (Rothe and Wied, 2013; Chernozhukov et al., 2013) with that of the IV-DR estimator (Wied, 2024), accounting for their dependence. Third, we construct a self-normalized test statistic in the spirit of Dette et al. (2025) that is asymptotically pivotal under $d = \Delta$, has asymptotic level α on the null $d \leq \Delta$, and is consistent. No bootstrap and no variance estimation are required.

Related literature. The practical-differences framework for testing distributional equality goes back to Dette et al. (1998) and is applied to conditional distributions in Dette et al. (2025), who study the *two-sample* problem $H_0 \colon \|F^1_{Y|X}(\cdot|x) - F^2_{Y|X}(\cdot|x)\|_2 \leq \Delta$ in a semiparametric DR setting. The present paper addresses the *one-sample* problem of comparing the naive and corrected estimators from the same data and instruments. The technical challenge is the non-trivial correlation between $\widehat{F}^{\mathrm{N}}$ and $\widehat{F}^{\mathrm{V}}$, which does not arise in the independent two-sample case. Our self-normalization proof adapts their Theorems A.1–A.4 to the one-sample correlated setting.

Wied (2024) (Remark 3.4 of Dette et al. (2025)) already explicitly flags the extension to endogeneity testing as an open problem. The present paper fills that gap.

Outline. Section 2 sets up the model and defines the two distributional objects. Section 3 defines the relevant hypotheses and discusses the threshold. Section 4 describes the estimation procedure and defines the test statistic. Section 5 contains the asymptotic theory. Section 6 discusses extensions. Appendix A collects background on Z-estimators and Donsker theory. Appendix B contains proofs.

2 Model and Target Objects

2.1 Setup

Let $(Y_i, X_i, Y_{2,i}, Z_i)_{i=1}^n$ be an i.i.d. sequence, where:

- $Y \in \mathbb{R}$ is the scalar outcome variable,
- $X \in \mathbb{R}^k$ is the vector of exogenous control variables,
- $Y_2 \in \mathbb{R}$ is the potentially endogenous scalar regressor,
- $Z \in \mathbb{R}^\ell$ is the vector of excluded instruments.

Write $W_i = (Y_i, X_i', Y_{2,i}, Z_i')'$ for the full data vector. Let $X_i^{\mathrm{f}} = (X_i', Y_{2,i})' \in \mathbb{R}^{k+1}$ denote the covariate vector that includes $Y_{2,i}$.

Let $\Lambda : \mathbb{R} \to (0,1)$ be a strictly increasing link function with derivative $\lambda = \Lambda'$ (e.g., the standard normal CDF Φ or the logistic CDF). Let $I = [a, b] \subset \mathbb{R}$ be a compact interval that captures the outcome range of interest.

2.2 Structural Model

We adopt the triangular simultaneous-equations structure of Wied (2024). For each threshold $y \in I$, let $\mathbb{1}_y = \mathbb{1}\{Y \leq y\}$ be the indicator variable. The structural model is:

$$\mathbb{1}_y^* = X'\beta_1(y) + Y_2\beta_2(y) + U(y), \tag{3}$$

$$Y_2 = X'\gamma_1 + Z'\gamma_2 + V, \tag{4}$$

where $\mathbb{1}_y = \mathbb{1}\{\mathbb{1}_y^* \geq 0\}$ (the latent-variable representation of the binary outcome). The error decomposition is:

$$U(y) = \alpha_1(y)\, V + \alpha_2(y)\, \varepsilon(y), \tag{5}$$

with the following conditions:

(i) $\varepsilon(y) \perp\!\!\!\perp (X, Y_2, V)$ with $\mathrm{Var}(\varepsilon(y)) = 1$;

(ii) $\mathbb{E}[V|X, Z] = 0$ and $\mathrm{Var}(U(y)) = 1$;

(iii) The distribution function of $U(y)$ (and of $\varepsilon(y)$) is Λ (up to scale normalization).

Endogeneity is captured by $\alpha_1(y) \neq 0$: whenever $\alpha_1(y) \neq 0$, $U(y)$ and V are correlated, so Y_2 is endogenous in the y-binary regression.

Remark 2.1 (Identification)**.** *The model is identified under standard rank conditions for the first stage:* $\mathbb{E}[A_i A_i']$ *has full rank* $k+\ell$*, where* $A_i = (X_i', Z_i')'$*, and* $\gamma_2 \neq 0$*. Identification*

of the structural parameters $\beta_1(y)$, $\beta_2(y)$, $\alpha_1(y)$ (up to the scaling $\alpha_2(y)$) follows from the control-function argument of Rivers and Vuong (1988). The scale parameter $\alpha_2(y)$ is a normalization that makes the latent error compatible with the chosen link distribution. The condition $\mathrm{Var}(U(y)) = 1$ by itself does not imply $\alpha_1^2(y) + \alpha_2^2(y) = 1$ unless one additionally normalizes $\mathrm{Var}(V) = 1$ and imposes zero covariance between V and $\varepsilon(y)$. Under such an additional normalization the identity may be used; otherwise $\alpha_2(y)$ should simply be interpreted as the scale normalization in the control-function binary choice model.

2.3 The Naive Distribution Regression Estimand

The *naive* DR model treats Y_2 as exogenous and fits, for each $y \in I$:

$$\mathbb{P}(Y \leq y | X = x, Y_2 = y_2) = \Lambda\big(x_{\mathrm{f}}' \beta^{\mathrm{N}}(y)\big)\,, \tag{6}$$

where $x_{\mathrm{f}} = (x', y_2)' \in \mathbb{R}^{k+1}$ and $\beta^{\mathrm{N}}(y) = (\beta_1^N(y)', \beta_2^N(y))' \in \mathbb{R}^{k+1}$. The parameter $\beta^{\mathrm{N}}(y)$ is uniquely identified by the moment condition $\mathbb{E}[\psi^N(W_i, \beta^{\mathrm{N}}(y), y)] = 0$ for each $y \in I$, where

$$\psi^N(W_i, \beta, y) = \frac{[\mathbb{1}\{Y_i \leq y\} - \Lambda(X_i^{\mathrm{f}'}\beta)]\,\lambda(X_i^{\mathrm{f}'}\beta)}{\Lambda(X_i^{\mathrm{f}'}\beta)\,[1 - \Lambda(X_i^{\mathrm{f}'}\beta)]}\, X_i^{\mathrm{f}} \tag{7}$$

is the score of the binary log-likelihood. The naive estimand is

$$F^{\mathrm{N}}(y | x, y_2) = \Lambda\big(x_{\mathrm{f}}' \beta^{\mathrm{N}}(y)\big)\,. \tag{8}$$

Remark 2.2 (What F^{N} estimates)**.** *Under the structural model (3)–(5), the naive DR estimand $F^{\mathrm{N}}(y|x, y_2)$ is not the structural conditional CDF. Rather, $\beta^{\mathrm{N}}(y)$ is the pseudo-true parameter that minimizes the Kullback–Leibler divergence between the binary response model $\Lambda(X^{\mathrm{f}'}\beta)$ and the true conditional probability $\mathbb{P}(Y \leq y | X, Y_2)$. When $\alpha_1(y) = 0$ for all y, Y_2 is exogenous and $F^{\mathrm{N}} = F^{\mathrm{V}}$ (see Proposition 2.1).*

2.4 The IV/Control-Function Distribution Regression Estimand

The IV-DR estimand follows Wied (2024). Define the rescaled parameters

$$\widetilde{\beta}_1(y) = \frac{\beta_1(y)}{\alpha_2(y)}, \quad \widetilde{\beta}_2(y) = \frac{\beta_2(y)}{\alpha_2(y)}, \quad \widetilde{\alpha}_1(y) = \frac{\alpha_1(y)}{\alpha_2(y)}, \tag{9}$$

and write $\theta(y) = (\widetilde{\beta}_1(y)', \widetilde{\beta}_2(y), \widetilde{\alpha}_1(y))' \in \mathbb{R}^{k+2}$. By substituting (5) into (3) and conditioning on V, the structural probability is

$$\mathbb{P}(Y \leq y | X = x, Y_2 = y_2, V = v) = \widetilde{\Lambda}\Big(x'\widetilde{\beta}_1(y) + y_2\widetilde{\beta}_2(y) + v\widetilde{\alpha}_1(y)\Big)\,,$$

where $\widetilde{\Lambda}$ is the distribution of $\varepsilon(y)$ (which may equal Λ, e.g. under the probit assumption). The *average structural distribution* (the target estimand of this paper) integrates out V:

$$F^{\mathrm{V}}(y|x, y_2) = \int \widetilde{\Lambda}\Big(x'\widetilde{\beta}_1(y) + y_2\widetilde{\beta}_2(y) + v\widetilde{\alpha}_1(y)\Big)\, dF_V(v), \tag{10}$$

where F_V is the marginal CDF of V.

Remark 2.3 (Normality)**.** *Under joint normality* $(U(y), V)'|(X, Z) \sim \mathcal{N}(\mathbf{0}, \Sigma)$ *with* $\Sigma_{11} = 1$ *(normalization) and* $\Sigma_{12} = \rho\sigma_V$*, one can show (see Wied (2024), Appendix A.1) that* $F^{\mathrm{V}}(y|x, y_2) = \Phi(x'\beta_1(y) + y_2\beta_2(y))$*, which is precisely what the naive probit DR estimates* if Y_2 were exogenous*. The structural CDF has the same functional form as the naive one, but with different parameters. The naive estimand* F^{N} *conflates structural slopes and endogeneity bias.*

2.5 The Endogeneity Distance Functional

For a fixed covariate profile $(x, y_2) \in \mathbb{R}^k \times \mathbb{R}$, define the *pointwise difference*

$$\delta(y|x, y_2) = F^{\mathrm{N}}(y|x, y_2) - F^{\mathrm{V}}(y|x, y_2), \quad y \in I, \tag{11}$$

and the *practical endogeneity distance*

$$d(x, y_2) = \|\delta(\cdot|x, y_2)\|_{2,I} = \left(\int_I \delta(y|x, y_2)^2\, dy\right)^{1/2}. \tag{12}$$

The squared distance $T(x, y_2) = d(x, y_2)^2 = \int_I \delta(y|x, y_2)^2\, dy$ is the population analog of the test statistic.

Proposition 2.1 (Exogeneity implies equality)**.** *Under the structural model* (3)–(5)*, if* $\alpha_1(y) = 0$ *for all* $y \in I$*, then*

$$\delta(y|x, y_2) = 0 \quad \textit{for all } y \in I \textit{ and all } (x, y_2).$$

Equivalently, exact exogeneity of Y_2 *in every threshold regression implies* $F^{\mathrm{N}}(\cdot|x, y_2) = F^{\mathrm{V}}(\cdot|x, y_2)$*.*

Proof. If $\alpha_1(y) = 0$, then $U(y) = \alpha_2(y)\varepsilon(y)$ is independent of V and, by the first-stage exogeneity condition, independent of the source of endogeneity in Y_2. Hence conditioning on the control function does not alter the conditional distribution of the binary response. The augmented control-function model and the ordinary distribution-regression model therefore identify the same conditional probability at each threshold y. Consequently, the naive score $\mathbb{E}[\psi^N(W_i, \beta^{\mathrm{N}}(y), y)] = 0$ identifies the structural binary choice parameter

(up to the adopted scale normalization), and $F^{\mathrm{N}}(y|x, y_2) = F^{\mathrm{V}}(y|x, y_2)$ for all $y \in I$ and all (x, y_2). □

Remark 2.4 (No converse without additional restrictions)**.** *The converse is not claimed. Equality of the two distributional objects at a fixed covariate profile, or even on a restricted set of profiles, need not imply $\alpha_1(y) = 0$ because differences can cancel through the pseudo-true naive parameter, weak relevance of Y_2, or special features of the distribution of (X, Y_2, V). A converse statement would require additional no-cancellation or completeness assumptions ensuring that equality of the induced conditional probabilities identifies the control-function coefficient. The test developed below does not require such a converse: it only measures the distance between F^{N} and F^{V}.*

3 Relevant-Endogeneity Hypotheses

3.1 The Testing Problem

The relevant-endogeneity hypotheses are, for a fixed covariate profile (x, y_2) and threshold $\Delta > 0$:

$$H_0 : \quad T(x, y_2) \leq \Delta^2, \tag{13}$$

$$H_1 : \quad T(x, y_2) > \Delta^2. \tag{14}$$

The null H_0 states that correcting for endogeneity shifts the conditional distribution over I by at most Δ (in L^2-norm). The alternative H_1 states that the shift is practically significant.

This formulation is a one-sided test on the functional $T(x, y_2)$. It differs from a standard two-sided equivalence test: we reject when endogeneity is *large*, not when it is *small*. The structure is analogous to power-of-a-test statements: we want the test to *reject* when $T > \Delta^2$ and to *not reject* when $T \leq \Delta^2$.

Remark 3.1 (Comparison with exact exogeneity tests)**.** *The exact-distance null is $H_0^{\mathrm{exact}} : \delta(y|x, y_2) = 0$ for all y, i.e., $T(x, y_2) = 0$. Proposition 2.1 shows that exact exogeneity of Y_2 is sufficient for this null, but the converse is not used here; see Remark 2.4. Our H_0 in* (13) *nests the exact-distance problem: if $\Delta = 0$, the two problems coincide. For $\Delta > 0$, the relevant-endogeneity H_0 is a composite null with non-trivial interior $\{T < \Delta^2\}$, which is where exact-distance tests have trivial asymptotic power.*

3.2 Choice of the Threshold Δ

The threshold Δ must be chosen before observing the data, on substantive grounds. We discuss two natural approaches.

Benchmark-shift calibration. Following Dette et al. (2025) (who set Δ via an inflation benchmark for income comparisons), we recommend calibrating Δ as the L^2-distance induced by an economically meaningful shift in the outcome distribution. Concretely, if the benchmark shift is $Y \mapsto Y + c$ (e.g., a c-log-point increase in log-wages), define

$$\Delta_c^2 = \int_I [F_Y(y) - F_Y(y-c)]^2 \, dy, \tag{15}$$

where F_Y is the marginal CDF of Y. This quantity can be estimated consistently from the data. Then endogeneity is practically relevant only if its effect on the conditional distribution exceeds the reference shift c.

Fraction of marginal variance. An alternative is to set

$$\Delta^2 = \eta \cdot \int_I F_Y(y)(1 - F_Y(y)) \, dy$$

for some small $\eta \in (0,1)$, e.g., $\eta = 0.01$. This normalizes the threshold as a fraction of the variance of a uniform random variable on I, making it comparable across different outcome scales.

Remark 3.2 (Data-driven lower bound)**.** *A data-driven summary of the magnitude of practical endogeneity can be obtained by inverting the rejection rule. Define* $\widehat{\Delta}_\alpha = \{(\widehat{T}_n - q_{1-\alpha}\widehat{V}_n) \vee 0\}^{1/2}$. *Then* $[\widehat{\Delta}_\alpha, \infty)$ *is best interpreted as a one-sided lower confidence set for* $d(x, y_2)$*, not as a new data-chosen relevance threshold. The substantive threshold* Δ *used in the test should still be fixed before looking at the data. See Section 6.3 for the corresponding coverage statement.*

4 Estimation and Test Statistic

4.1 Naive DR Estimation

For each $y \in I$, the naive DR parameter $\beta^{\mathrm{N}}(y)$ is estimated by the MLE

$$\widehat{\beta}^{\mathrm{N}}(y) = \arg\max_{\beta \in \mathcal{B}} \sum_{i=1}^{n} \left[\mathbb{1}\{Y_i \le y\} \log \Lambda(X_i^{\mathrm{f}\prime}\beta) + \mathbb{1}\{Y_i > y\} \log(1 - \Lambda(X_i^{\mathrm{f}\prime}\beta)) \right], \tag{16}$$

where $\mathcal{B} \subset \mathbb{R}^{k+1}$ is compact. The naive DR estimator of the conditional CDF is:

$$\widehat{F}^{\mathrm{N}}(y|x, y_2) = \Lambda\Big(x_{\mathrm{f}}' \widehat{\beta}^{\mathrm{N}}(y) \Big). \tag{17}$$

4.2 IV-DR Estimation: Three Steps

Step 1 (First stage). Estimate $\Gamma_0 = (\gamma_1', \gamma_2')'$ by OLS of $Y_{2,i}$ on $A_i = (X_i', Z_i')'$:

$$\widehat{\Gamma} = \left(\sum_{i=1}^{n} A_i A_i' \right)^{-1} \sum_{i=1}^{n} A_i Y_{2,i}. \tag{18}$$

Compute the estimated residuals $\widehat{V}_i = Y_{2,i} - A_i' \widehat{\Gamma}$.

Step 2 (Control-function binary choice). For each $y \in I$, estimate $\theta_0(y) = (\widetilde{\beta}_1(y)', \widetilde{\beta}_2(y), \widetilde{\alpha}_1(y))'$ by MLE in the augmented binary choice model that treats $\widehat{V}_i$ as an additional regressor:

$$\widehat{\theta}(y) = \arg\max_{\theta \in \Theta} \sum_{i=1}^{n} \Big[\mathbb{1}\{Y_i \le y\} \log \widetilde{\Lambda}(\xi_i' \theta) + \mathbb{1}\{Y_i > y\} \log(1 - \widetilde{\Lambda}(\xi_i' \theta)) \Big], \tag{19}$$

where $\xi_i = \xi_i(\widehat{V}_i) = (X_i', Y_{2,i}, \widehat{V}_i)' \in \mathbb{R}^{k+2}$ and $\Theta \subset \mathbb{R}^{k+2}$ is compact.

Step 3 (Average over residual distribution). Integrate out the estimated residuals to obtain:

$$\widehat{F}^{\mathrm{V}}(y|x, y_2) = \frac{1}{n} \sum_{i=1}^{n} \widetilde{\Lambda}\Big(x' \widehat{\widetilde{\beta}}_1(y) + y_2 \widehat{\widetilde{\beta}}_2(y) + \widehat{V}_i \widehat{\widetilde{\alpha}}_1(y) \Big). \tag{20}$$

This nonparametrically averages out the unobserved heterogeneity V by plugging in the empirical distribution $\widehat{F}_V(v) = n^{-1} \sum_i \mathbb{1}\{\widehat{V}_i \le v\}$ as an estimate of F_V.

Remark 4.1 (Relation to average structural function)**.** *Step 3 consistently estimates* $F^{\mathrm{V}}(y|x, y_2) = \int \widetilde{\Lambda}(x' \widetilde{\beta}_1(y) + y_2 \widetilde{\beta}_2(y) + v \widetilde{\alpha}_1(y))\, dF_V(v)$*, the average structural distribution defined in* (10)*. The three-step estimator is the semiparametric analog of Blundell-Powell's average structural function; see Wied (2024) for discussion.*

4.3 Sequential Estimators

To construct the self-normalizing test statistic we need sequential (partial-sample) versions of both estimators. For $t \in [\varepsilon, 1]$ with fixed $\varepsilon > 0$ small, let $n_t = \lfloor nt \rfloor$.

Sequential naive DR. Let $\widehat{\beta}^{\mathrm{N}}(t, y)$ be the MLE (16) based on the first n_t observations. Set

$$\widehat{F}^{\mathrm{N}}(t, y|x, y_2) = \Lambda\Big(x_{\mathrm{f}}' \widehat{\beta}^{\mathrm{N}}(t, y) \Big). \tag{21}$$

Sequential IV-DR. The self-normalized statistic requires a genuinely sequential version of all estimated components. Therefore the first stage is also recomputed on the first

n_t observations:

$$\widehat{\Gamma}(t) = \left(\sum_{i=1}^{n_t} A_i A_i'\right)^{-1} \sum_{i=1}^{n_t} A_i Y_{2,i}, \qquad \widehat{V}_i(t) = Y_{2,i} - A_i'\widehat{\Gamma}(t), \quad i \le n_t. \tag{22}$$

For each $y \in I$, let $\widehat{\theta}(t, y)$ be the MLE in (19) based only on observations $i \le n_t$ and residuals $\widehat{V}_i(t)$. The sequential IV-DR estimator is

$$\widehat{F}^{\mathrm{V}}(t, y|x, y_2) = \frac{1}{n_t} \sum_{i=1}^{n_t} \widetilde{\Lambda}\left(x'\widehat{\widetilde{\beta}}_1(t, y) + y_2\widehat{\widetilde{\beta}}_2(t, y) + \widehat{V}_i(t)\widehat{\widetilde{\alpha}}_1(t, y)\right). \tag{23}$$

Remark 4.2 (Why the first stage is sequential)**.** *Using a full-sample first stage inside the partial-sample statistic would introduce a Brownian component evaluated at $t = 1$ into every partial-sample estimator. The resulting covariance in t would not be the Brownian-motion covariance used by the self-normalized limit. Recomputing the first stage on the same first n_t observations is the clean construction assumed in Assumption 5.5.*

4.4 Test Statistic and Decision Rule

Sequential difference and primary statistic. Define the sequential estimated difference:

$$\widehat{\delta}(t, y|x, y_2) = \widehat{F}^{\mathrm{N}}(t, y|x, y_2) - \widehat{F}^{\mathrm{V}}(t, y|x, y_2), \quad (t, y) \in [\varepsilon, 1] \times I. \tag{24}$$

The primary test statistic is the full-sample integrated squared difference:

$$\widehat{T}_n(x, y_2) = \int_I \widehat{\delta}(1, y|x, y_2)^2 \, dy. \tag{25}$$

Self-normalizing variance estimator. Motivated by Dette et al. (2025), the self-normalizing term is:

$$\widehat{V}_n^2(x, y_2) = \int_\varepsilon^1 \left[\int_I \widehat{\delta}(t, y|x, y_2)^2 \, dy - \int_I \widehat{\delta}(1, y|x, y_2)^2 \, dy\right]^2 dt. \tag{26}$$

Decision rule. Let $W = \mathbb{B}(1)/\left(\int_\varepsilon^1 (\mathbb{B}(t)/t - \mathbb{B}(1))^2 dt\right)^{1/2}$ where $\mathbb{B}$ is a standard Brownian motion on $[0, 1]$, and let $q_{1-\alpha}$ denote the $(1-\alpha)$-quantile of W (simulatable, cf. Dette et al. (2025), Table 1: $q_{0.95} \approx 1.755$ for $\varepsilon = 0.1$). We reject H_0 at level α if

$$\boxed{\widehat{T}_n(x, y_2) > \Delta^2 + q_{1-\alpha}\, \widehat{V}_n(x, y_2).} \tag{27}$$

Remark 4.3 (Interpretation of the self-normalizer)**.** *The term $\int_I \widehat{\delta}(t, y)^2 dy - \int_I \widehat{\delta}(1, y)^2 dy$ measures how much the integrated squared difference at "time" t deviates from its full-*

sample value. Under the null, both $\widehat{F}^{\mathrm{N}}(t,\cdot)$ and $\widehat{F}^{\mathrm{V}}(t,\cdot)$ are $\sqrt{nt}$-consistent for the same underlying object $\delta(\cdot)$, so this deviation is of order $O_p(1/\sqrt{n})$ and drives the normalization. Under the alternative, $\widehat{T}_n$ grows faster than $\widehat{V}_n$, giving power.

5 Asymptotic Theory

5.1 Assumptions

We collect the regularity conditions required for our main results.

Assumption 5.1 (Naive DR regularity)**.**

(i) The parameter space $\mathcal{B} \subset \mathbb{R}^{k+1}$ is compact and convex. The true parameter $\beta_0^{\mathrm{N}}(y) = (\beta_{1,0}^N(y)', \beta_{2,0}^N(y))'$ lies in the interior of $\mathcal{B}$ for each $y \in I$.

(ii) The function class $\mathcal{F}^N = \{\psi^N(w, \beta, y) : \beta \in \mathcal{B}, y \in I\}$ is P-Donsker with square-integrable envelope.

(iii) The Jacobian $C^N(y) = \mathbb{E}[\partial\psi^N(W_i, \beta_0^{\mathrm{N}}(y), y)/\partial\beta']$ exists, is continuous in y, and satisfies $\inf_{y\in I} \sigma_{\min}(C^N(y)) > 0$, where $\sigma_{\min}$ denotes the smallest singular value.

(iv) The link function Λ is twice continuously differentiable with bounded derivative $\lambda = \Lambda'$.

(v) The conditional density $f_{Y|X,Y_2}(y|\cdot)$ exists, is continuous in y, and is bounded uniformly over I.

(vi) $\mathbb{E}[\|X_i^{\mathrm{f}}\|^2] < \infty$.

Assumption 5.2 (First-stage regularity)**.**

(i) The matrix $M_{AA} = \mathbb{E}[A_i A_i']$ has full rank $k + \ell$.

(ii) The first stage (4) is correctly specified and $\mathbb{E}[V_i|A_i] = 0$.

(iii) $\mathbb{E}[\|A_i V_i\|^2] < \infty$ and $\mathbb{E}[\|A_i\|^2 V_i^2] < \infty$.

(iv) The sequential OLS estimator based on the first $n_t = \lfloor nt \rfloor$ observations satisfies, uniformly in $t \in [\varepsilon, 1]$,

$$\sqrt{n}\{\widehat{\Gamma}(t) - \Gamma_0\} = M_{AA}^{-1} \frac{\sqrt{n}}{n_t} \sum_{i=1}^{n_t} A_i V_i + o_p(1).$$

Assumption 5.3 (IV-DR regularity)**.**

(i) The parameter space $\Theta \subset \mathbb{R}^{k+2}$ is compact and convex. The true parameter $\theta_0(y)$ lies in the interior of Θ for each $y \in I$.

(ii) The function class $\mathcal{F}^V = \{\psi^{V*}(w, \theta, y, v) : \theta \in \Theta, y \in I\}$ is P-Donsker with square-integrable envelope, where ψ^{V*} is the score of the step-2 binary choice model evaluated at true residuals V_i:

$$\psi^{V*}(W_i, \theta, y, V_i) = \frac{[\mathbb{1}\{Y_i \le y\} - \widetilde{\Lambda}(\xi_i^{*\prime}\theta)]\,\widetilde{\lambda}(\xi_i^{*\prime}\theta)}{\widetilde{\Lambda}(\xi_i^{*\prime}\theta)\,[1 - \widetilde{\Lambda}(\xi_i^{*\prime}\theta)]}\,\xi_i^*, \tag{28}$$

with $\xi_i^* = (X_i', Y_{2,i}, V_i)' \in \mathbb{R}^{k+2}$.

(iii) The Jacobian $C^V(y) = \mathbb{E}[\partial\psi^{V*}(W_i, \theta_0(y), y, V_i)/\partial\theta']$ exists, is continuous in y, and satisfies $\inf_{y\in I} \sigma_{\min}(C^V(y)) > 0$.

(iv) The matrix $M^V(y) = \mathbb{E}[(\partial\psi^{V*}(W_i, \theta_0(y), y, V_i)/\partial v) \cdot A_i']$ exists and is continuous in y.

(v) The vector

$$r^V(y) = \mathbb{E}\Big[\widetilde{\lambda}(\xi_i^{*\prime}\theta_0(y))\,\widetilde{\alpha}_{1,0}(y)\,A_i\Big]$$

exists and is continuous in y.

(vi) $\mathbb{E}[\|A_i\|^2\|\xi_i^*\|^2] < \infty$.

Assumption 5.4 (Joint moment conditions)**.**

(i) The function class $\mathcal{H} = \{(\varphi_i^N(y), \varphi_i^V(y)) : y \in I\}$ (defined in Section 5.3 below) is P-Donsker in $\ell^\infty(I) \times \ell^\infty(I)$ with square-integrable envelope.

(ii) The covariance kernel $\Sigma(y_1, y_2) = \mathbb{E}[\xi_i(y_1)\xi_i(y_2)]$ is continuous on $I \times I$, where $\xi_i(y) = \varphi_i^N(y) - \varphi_i^V(y)$.

(iii) $\int_I \int_I \delta(y_1)\delta(y_2)\Sigma(y_1, y_2)\,dy_1\,dy_2 > 0$ (non-degenerate variance at the boundary $T = \Delta^2$).

Assumption 5.5 (Sequential asymptotic linearity)**.** For the genuinely sequential estimators defined in Section 4.3, with the first stage recomputed on the same first $n_t = \lfloor nt \rfloor$ observations, the following expansion holds uniformly in $(t, y) \in [\varepsilon, 1] \times I$:

$$\sqrt{n}\{\widehat{\delta}(t, y|x, y_2) - \delta(y|x, y_2)\} = \frac{\sqrt{n}}{n_t}\sum_{i=1}^{n_t} \xi_i(y) + o_p(1), \qquad \xi_i(y) = \varphi_i^N(y) - \varphi_i^V(y). \tag{29}$$

This is a high-level uniform linearization assumption. It collects the technical empirical-process remainder conditions needed for the sequential naive DR estimator, the sequential

first-stage OLS estimator, the generated residuals, the sequential control-function binary-choice estimator, and the empirical averaging over the residual distribution. Primitive sufficient conditions for this high-level expansion are given in Proposition 5.1. The proposition imposes somewhat strong regularity conditions, but it shows that the sequential linearity assumption follows from standard empirical-process, smoothness, moment, and nonsingularity assumptions adapted to the generated-residual IV-DR setting.

5.2 Primitive Sufficient Conditions for Sequential Linearity

The main theorem above is stated under the high-level sequential asymptotic linearity condition in Assumption 5.5. This subsection gives a set of primitive sufficient conditions under which that assumption holds. The conditions are intentionally somewhat stronger than necessary, but they are convenient and transparent for the generated-regressor IV-DR problem.

For $t \in [\varepsilon, 1]$, write $n_t = \lfloor nt \rfloor$ and define the sequential empirical average

$$\mathbb{P}_{n,t} f = \frac{1}{n_t} \sum_{i=1}^{n_t} f(W_i).$$

Let $\xi_i^* = (X_i', Y_{2,i}, V_i)'$ and let

$$\eta_i(y) = \xi_i^{*\prime} \theta_0(y).$$

For the IV/control-function score, write

$$\psi_i^{V*}(y) = \psi^{V*}(W_i, \theta_0(y), y, V_i).$$

Furthermore define the derivative of the score with respect to the residual argument by

$$\dot{\psi}_{v,i}^{V*}(y) = \frac{\partial}{\partial v} \psi^{V*}(W_i, \theta_0(y), y, v)\Big|_{v=V_i}.$$

Assumption 5.6 (Primitive conditions for sequential IV-DR expansion)**.** The following conditions hold.

(P1) **Sampling and moments.** The observations $(Y_i, X_i, Y_{2,i}, Z_i)$ are i.i.d. The first-stage equation

$$Y_{2,i} = A_i' \Gamma_0 + V_i, \qquad A_i = (X_i', Z_i')',$$

satisfies $\mathbb{E}[V_i | A_i] = 0$. Moreover, for some $q > 4$,

$$\mathbb{E}\|A_i\|^q < \infty, \qquad \mathbb{E}|V_i|^q < \infty, \qquad \mathbb{E}\|\xi_i^*\|^q < \infty.$$

(P2) **First-stage rank.** The matrix

$$M_{AA} = \mathbb{E}[A_i A_i']$$

is nonsingular. Moreover,

$$\sup_{t\in[\varepsilon,1]} \left\| (\mathbb{P}_{n,t} A_i A_i')^{-1} - M_{AA}^{-1} \right\| = o_p(1).$$

(P3) **Compactness and interiority.** The index set $I \subset \mathbb{R}$ is compact. The parameter space $\Theta \subset \mathbb{R}^{k+2}$ is compact and convex, and $\theta_0(y)$ lies in the interior of Θ for every $y \in I$.

(P4) **Smoothness and boundedness of the link.** The link function $\widetilde{\Lambda}$ is three times continuously differentiable. For all relevant η,

$$0 < c \leq \widetilde{\Lambda}(\eta) \leq 1 - c < 1,$$

and the first three derivatives of $\widetilde{\Lambda}$ are uniformly bounded.

(P5) **Uniform nonsingularity.** The Jacobian

$$C^V(y) = \mathbb{E}\left[\frac{\partial}{\partial \theta'} \psi^{V*}(W_i, \theta_0(y), y, V_i) \right]$$

exists, is continuous in y, and satisfies

$$\inf_{y\in I} \sigma_{\min}(C^V(y)) > 0.$$

(P6) **Uniform differentiability in the generated residual.** The score map

$$v \mapsto \psi^{V*}(W_i, \theta_0(y), y, v)$$

is twice continuously differentiable uniformly in $y \in I$. There exists an integrable envelope D_i such that, for all $y \in I$ and all v in a neighborhood of V_i,

$$\left\| \frac{\partial^2}{\partial v^2} \psi^{V*}(W_i, \theta_0(y), y, v) \right\| \leq D_i, \qquad \mathbb{E}[D_i] < \infty.$$

(P7) **Donsker and stochastic equicontinuity conditions.** The function classes

$$\left\{ \psi^{V*}(W_i, \theta, y, V_i) : \theta \in \Theta,\ y \in I \right\},$$

$$\left\{ \frac{\partial}{\partial \theta'} \psi^{V*}(W_i, \theta, y, V_i) : \theta \in \Theta,\ y \in I \right\},$$

and

$$\left\{ \frac{\partial}{\partial v} \psi^{V*}(W_i, \theta_0(y), y, V_i) : y \in I \right\}$$

are P-Donsker and Glivenko–Cantelli with square-integrable envelopes.

(P8) **Smoothness of the final averaging map.** Define

$$g_i(y, \theta, v) = \widetilde{\Lambda}(x'\theta_1 + y_2\theta_2 + v\theta_3),$$

where $\theta = (\theta_1', \theta_2, \theta_3)'$. The function class

$$\{g_i(y, \theta, v) : y \in I, \theta \in \Theta, v \in \mathcal{V}\}$$

and its first derivatives with respect to (θ, v) are Donsker and Glivenko–Cantelli with square-integrable envelopes. The second derivatives are dominated by an integrable envelope.

Proposition 5.1 (Primitive sufficient conditions for Assumption 5.5)**.** *Suppose Assumptions 5.1, 5.2, 5.3, 5.4, and 5.6 hold. Then the sequential IV/control-function DR estimator satisfies, uniformly in* $(t, y) \in [\varepsilon, 1] \times I$*,*

$$\sqrt{n}\left\{ \widehat{F}^{\mathrm{V}}(t, y|x, y_2) - F^{\mathrm{V}}(y|x, y_2) \right\} = \frac{\sqrt{n}}{n_t} \sum_{i=1}^{n_t} \varphi_i^V(y) + o_p(1), \tag{30}$$

where $\varphi_i^V(y)$ *is defined in* (46)*. If, in addition, the analogous sequential linear expansion for the naive DR estimator holds, namely*

$$\sqrt{n}\left\{ \widehat{F}^{\mathrm{N}}(t, y|x, y_2) - F^{\mathrm{N}}(y|x, y_2) \right\} = \frac{\sqrt{n}}{n_t} \sum_{i=1}^{n_t} \varphi_i^N(y) + o_p(1), \tag{31}$$

uniformly in $(t, y) \in [\varepsilon, 1] \times I$*, then Assumption 5.5 holds with*

$$\xi_i(y) = \varphi_i^N(y) - \varphi_i^V(y).$$

Proof. We prove (30); the naive expansion (31) follows from the same sequential Z-estimator argument without the generated-regressor terms.

Step 1: Sequential first-stage expansion. The first-stage OLS estimator based on the first n_t observations is

$$\widehat{\Gamma}(t) = \left(\mathbb{P}_{n,t} A_i A_i'\right)^{-1} \mathbb{P}_{n,t} A_i Y_{2,i}.$$

Using $Y_{2,i} = A_i'\Gamma_0 + V_i$ gives

$$\widehat{\Gamma}(t) - \Gamma_0 = \left(\mathbb{P}_{n,t} A_i A_i'\right)^{-1} \mathbb{P}_{n,t} A_i V_i.$$

By Assumption 5.6(P2), uniformly in $t \in [\varepsilon, 1]$,

$$\sqrt{n}\{\widehat{\Gamma}(t) - \Gamma_0\} = M_{AA}^{-1} \frac{\sqrt{n}}{n_t} \sum_{i=1}^{n_t} A_i V_i + o_p(1).$$

Moreover,

$$\sup_{t \in [\varepsilon,1]} \|\widehat{\Gamma}(t) - \Gamma_0\| = O_p(n^{-1/2}),$$

because n_t/n is bounded away from zero.

Step 2: Expansion of the generated residuals in the score. For $i \leq n_t$,

$$\widehat{V}_i(t) - V_i = -A_i'\{\widehat{\Gamma}(t) - \Gamma_0\}.$$

By a Taylor expansion in the residual argument, uniformly in (t, y),

$$\begin{aligned}\mathbb{P}_{n,t}\psi^{V*}(W_i, \theta_0(y), y, \widehat{V}_i(t)) = \mathbb{P}_{n,t}\psi_i^{V*}(y) - \mathbb{P}_{n,t}\left[\dot{\psi}_{v,i}^{V*}(y) A_i'\right]\{\widehat{\Gamma}(t) - \Gamma_0\} \\ + R_{n,t}^{\psi}(y),\end{aligned} \tag{32}$$

where the remainder satisfies

$$\sup_{(t,y) \in [\varepsilon,1] \times I} \sqrt{n} \|R_{n,t}^{\psi}(y)\| = o_p(1).$$

This follows from Assumption 5.6(P6), the uniform $O_p(n^{-1/2})$ rate of $\widehat{\Gamma}(t) - \Gamma_0$, and the integrable envelope for the second derivative.

By the sequential uniform law of large numbers implied by Assumption 5.6(P7),

$$\sup_{(t,y)} \left\| \mathbb{P}_{n,t}\left[\dot{\psi}_{v,i}^{V*}(y) A_i'\right] - M^V(y) \right\| = o_p(1).$$

Therefore

$$\mathbb{P}_{n,t}\psi^{V*}(W_i, \theta_0(y), y, \widehat{V}_i(t)) = \mathbb{P}_{n,t}\psi_i^{V*}(y) - M^V(y)\{\widehat{\Gamma}(t) - \Gamma_0\} + o_p(n^{-1/2}), \tag{33}$$

uniformly in (t, y).

Step 3: Sequential Z-estimator expansion. The sequential step-2 estimator satisfies

$$\mathbb{P}_{n,t}\psi^{V*}(W_i, \widehat{\theta}(t,y), y, \widehat{V}_i(t)) = 0.$$

Expanding the score around $\theta_0(y)$ yields

$$0 = \mathbb{P}_{n,t}\psi^{V*}(W_i, \theta_0(y), y, \widehat{V}_i(t)) + C^V(y)\{\widehat{\theta}(t,y) - \theta_0(y)\} + R^{\theta}_{n,t}(y), \tag{34}$$

where

$$\sup_{(t,y)} \sqrt{n}\|R^{\theta}_{n,t}(y)\| = o_p(1).$$

The remainder control follows from the stochastic equicontinuity and Glivenko–Cantelli conditions in Assumption 5.6(P7), the compactness of Θ, the uniform nonsingularity in (P5), and standard Z-estimator arguments applied uniformly over $t \in [\varepsilon, 1]$ and $y \in I$.

Combining (33) and (34) gives

$$\begin{aligned}\sqrt{n}\{\widehat{\theta}(t,y) - \theta_0(y)\} = -\{C^V(y)\}^{-1}\frac{\sqrt{n}}{n_t}\sum_{i=1}^{n_t}\psi_i^{V*}(y) \\ + \{C^V(y)\}^{-1}M^V(y)\sqrt{n}\{\widehat{\Gamma}(t) - \Gamma_0\} + o_p(1),\end{aligned} \tag{35}$$

uniformly in (t, y).

Substituting the sequential first-stage expansion from Step 1,

$$\begin{aligned}\sqrt{n}\{\widehat{\theta}(t,y) - \theta_0(y)\} = -\{C^V(y)\}^{-1}\frac{\sqrt{n}}{n_t}\sum_{i=1}^{n_t}\psi_i^{V*}(y) \\ + \{C^V(y)\}^{-1}M^V(y)M_{AA}^{-1}\frac{\sqrt{n}}{n_t}\sum_{i=1}^{n_t}A_iV_i + o_p(1).\end{aligned} \tag{36}$$

Step 4: Expansion of the final averaging step. The sequential IV-DR estimator is

$$\widehat{F}^{\mathrm{V}}(t, y|x, y_2) = \mathbb{P}_{n,t}\widetilde{\Lambda}\left(x'\widehat{\widetilde{\beta}}_1(t,y) + y_2\widehat{\widetilde{\beta}}_2(t,y) + \widehat{V}_i(t)\widehat{\widetilde{\alpha}}_1(t,y)\right).$$

Using the notation $g_i(y) = \widetilde{\Lambda}(\xi_i^{*\prime}\theta_0(y))$, a Taylor expansion around $(\theta_0(y), V_i)$ gives, uniformly in (t, y),

$$\begin{aligned}\widehat{F}^{\mathrm{V}}(t, y|x, y_2) - F^{\mathrm{V}}(y|x, y_2) = \mathbb{P}_{n,t}\{g_i(y) - F^{\mathrm{V}}(y|x, y_2)\} \\ + m^V(y)'\{\widehat{\theta}(t,y) - \theta_0(y)\} \\ - r^V(y)'\{\widehat{\Gamma}(t) - \Gamma_0\} + o_p(n^{-1/2}),\end{aligned} \tag{37}$$

where

$$m^V(y) = \mathbb{E}\left[\widetilde{\lambda}(\xi_i^{*\prime}\theta_0(y))\xi_i^*\right],$$

and

$$r^V(y) = \mathbb{E}\left[\widetilde{\lambda}(\xi_i^{*\prime}\theta_0(y))\widetilde{\alpha}_{1,0}(y)A_i\right].$$

The term involving $r^V(y)$ appears because

$$\widehat{V}_i(t) - V_i = -A_i'\{\widehat{\Gamma}(t) - \Gamma_0\}.$$

The uniform negligibility of the Taylor remainder follows from Assumption 5.6(P8), the uniform $O_p(n^{-1/2})$ rates for $\widehat{\theta}(t,y)-\theta_0(y)$ and $\widehat{\Gamma}(t)-\Gamma_0$, and the bounded-away-from-zero condition on the link probabilities.

Multiplying (37) by $\sqrt{n}$ and substituting (36) and the first-stage expansion gives

$$\begin{aligned}\sqrt{n}\left\{\widehat{F}^{\mathrm{V}}(t,y|x,y_2) - F^{\mathrm{V}}(y|x,y_2)\right\} = \frac{\sqrt{n}}{n_t}\sum_{i=1}^{n_t}\{g_i(y) - F^{\mathrm{V}}(y|x,y_2)\} \\ - m^V(y)'\{C^V(y)\}^{-1}\frac{\sqrt{n}}{n_t}\sum_{i=1}^{n_t}\psi_i^{V*}(y) \\ + \left[m^V(y)'\{C^V(y)\}^{-1}M^V(y) - r^V(y)'\right]M_{AA}^{-1}\frac{\sqrt{n}}{n_t}\sum_{i=1}^{n_t}A_iV_i + o_p(1).\end{aligned}$$

The right-hand side is exactly

$$\frac{\sqrt{n}}{n_t}\sum_{i=1}^{n_t}\varphi_i^V(y) + o_p(1),$$

with $\varphi_i^V(y)$ as defined in (46). This proves (30).

Finally, subtracting the IV expansion from the corresponding naive expansion (31) yields

$$\sqrt{n}\{\widehat{\delta}(t,y|x,y_2) - \delta(y|x,y_2)\} = \frac{\sqrt{n}}{n_t}\sum_{i=1}^{n_t}\{\varphi_i^N(y) - \varphi_i^V(y)\} + o_p(1),$$

uniformly in $(t,y) \in [\varepsilon,1]\times I$. Hence Assumption 5.5 holds. □

5.3 Influence Function Representations

We now state the influence-function representations of both estimators, which are the key building blocks for the joint weak convergence result.

5.3.1 Naive DR Influence Function

Under Assumption 5.1, the Z-estimator theory of Chernozhukov et al. (2013) (Lemma E.3) gives uniformly in $y \in I$:

$$\sqrt{n}\left(\widehat{\beta}^{\mathrm{N}}(y) - \beta_0^{\mathrm{N}}(y)\right) = -(C^N(y))^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi^N(W_i, \beta_0^{\mathrm{N}}(y), y) + o_p(1). \tag{38}$$

Applying the delta method to $\widehat{F}^{\mathrm{N}}(y|x, y_2) = \Lambda(x_{\mathrm{f}}'\widehat{\beta}^{\mathrm{N}}(y))$:

$$\sqrt{n}\left(\widehat{F}^{\mathrm{N}}(y|x, y_2) - F^{\mathrm{N}}(y|x, y_2)\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi_i^N(y) + o_p(1), \tag{39}$$

uniformly in $y \in I$, where the naive influence function is

$$\varphi_i^N(y) = -\lambda(x_{\mathrm{f}}'\beta_0^{\mathrm{N}}(y)) \cdot x_{\mathrm{f}}' \cdot (C^N(y))^{-1} \cdot \psi^N(W_i, \beta_0^{\mathrm{N}}(y), y). \tag{40}$$

Since $\mathbb{E}[\psi^N(W_i, \beta_0^{\mathrm{N}}(y), y)] = 0$, we have $\mathbb{E}[\varphi_i^N(y)] = 0$.

5.3.2 IV-DR Influence Function

The derivation for $\widehat{F}^{\mathrm{V}}$ is more involved because the estimator uses the generated regressors $\widehat{V}_i$ from Step 1. We follow Wied (2024) (Appendix B) and decompose the error into three components.

Step-2 score contribution. Define the population score at the true residuals: $\Psi^V(\theta, y) = \mathbb{E}[\psi^{V*}(W_i, \theta, y, V_i)]$. Since $\widehat{V}_i - V_i = -A_i'(\widehat{\Gamma} - \Gamma_0)$, a first-order expansion of the step-2 score around $(\theta_0(y), V_i)$ gives, uniformly in $y \in I$,

$$\begin{aligned} 0 = \frac{1}{n}\sum_{i=1}^{n}\psi^{V*}(W_i, \theta_0(y), y, V_i) + C^V(y)\{\widehat{\theta}(y) - \theta_0(y)\} \\ - M^V(y)\{\widehat{\Gamma} - \Gamma_0\} + o_p(n^{-1/2}), \end{aligned} \tag{41}$$

where $M^V(y) = \mathbb{E}[(\partial\psi^{V*}(W_i, \theta_0(y), y, V_i)/\partial v)A_i']$. Consequently,

$$\begin{aligned} \sqrt{n}\left(\widehat{\theta}(y) - \theta_0(y)\right) = -\{C^V(y)\}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi^{V*}(W_i, \theta_0(y), y, V_i) \\ + \{C^V(y)\}^{-1}M^V(y)\sqrt{n}(\widehat{\Gamma} - \Gamma_0) + o_p(1). \end{aligned} \tag{42}$$

Using the OLS expansion

$$\sqrt{n}(\widehat{\Gamma} - \Gamma_0) = M_{AA}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}A_iV_i + o_p(1), \tag{43}$$

the first-stage component in (42) can be written as a sample average.

Averaging and generated-residual terms. For notational compactness, define

$$\begin{aligned} g_i(y) &= \widetilde{\Lambda}(\xi_i^{*\prime}\theta_0(y)), \\ m^V(y) &= \mathbb{E}\Big[\widetilde{\lambda}(\xi_i^{*\prime}\theta_0(y))\xi_i^*\Big] \in \mathbb{R}^{k+2}, \\ r^V(y) &= \mathbb{E}\Big[\widetilde{\lambda}(\xi_i^{*\prime}\theta_0(y))\widetilde{\alpha}_{1,0}(y)A_i\Big] \in \mathbb{R}^{k+\ell}. \end{aligned}$$

The first term $g_i(y) - F^{\mathrm{V}}(y|x, y_2)$ is the empirical averaging contribution. The vector $m^V(y)$ is the derivative of the average structural distribution with respect to θ, while $r^V(y)$ captures the direct effect of using $\widehat{V}_i$ rather than V_i in the final averaging step. Indeed, since $\widehat{V}_i - V_i = -A_i'(\widehat{\Gamma} - \Gamma_0)$,

$$\frac{1}{n}\sum_{i=1}^{n}\widetilde{\lambda}(\xi_i^{*\prime}\theta_0(y))\widetilde{\alpha}_{1,0}(y)(\widehat{V}_i - V_i) = -r^V(y)'(\widehat{\Gamma} - \Gamma_0) + o_p(n^{-1/2}). \tag{44}$$

Linearization. Combining the empirical averaging term, the second-stage parameter effect, the first-stage effect through $\widehat{\theta}(y)$, and the direct generated-residual effect in Step 3, the full influence-function expansion of $\widehat{F}^{\mathrm{V}}$ is, uniformly in $y \in I$,

$$\sqrt{n}\left(\widehat{F}^{\mathrm{V}}(y|x, y_2) - F^{\mathrm{V}}(y|x, y_2)\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi_i^V(y) + o_p(1), \tag{45}$$

where the IV influence function is

$$\begin{aligned} \varphi_i^V(y) = {} & g_i(y) - F^{\mathrm{V}}(y|x, y_2) \\ & - m^V(y)'\{C^V(y)\}^{-1}\psi^{V*}(W_i, \theta_0(y), y, V_i) \\ & + \left[m^V(y)'\{C^V(y)\}^{-1}M^V(y) - r^V(y)'\right]M_{AA}^{-1}A_iV_i. \end{aligned} \tag{46}$$

The three lines correspond to: (i) empirical averaging over the residual law, (ii) second-stage parameter estimation, and (iii) first-stage estimation, including both its indirect effect through $\widehat{\theta}(y)$ and its direct effect in the final average. Under the maintained moment conditions, $\mathbb{E}[\varphi_i^V(y)] = 0$.

5.4 Primitive Sufficient Conditions for Sequential Linearity

The main theorem is stated under the high-level sequential asymptotic linearity condition in Assumption 5.5. This subsection gives strong but transparent sufficient conditions under which that condition holds for the generated-residual IV-DR estimator. The conditions are stronger than necessary, but they are useful because they make all uniform

Taylor remainders and stochastic-equicontinuity steps explicit.

Throughout this subsection, fix the covariate profile (x, y_2) at which the test is evaluated. Define

$$z_i^0 = z_i^0(x, y_2) = (x', y_2, V_i)' \in \mathbb{R}^{k+2}, \qquad \xi_i^* = (X_i', Y_{2,i}, V_i)' \in \mathbb{R}^{k+2}.$$

The vector ξ_i^* is used in the second-stage binary choice score, while z_i^0 is used in the final average structural distribution evaluated at the fixed profile (x, y_2). Thus,

$$g_i^0(y) = \widetilde{\Lambda}(z_i^{0\prime}\theta_0(y)),$$

and

$$F^{\mathrm{V}}(y|x, y_2) = \mathbb{E}\left[g_i^0(y)\right] = \mathbb{E}\left[\widetilde{\Lambda}\{x'\widetilde{\beta}_{1,0}(y) + y_2\widetilde{\beta}_{2,0}(y) + V_i\widetilde{\alpha}_{1,0}(y)\}\right].$$

Correspondingly, define

$$m^V(y) = \mathbb{E}\left[\widetilde{\lambda}(z_i^{0\prime}\theta_0(y))z_i^0\right],$$

and

$$r^V(y) = \mathbb{E}\left[\widetilde{\lambda}(z_i^{0\prime}\theta_0(y))\widetilde{\alpha}_{1,0}(y)A_i\right].$$

With this notation, the IV influence function at the fixed profile (x, y_2) is

$$\begin{aligned}\varphi_i^V(y) = {} & g_i^0(y) - F^{\mathrm{V}}(y|x, y_2) \\ & - m^V(y)'\{C^V(y)\}^{-1}\psi^{V*}(W_i, \theta_0(y), y, V_i) \\ & + \left[m^V(y)'\{C^V(y)\}^{-1}M^V(y) - r^V(y)'\right]M_{AA}^{-1}A_iV_i.\end{aligned} \tag{47}$$

This is the same form as (46), but with the final averaging objects evaluated at the fixed target profile rather than at the random regressors $(X_i, Y_{2,i})$.

For $t \in [\varepsilon, 1]$, let $n_t = \lfloor nt \rfloor$ and

$$\mathbb{P}_{n,t}f = \frac{1}{n_t}\sum_{i=1}^{n_t} f(W_i).$$

Assumption 5.7 (Strong primitive conditions for sequential IV-DR)**.** The following conditions hold.

(S1) **Sampling and compact support.** The observations $(Y_i, X_i, Y_{2,i}, Z_i)$ are i.i.d. The vectors $A_i = (X_i', Z_i')'$, $\xi_i^* = (X_i', Y_{2,i}, V_i)'$, and the residual V_i have compact support. In particular, there exists a constant $K < \infty$ such that almost surely

$$\|A_i\| \le K, \qquad \|\xi_i^*\| \le K, \qquad |V_i| \le K.$$

(S2) **Correct first stage and rank.** The first-stage equation satisfies

$$Y_{2,i} = A_i'\Gamma_0 + V_i, \qquad \mathbb{E}[V_i|A_i] = 0,$$

and

$$M_{AA} = \mathbb{E}[A_i A_i']$$

is nonsingular.

(S3) **Compact threshold and parameter spaces.** The outcome interval $I \subset \mathbb{R}$ is compact. The parameter space $\Theta \subset \mathbb{R}^{k+2}$ is compact and convex, and $\theta_0(y)$ lies in the interior of Θ for every $y \in I$. Moreover, the map $y \mapsto \theta_0(y)$ is Lipschitz continuous on I.

(S4) **Smooth link and trimming away from zero and one.** The link function $\widetilde{\Lambda}$ is three times continuously differentiable. Its first three derivatives are uniformly bounded. Moreover, there exists a constant $c \in (0, 1/2)$ such that, for all relevant (w, θ, y, v) in a neighborhood of the true values,

$$c \leq \widetilde{\Lambda}(\xi(v)'\theta) \leq 1 - c,$$

where $\xi(v) = (X', Y_2, v)'$. The same bound holds for the final averaging index $z(v) = (x', y_2, v)'$.

(S5) **Uniform identification and nonsingularity.** For each $y \in I$, the equation

$$\mathbb{E}\left[\psi^{V*}(W_i, \theta, y, V_i)\right] = 0$$

has the unique solution $\theta = \theta_0(y)$. The Jacobian

$$C^V(y) = \mathbb{E}\left[\frac{\partial}{\partial \theta'}\psi^{V*}(W_i, \theta_0(y), y, V_i)\right]$$

exists, is continuous in y, and satisfies

$$\inf_{y \in I} \sigma_{\min}\{C^V(y)\} > 0.$$

(S6) **Uniform smoothness of the score.** The score map

$$(\theta, v) \mapsto \psi^{V*}(W_i, \theta, y, v)$$

is twice continuously differentiable in (θ, v), uniformly in $y \in I$. All first and second derivatives with respect to (θ, v) are uniformly bounded by an integrable envelope.

Under compact support, this envelope may be taken to be a finite constant.

(S7) **Uniform empirical-process conditions.** The classes consisting of the score, its first derivatives, and its second derivatives,

$$\left\{\psi^{V*}(W_i,\theta,y,V_i):\theta\in\Theta,\ y\in I\right\},$$

$$\left\{\frac{\partial}{\partial\theta'}\psi^{V*}(W_i,\theta,y,V_i):\theta\in\Theta,\ y\in I\right\},\qquad \left\{\frac{\partial}{\partial v}\psi^{V*}(W_i,\theta_0(y),y,V_i):y\in I\right\},$$

and their products with A_i are P-Donsker and Glivenko–Cantelli. Moreover, the corresponding sequential empirical-process bounds hold: for every such class $\mathcal{F}$,

$$\sup_{t\in[\varepsilon,1]}\sup_{f\in\mathcal{F}}|\mathbb{P}_{n,t}f-\mathbb{E}f|=O_p(n^{-1/2}).$$

(S8) **Uniform smoothness of the final averaging map.** For

$$h_i(y,\theta,\Gamma)=\widetilde{\Lambda}(x'\theta_1+y_2\theta_2+(Y_{2,i}-A_i'\Gamma)\theta_3),$$

the map $(\theta,\Gamma)\mapsto h_i(y,\theta,\Gamma)$ is twice continuously differentiable uniformly in $y\in I$. Its first and second derivatives are bounded by an integrable envelope. The classes formed by h_i and its first and second derivatives, indexed by (y,θ,Γ) in neighborhoods of the true values, are Glivenko–Cantelli and satisfy the sequential empirical bound in (S7).

(S9) **Uniform consistency and rate of the sequential Z-estimator.** The sequential second-stage estimator satisfies

$$\sup_{(t,y)\in[\varepsilon,1]\times I}\|\widehat{\theta}(t,y)-\theta_0(y)\|=O_p(n^{-1/2}).$$

This rate follows from (S3)–(S7) by the standard uniform Z-estimator argument under compactness, uniform identification, and uniform nonsingularity; it is stated here explicitly to make the proof of the linear representation self-contained.

Proposition 5.2 (Primitive sequential expansion of IV-DR)**.** *Suppose Assumption 5.7 holds. Then the sequential IV/control-function DR estimator satisfies, uniformly in* $(t,y)\in[\varepsilon,1]\times I$*,*

$$\sqrt{n}\left\{\widehat{F}^{\mathrm{V}}(t,y|x,y_2)-F^{\mathrm{V}}(y|x,y_2)\right\}=\frac{\sqrt{n}}{n_t}\sum_{i=1}^{n_t}\varphi_i^V(y)+o_p(1),\tag{48}$$

where $\varphi_i^V(y)$ *is given in* (47)*. If the analogous sequential linear expansion for the naive*

DR estimator holds,

$$\sqrt{n}\left\{\widehat{F}^{\mathrm{N}}(t,y|x,y_2)-F^{\mathrm{N}}(y|x,y_2)\right\}=\frac{\sqrt{n}}{n_t}\sum_{i=1}^{n_t}\varphi_i^N(y)+o_p(1),$$

uniformly in $(t,y)\in[\varepsilon,1]\times I$*, then Assumption 5.5 holds with*

$$\xi_i(y)=\varphi_i^N(y)-\varphi_i^V(y).$$

Proof. We prove the IV expansion (48). All $o_p(1)$ terms below are uniform in $(t,y)\in[\varepsilon,1]\times I$.

Step 1: Sequential first-stage expansion. The sequential first-stage estimator is

$$\widehat{\Gamma}(t)=\left(\mathbb{P}_{n,t}A_iA_i'\right)^{-1}\mathbb{P}_{n,t}A_iY_{2,i}.$$

Using $Y_{2,i}=A_i'\Gamma_0+V_i$,

$$\widehat{\Gamma}(t)-\Gamma_0=\left(\mathbb{P}_{n,t}A_iA_i'\right)^{-1}\mathbb{P}_{n,t}A_iV_i.$$

By the sequential empirical-process bound in Assumption 5.7,

$$\sup_{t\in[\varepsilon,1]}\|\mathbb{P}_{n,t}A_iA_i'-M_{AA}\|=O_p(n^{-1/2}).$$

Since M_{AA} is nonsingular, the inverse map is continuous in a neighborhood of M_{AA}, and therefore

$$\sup_{t\in[\varepsilon,1]}\left\|\left(\mathbb{P}_{n,t}A_iA_i'\right)^{-1}-M_{AA}^{-1}\right\|=O_p(n^{-1/2}).$$

Also,

$$\sup_{t\in[\varepsilon,1]}\left\|\frac{\sqrt{n}}{n_t}\sum_{i=1}^{n_t}A_iV_i\right\|=O_p(1).$$

Thus,

$$\sqrt{n}\{\widehat{\Gamma}(t)-\Gamma_0\}=M_{AA}^{-1}\frac{\sqrt{n}}{n_t}\sum_{i=1}^{n_t}A_iV_i+o_p(1). \tag{49}$$

In particular,

$$\sup_{t\in[\varepsilon,1]}\|\widehat{\Gamma}(t)-\Gamma_0\|=O_p(n^{-1/2}).$$

Step 2: Generated-residual expansion of the step-2 score. For $i\le n_t$,

$$\widehat{V}_i(t)-V_i=-A_i'\{\widehat{\Gamma}(t)-\Gamma_0\}.$$

By a second-order Taylor expansion in the residual argument v,

$$\begin{aligned}\mathbb{P}_{n,t}\psi^{V*}(W_i,\theta_0(y),y,\widehat{V}_i(t)) = \mathbb{P}_{n,t}\psi^{V*}(W_i,\theta_0(y),y,V_i) \\ - \mathbb{P}_{n,t}\left[\frac{\partial}{\partial v}\psi^{V*}(W_i,\theta_0(y),y,V_i)A_i'\right]\{\widehat{\Gamma}(t)-\Gamma_0\} \\ + R^{\psi}_{n,t}(y). \end{aligned} \tag{50}$$

By the bounded second derivative assumption,

$$\sup_{(t,y)}\|R^{\psi}_{n,t}(y)\| \le C\sup_t\|\widehat{\Gamma}(t)-\Gamma_0\|^2 = O_p(n^{-1}).$$

Hence

$$\sup_{(t,y)}\sqrt{n}\|R^{\psi}_{n,t}(y)\| = o_p(1).$$

By the sequential empirical-process bound,

$$\sup_{(t,y)}\left\|\mathbb{P}_{n,t}\left[\frac{\partial}{\partial v}\psi^{V*}(W_i,\theta_0(y),y,V_i)A_i'\right] - M^V(y)\right\| = O_p(n^{-1/2}).$$

Combining this with $\sup_t\|\widehat{\Gamma}(t)-\Gamma_0\| = O_p(n^{-1/2})$ gives

$$\sup_{(t,y)}\sqrt{n}\left\|\left[\mathbb{P}_{n,t}\dot{\psi}^{V*}_{v,i}(y)A_i' - M^V(y)\right]\{\widehat{\Gamma}(t)-\Gamma_0\}\right\| = o_p(1).$$

Therefore

$$\mathbb{P}_{n,t}\psi^{V*}(W_i,\theta_0(y),y,\widehat{V}_i(t)) = \mathbb{P}_{n,t}\psi^{V*}_i(y) - M^V(y)\{\widehat{\Gamma}(t)-\Gamma_0\} + o_p(n^{-1/2}). \tag{51}$$

Step 3: Sequential Z-estimator expansion. The sequential step-2 estimator satisfies

$$\mathbb{P}_{n,t}\psi^{V*}(W_i,\widehat{\theta}(t,y),y,\widehat{V}_i(t)) = 0.$$

Expanding the score around $\theta_0(y)$ gives

$$\begin{aligned}0 = \mathbb{P}_{n,t}\psi^{V*}(W_i,\theta_0(y),y,\widehat{V}_i(t)) \\ + \mathbb{P}_{n,t}\left[\frac{\partial}{\partial\theta'}\psi^{V*}(W_i,\theta_0(y),y,\widehat{V}_i(t))\right]\{\widehat{\theta}(t,y)-\theta_0(y)\} + R^{\theta}_{n,t}(y).\end{aligned} \tag{52}$$

By the bounded second derivatives and the uniform rate $\sup_{(t,y)}\|\widehat{\theta}(t,y)-\theta_0(y)\| = O_p(n^{-1/2})$,

$$\sup_{(t,y)}\sqrt{n}\|R^{\theta}_{n,t}(y)\| = o_p(1).$$

Moreover, by the same generated-residual Taylor argument and the sequential empirical-process bound,

$$\sup_{(t,y)} \left\| \mathbb{P}_{n,t} \left[\frac{\partial}{\partial \theta'} \psi^{V*}(W_i, \theta_0(y), y, \widehat{V}_i(t)) \right] - C^V(y) \right\| = o_p(1).$$

Using uniform nonsingularity of $C^V(y)$, equations (51) and (52) imply

$$\begin{aligned} \sqrt{n}\{\widehat{\theta}(t,y) - \theta_0(y)\} = -\{C^V(y)\}^{-1} \frac{\sqrt{n}}{n_t} \sum_{i=1}^{n_t} \psi_i^{V*}(y) \\ + \{C^V(y)\}^{-1} M^V(y) \sqrt{n}\{\widehat{\Gamma}(t) - \Gamma_0\} + o_p(1). \end{aligned} \tag{53}$$

Substituting the first-stage expansion (49) yields

$$\begin{aligned} \sqrt{n}\{\widehat{\theta}(t,y) - \theta_0(y)\} = -\{C^V(y)\}^{-1} \frac{\sqrt{n}}{n_t} \sum_{i=1}^{n_t} \psi_i^{V*}(y) \\ + \{C^V(y)\}^{-1} M^V(y) M_{AA}^{-1} \frac{\sqrt{n}}{n_t} \sum_{i=1}^{n_t} A_i V_i + o_p(1). \end{aligned} \tag{54}$$

Step 4: Expansion of the final averaging step. The sequential IV-DR estimator can be written as

$$\widehat{F}^{\mathrm{V}}(t, y|x, y_2) = \mathbb{P}_{n,t} h_i(y, \widehat{\theta}(t,y), \widehat{\Gamma}(t)),$$

where

$$h_i(y, \theta, \Gamma) = \widetilde{\Lambda}\left(x'\theta_1 + y_2\theta_2 + (Y_{2,i} - A_i'\Gamma)\theta_3\right).$$

At the true values,

$$h_i(y, \theta_0(y), \Gamma_0) = g_i^0(y).$$

A second-order Taylor expansion around $(\theta_0(y), \Gamma_0)$ gives

$$\begin{aligned} \widehat{F}^{\mathrm{V}}(t, y|x, y_2) - F^{\mathrm{V}}(y|x, y_2) = \mathbb{P}_{n,t}\{g_i^0(y) - F^{\mathrm{V}}(y|x, y_2)\} \\ + \mathbb{P}_{n,t} \left[\frac{\partial}{\partial \theta} h_i(y, \theta_0(y), \Gamma_0) \right]' \{\widehat{\theta}(t,y) - \theta_0(y)\} \\ + \mathbb{P}_{n,t} \left[\frac{\partial}{\partial \Gamma} h_i(y, \theta_0(y), \Gamma_0) \right]' \{\widehat{\Gamma}(t) - \Gamma_0\} + R_{n,t}^h(y). \end{aligned} \tag{55}$$

The derivatives at the true values are

$$\frac{\partial}{\partial \theta} h_i(y, \theta_0(y), \Gamma_0) = \widetilde{\lambda}(z_i^{0\prime}\theta_0(y)) z_i^0,$$

and

$$\frac{\partial}{\partial \Gamma} h_i(y, \theta_0(y), \Gamma_0) = -\widetilde{\lambda}(z_i^{0\prime}\theta_0(y)) \widetilde{\alpha}_{1,0}(y) A_i.$$

Thus their expectations are $m^V(y)$ and $-r^V(y)$, respectively. By the sequential empirical-process bound,

$$\sup_{(t,y)} \left\| \mathbb{P}_{n,t} \frac{\partial}{\partial \theta} h_i(y, \theta_0(y), \Gamma_0) - m^V(y) \right\| = O_p(n^{-1/2}),$$

and similarly for the derivative with respect to Γ. Multiplying these empirical deviations by the $O_p(n^{-1/2})$ rates of $\widehat{\theta}(t,y) - \theta_0(y)$ and $\widehat{\Gamma}(t) - \Gamma_0$ gives terms that are $o_p(n^{-1/2})$ uniformly. The second-order remainder satisfies

$$\sup_{(t,y)} |R^h_{n,t}(y)| \leq C \left[\sup_{(t,y)} \|\widehat{\theta}(t,y) - \theta_0(y)\|^2 + \sup_t \|\widehat{\Gamma}(t) - \Gamma_0\|^2 \right] = O_p(n^{-1}),$$

and therefore $\sup_{(t,y)} \sqrt{n} |R^h_{n,t}(y)| = o_p(1)$.

Consequently,

$$\begin{aligned} \widehat{F}^{\mathrm{V}}(t,y|x,y_2) - F^{\mathrm{V}}(y|x,y_2) = \mathbb{P}_{n,t}\{g_i^0(y) - F^{\mathrm{V}}(y|x,y_2)\} \\ + m^V(y)'\{\widehat{\theta}(t,y) - \theta_0(y)\} - r^V(y)'\{\widehat{\Gamma}(t) - \Gamma_0\} + o_p(n^{-1/2}). \end{aligned} \tag{56}$$

Step 5: Combine the expansions. Multiplying (56) by $\sqrt{n}$ and substituting (54) and (49) gives

$$\begin{aligned} \sqrt{n}\{\widehat{F}^{\mathrm{V}}(t,y|x,y_2) - F^{\mathrm{V}}(y|x,y_2)\} = \frac{\sqrt{n}}{n_t} \sum_{i=1}^{n_t} \{g_i^0(y) - F^{\mathrm{V}}(y|x,y_2)\} \\ - m^V(y)'\{C^V(y)\}^{-1} \frac{\sqrt{n}}{n_t} \sum_{i=1}^{n_t} \psi_i^{V*}(y) \\ + \left[m^V(y)'\{C^V(y)\}^{-1} M^V(y) - r^V(y)' \right] M_{AA}^{-1} \frac{\sqrt{n}}{n_t} \sum_{i=1}^{n_t} A_i V_i + o_p(1). \end{aligned}$$

The right-hand side is precisely

$$\frac{\sqrt{n}}{n_t} \sum_{i=1}^{n_t} \varphi_i^V(y) + o_p(1),$$

with $\varphi_i^V(y)$ defined in (47). This proves (48).

Finally, subtracting the corresponding sequential expansion of the naive DR estimator gives

$$\sqrt{n}\{\widehat{\delta}(t,y|x,y_2) - \delta(y|x,y_2)\} = \frac{\sqrt{n}}{n_t} \sum_{i=1}^{n_t} \{\varphi_i^N(y) - \varphi_i^V(y)\} + o_p(1),$$

uniformly in $(t,y) \in [\varepsilon, 1] \times I$. Therefore Assumption 5.5 holds with $\xi_i(y) = \varphi_i^N(y) -$

$\varphi_i^V(y)$. □

5.5 Joint Weak Convergence

We now establish the joint weak convergence of $(\widehat{F}^{\mathrm{N}}, \widehat{F}^{\mathrm{V}})$, which is the core theoretical result of this paper.

Theorem 5.3 (Joint influence function convergence)**.** *Under Assumptions 5.1–5.4, as $n \to \infty$:*

$$\sqrt{n}\begin{pmatrix}\widehat{F}^{\mathrm{N}}(\cdot|x,y_2) - F^{\mathrm{N}}(\cdot|x,y_2)\\ \widehat{F}^{\mathrm{V}}(\cdot|x,y_2) - F^{\mathrm{V}}(\cdot|x,y_2)\end{pmatrix} \rightsquigarrow \begin{pmatrix}\mathbb{G}^N(\cdot)\\ \mathbb{G}^V(\cdot)\end{pmatrix} \quad \text{in } \ell^\infty(I)\times\ell^\infty(I), \tag{57}$$

where $(\mathbb{G}^N, \mathbb{G}^V)$ is a bivariate mean-zero Gaussian process with covariance structure:

$$\mathrm{Cov}(\mathbb{G}^N(y_1), \mathbb{G}^N(y_2)) = \mathbb{E}[\varphi_i^N(y_1)\varphi_i^N(y_2)], \tag{58}$$

$$\mathrm{Cov}(\mathbb{G}^V(y_1), \mathbb{G}^V(y_2)) = \mathbb{E}[\varphi_i^V(y_1)\varphi_i^V(y_2)], \tag{59}$$

$$\mathrm{Cov}(\mathbb{G}^N(y_1), \mathbb{G}^V(y_2)) = \mathbb{E}[\varphi_i^N(y_1)\varphi_i^V(y_2)]. \tag{60}$$

Proof. The claim follows from showing that the joint empirical process $n^{-1/2}\sum_{i=1}^n(\varphi_i^N(y), \varphi_i^V(y))_{y\in I}$ converges weakly to a Gaussian process in $\ell^\infty(I)^2$.

By Assumption 5.4(i), the function class $\mathcal{H}$ is P-Donsker. The central limit theorem in $\ell^\infty(I)^2$ (van der Vaart and Wellner, 1996, Theorem 2.1.20) then gives $(1/\sqrt{n})\sum_{i=1}^n(\varphi_i^N(y), \varphi_i^V(y)) \rightsquigarrow (\mathbb{G}^N(y), \mathbb{G}^V(y))$. The remainder terms in (39) and (45) are $o_p(1)$ uniformly in y by Assumptions 5.1–5.3 and the argument in Wied (2024) (Appendix B). Tightness is inherited from the Donsker property of $\mathcal{H}$. □

Corollary 5.4 (Weak convergence of the difference)**.** *Under the conditions of Theorem 5.3,*

$$\sqrt{n}\left(\widehat{\delta}(\cdot|x,y_2) - \delta(\cdot|x,y_2)\right) \rightsquigarrow \mathbb{G}(\cdot) \quad \text{in } \ell^\infty(I), \tag{61}$$

where $\mathbb{G} = \mathbb{G}^N - \mathbb{G}^V$ is a mean-zero Gaussian process with covariance kernel

$$\Sigma(y_1, y_2) = \mathbb{E}[\xi_i(y_1)\xi_i(y_2)], \quad \xi_i(y) = \varphi_i^N(y) - \varphi_i^V(y). \tag{62}$$

Proof. The result is immediate from Theorem 5.3 by the continuous mapping theorem applied to the difference map $(f, g) \mapsto f - g$. □

Remark 5.1 (The covariance cross-term matters)**.** *The crucial difference from the two-sample setting of Dette et al. (2025) is the cross-covariance* (60)*. In their case, $\mathbb{G}^N$ and $\mathbb{G}^V$ are independent (samples from two populations), so the covariance of $\mathbb{G} = \mathbb{G}^N - \mathbb{G}^V$ is simply* $\mathrm{Cov}(\mathbb{G}^N(y_1), \mathbb{G}^N(y_2)) + \mathrm{Cov}(\mathbb{G}^V(y_1), \mathbb{G}^V(y_2))$*. Here the cross-term $\mathbb{E}[\varphi_i^N(y_1)\varphi_i^V(y_2)]$ is generally non-zero because both estimators use the same data $\{W_i\}$.*

The self-normalization device below remains valid because Σ only appears as an overall scale factor τ^2 (see (67)), which cancels in the pivotal ratio.

5.6 Sequential Process and Self-Normalization

We now extend the convergence to the sequential estimators and establish the key functional convergence that underpins the self-normalization.

Theorem 5.5 (Sequential FCLT)**.** *Under Assumptions 5.1–5.4 and 5.5, as $n \to \infty$:*

$$\left\{\sqrt{n}\left(\widehat{\delta}(t, y|x, y_2) - \delta(y|x, y_2)\right)\right\}_{(t,y)\in[\varepsilon,1]\times I} \rightsquigarrow \{H(t, y)\}_{(t,y)\in[\varepsilon,1]\times I} \quad \text{in } \ell^\infty([\varepsilon, 1] \times I), \tag{63}$$

where H is a centered Gaussian process with covariance

$$\mathrm{Cov}(H(t_1, y_1), H(t_2, y_2)) = \frac{t_1 \wedge t_2}{t_1 t_2}\,\Sigma(y_1, y_2). \tag{64}$$

Equivalently, $H(t, y) = (1/t)\,\mathbb{G}(t, y)$, where $\mathbb{G}(t, y)$ is a Gaussian process with covariance $(t_1 \wedge t_2)\,\Sigma(y_1, y_2)$ (i.e., with "Brownian-motion-type" covariance in t for fixed y).

Proof. By Assumption 5.5, uniformly in $(t, y) \in [\varepsilon, 1] \times I$,

$$\sqrt{n}\{\widehat{\delta}(t, y) - \delta(y)\} = \frac{\sqrt{n}}{n_t}\sum_{i=1}^{n_t} \xi_i(y) + o_p(1).$$

Since $\{\xi_i(y) : y \in I\}$ is P-Donsker by Assumption 5.4(i) and has mean zero, the sequential empirical-process central limit theorem gives

$$\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n_t} \xi_i(y)\right\}_{(t,y)} \rightsquigarrow \{\mathbb{G}(t, y)\}_{(t,y)} \quad \text{in } \ell^\infty([\varepsilon, 1] \times I),$$

where $\mathbb{G}$ is a centered Gaussian process with covariance $\mathrm{Cov}(\mathbb{G}(t_1, y_1), \mathbb{G}(t_2, y_2)) = (t_1 \wedge t_2)\Sigma(y_1, y_2)$. Because $n_t/n \to t$ uniformly on $[\varepsilon, 1]$, the continuous mapping $f(t, y) \mapsto f(t, y)/t$ yields $H(t, y) = \mathbb{G}(t, y)/t$. Hence

$$\mathrm{Cov}(H(t_1, y_1), H(t_2, y_2)) = \frac{t_1 \wedge t_2}{t_1 t_2}\Sigma(y_1, y_2),$$

which proves the claim. □

5.6.1 Implications for the test statistic components

Define the integrated process:

$$D_n(t) = 2\int_I \delta(y|x, y_2)\left(\widehat{\delta}(t, y|x, y_2) - \delta(y|x, y_2)\right) dy. \tag{65}$$

This is the linearized difference $\widehat{T}_n^{(t)} - T(x, y_2) + o_p(n^{-1/2})$, where $\widehat{T}_n^{(t)} = \int_I \widehat{\delta}(t, y)^2 dy$.

Corollary 5.6 (Convergence of linearized process)**.** *Under Assumptions 5.1–5.4 and 5.5, when $T(x, y_2) > 0$:*

$$\left\{\sqrt{n}\, D_n(t)\right\}_{t\in[\varepsilon,1]} \rightsquigarrow \left\{2\int_I \delta(y|x, y_2)H(t, y)\, dy\right\}_{t\in[\varepsilon,1]} = \left\{\frac{\tau}{t}\,\mathbb{B}(t)\right\}_{t\in[\varepsilon,1]}, \tag{66}$$

in $\ell^\infty([\varepsilon, 1])$, where $\mathbb{B}$ is a standard Brownian motion and

$$\tau^2 = 4\int_I\int_I \delta(y_1|x, y_2)\delta(y_2|x, y_2)\,\Sigma(y_1, y_2)\, dy_1\, dy_2. \tag{67}$$

Proof. By the continuous mapping theorem applied to Theorem 5.5 and the integral map $f \mapsto 2\int_I \delta(y)f(y)\, dy$:

$$\sqrt{n}\, D_n(t) \rightsquigarrow 2\int_I \delta(y)\, H(t, y)\, dy =: \tilde{\mathbb{G}}(t).$$

By (64), $\tilde{\mathbb{G}}$ is a centered Gaussian process with

$$\mathrm{Cov}(\tilde{\mathbb{G}}(t_1), \tilde{\mathbb{G}}(t_2)) = 4\int_I\int_I \delta(y_1)\delta(y_2)\,\frac{t_1 \wedge t_2}{t_1 t_2}\,\Sigma(y_1, y_2)\, dy_1\, dy_2 = \frac{t_1 \wedge t_2}{t_1 t_2}\,\tau^2.$$

This is the covariance of $(\tau/t)\mathbb{B}(t)$, identifying $\tilde{\mathbb{G}}(t) = (\tau/t)\mathbb{B}(t)$ (up to distributional equality). □

5.7 Main Theorem

We now state and prove the main result of the paper.

Theorem 5.7 (Main result)**.** *Suppose Assumptions 5.1–5.4 and 5.5 hold. Then:*

(i) ***(Asymptotic level.)*** *For any $\alpha \in (0, 1)$, the rejection rule (27) has asymptotic level α on the boundary $\{T(x, y_2) = \Delta^2\}$ of the null: under $T(x, y_2) = \Delta^2$,*

$$\lim_{n\to\infty} \mathbb{P}\Big(\widehat{T}_n(x, y_2) > \Delta^2 + q_{1-\alpha}\,\widehat{V}_n(x, y_2)\Big) = \alpha. \tag{68}$$

(ii) ***(Interior null.)*** *For every fixed data-generating process satisfying $T(x, y_2) < \Delta^2$,*

$$\lim_{n\to\infty} \mathbb{P}\Big(\widehat{T}_n(x, y_2) > \Delta^2 + q_{1-\alpha}\,\widehat{V}_n(x, y_2)\Big) = 0. \tag{69}$$

Hence the boundary $T(x, y_2) = \Delta^2$ is least favorable, and the test has asymptotic size α in the sense that the limiting rejection probability is at most α over the null $T(x, y_2) \le \Delta^2$.

(iii) ***(Consistency.)*** *For* $T(x, y_2) > \Delta^2$,

$$\lim_{n\to\infty} \mathbb{P}\Big(\widehat{T}_n(x, y_2) > \Delta^2 + q_{1-\alpha}\,\widehat{V}_n(x, y_2)\Big) = 1. \tag{70}$$

Proof. **Step 1: Expansion of $\widehat{T}_n$.** Write $\widehat{T}_n = \int_I \widehat{\delta}(1, y)^2 dy$. By Corollary 5.4 and the functional delta method applied to $\phi(f) = \int_I f(y)^2 dy$:

$$\widehat{T}_n - T(x, y_2) = D_n(1) + o_p(n^{-1/2}), \tag{71}$$

where $D_n(1) = 2\int_I \delta(y)(\widehat{\delta}(1, y) - \delta(y))dy$.

Step 2: Convergence of (T_n, V_n). By Corollary 5.6:

$$\sqrt{n}\begin{pmatrix} \widehat{T}_n - T(x, y_2) \\ \widehat{V}_n \end{pmatrix} \rightsquigarrow \tau \begin{pmatrix} \mathbb{B}(1) \\ \left(\int_\varepsilon^1 \left(\frac{\mathbb{B}(t)}{t} - \mathbb{B}(1)\right)^2 dt\right)^{1/2} \end{pmatrix}. \tag{72}$$

To see this for $\widehat{V}_n$: by Corollary 5.6, $\sqrt{n}(\int_I \widehat{\delta}(t, y)^2 dy - T) \approx D_n(t)\cdot\sqrt{n}$ and $\sqrt{n}(\int_I \widehat{\delta}(1, y)^2 dy - T) \approx D_n(1)\cdot\sqrt{n}$, so

$$\begin{aligned}\sqrt{n}\,\widehat{V}_n &= \sqrt{n}\left(\int_\varepsilon^1 \left[\int_I \widehat{\delta}^2(t)\,dy - \int_I \widehat{\delta}^2(1)\,dy\right]^2 dt\right)^{1/2} \\ &\rightsquigarrow \tau\left(\int_\varepsilon^1 \left(\frac{\mathbb{B}(t)}{t} - \mathbb{B}(1)\right)^2 dt\right)^{1/2}.\end{aligned}$$

Step 3: Pivotal ratio. Dividing (72):

$$\frac{\sqrt{n}(\widehat{T}_n - T)}{\sqrt{n}\,\widehat{V}_n} = \frac{\widehat{T}_n - T}{\widehat{V}_n} \rightsquigarrow \frac{\mathbb{B}(1)}{\left(\int_\varepsilon^1 (\mathbb{B}(t)/t - \mathbb{B}(1))^2 dt\right)^{1/2}} =: W. \tag{73}$$

The limiting random variable W has a distribution free of nuisance parameters (in particular, free of τ, Σ, and all model parameters), so (73) provides a pivotal statistic.

Step 4: Proof of (i)–(iii). *Boundary case $T = \Delta^2$:* The rejection event is $\widehat{T}_n > \Delta^2 + q_{1-\alpha}\widehat{V}_n$, equivalently, $(\widehat{T}_n - T)/\widehat{V}_n > q_{1-\alpha}$. By (73), this converges to $\mathbb{P}(W > q_{1-\alpha}) = \alpha$ by definition of $q_{1-\alpha}$. This gives (68).

Interior null $T < \Delta^2$: We have $\widehat{T}_n \to T$ in probability, so $\widehat{T}_n - \Delta^2 \to T - \Delta^2 < 0$. Simultaneously $q_{1-\alpha}\widehat{V}_n \geq 0$. Hence $\widehat{T}_n - \Delta^2 - q_{1-\alpha}\widehat{V}_n \to T - \Delta^2 < 0$ and the rejection probability converges to 0. This gives (69).

Alternative $T > \Delta^2$: Now $\widehat{T}_n \to T > \Delta^2$. Since $\widehat{V}_n = O_p(n^{-1/2})$, we have $q_{1-\alpha}\widehat{V}_n = O_p(n^{-1/2}) \to 0$ in probability. Hence $\widehat{T}_n - \Delta^2 - q_{1-\alpha}\widehat{V}_n \to T - \Delta^2 > 0$, so the rejection probability converges to 1. This gives (70). □

Remark 5.2 (Why self-normalization works)**.** *The key observation is that both* $\sqrt{n}(\widehat{T}_n - T)$ *and* $\sqrt{n}\widehat{V}_n$ *converge to* τ *times a Brownian-motion functional. The* τ *cancels in the ratio* $(\widehat{T}_n - T)/\widehat{V}_n$*, giving the pivotal limit* W*. This mirrors exactly the argument in Dette et al. (2025) (their Theorem 3.1), but with the one-sample* τ^2 *of* (67) *instead of their two-sample variance. The pivot* W *is the same in both cases.*

Corollary 5.8 (Local alternatives)**.** *Suppose the data-generating process is indexed by* n *and satisfies*

$$T^{(n)}(x, y_2) = \Delta^2 + \frac{2c}{\sqrt{n}} + o(n^{-1/2}) \qquad \text{for some } c > 0, \tag{74}$$

while the sequential expansion in Assumption 5.5 holds with the same non-degenerate scale τ*. Then*

$$\lim_{n\to\infty} \mathbb{P}\Big(\widehat{T}_n > \Delta^2 + q_{1-\alpha}\widehat{V}_n\Big) = \mathbb{P}\left(\frac{\mathbb{B}(1) + 2c/\tau}{\left(\int_\varepsilon^1 (\mathbb{B}(t)/t - \mathbb{B}(1))^2 dt\right)^{1/2}} > q_{1-\alpha}\right). \tag{75}$$

In particular, for $c > 0$ *this limiting rejection probability is larger than* α *whenever the limiting distribution is continuous at* $q_{1-\alpha}$*.*

Proof. Under (74), the expansion (71) becomes

$$\widehat{T}_n - \Delta^2 = \frac{2c}{\sqrt{n}} + D_n(1) + o_p(n^{-1/2}).$$

Multiplying by $\sqrt{n}$ gives

$$\sqrt{n}(\widehat{T}_n - \Delta^2) \rightsquigarrow 2c + \tau\mathbb{B}(1).$$

The self-normalizer has the same first-order limit as in (72). Therefore

$$\frac{\widehat{T}_n - \Delta^2}{\widehat{V}_n} \rightsquigarrow \frac{\tau\mathbb{B}(1) + 2c}{\tau\left(\int_\varepsilon^1 (\mathbb{B}(t)/t - \mathbb{B}(1))^2 dt\right)^{1/2}} = \frac{\mathbb{B}(1) + 2c/\tau}{\left(\int_\varepsilon^1 (\mathbb{B}(t)/t - \mathbb{B}(1))^2 dt\right)^{1/2}},$$

and the result follows from the continuous mapping theorem. □

6 Extensions

6.1 Aggregated Version

The pointwise test (13) tests at a single covariate profile (x, y_2). A stronger statement aggregates over all covariate profiles:

$$D^2 = \int_{\mathcal{X}\times\mathcal{Y}_2} \int_I \left[F^{\mathrm{N}}(y|x, y_2) - F^{\mathrm{V}}(y|x, y_2)\right]^2 dy\, d\mu(x, y_2), \tag{76}$$

where μ is a weight measure on the covariate space. Natural choices:

(i) *Empirical measure:* $d\mu = n^{-1}\sum_i \delta_{(X_i,Y_{2,i})}$. This yields the average squared endogeneity bias over observed covariates.

(ii) *Policy-relevant measure:* $d\mu$ concentrated on a finite grid of economically relevant profiles (e.g., workers by education group).

(iii) *Population measure:* $d\mu = dF_{X,Y_2}$, estimated by the empirical distribution.

The aggregated test statistic is $\hat{D}_n^2 = \int \hat{T}_n(x, y_2)\, d\mu(x, y_2)$, which inherits weak convergence from the corresponding process indexed by (x, y_2) under additional stochastic-equicontinuity and integrability conditions. The self-normalization approach can be extended, but the same pivotal limit W follows only after verifying that the aggregated linearized sequential process has covariance of the form $((s \wedge t)/(st))\tau_D^2$ for some scalar $\tau_D^2 > 0$. Thus this extension is conceptually straightforward but requires its own influence-function and sequential-FCLT argument.

6.2 Monotonicity Correction

The DR estimators $\widehat{F}^{\mathrm{N}}(\cdot|x, y_2)$ and $\widehat{F}^{\mathrm{V}}(\cdot|x, y_2)$ are not guaranteed to be monotone in y in finite samples, since each y-binary regression is fit independently. Two standard corrections are available: monotone rearrangement (Chernozhukov et al., 2010) and isotonic regression (Wied, 2024).

Monotone rearrangement. Define the rearranged estimator:

$$\widetilde{F}^N(y|x, y_2) = \int_0^1 \mathbb{1}\Big\{\widehat{Q}^N_{Y|X,Y_2}(u|x, y_2) \le y\Big\}\, du,$$

where $\widehat{Q}^N(u|x, y_2) = \inf_y\{\widehat{F}^{\mathrm{N}}(y|x, y_2) \ge u\}$ is the empirical quantile function. The rearrangement is Hadamard differentiable at strictly increasing functions (Chernozhukov et al., 2010), so the $\sqrt{n}$-convergence rate and the Gaussian process limit of Theorem 5.3 carry over to $\widetilde{F}^N$ and $\widetilde{F}^V$, and all asymptotic results remain valid.

Isotonic regression. Alternatively, one may apply isotonic regression (pool-adjacent-violators) to the raw estimates $\{\widehat{F}^{\mathrm{N}}(Y_i|x, y_2)\}_{i=1}^n$ at the observed outcome values. Wied (2024) (Section 4.2) verifies that this correction does not alter the $\sqrt{n}$-asymptotic distribution.

Practical recommendation. For the main theoretical results, we use the raw (potentially non-monotone) estimators because the asymptotics are cleaner. For finite-sample

implementation, we recommend the monotone-rearrangement correction. Both versions should be reported in simulations.

6.3 Confidence Sets for $d(x, y_2)$

Inverting the test (27) gives a one-sided lower confidence set for the practical endogeneity distance $d(x, y_2)$. Define:

$$\widehat{\Delta}_\alpha = \left\{ (\widehat{T}_n(x, y_2) - q_{1-\alpha} \widehat{V}_n(x, y_2)) \vee 0 \right\}^{1/2}. \tag{77}$$

Then $[\widehat{\Delta}_\alpha, \infty)$ is interpreted as a lower confidence set: under the non-degeneracy condition in Assumption 5.4(iii),

$$\mathbb{P}\{d(x, y_2) \geq \widehat{\Delta}_\alpha\} \to 1 - \alpha$$

at regular boundary points of the squared-distance functional. For fixed interior values with $d(x, y_2) > 0$, the coverage is at least $1-\alpha$ asymptotically. This confidence set should not be confused with the substantive threshold Δ, which is chosen before estimation.

6.4 Multiple Endogenous Regressors

The framework extends to $Y_2 \in \mathbb{R}^m$ with $m > 1$ endogenous regressors and $\ell \geq m$ instruments. The first stage (4) becomes a system of linear equations, and the control-function augmented binary choice in Step 2 includes m generated residuals $(\widehat{V}_{i,1}, \ldots, \widehat{V}_{i,m})$. The IV influence function (46) acquires multivariate first-stage correction terms analogous to the last line, including the direct effect of the generated residual vector in the final average. The asymptotic results carry over under the corresponding multivariate generalizations of Assumptions 5.2, 5.3, and 5.5.

7 Conclusion

We have proposed and analyzed an asymptotically pivotal test for the practical relevance of endogeneity in semiparametric distribution regression. The key technical contributions are: (i) the joint influence-function representation of the naive and IV-DR estimators estimated from the same data, (ii) the self-normalized sequential limit under a high-level sequential asymptotic linearity condition, and (iii) the adaptation of the self-normalization approach of Dette et al. (2025) to the one-sample correlated setting. The resulting test has exact asymptotic level α at the boundary, is consistent, and is pivotal without bootstrap or variance estimation. Simulation evidence and empirical application (education endogeneity in the wage distribution) will be presented in subsequent work.

A Mathematical Background

A.1 Z-Estimators in Function Spaces

We collect the key result on Z-estimators indexed by a compact set, which underpins both the naive DR and IV-DR expansions. Let $\mathcal{U}$ be a compact subset of $\mathbb{R}$ and $\Theta \subset \mathbb{R}^p$ compact. Suppose that for each $u \in \mathcal{U}$, $\theta_0(u)$ is identified by $\Psi(\theta_0(u), u) = \mathbb{E}[\psi(W_i, \theta_0(u), u)] = 0$. Let $\hat{\Psi}_n(\theta, u) = n^{-1} \sum_i \psi(W_i, \theta, u)$ and suppose $\hat{\theta}(u)$ satisfies $\|\hat{\Psi}_n(\hat{\theta}(u), u)\| = o_p(n^{-1/2})$.

Lemma A.1 (Chernozhukov et al. 2013, Lemma E.3)**.** *If $\psi(W_i, \cdot, \cdot)$ is continuous with probability one, the class $\{\psi(w, \theta, u) : \theta \in \Theta, u \in \mathcal{U}\}$ is P-Donsker, the Jacobian $\dot{\Psi}_{\theta_0(u),u} = \partial\Psi/\partial\theta'|_{\theta_0(u)}$ is non-singular uniformly in u, and $\hat{\theta}(u) \to \theta_0(u)$ uniformly in probability, then:*

$$\sqrt{n}(\hat{\theta}(\cdot) - \theta_0(\cdot)) = -\dot{\Psi}^{-1}_{\theta_0(\cdot),\cdot} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(W_i, \theta_0(\cdot), \cdot) + o_p(1) \quad \text{in } \ell^\infty(\mathcal{U}).$$

A.2 Functional Delta Method

Let $\phi : \mathbb{D} \to \mathbb{E}$ be Hadamard differentiable at f tangentially to $\mathbb{D}_0 \subset \mathbb{D}$, with derivative $\dot{\phi}_f$.

Lemma A.2 (Functional delta method, van der Vaart and Wellner 1996, Theorem 3.9.4)**.** *If $\sqrt{n}(f_n - f) \rightsquigarrow G$ in $\mathbb{D}$ and $f_n \in \mathbb{D}$ for all n, then $\sqrt{n}(\phi(f_n) - \phi(f)) \rightsquigarrow \dot{\phi}_f(G)$ in $\mathbb{E}$.*

The map $\phi(f) = \int_I f(y)^2 dy$ is Hadamard differentiable at any $f \in L^2(I)$ with $\dot{\phi}_f(h) = 2\int_I f(y)h(y)dy$.

A.3 Pivotal Distribution of W

The random variable

$$W = \frac{\mathbb{B}(1)}{\left(\int_\varepsilon^1 \left(\frac{\mathbb{B}(t)}{t} - \mathbb{B}(1) \right)^2 dt \right)^{1/2}}$$

has a distribution free of nuisance parameters, which can be simulated to arbitrary precision. For $\varepsilon = 0.1$, Dette et al. (2025) report: $q_{0.90} \approx 1.317$, $q_{0.95} \approx 1.755$, $q_{0.99} \approx 2.636$.

B Proof Details

B.1 Proof of Theorem 5.3

We verify the conditions for applying the joint CLT in $\ell^\infty(I)^2$.

Mean zero: By construction, $\mathbb{E}[\varphi_i^N(y)] = 0$ because $\mathbb{E}[\psi^N(W_i, \beta_0^{\mathrm{N}}(y), y)] = 0$. Similarly, $\mathbb{E}[\varphi_i^V(y)] = 0$ because both the averaging term, the second-stage score, and the first-stage correction have mean zero: $\mathbb{E}[\widetilde{\Lambda}(\xi_i^{*\prime}\theta_0(y))] = F^{\mathrm{V}}(y|x, y_2)$, $\mathbb{E}[\psi^{V*}(W_i, \theta_0(y), y, V_i)] = 0$, and $\mathbb{E}[A_i V_i] = 0$.

Donsker property: By Assumption 5.4(i), the class $\mathcal{H} = \{(\varphi_i^N(y), \varphi_i^V(y)) : y \in I\}$ is P-Donsker. This is verified by noting that $\varphi_i^N(y)$ is a smooth function of $\psi^N(W_i, \beta_0^{\mathrm{N}}(y), y)$ (which is Donsker by Assumption 5.1(ii)), and $\varphi_i^V(y)$ is a smooth function of $\psi^{V*}(W_i, \theta_0(y), y, V_i)$ and $A_i V_i$ (Donsker by Assumptions 5.2 and 5.3(ii)). Permanence of the Donsker property under Lipschitz transformations and finite unions gives the result.

Remainder terms: The $o_p(1)$ terms in (39) and (45) follow from (i) consistency of $\widehat{\beta}^{\mathrm{N}}$ and $\widehat{\theta}$ in $\ell^\infty(I)$ (standard under the Z-estimator conditions), and (ii) the first-stage approximation error $\widehat{V}_i - V_i = -A_i'(\widehat{\Gamma} - \Gamma_0) = O_p(n^{-1/2})$ uniformly, which implies that the first-stage and direct generated-residual correction terms in (45) have negligible remainders.

B.2 Proof of Theorem 5.5

Theorem 5.5 is proved in the main text from Assumption 5.5 and the sequential empirical-process central limit theorem. The role of Assumption 5.5 is to avoid claiming, without a lengthy generated-regressor proof, that the sequential IV-DR estimator admits a uniform partial-sum expansion over both t and y. This is the only additional high-level condition beyond the pointwise and uniform weak convergence assumptions used for the full-sample influence-function representation.

B.3 Computation of τ^2

In practice, τ^2 need not be estimated (it cancels in the pivotal ratio), but it provides a useful measure of the asymptotic uncertainty. A consistent estimator is:

$$\hat{\tau}^2 = 4 \int_I \int_I \widehat{\delta}(1, y_1)\widehat{\delta}(1, y_2)\hat{\Sigma}(y_1, y_2)\, dy_1\, dy_2,$$

where $\hat{\Sigma}(y_1, y_2) = n^{-1}\sum_{i=1}^n \hat{\xi}_i(y_1)\hat{\xi}_i(y_2)$ and $\hat{\xi}_i(y) = \hat{\varphi}_i^N(y) - \hat{\varphi}_i^V(y)$ are the estimated influence functions obtained by substituting $(\widehat{\beta}^{\mathrm{N}}, \widehat{\theta}, \widehat{\Gamma})$ for their population counterparts in (40)–(46).

C Monte Carlo Evidence

This section provides a small Monte Carlo experiment to illustrate the finite-sample behavior of the proposed relevant-endogeneity test. The goal is not to provide an exhaustive

simulation study, but rather to verify that the test behaves in line with the asymptotic theory derived in Section 5.

C.1 Data-generating process

We simulate data from a triangular model with one exogenous regressor X, one instrument Z, one endogenous regressor Y_2, and a continuous outcome Y:

$$Y_2 = \pi_x X + \pi_z Z + V, \tag{78}$$

$$Y = \beta_0 + \beta_x X + \beta_2 Y_2 + U, \tag{79}$$

$$U = \rho V + \sqrt{1-\rho^2}\,\varepsilon, \tag{80}$$

where X, Z, V, ε are mutually independent standard normal random variables. The parameter $\rho \in [0, 1)$ controls the strength of endogeneity: if $\rho = 0$, then U is independent of V and Y_2 is exogenous; if $\rho > 0$, then U and V are correlated and Y_2 is endogenous.

Throughout the simulation we set

$$\beta_0 = 0, \qquad \beta_x = 0.5, \qquad \beta_2 = 1, \qquad \pi_x = 0.5, \qquad \pi_z = 1.$$

The test is evaluated at the covariate profile $(x, y_2) = (0, 0)$. Distribution regression is implemented using a probit link. For each threshold y, the naive distribution-regression estimator is obtained from a probit regression of $\mathbb{1}\{Y_i \le y\}$ on $(X_i, Y_{2,i})$. The IV/control-function estimator is obtained by first regressing $Y_{2,i}$ on (X_i, Z_i), then using the first-stage residuals $\widehat{V}_i$ in a probit regression of $\mathbb{1}\{Y_i \le y\}$ on $(X_i, Y_{2,i}, \widehat{V}_i)$, and finally averaging over the empirical residual distribution.

The L^2-integrals are approximated numerically over a grid of empirical quantiles of Y. The self-normalizer is computed from sequential estimates based on the first $\lfloor nt \rfloor$ observations. The nominal level is $\alpha = 0.05$.

C.2 Interior null and increasing endogeneity

Table 1 reports rejection frequencies for $\Delta = 0.05$, sample size $n = 1000$, and $B = 100$ Monte Carlo replications. The estimated distance $\widehat{d} = \sqrt{\widehat{T}_n}$ increases with the endogeneity parameter ρ, as expected. However, even for $\rho = 0.5$, the average estimated distance is only about 0.026, which remains below the relevance threshold $\Delta = 0.05$. Consequently, the test does not reject.

This result illustrates the difference between exact and relevant endogeneity. Although $\rho = 0.5$ represents substantial correlation between the structural error and the first-stage error, the resulting distributional distance at the chosen covariate profile is still smaller than the threshold $\Delta = 0.05$. The proposed test therefore correctly does not reject the

Table 1: Monte Carlo results for $\Delta = 0.05$

ρ	Δ	n	Rejection rate	Mean $\widehat{T}_n$	Mean $\widehat{d}$
0.0	0.05	1000	0.00	$2.82 \cdot 10^{-6}$	0.0014
0.2	0.05	1000	0.00	$3.95 \cdot 10^{-5}$	0.0057
0.5	0.05	1000	0.00	$7.23 \cdot 10^{-4}$	0.0263

null hypothesis of practical irrelevance.

C.3 Power as the threshold varies

Next, we fix the endogeneity strength at $\rho = 0.5$ and vary the relevance threshold Δ. Table 2 shows that the rejection probability decreases as Δ increases. This is exactly the behavior predicted by the theory: for small thresholds, the estimated distance is often larger than Δ, while for thresholds above the population distance the test rarely rejects.

Table 2: Power as the relevance threshold varies, $\rho = 0.5$

ρ	Δ	n	Rejection rate	Mean $\widehat{T}_n$	Mean $\widehat{d}$
0.5	0.005	1000	0.65	$8.41 \cdot 10^{-4}$	0.0284
0.5	0.010	1000	0.62	$7.56 \cdot 10^{-4}$	0.0268
0.5	0.020	1000	0.21	$7.83 \cdot 10^{-4}$	0.0274
0.5	0.030	1000	0.01	$7.52 \cdot 10^{-4}$	0.0269
0.5	0.040	1000	0.00	$8.13 \cdot 10^{-4}$	0.0279
0.5	0.050	1000	0.00	$8.11 \cdot 10^{-4}$	0.0277

The transition occurs around $\Delta \approx 0.027$, which is close to the average estimated distance. This confirms that the test is sensitive to the practical-relevance threshold rather than merely to the presence of nonzero endogeneity.

C.4 Power as endogeneity increases

Table 3 fixes the relevance threshold at $\Delta = 0.02$ and varies the endogeneity parameter ρ. The estimated distance increases monotonically with ρ, and the rejection frequency increases accordingly.

For $\rho = 0$ and $\rho = 0.2$, the estimated distance remains below $\Delta = 0.02$, and the test does not reject. For $\rho = 0.5$, the estimated distance exceeds the threshold and the rejection probability becomes positive. For larger values of ρ, the estimated distance is far above the threshold and the rejection probability becomes large.

Table 3: Power as endogeneity strength varies, $\Delta = 0.02$

ρ	Δ	n	Rejection rate	Mean $\widehat{T}_n$	Mean $\widehat{d}$
0.0	0.02	1000	0.00	$2.75 \cdot 10^{-6}$	0.0014
0.2	0.02	1000	0.00	$4.42 \cdot 10^{-5}$	0.0060
0.5	0.02	1000	0.23	$7.91 \cdot 10^{-4}$	0.0274
0.7	0.02	1000	0.77	$3.03 \cdot 10^{-3}$	0.0543
0.9	0.02	1000	0.88	$9.10 \cdot 10^{-3}$	0.0946

C.5 Boundary behavior

Finally, we examine the behavior of the test close to the boundary $T(x, y_2) = \Delta^2$. The previous simulations suggest that for $\rho = 0.5$ the distance is approximately $d(x, y_2) \approx 0.027$. We therefore set $\Delta = 0.027$ and run $B = 300$ replications. The results are reported in Table 4.

Table 4: Boundary behavior for $\rho = 0.5$ and $\Delta = 0.027$

ρ	Δ	n	Rejection rate	Mean $\widehat{T}_n$	Mean $\widehat{d}$
0.5	0.027	1000	0.043	$7.75 \cdot 10^{-4}$	0.0273
0.5	0.027	2000	0.043	$6.90 \cdot 10^{-4}$	0.0260

The rejection frequencies are close to the nominal level $\alpha = 0.05$. This is consistent with Theorem 5.7, which states that the rejection probability converges to α on the boundary $T(x, y_2) = \Delta^2$.

C.6 Summary of simulation evidence

The simulation results support the qualitative predictions of the asymptotic theory. First, when the distance between the naive and IV-DR estimands is below the relevance threshold, the test does not reject. Second, for fixed endogeneity strength, the rejection probability decreases as Δ increases. Third, for fixed Δ, the rejection probability increases as the endogeneity parameter ρ increases. Finally, when the threshold is chosen close to the estimated distance, the rejection probability is close to the nominal level. Overall, the Monte Carlo evidence suggests that the proposed test captures practical rather than merely statistical endogeneity.

D Empirical Application: Returns to Education in the Card Data

This section illustrates the proposed relevant-endogeneity test using the returns-to-schooling data of Card (1995), as provided in the `wooldridge` R package. The empirical design

is closely related to the Mincer-type income application in Wied (2024). The outcome variable is log wages, education is treated as potentially endogenous, and an excluded instrument is used to construct the control-function version of distribution regression.

The purpose of this section is not to provide a definitive empirical contribution on the returns to education. Rather, the goal is to demonstrate how the proposed test can be implemented in a real-data application and how its results should be interpreted.

D.1 Data and empirical specification

The data contain $n = 3010$ observations. The dependent variable is log wage, denoted by $\log(wage_i)$. The potentially endogenous regressor is years of education, denoted by $educ_i$. The excluded instrument is $nearc4_i$, an indicator for whether the individual grew up near a four-year college. The baseline set of controls is

$$X_i = (exper_i, exper_i^2, black_i, smsa_i, south_i)'.$$

The first-stage equation is therefore

$$educ_i = \pi_0 + \pi_1 nearc4_i + \pi_2 exper_i + \pi_3 exper_i^2 + \pi_4 black_i + \pi_5 smsa_i + \pi_6 south_i + V_i. \quad (81)$$

The corresponding linear wage equation used as a preliminary diagnostic is

$$\log(wage_i) = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + \beta_3 exper_i^2 + \beta_4 black_i + \beta_5 smsa_i + \beta_6 south_i + \varepsilon_i. \quad (82)$$

The first-stage coefficient on $nearc4_i$ is positive and statistically significant. In the baseline specification, the estimated coefficient is 0.3373 with a t-statistic of 4.089. The corresponding weak-instrument diagnostic from the linear IV regression is 16.718. Thus, the instrument is not obviously weak in this specification.

As a benchmark, Table 5 reports the OLS and IV estimates of the linear wage equation. The OLS coefficient on education is 0.0740, while the IV coefficient is 0.1323. However, the Wu–Hausman test does not reject exogeneity at conventional significance levels, with a p-value of 0.215. This provides a useful background for the distributional analysis: the usual mean-based Hausman test does not find strong evidence against exogeneity, but the proposed method asks a different question, namely whether the distributional effect of the IV correction is practically relevant at specific covariate profiles.

Following the empirical strategy of Wied (2024), the distributional analysis is conducted at three covariate profiles. These profiles correspond to low, middle, and high values of education and experience. The binary covariates are set equal to their sample means.

For each threshold y, the naive DR estimator is obtained from a probit regression of

Table 5: Linear OLS and IV estimates in the Card data

Variable	OLS	IV
Intercept	4.7337	3.7528
Education	0.0740	0.1323
Experience	0.0836	0.1075
Experience squared	-0.0022	-0.0023
Black	-0.1896	-0.1308
SMSA	0.1614	0.1313
South	-0.1249	-0.1049
Weak-instrument statistic	–	16.718
Wu–Hausman p-value	–	0.215

Table 6: Covariate profiles used in the Card-data application

Profile	*educ*	*exper*	$exper^2$	*black*	*smsa*	*south*
Low	10	4	16	0.234	0.713	0.404
Middle	13	8	64	0.234	0.713	0.404
High	17	15	225	0.234	0.713	0.404

$\mathbb{1}\{\log(wage_i) \le y\}$ on $(educ_i, X_i')'$. The IV/control-function DR estimator first estimates (81), then includes the first-stage residual $\widehat{V}_i$ in the threshold-specific probit regression, and finally averages over the empirical distribution of $\widehat{V}_i$. The L^2-integrals are approximated over empirical quantile grids of log wages.

D.2 Estimated naive and IV-DR distribution functions

Figure 1 displays the estimated conditional distribution functions for the three profiles in the baseline specification. The solid line is the naive DR estimator and the dashed line is the IV/control-function DR estimator.

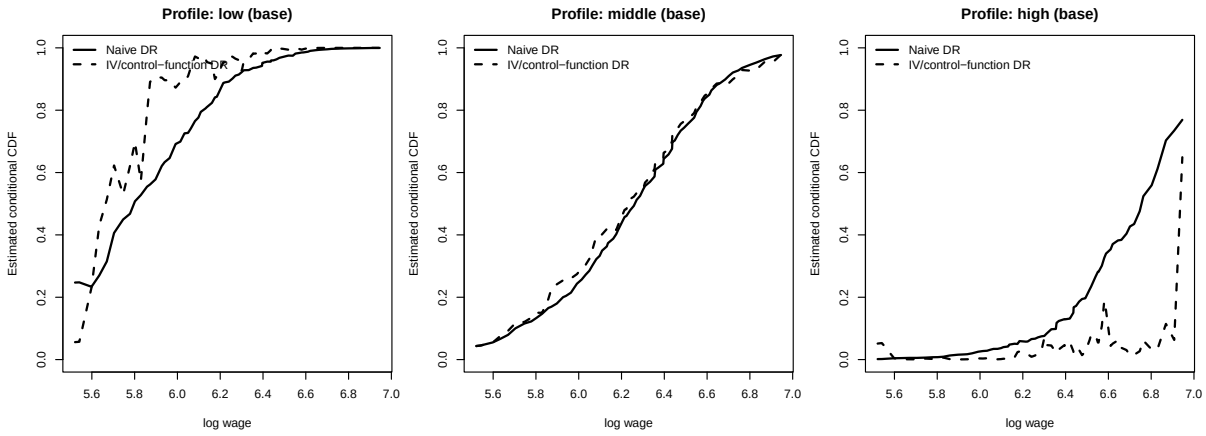

Figure 1: Naive and IV/control-function DR estimates, baseline specification

The curves are closest for the middle profile. For the low and high profiles, the estimated IV/control-function distribution differs more visibly from the naive distribution. This suggests that the distributional impact of correcting for endogeneity is heterogeneous across covariate profiles. In particular, the largest visual differences appear away from the middle profile, consistent with the motivation for analyzing the entire conditional distribution rather than only the conditional mean.

Figure 2 repeats the same exercise for a richer specification that additionally includes $smsa66$ and region indicators. The qualitative pattern remains similar: the middle profile displays the smallest difference, while the low and high profiles show larger discrepancies between the naive and IV/control-function distribution functions.

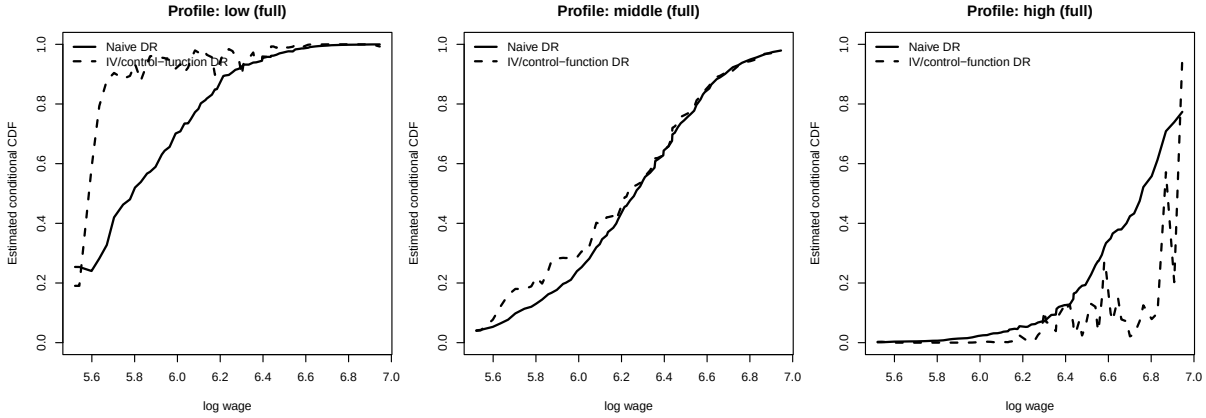

Figure 2: Naive and IV/control-function DR estimates, full-control specification

D.3 Relevant-endogeneity test in the baseline specification

The proposed test is applied to the null hypothesis

$$H_0 : d(x, y_2) \leq \Delta \qquad \text{against} \qquad H_1 : d(x, y_2) > \Delta,$$

where

$$d(x, y_2) = \left(\int_I \left[\widehat{F}^N(y|x, y_2) - \widehat{F}^V(y|x, y_2) \right]^2 dy \right)^{1/2}.$$

The baseline implementation uses $\varepsilon = 0.2$ in the self-normalizer and a central 10%–90% log-wage grid with 31 grid points. The simulated Brownian critical value for this choice is approximately $q_{0.95} = 2.69$.

Table 7 reports the test results for relevance thresholds $\Delta \in \{0.01, 0.02, 0.03, 0.05\}$. The estimated distance is largest for the high profile, where $\widehat{d} = 0.195$. The baseline test rejects practical irrelevance for the high profile for all displayed values of Δ. The low profile also has a sizeable estimated distance, $\widehat{d} = 0.148$, but its self-normalizer is large

enough that the test does not reject. The middle profile has a much smaller estimated distance and does not reject.

Table 7: Baseline relevant-endogeneity test results

Profile	Δ	$\widehat{T}_n$	$\widehat{d}$	Critical value	Reject
Low	0.01	0.0219	0.1479	0.0254	No
Low	0.02	0.0219	0.1479	0.0257	No
Low	0.03	0.0219	0.1479	0.0262	No
Low	0.05	0.0219	0.1479	0.0278	No
Middle	0.01	0.0009	0.0304	0.0129	No
Middle	0.02	0.0009	0.0304	0.0132	No
Middle	0.03	0.0009	0.0304	0.0137	No
Middle	0.05	0.0009	0.0304	0.0153	No
High	0.01	0.0380	0.1949	0.0353	Yes
High	0.02	0.0380	0.1949	0.0356	Yes
High	0.03	0.0380	0.1949	0.0361	Yes
High	0.05	0.0380	0.1949	0.0377	Yes

The baseline result therefore suggests that the IV correction is practically relevant for the high education/high experience profile, but not for the low or middle profiles. Importantly, this conclusion is distributional and profile-specific. It differs from the usual linear Wu–Hausman test, which summarizes the evidence in a single mean-based statistic.

D.4 Robustness checks

This subsection examines the sensitivity of the empirical findings to three implementation choices: the trimming parameter ε, the grid used for numerical integration over log wages, and the set of control variables.

Choice of the sequential trimming parameter. Table 8 reports results for $\varepsilon \in \{0.1, 0.2, 0.3\}$, using the baseline control set and $\Delta = 0.05$. The estimated distances $\widehat{d}$ are unchanged because they depend only on the full-sample naive and IV-DR estimates. However, the self-normalizer changes with the sequential grid. The rejection for the high profile occurs for $\varepsilon = 0.2$, but not for $\varepsilon = 0.1$ or $\varepsilon = 0.3$. This indicates that the formal rejection decision is sensitive to the sequential tuning parameter.

Choice of the log-wage integration grid. Table 9 reports robustness checks for different grids used to approximate the L^2-integral. For compactness, the table reports results for $\Delta = 0.05$. The low and middle profiles do not reject for any grid. The high profile rejects for the wider 5%–95% grid and is borderline for the central 10%–90% grids. Thus, the qualitative ranking of the estimated distances is stable: the middle profile is

Table 8: Robustness to the trimming parameter ε, $\Delta = 0.05$

ε	Profile	$\widehat{T}_n$	$\widehat{d}$	$\widehat{V}_n$	Critical value	Reject
0.1	Low	0.0219	0.1479	0.0492	0.1004	No
0.1	Middle	0.0009	0.0304	0.0063	0.0150	No
0.1	High	0.0380	0.1949	0.0626	0.1271	No
0.2	Low	0.0219	0.1479	0.0094	0.0277	No
0.2	Middle	0.0009	0.0304	0.0048	0.0152	No
0.2	High	0.0380	0.1949	0.0131	0.0376	Yes
0.3	Low	0.0219	0.1479	0.0085	0.0324	No
0.3	Middle	0.0009	0.0304	0.0033	0.0140	No
0.3	High	0.0380	0.1949	0.0110	0.0412	No

smallest, the low profile is larger, and the high profile is largest. The formal rejection decision for the high profile is, however, close to the boundary for the central grids.

Table 9: Robustness to the log-wage integration grid, $\Delta = 0.05$

Grid	Profile	$\widehat{T}_n$	$\widehat{d}$	$\widehat{V}_n$	Critical value	Reject
10–90, 21 points	Low	0.0184	0.1356	0.0102	0.0301	No
10–90, 21 points	Middle	0.0010	0.0311	0.0048	0.0157	No
10–90, 21 points	High	0.0389	0.1973	0.0127	0.0370	Yes
10–90, 31 points	Low	0.0219	0.1479	0.0094	0.0281	No
10–90, 31 points	Middle	0.0009	0.0304	0.0048	0.0154	No
10–90, 31 points	High	0.0380	0.1949	0.0131	0.0382	No
10–90, 51 points	Low	0.0219	0.1479	0.0094	0.0280	No
10–90, 51 points	Middle	0.0010	0.0319	0.0046	0.0150	No
10–90, 51 points	High	0.0379	0.1947	0.0134	0.0391	No
5–95, 51 points	Low	0.0269	0.1641	0.0139	0.0402	No
5–95, 51 points	Middle	0.0011	0.0331	0.0054	0.0171	No
5–95, 51 points	High	0.0845	0.2907	0.0254	0.0717	Yes

Choice of control variables. Finally, Table 10 compares the baseline specification with richer control sets. The richer specifications add *smsa*66, region indicators, or both. The table again reports results for $\Delta = 0.05$.

The estimated distances remain sizeable for the low and high profiles. However, the richer specifications also produce larger self-normalizers, and the formal rejection for the high profile disappears. This suggests that the evidence for practically relevant endogeneity is not fully robust to the choice of controls.

Table 10: Robustness to control variables, $\Delta = 0.05$

Controls	Profile	$\widehat{T}_n$	$\widehat{d}$	$\widehat{V}_n$	Critical value	Reject
Base	Low	0.0219	0.1479	0.0094	0.0278	No
Base	Middle	0.0009	0.0304	0.0048	0.0153	No
Base	High	0.0380	0.1949	0.0131	0.0377	Yes
Base + *smsa*66	Low	0.0401	0.2003	0.0196	0.0551	No
Base + *smsa*66	Middle	0.0012	0.0346	0.0041	0.0135	No
Base + *smsa*66	High	0.0270	0.1644	0.0101	0.0298	No
Base + regions	Low	0.0471	0.2170	0.0247	0.0688	No
Base + regions	Middle	0.0021	0.0462	0.0024	0.0089	No
Base + regions	High	0.0370	0.1924	0.0220	0.0616	No
Full	Low	0.0569	0.2386	0.0299	0.0827	No
Full	Middle	0.0028	0.0527	0.0024	0.0090	No
Full	High	0.0283	0.1682	0.0153	0.0435	No

D.5 Discussion

The Card-data application illustrates three important features of the proposed test.

First, the estimated difference between naive and IV/control-function DR curves is strongly profile-dependent. The middle profile shows only small differences, while the low and high profiles show substantially larger estimated distances. This would be missed by a purely mean-based comparison.

Second, the formal test decision depends not only on the estimated distance $\widehat{d}$, but also on the sequential self-normalizer $\widehat{V}_n$. The low profile has a sizeable estimated distance, but the sequential variability is large enough that the test does not reject. This demonstrates the role of self-normalization: the test does not simply reject whenever the raw distance is large.

Third, the empirical evidence is not fully robust. The baseline specification rejects practical irrelevance for the high profile, while richer control sets do not. Similarly, the rejection for the high profile is sensitive to the choice of the sequential trimming parameter. The most cautious interpretation is therefore not that the Card data provide conclusive evidence of practically relevant endogeneity. Rather, the application shows that the proposed method can reveal heterogeneous distributional differences between naive and IV/control-function DR, and that these differences can be formally evaluated against economically meaningful relevance thresholds.

Overall, the empirical exercise supports the usefulness of the proposed procedure as a diagnostic tool. It complements classical Hausman-type tests by asking whether the IV correction changes the estimated conditional distribution by a practically relevant amount, rather than whether exact exogeneity can be rejected.

References

Berkson, J. (1938). Some difficulties of interpretation encountered in the application of the chi-square test. *Journal of the American Statistical Association*, 33(203):526–536.

Card, D. (1995). Using geographic variation in college proximity to estimate the return to schooling. *Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp*, pages 201–222.

Chernozhukov, V., Fernández-Val, I., and Galichon, A. (2010). Quantile and probability curves without crossing. *Econometrica*, 78(3):1093–1125.

Chernozhukov, V., Fernández-Val, I., and Melly, B. (2013). Inference on counterfactual distributions. *Econometrica*, 81(6):2205–2268.

Choi, S. and Hall, P. (2000). Bootstrap confidence regions computed from autoregressions of arbitrary order. *Journal of the Royal Statistical Society, Series B*, 62(2):461–477. (distributional regression context).

Dette, H., Möllenhoff, K., and Wied, D. (2025). Practically significant differences between conditional distribution functions. arXiv:2506.06545.

Dette, H., Munk, A., and Wagner, T. (1998). Estimating the variance in nonparametric regression—what is a reasonable choice? *Journal of the Royal Statistical Society, Series B*, 60(4):751–764. (cited for the relevant-differences framework).

Foresi, S. and Peracchi, F. (1995). The conditional distribution of excess returns: An empirical analysis. *Journal of the American Statistical Association*, 90(430):451–466.

Hodges, J. L. and Lehmann, E. L. (1954). Testing the approximate validity of statistical hypotheses. *Journal of the Royal Statistical Society, Series B*, 16(2):261–268.

Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, 8(4):355–362.

Rivers, D. and Vuong, Q. H. (1988). Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of Econometrics*, 39(3):347–366.

Rothe, C. and Wied, D. (2013). Misspecification testing in a class of conditional distributional models. *Journal of the American Statistical Association*, 108(501):314–324.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.

Wellek, S. (2010). *Testing Statistical Hypotheses of Equivalence and Noninferiority*. Chapman and Hall/CRC, 2nd edition.

Wied, D. (2024). Semiparametric distribution regression with instruments and monotonicity. *Labour Economics*, 90:102565.