



Correcting for Endogeneity in Models with Bunching

Carolina Caetano^a, Gregorio Caetano^a, and Eric Nielsen^b

^aUniversity of Georgia, Athens, GA; ^bFederal Reserve Board, Washington, DC

ABSTRACT

We develop a novel control function approach in models where the treatment variable has bunching at one corner of its support. This situation typically arises when the treatment variable is a constrained choice and some observations choose the corner solution. The method exploits distributional shape restrictions but makes no exclusion restrictions. We provide estimators and establish their asymptotic behavior, prove the convergence of the bootstrap, and develop tests of the identification assumptions. An application reveals that watching television has no effect on cognitive skills and a negative effect on noncognitive skills in children.

ARTICLE HISTORY

Received November 2021
Accepted August 2023

KEYWORDS

Bunching; Constrained choice; Control function

1. Introduction

When the treatment variable is constrained to be above or below a certain threshold, bunching of observations is often seen at the threshold itself. Caetano (2015) develops a test of exogeneity in these situations based on the idea that unobservables vary discontinuously at the threshold. In this article, we show that the same idea can be leveraged further to build a correction for endogeneity, provided that further structure is imposed. Specifically, we impose a restriction on the shape of the distribution of the confounders conditional on the controls, but we allow the parameters of that distribution to be nonparametric functions of the controls. In particular, all of the controls may be endogenous. The approach does not require exclusion restrictions or specific data structures (e.g., a panel), so it can be useful when none of the well-established selection-on-unobservables identification strategies are applicable, either because they are infeasible or because they do not identify the parameter of interest.

In a linear model, the method translates to adding a generated control to the original regression. The entire approach in this case can be implemented with packaged software. We derive the asymptotic behavior of the estimator of the treatment effect coefficient, provide a consistent estimator of the standard errors, and prove the consistency of the bootstrap. We also develop tests for all the assumptions in the model. Finally, we extend the approach beyond the linear model, showing that it can be applied in many widely-used nonlinear models. Examples include correlated random effects models, partially linear models, some types of nonparametric nonseparable models, and probit models with endogeneity.

We apply the approach to estimate the effect of time spent watching television on children's skills using time diary data from the Panel Study of Income Dynamics Child Development Supplement (PSID-CDS). We find strong evidence of selection on unobservables in this application. Our correction approach

reveals that television time has insignificant, positive effects on children's cognitive skills and significant, negative effects on their noncognitive skills. We test all our identifying assumptions using the tests we propose.

Our method can be understood as a control function approach (e.g., Heckman and Robb 1985; Navarro 2010). However, unlike typical control function approaches, our method does not require exclusion restrictions, exploiting shape restrictions instead. There are alternative approaches in the literature that identify treatment effects without exclusion restrictions, but the assumptions in those methods apply to different contexts. For instance, Klein and Vella (2010) and Lewbel (2012) show that heteroscedasticity can be used to achieve identification in linear models. D'Haultfoeuille, Hoderlein, and Sasaki (2021) show that if the instrument satisfies a local irrelevance condition, then it is possible to identify the causal effect of interest in nonseparable models without an exclusion restriction. For models with binary treatment, see Millimet and Tchernis (2013) and the references therein for approaches that achieve identification without exclusion restrictions.

Our model is also related to the sample selection correction literature following Heckman (1979). That literature typically exploits distributional assumptions, as in our case. However, the key distinction is that our setup is not a sample selection model—no variable in the structural equation is censored in our context. In particular, both the outcome and the treatment variable are observed for the whole sample.

There is a large applied literature using bunching in the outcome variable (rather than in the treatment variable) to infer treatment effect parameters. Recently, there have been advancements in the theoretical treatment of identification in that context, notably in Blomquist et al. (2019) and Bertanha, McCallum, and Seeger (2020), on the identification of elasticities with respect to changes in a schedule of incentives (e.g., taxable income elasticity), and in Goff's (2020) generalization to

models where the outcome is a choice subject to a budget set with a kink. Blomquist et al. (2019) and Bertanha, McCallum, and Seegert (2020) establish that identification in that context is impossible without restrictions on the conditional distribution of the unobservables similar to those imposed in our approach. However, differently from this literature, it is not necessary in our setting to specify how the bunched variable is chosen: we specify neither the optimization function nor the budget constraint.

Bunching at one of the extremes of the distribution of the treatment variable is often observed when the treatment is a choice constrained to be nonnegative, as in the case of demand or inputs to production. Examples include behavioral variables like the consumption of vitamin supplements, cigarettes, alcohol, and coffee;¹ financial variables such as credit card debt, credit access, expenditure on advertisements, and bequests;² variables quantifying different uses of time such as exercising, working, doing homework, volunteering, and using social media;³ and count data such as the number of children, the frequency of doctor visits, and the crime rate.⁴ Moreover, our approach is a natural solution whenever exogeneity is rejected by Caetano's (2015) test, which has been applied in a variety of settings in economics, political science, and finance.⁵

We develop a general inference framework that makes extensive use of Chen, Linton, and Van Keilegom's (2003) work for extremum estimators with possibly infinite dimensional nuisance parameters. However, the primitive results for stochastic equicontinuity in that article, which would greatly simplify the analysis, cannot be applied in our context. Instead we must prove the stochastic equicontinuity directly, and for this we use results in Pakes and Pollard (1989) and Andrews (1994). We also relax Chen, Linton, and Van Keilegom's (2003) almost sure stochastic equicontinuity requirement for the bootstrap, which has no applicable primitive conditions in the literature, making it hard to defend in practice. We substitute it by stochastic equicontinuity in probability requirements which are standard and for which primitives are well known (generally Donsker classes, e.g., smooth functions, limited variation functions, etc.).

To establish the convergence of the estimator in the specific tail symmetry case we use in our application, we develop new results for the estimation of quantiles at random

(estimated) probabilities, and trimmed means at random (estimated) trimming points. Since these results may be of interest beyond the matters studied in this article, we present them in general notation which can stand alone (see Lemmas G.1, G.2, and G.3 in Appendix G.6). The trimmed mean results apply Fang and Santos's (2019) recent findings on Hadamard directional differentiability.

The article is laid out as follows. Sections 2 and 3 present the identification approach, and Sections 4 and 5 present estimators and asymptotic results. Section 6 develops tests aimed at detecting violations of all the identifying assumptions. In Section 7, we estimate the effect of time spent watching television (TV) on children's cognitive and noncognitive skills using our method. Other applications of this method can be seen in Caetano, Caetano, and Nielsen (2022) and Caetano et al. (2022). Section 8 concludes. The Appendix develops extensions to nonlinear models, presents further material on estimation and on our application, an empirical Monte Carlo study, and proofs. The code for implementing the method is available at <https://github.com/GregorioCaetano/Bunching>.

2. Correction Strategy

The main idea behind our approach can be contextualized using the example of our application, discussed in more detail in Section 7. We are interested in estimating the effect of the treatment X , hours per week the child watches TV, on the outcome Y , the child's cognitive or noncognitive skills. Figure 1 shows the unconditional cdf (left panel) and the empirical distribution (right panel) of X . About 5% of the sample is bunched at $X = 0$.

Why would this bunching occur? One explanation may be given if X is chosen in an optimization problem. Suppose that the amount of TV watching is chosen taking the family and child's characteristics, preferences and constraints into account. One of the constraints of this problem is that the amount of time spent watching TV cannot be negative. We denote the solution of the optimization problem where the nonnegativity constraint is removed as X^* . In this setting, an explanation for the bunching at $X = 0$ is that some individuals find the nonnegativity constraint binding (i.e., $X^* < 0$), so they choose a "corner solution."⁶ The variable X^* indexes all factors that affect the demand for watching TV, such as preferences for TV watching, preferences for alternative activities, and any additional constraints (e.g., all activities must add up to 24 hr a day). Note that this conceptualization of X^* motivates our approach, but it is not necessary as long as the model below holds.

¹Meta-analyses of studies estimating the effects of these variables on health outcomes include Fawzi et al. (1993), Hernán et al. (2002), Reynolds et al. (2003), Noordzij et al. (2005), Bischoff-Ferrari et al. (2005), Oken, Levitan, and Gillman (2008), and Richardson, Elliott, and Roberts (2013).

²See, for example, Peek, Rosengren, and Tootell (2003), Ekici and Dunn (2010), Bertrand et al. (2010), Brown, Coile, and Weisbenner (2010), Melzer (2011), Boserup, Kopczuk, and Kreiner (2016), and Erixson (2017).

³See, for example, Luoh and Herzog (2002), James-Burdumy (2005), Eren and Henderson (2011), Bhutani et al. (2013), Holt et al. (2013), and Boulianne (2015).

⁴See, for example, McDuffie et al. (1996), Black, Devereux, and Salvanes (2005), and Cohen (2008).

⁵See, for example, Rozenas, Schutte, and Zhukov (2017), Erhardt (2017), Pang (2017), Bleemer (2018a), Bleemer (2018b), Ferreira, Ferreira, and Mariano (2018), Lavetti and Schmutte (2018), Caetano and Maheshri (2018), De Vito, Jacob, and Müller (2019), Caetano, Kinsler, and Teng (2019), Caetano, Caetano, and Nielsen (2022), and Caetano et al. (2022).

⁶It may be difficult to conceptualize the idea that X^* can be negative, as it would mean that someone would want to choose negative amounts of TV watching. It may be easier to think of X^* at $X = 0$ as a measure of the distance from exact indifference between watching some TV and alternative activities. For instance, those at $X^* = -0.1$ have characteristics, preferences and constraints that led them to be nearly indifferent between watching TV and another activity at $X = 0$, while those at $X^* = -3$ have characteristics, preferences and constraints that led them to be farther from indifference at $X = 0$ (e.g., this family could be equal in every way to a family of type $X^* = 0$, except for having a higher relative preference for playing sports versus watching TV).

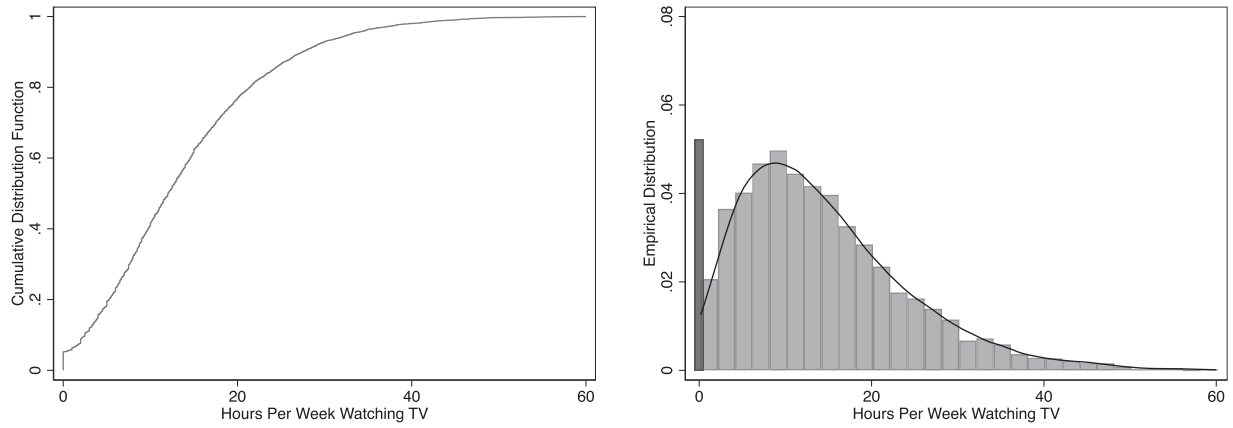


Figure 1. Unconditional distribution of X .

NOTE: The left panel shows the cumulative distribution function of X (TV time). The right panel shows the kernel density estimate along with the histogram for $X > 0$ (bandwidth equals to 2). The darker bar is the proportion of observations with $X = 0$. See Section 7 for more details about the application.

The treatment variable X is related to X^* according to⁷

$$X = \max\{0, X^*\}, \quad \text{with} \quad 0 < \mathbb{P}(X^* < 0) < 1. \quad (1)$$

The model's key requirement, $0 < \mathbb{P}(X^* < 0) < 1$ in (1), is that the nonnegativity constraint is binding for part of the population. For those observations, X^* is different from the actual choice X .

We decompose X^* into a part that is determined by the characteristics we observe in the data, Z , and a remainder η , that is,

$$X^* = Z'\pi + \eta. \quad (2)$$

We now consider the outcome equation, which is where we impose structure.

$$Y = \beta X + Z'\gamma + \delta\eta + \varepsilon, \quad \text{where} \quad \mathbb{E}[\varepsilon|X, Z, \eta] = 0. \quad (3)$$

This equation specifies the unobservable determinant of the outcome as $\delta\eta + \varepsilon$. Here, η is a sufficient index of all confounders, in the sense that, if we were able to observe and control for it, there would be no endogeneity problem. If $\delta = 0$, then X is exogenous (i.e., $\mathbb{E}[\delta\eta + \varepsilon|X, Z] = 0$), and thus β is identifiable as in the standard selection-on-observables model. If $\delta \neq 0$, then X is endogenous. We keep the linear specification here for concreteness, as linearity is often assumed by researchers in practice, but we show in Appendix A that (2) and (3) can be generalized substantially, including to some nonparametric nonseparable models.

By (1) and (2), $\mathbb{E}[\eta|X, Z] = X + \mathbb{E}[X^*|X^* \leq 0, Z]\mathbf{1}(X = 0) - Z'\pi$. This implies that, if $\mathbb{E}[X^*|X^* \leq 0, Z = z]$ can be identified for all $z \in \mathcal{Z}|X = 0$, then Z and $X + \mathbb{E}[X^*|X^* \leq 0, Z]\mathbf{1}(X = 0)$ are a proxy for the omitted variable η . Since Z is already a control in the structural equation, we would only need one additional

regressor, $X + \mathbb{E}[X^*|X^* \leq 0, Z]\mathbf{1}(X = 0)$, in order to proxy for the omitted variable. The identification equation is therefore

$$\mathbb{E}[Y|X, Z] = \beta X + Z'(\gamma - \pi\delta) + \delta(X + \mathbb{E}[X^*|X^* \leq 0, Z]\mathbf{1}(X = 0)). \quad (4)$$

As long as $\mathbb{E}[X^*|X^* \leq 0, Z]$ can be identified, and X, Z and the correction term $X + \mathbb{E}[X^*|X^* \leq 0, Z]\mathbf{1}(X = 0)$ are linearly independent (see Remark 2.1), β is identifiable.

Correcting for endogeneity thus depends on the identification of $\mathbb{E}[X^*|X^* \leq 0, Z]$. However, since X^* is not observed when it is negative, $\mathbb{E}[X^*|X^* \leq 0, Z]$ is not identifiable without additional assumptions, such as shape or parametric restrictions on the distribution of $X^*|Z$. We examine some of these assumptions and derive the corresponding expressions for $\mathbb{E}[X^*|X^* \leq 0, Z]$ in Section 3.

Our approach therefore exploits bunching and two assumptions: the linearity assumption in (3) (i.e., that it is sufficient to control for Z and η linearly), and whichever shape or distributional assumption we make for the identification of $\mathbb{E}[X^*|X^* \leq 0, Z]$. With these assumptions, we can identify $\beta, (\gamma - \pi\delta)'$ and δ in equation (4). In Section 6, we present several tests for both of these assumptions. In Appendix A, we relax the linearity assumption and show how the correction may be implemented in some nonlinear and nonparametric models.

In Section 4, we prove the asymptotic normality of the corrected regression coefficients $(\hat{\beta}, (\hat{\gamma} - \pi\hat{\delta})', \hat{\delta})'$ when the correction uses any consistent estimator $\hat{\mathbb{E}}[X^*|X^* \leq 0, Z]$ satisfying some general conditions. We also derive a consistent estimator of the standard errors and prove the consistency of the bootstrap. In Section 5, we discuss how to estimate $\hat{\mathbb{E}}[X^*|X^* \leq 0, Z]$.

Remark 2.1 (On the Rank Condition and the Probability of Bunching). The rank condition for the identification of β in (4) holds if X, Z and $X + \mathbb{E}[X^*|X^* \leq 0, Z]\mathbf{1}(X = 0)$ are linearly independent. The linear independence of X and $X + \mathbb{E}[X^*|X^* \leq 0, Z]\mathbf{1}(X = 0)$ is guaranteed if, and only if, $\mathbb{P}(X^* < 0) > 0$ and $\mathbb{P}(X > 0) > 0$. This means that the approach cannot be applied in standard models without bunching, as bunching is a necessary condition for identification. The linear independence of Z, X and $X + \mathbb{E}[X^*|X^* \leq 0, Z]\mathbf{1}(X = 0)$ imposes the constraint that no linear combination of the latter two terms be included in Z .

⁷All equations and results involving random variables should be read as holding almost surely. \mathbb{P} denotes the probability, and details about the implied probability spaces and conditional sigma-algebras are self-evident and thus omitted. \mathcal{Z} denotes the support of the distribution of Z , and $\mathcal{Z}|A$ denotes the support of the conditional distribution of Z given A . For brevity, we often mention the support of the variable V when we mean the support of the distribution of V . Finally, the expectation \mathbb{E} is assumed to exist wherever written.

In Appendix B.1, we derive the expression of the rank condition as a function of the bunching probability for the non-extreme cases (i.e., $0 < \mathbb{P}(X = 0) < 1$), and we show that it is not generally possible to separate the effect of the probability of bunching from other functions of the distribution of X^* which also affect the rank condition. Thus, any study of how the rank condition changes with $\mathbb{P}(X = 0)$ will depend on how the distribution of X^* is modified in order to achieve a shift in $\mathbb{P}(X = 0)$.

In Appendix F.3, we simulate changes in $\mathbb{P}(X = 0)$ inside our Monte Carlo experiments by changing $\mathbb{E}[X^*|Z]$, allowing all quantities in the rank condition to vary as a consequence. We find that the variance of our estimators is stable for bunching rates away from extremes, and increases sharply near extreme bunching probabilities.

Remark 2.2 (On the Role of Controls). The approach can be implemented without controls, provided the assumptions are valid (i.e., (3) holds with $Z \equiv 1$, and $\mathbb{E}[X^*|X^* \leq 0]$ is identified, see Section 3). In this case, there are two constants π_0 and γ_0 such that $\mathbb{E}[\eta|X, Z] = X + \mathbb{E}[X^*|X^* \leq 0]\mathbf{1}(X = 0) - \pi_0$, and therefore $\mathbb{E}[Y|X] = (\gamma_0 - \pi_0\delta) + \beta X + \delta(X + \mathbb{E}[X^*|X^* \leq 0]\mathbf{1}(X = 0))$. If $\mathbb{P}(X^* < 0) > 0$ and $\mathbb{P}(X > 0) > 0$, the correction term is linearly independent of X , and so β is identifiable.

If (3) holds for a given Z , what is the consequence of using the vector of controls $\tilde{Z} \neq Z$? In Appendix B.2, we study this situation in detail. In general, using a different control vector impacts the identification of β through the parts of ε and of the omitted controls (the elements of Z not included in \tilde{Z}) that are not linearly predicted by X^* and the included controls \tilde{Z} .

Remark 2.3 (On the Effect of Misidentification of $\mathbb{E}[X^*|X^* \leq 0, Z]$). What is the consequence of using a different function $\tilde{e}(Z)$, instead of the true expectation $\mathbb{E}[X^*|X^* \leq 0, Z]$? In Appendix B.3, we show that this introduces the omitted variable $(\tilde{e}(Z) - \mathbb{E}[X^*|X^* \leq 0, Z])\mathbf{1}(X = 0)$ into the model, and study its consequences in detail. In general, using a mistaken expectation in the correction impacts the identification of β through δ and the part of the omitted variable which is not predicted by Z and the correction term $X + \tilde{e}(Z)\mathbf{1}(X = 0)$. We also show that, if the rank condition holds when using the true correction term $X + \mathbb{E}[X^*|X^* \leq 0, Z]\mathbf{1}(X = 0)$, then the identification of β is generally robust to small mistakes in the identification of the expectation.

3. Identification of $\mathbb{E}[X^*|X^* \leq 0, Z]$

Since we only observe X^* when it is positive, $\mathbb{E}[X^*|X^* \leq 0, Z]$ is an out-of-sample moment, and thus it cannot be nonparametrically identified. Nevertheless, if $X^*|Z$ has a stable distribution that maintains certain properties on its entire support, then the fact that we do observe the distribution of $X^*|Z$ when $X^* > 0$ may be sufficient for the identification of this out-of-sample expectation.

More specifically, identification of the expectation can be achieved by relying on properties of the distribution of $X^*|Z = z$ that hold both when X^* is negative and for at least some positive values of X^* . The following theorem, proved in Appendix G.1,

states formally a sufficient condition for identification of $\mathbb{E}[X^*|X^* \leq 0, Z = z]$. Define $F_{X|Z=z}$ and $F_{X^*|Z=z}$ as the cumulative distribution functions (cdf's) of $X|Z = z$ and $X^*|Z = z$, respectively, and \mathcal{F} the space of cdf's in \mathbb{R} .

Theorem 3.1. For a given value $z \in \mathcal{Z}$, if $F_{X^*|Z=z}(x) = G(x, F_{X|Z=z})$ for an identifiable functional $G: \mathbb{R} \times \mathcal{F} \rightarrow [0, 1]$, then $\mathbb{E}[X^*|X^* \leq 0, Z = z]$ is identified.

Parametric families usually satisfy the condition in [Theorem 3.1](#), provided all the parameters that determine $\mathbb{E}[X^*|X^* \leq 0, Z = z]$ also affect the distribution of $X|Z = z$.⁸

Next, we consider two large classes of distributions that satisfy [Theorem 3.1](#), and for which the identification of $\mathbb{E}[X^*|X^* \leq 0, Z = z]$ can be obtained in closed-form. Note that the identification of $\mathbb{E}[X^*|X^* \leq 0, Z = z]$ for a specific z is not necessarily tied to the identification of the expectation for a different value z' . Therefore, different identification assumptions may be made for different values of Z . Below, we denote the generalized inverse of a cdf $F \in \mathcal{F}$ as $F^{-1}(p) = \inf\{x; F(x) \geq p\}$.

Tail Symmetric Distributions

The class of distributions that are symmetric in the tail is formally characterized by the following condition:

Assumption 1. (Conditional Tail Symmetry) Suppose that $\mathbb{P}(X = 0|Z = z) \leq 0.5$ and that, for $x \leq 0$, $F_{X^*|Z=z}(x) = 1 - F_{X^*|Z=z}(F_{X^*|Z=z}^{-1}(1 - F_{X^*|Z=z}(0)) - x)$.

[Figure 2](#) shows an example of a distribution that is symmetric in the tails. Following [Powell \(1986\)](#), conditional symmetry assumptions have been used extensively in the censoring literature. For a large list of citations in the applied and theoretical literature which use conditional symmetry assumptions, see [Chen and Tripathi \(2017\)](#). This class includes many well-known distribution families, such as Gaussian, Student's- t , logistic, and uniform. Tail symmetry is a weaker assumption than symmetry, since it also allows distributions that are not symmetric everywhere, as depicted in [Figure 2](#).

Since for $x < 0$, $F_{X^*|Z=z}^{-1}(1 - F_{X^*|Z=z}(0)) - x > 0$, it follows that $F_{X^*|Z=z}(F_{X^*|Z=z}^{-1}(1 - F_{X^*|Z=z}(0)) - x) = F_{X|Z=z}(F_{X|Z=z}^{-1}(1 - F_{X|Z=z}(0)) - x)$, and thus Tail Symmetric distributions satisfy the condition of [Theorem 3.1](#) with functional $G(x, F) = 1 - F(F^{-1}(1 - F(0)) - x)$ for any distribution function F . The conditional expectation can be identified using the formula

$$\begin{aligned} \mathbb{E}[X^*|X^* \leq 0, Z = z] &= F_{X|Z=z}^{-1}(1 - F_{X|Z=z}(0)) \\ &\quad - \mathbb{E}[X|X \geq F_{X|Z=z}^{-1}(1 - F_{X|Z=z}(0)), Z = z]. \end{aligned} \quad (5)$$

⁸This is not satisfied if elements of the parameter vector affect $F_{X^*|Z=z}$ only when $x \leq 0$. One example is the piecewise-linear distribution family with a kink at zero, which has cdf $F_{X^*|Z=z}(x) = \omega_2 + \omega_1 \cdot x \cdot \mathbf{1}\left(-\frac{\omega_2}{\omega_1} \leq x \leq 0\right) + \omega_2 \cdot x \cdot \mathbf{1}\left(0 \leq x \leq \frac{1}{\omega_2} - 1\right) + \mathbf{1}\left(z \geq \frac{1}{\omega_2} - 1\right)$. The expectation satisfies $\mathbb{E}[X^*|X^* \leq 0, Z = z] = -\frac{1}{3} \frac{\omega_2}{\omega_1}$, and therefore varies with ω_1 . However, $F_{X^*|Z=z}(x)$ varies with ω_1 only when $x \leq 0$, so $F_{X|Z=z}(x) = \omega_2 \cdot \mathbf{1}(x \geq 0) + \omega_2 \cdot x \cdot \mathbf{1}\left(0 \leq x \leq \frac{1}{\omega_2} - 1\right) + \mathbf{1}\left(z \geq \frac{1}{\omega_2} - 1\right)$ cannot identify ω_1 . The expectation is not identifiable in this family.

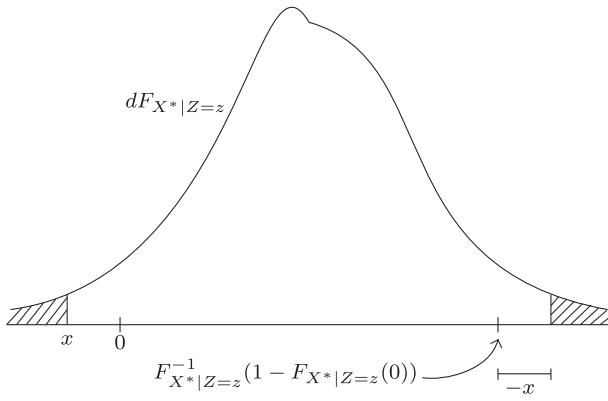


Figure 2. Tail symmetric density.

Location-Scale Distributions

The conditional location-scale class generated by the function H is characterized by the following condition.

Assumption 2. (Conditional Location-Scale) Suppose that $F_{X^*|Z=z}(x) = H\left(\frac{x-\mu_z}{\sigma_z}\right)$ for $H \in \mathcal{F}$ identifiable, and some pair (μ_z, σ_z) , where $\sigma_z > 0$.

The location-scale family is a well-known class of distributions studied at length in statistics, finance and decision theory (see e.g., Meyer 1987; Wong and Ma 2008; Hazra et al. 2017). All parametric symmetric families, such as Gaussian, Student's- t , logistic, and uniform are location-scale, as well as some parametric asymmetric families, such as the two-parameter exponential. Some nonparametric families also satisfy this condition, such as the class of symmetric distributions with nonnegative median.⁹ Because $-\mu_z/\sigma_z = H^{-1}(F_{X|Z=z}(0))$ and $\sigma_z = (H^{-1}(F_{X|Z=z}(1)) - H^{-1}(F_{X|Z=z}(0)))^{-1}$, location-scale families satisfy the condition of Theorem 3.1 with functional $G(x, F) = H(H^{-1}(F(1))x + H^{-1}(F(0))(1 - x))$.

Define the terms $\mu_H(z) = \int v dH(v)$, $\tau_H(z) = H^{-1}(F_{X|Z=z}(0))$, and $\lambda_H(z) = -F_{X|Z=z}(0)^{-1} \cdot \int_{-\infty}^{\tau_H(z)} v dH(v)$. Then, the conditional expectation can be identified using the formula

$$\begin{aligned} \mathbb{E}[X^*|X^* \leq 0, Z = z] &= \frac{-\mathbb{E}[X|Z = z](\tau_H(z) + \lambda_H(z))}{F_{X|Z=z}(0)(\tau_H(z) + \lambda_H(z)) - (\tau_H(z) - \mu_H(z))}. \end{aligned} \quad (6)$$

Equation (6) is a deterministic function of H , except for the conditional expected treatment $\mathbb{E}[X|Z = z]$, and the conditional probability of bunching $F_{X|Z=z}(0)$. Note that this equation is well defined because we only need to identify the expectation for $z \in \mathcal{Z}|X = 0$, which implies that $F_{X^*|Z=z}(0) > 0$ (this follows from a more general claim proved in Appendix G.1).

Example 3.1 (The Gaussian Family). Let Φ be the standard normal cdf, ϕ its derivative, and $\lambda = \phi/\Phi$. The Gaussian family is the set of distribution functions which satisfy $F_{X^*|Z=z}(x) = \Phi\left(\frac{x-\mu_z}{\sigma_z}\right)$ for some pair (μ_z, σ_z) , where $\sigma_z > 0$.

⁹The symmetric family is location-scale with function $H(u) = 1 - F_{X|Z=z}(-u)$ if $u \geq 0$, and $H(u) = F_{X|Z=z}(u + 2\text{med}(X|Z = z))$, if $u > 0$, where $\text{med}(X|Z = z)$ is the median of $X|Z = z$. In this case, $\mu_z = 2\text{med}(X|Z = z)$ and $\sigma_z = 1$.

The Gaussian family satisfies Assumption 2 with $H = \Phi$, and is therefore location-scale. The expectation may thus be identified using (6), which in this case gives the formula

$$\begin{aligned} \mathbb{E}[X^*|X^* \leq 0, Z = z] &= -\frac{\mathbb{E}[X|Z = z]}{1 - F_{X|Z=z}(0)} \\ &\cdot \frac{\Phi^{-1}(F_{X|Z=z}(0)) - \lambda(\Phi^{-1}(F_{X|Z=z}(0)))}{\Phi^{-1}(F_{X|Z=z}(0)) - \lambda(-\Phi^{-1}(F_{X|Z=z}(0)))}. \end{aligned} \quad (7)$$

If $\mathbb{P}(X = 0|Z = z) \leq 0.5$, the Gaussian family also satisfies Assumption 1, and is therefore tail symmetric. In this case, the expectation may be also identified using (5).

4. General Estimation and Asymptotic Results

For a given estimator of the expectation $\hat{\mathbb{E}}[X^*|X^* \leq 0, Z]$, estimation of β follows (4) and consists of an OLS regression of Y onto X, Z and the estimated correction term, $X + \hat{\mathbb{E}}[X^*|X^* \leq 0, Z]\mathbf{1}(X = 0)$. The coefficient of X is $\hat{\beta}$; the coefficient of the correction term is $\hat{\delta}$.

We prove in this section the asymptotic normality of $\hat{\beta}$ and $\hat{\delta}$, propose an estimator of the asymptotic variance, and prove its consistency. We also prove the consistency of the bootstrap. Supplementary details and further results are in Appendix C.1. Our theorems allow for broad classes of estimators of $\mathbb{E}[X^*|X^* \leq 0, Z]$, including estimators not discussed in this paper. In Section 5, we examine the estimation of $\mathbb{E}[X^*|X^* \leq 0, Z]$, and in Appendix C.2 we prove that the specific estimators used in the empirical application satisfy the conditions of the general theorems in this section.

Assumption 3. Denote $\psi_0 = \mathbb{E}[X^*|X^* \leq 0, Z = \cdot]$ and $\hat{\psi} = \hat{\mathbb{E}}[X^*|X^* \leq 0, Z = \cdot]$. Let the vector of regressors be $W = (X, Z', X + \psi_0(Z)\mathbf{1}(X = 0))'$ with W_i an observation of this vector. The corresponding parameter vector is denoted $\theta_0 = (\beta, (\gamma - \pi\delta)', \delta)'$. Suppose that Assumption 4 in Appendix C.1 holds, and

- (i) The observations $\{(Y_i, X_i, Z_i')\}_{i=1}^n$ are independent.
- (ii) $\sup_{z \in \mathcal{Z}} \left| \hat{\psi}(z) - \psi_0(z) \right| = o_p(n^{-1/4})$.
- (iii) For some finite matrix Ω , $\sqrt{n}\mathbb{E}[(\hat{\psi}(Z_i) - \psi_0(Z_i))\mathbf{1}(X_i = 0)W_i] \rightarrow_d \mathcal{N}(0, \Omega)$.¹⁰

Assumption 4 in Appendix C.1 includes the technical requirements that several moments are bounded (equivalent to White 1980), that θ_0 belongs to a compact set, and that

¹⁰The use of \mathbb{E} in this expression gives the impression that the left hand side is constant. However, the expectation here is taken with respect to X_i and Z_i , but conditional on the data that generated $\hat{\psi}$. Thus, the left hand side term is a random variable. This notation can be confusing, but it is standard in the empirical process literature (e.g., Chen, Linton, and Van Keilegom 2003). To clarify the notation further, we give an example for two variables V and Q , where V is discrete, assuming values $\{v_1, \dots, v_M\}$ with probabilities p_1, \dots, p_M . Suppose that $\hat{\xi}(v_m) = \frac{1}{n} \sum_{j=1}^n Q_j \mathbf{1}(V_j = v_m)$ and $\xi(v) = \mathbb{E}[Q\mathbf{1}(V = v_m)]$. Then, the notation $\sqrt{n}\mathbb{E}[\hat{\xi}(V_i) - \xi(V_i)]$ is equivalent to $\sqrt{n} \frac{1}{n} \sum_{j=1}^n (Q_j \mathbb{P}(V = v_j) - \mathbb{E}[\mathbb{E}[Q\mathbf{1}(V = v_j)]])$ = $\sqrt{n} \frac{1}{n} \sum_{j=1}^n \sum_{m=1}^M \mathbf{1}(V_j = v_m)(Q_j p_m - \mathbb{E}[Q|V = v_m] p_m)$ = $\sum_{m=1}^M p_m \sqrt{n} \frac{1}{n} \sum_{j=1}^n (Q_j - \mathbb{E}[Q|V = v_m]) \mathbf{1}(V_j = v_m)$. Supposing that the data is iid and $\text{var}(Q|V = v_m) = \zeta_m^2$, then $\Omega = \sum_{m=1}^M p_m (\zeta_m^2 p_m)$.

the expectation and its estimator are well behaved functions. [Assumption 3\(i\)](#) allows the data to not be identically distributed. [Assumption 3\(ii\)](#) allows the expectation to be estimated at non-parametric rates, provided the estimator is uniformly consistent at a minimum rate of $n^{1/4}$. Given [Assumptions 3\(ii\)](#) and [4\(iii\)](#), [Assumption 3\(iii\)](#) may be substituted by $\frac{1}{\sqrt{n}} \sum_{i=1}^n (\hat{\psi}(Z_i) - \psi_0(Z_i)) \mathbf{1}(X_i = 0) W_i \rightarrow_d \mathcal{N}(0, \Omega)$ (this follows from Lemma 19.24 in Van der Vaart 1998). Note that the first element of $(\hat{\psi}(Z_i) - \psi_0(Z_i)) \mathbf{1}(X_i = 0) W_i$ is equal to zero, and therefore both the first row and the first column of Ω are equal to zero. Other rows and columns of Ω may also be equal to zero, for example if Z includes X^2 . The notation $\mathcal{N}(0, \Omega)$ thus should be understood as a random variable with a degenerate normal distribution with (singular) covariance matrix Ω .

[Theorem 4.1](#) establishes that the coefficients estimated with the correction approach are asymptotically normal at parametric rates.

Theorem 4.1. If equations (1), (2), and (3), and [Assumption 3](#) hold, then

$$\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow_d \mathcal{N}(0, \Sigma + \delta^2 \mathbb{E}[WW']^{-1} \Omega \mathbb{E}[WW']^{-1}),$$

where Σ is the asymptotic covariance matrix of the coefficients of a regression of Y onto X, Z , and $X + \mathbb{E}[X^* | X^* \leq 0, Z] \mathbf{1}(X = 0)$.¹¹

The proof ([Appendix G.2](#)) applies [Theorem 2](#) in Chen, Linton, and Van Keilegom (2003). Establishing the stochastic equicontinuity condition 2.5' in Chen, Linton, and Van Keilegom (2003) is challenging. For this, we show that [Lemma 2.17](#) in Pakes and Pollard (1989) can be directly applied in our context.

Next, we present an estimator of the asymptotic variance. Let \hat{W} be the matrix of regressors, with rows equal to $\hat{W}_i := (X_i, Z_i', X_i + \hat{\psi}(Z_i) \mathbf{1}(X_i = 0))'$. Let $\hat{D} = \text{diag}\{(Y_i - \hat{W}_i' \hat{\theta})^2\}_{i=1}^n$ be the diagonal matrix of the square of the residuals. Finally, let \hat{V} be a matrix with row i equal to $(\hat{C}_{i1} \mathbf{1}(X_i = 0, X_i = 0), \dots, \hat{C}_{in} \mathbf{1}(X_i = 0, X_i = 0))'$, where the \hat{C}_{ij} are estimators of the asymptotic covariance of $\hat{\mathbb{E}}[X^* | X^* \leq 0, Z = Z_i]$ and $\hat{\mathbb{E}}[X^* | X^* \leq 0, Z = Z_j]$, precisely defined in [Assumption 5](#) in [Appendix C.1](#). Then, an estimator of the asymptotic variance of $\hat{\theta}$ is

$$\begin{aligned} \hat{V}_\theta &= \left(\frac{\hat{W}' \hat{W}}{n} \right)^{-1} \left(\frac{\hat{W}' \hat{D} \hat{W}}{n} \right) \left(\frac{\hat{W}' \hat{W}}{n} \right)^{-1} \\ &+ \delta^2 \left(\frac{\hat{W}' \hat{W}}{n} \right)^{-1} \left(\frac{\hat{W}' \hat{V} \hat{W}}{n^2} \right) \left(\frac{\hat{W}' \hat{W}}{n} \right)^{-1}. \end{aligned}$$

The first term is simply the Eicker-White covariance estimator in a regression of Y onto X, Z and $X + \hat{\psi}(Z) \mathbf{1}(X = 0)$. The second term is the penalty resulting from the fact that we are using an estimate of the expectation, instead of the true value $\psi_0(Z)$.

Theorem 4.2. If equations (1) and (2), and [Assumptions 3](#) and [5](#) (in [Appendix C.1](#)) hold, then

$$\hat{V}_\theta \rightarrow_p \Sigma + \delta^2 \mathbb{E}[WW']^{-1} \Omega \mathbb{E}[WW']^{-1}.$$

The proof of this theorem is in [Appendix G.3](#). It uses a special Strong Law of Large numbers for U-statistics when the

¹¹Precisely, $\Sigma = \mathbb{E}[WW']^{-1} \lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{1}{n} \mathbb{E}[\text{var}(Y_i | X_i, Z_i) W_i W_i'] \mathbb{E}[WW']^{-1}$. The limit is well defined by [Assumption 3\(i\)](#) and [Assumption 4\(i\)](#), and is established in [White \(1980\)](#).

data is independent but not identically distributed. [Assumption 5](#) is presented in [Appendix C.1](#) for brevity. It requires further bounded moments, as well as that the \hat{C}_{ij} are uniformly consistent.

When the expectation estimator converges weakly at the rate \sqrt{n} , both [Assumption 3\(iii\)](#) in [Theorem 4.1](#) and the uniform consistency of the \hat{C}_{ij} in [Theorem 4.2](#) ([Assumption 5\(iii\)](#) in [Appendix C.1](#)) do not need to be verified because they always hold; see [Remark C.1](#) in [Appendix C.1](#). When Z has a discrete or discretized support, the conditions of [Theorems 4.1](#) and [4.2](#) are further simplified; see [Remark C.2](#) in [Appendix C.1](#).

Letting \hat{V}_β be the first element in the matrix \hat{V}_θ , the standard error of $\hat{\beta}$ is $\sqrt{\hat{V}_\beta/n}$. The first term in \hat{V}_β can be obtained directly from packaged software as the square of the Eicker-White standard errors of $\hat{\beta}$ on a regression of Y onto X, Z and $X + \hat{\psi}(Z) \mathbf{1}(X = 0)$. The second term in \hat{V}_β is the first element in the second term in \hat{V}_θ divided by n .

Now we establish that the ordinary nonparametric bootstrap can consistently estimate the distribution of $\sqrt{n}(\hat{\theta} - \theta_0)$ for an iid sample for a wide class of estimators $\hat{\psi}(Z)$. Denote the estimator of θ using a bootstrap sample as $\hat{\theta}^b$.

Theorem 4.3. If (1), (2), and (3), and [Assumptions 3](#) and [6](#) (in [Appendix C.1](#)) hold, then

$$\begin{aligned} \sqrt{n}(\hat{\theta}^b - \hat{\theta}) &\rightarrow_d \mathcal{N}(0, \Sigma + \delta^2 \mathbb{E}[WW']^{-1} \Omega \mathbb{E}[WW']^{-1}) \\ &\text{in } \mathbb{P}^b\text{-probability.} \end{aligned}$$

This theorem follows from [Theorems 3](#) and [B](#) in Chen, Linton, and Van Keilegom (2003), and its proof is presented in [Appendix G.4](#). [Theorem B](#) in Chen, Linton, and Van Keilegom (2003) makes a requirement of almost sure stochastic equicontinuity.¹² This is a difficult condition to justify in applications, which is why we would like to avoid requiring it here.¹³ Instead, we bypass the need for almost sure stochastic equicontinuity and prove that, in our context and given the other assumptions of [Theorem 4.3](#), this condition may be substituted by two standard stochastic equicontinuity in probability conditions ([Assumptions 6\(iv\)](#) and [6\(vb\)](#) in [Appendix C.1](#)), for which well-known primitive conditions exist (e.g., Donsker classes).

5. Estimation of $\mathbb{E}[X^* | X^* \leq 0, Z]$

5.1. Parametric Estimation

If the identification of $\mathbb{E}[X^* | X^* \leq 0, Z]$ relies on a parametric assumption on the joint distribution of X^* and Z , that is, $X^* | Z = z \sim G(x, z; \kappa)$ for a known function G and finite parameter vector κ , then κ can be estimated by censored maximum likelihood.

¹²Specifically, [Theorem B](#) requires that [Assumption 2.5'](#) in Chen, Linton, and Van Keilegom (2003) holds with "in probability" replaced by "almost surely." The direct translation of this condition to our context is expressed in [Footnote 25](#) in [Appendix C.1](#).

¹³Pötscher and Prucha (1994) and Jenish and Prucha (2009) discuss almost sure stochastic equicontinuity for establishing Uniform Laws of Large Numbers. The primitives explored in those papers are not sufficient to establish almost sure stochastic equicontinuity with a \sqrt{n} denominator as required by Chen, Linton, and Van Keilegom (2003). We did not find other references of primitives of almost sure stochastic equicontinuity, and a general treatment of this condition in the lines of Pakes and Pollard (1989) is an open question.

Supposing G admits a density $g(x, z; \kappa)$, the conditional log-likelihood function is

$$L(\kappa) = \sum_{i=1}^n \left(\mathbf{1}(X_i > 0) \log(g(X_i, Z_i; \kappa)) + \mathbf{1}(X_i = 0) \log \left(\int_{-\infty}^0 g(x, Z_i; \kappa) dx \right) \right),$$

and the expectation estimator is

$$\hat{\mathbb{E}}[X^* | X^* \leq 0, Z] = \left(\int_{-\infty}^0 x g(x, Z; \hat{\kappa}) dx \right) / \int_{-\infty}^0 g(x, Z; \hat{\kappa}) dx.$$

The asymptotic properties of this estimator may be derived by applying the Delta Method and the Continuous Mapping Theorem to standard results on the asymptotic behavior of the censored maximum likelihood estimator (e.g., van der Vaart 1994).

Example 5.1 (Homoscedastic Tobit Estimator). Suppose that $F_{X^*|Z}$ belongs to the Gaussian family (Example 3.1) with $\mu_Z = Z'\mu$ and $\sigma_Z^2 = \sigma^2$. This condition and (1) specify a standard homoscedastic Tobit model (Tobin 1958). Then,

$$\mathbb{E}[X^* | X^* \leq 0, Z = z_l] = Z'\mu - \sigma \lambda \left(-\frac{Z'\mu}{\sigma} \right). \quad (8)$$

The parameters μ and σ^2 can be estimated by a Tobit regression of X onto Z with censoring below zero. This can be implemented with any packaged Tobit software. Letting $\hat{\mu}$ and $\hat{\sigma}$ be the resulting estimators of the coefficients of Z and standard deviation, respectively, then $\hat{\mathbb{E}}[X^* | X^* \leq 0, Z]$ is obtained by substituting these into (8). Details regarding the asymptotic normality of this estimator and consistency of the standard errors can be found in Footnote 27 in Appendix C.2.1.

5.2. Nonparametric Estimation with Discrete Z

Suppose that $F_{X^*|Z=z} = G(x, F_{X|Z=z})$, as in Theorem 3.1, for a known G . If Z has a discrete support $\mathcal{Z} = \{z_1, \dots, z_L\}$, then it is possible to estimate the distribution of $X^*|Z = z_l$ as $\hat{F}_{X^*|Z=z_l}(x) = G(x, \hat{F}_{X|Z=z_l})$, where $\hat{F}_{X|Z=z_l}(\cdot) = \left(\frac{1}{n} \sum_{i=1}^n \mathbf{1}(X_i \leq \cdot, Z_i = z_l) \right) / \left(\frac{1}{n} \sum_{i=1}^n \mathbf{1}(Z_i = z_l) \right)$. Then,

$$\hat{\mathbb{E}}[X^* | X^* \leq 0, Z = z_l] = -\hat{F}_{X|Z=z_l}(0)^{-1} \int_{-\infty}^0 G(x, \hat{F}_{X|Z=z_l}) dx. \quad (9)$$

The asymptotic properties of this estimator can be derived using empirical process theory, observing that $F_{X|Z=z_l}$ is a function of the joint distribution of X and Z (see Step 2 in Appendix G.6), and therefore $\mathbb{E}[X^* | X^* \leq 0, Z = z]$ can be written as an empirical process.

Example 5.2 (Semiparametric Tobit Estimator). Suppose that $F_{X^*|Z}$ belongs to the Gaussian family (Example 3.1). For a given z_l , this condition and (1) imply that $X^*|Z = z_l$ satisfies a standard Tobit regression model with constant mean μ_{z_l} and constant variance $\sigma_{z_l}^2$ (Tobin 1958). Then,

$$\mathbb{E}[X^* | X^* \leq 0, Z = z_l] = \mu_{z_l} - \sigma_{z_l} \lambda \left(-\frac{\mu_{z_l}}{\sigma_{z_l}} \right). \quad (10)$$

Because $\mathbb{P}(Z = z_l) > 0$, the parameters μ_{z_l} and $\sigma_{z_l}^2$ can be estimated by a Tobit regression of X onto a constant with censoring below zero, using only observations such that $Z = z_l$. This can be implemented with any packaged Tobit software. Letting $\hat{\mu}_{z_l}$ and $\hat{\sigma}_{z_l}$ be the resulting estimators of the mean and standard deviation, $\hat{\mathbb{E}}[X^* | X^* \leq 0, Z = z_l]$ is then obtained by substituting these estimates into (10). Details regarding the asymptotic normality of this estimator and consistency of the standard errors can be found in Appendix C.2.1.

Example 5.3 (Tail Symmetry Estimator). Supposing that $F_{X^*|Z=z_l}$ satisfies Assumption 1, the expectation identification formula in (5) holds. That formula has three components which have to be estimated and then substituted into (5). First, we estimate the bunching probability using sample frequencies:

$$\hat{F}_{X|Z=z_l}(0) = \frac{\sum_{i=1}^n \mathbf{1}(X_i = 0, Z_i = z_l)}{\sum_{i=1}^n \mathbf{1}(Z_i = z_l)}.$$

Second, we estimate the quantile $F_{X|Z=z_l}^{-1}(1 - \hat{F}_{X|Z=z_l}(0))$ by substituting the empirical quantile of the distribution of X among the observations such that $Z = z_l$. Finally, we estimate the conditional trimmed mean using

$$\begin{aligned} \hat{\mathbb{E}}[X | X \geq \hat{F}_{X|Z=z_l}^{-1}(1 - \hat{F}_{X|Z=z_l}(0)), Z = z_l] &= \\ &= \frac{\sum_{i=1}^n X_i \cdot \mathbf{1}(X_i \geq \hat{F}_{X|Z=z_l}^{-1}(1 - \hat{F}_{X|Z=z_l}(0)), Z_i = z_l)}{\sum_{i=1}^n \mathbf{1}(X_i \geq \hat{F}_{X|Z=z_l}^{-1}(1 - \hat{F}_{X|Z=z_l}(0)), Z_i = z_l)}. \quad (11) \end{aligned}$$

Theorems establishing the asymptotic normality of this estimator and consistency of the standard errors can be found in Appendix C.2.2. Although the estimator itself is very simple, the asymptotic result is very challenging to prove. The proof can be found in Appendix G.6 which, although long, can be well understood by reading the titles of the several steps. We would like to call the reader's attention, in particular, to the intermediate lemmas (Lemmas G.1 and G.2) and surrounding discussion, as these are of general interest for the estimation of quantiles at random (e.g., estimated) probabilities and trimmed means at random (e.g., estimated) trimming locations. These results are stated and proven in self-contained notation, and thus they should be understandable (and useful) independently of this paper's concerns. Lemma G.3 may also be useful, as it derives the Hadamard derivative of the conditional distribution, which allows one to transform an estimator that is a functional of the estimated conditional distribution into an empirical process estimator.

5.3. Nonparametric Estimation (General Case)

Estimation of the expectation in the general case requires the use of some multivariate nonparametric estimation technique. If G in Theorem 3.1 is known, a general framework for estimation can follow (9), substituting z_l with z . In this case, $F_{X|Z=z}(\cdot)$ may be estimated with a specialized conditional probability estimator (e.g., Durrieu et al. 2015 and the citations therein). Another option, since $F_{X|Z=z}(\cdot) = \mathbb{E}[\mathbf{1}(X \leq \cdot) | Z = z]$, is to estimate this quantity as a nonparametric regression of $\mathbf{1}(X \leq \cdot)$ onto Z at z . This opens up a number of possibilities, starting from the

basic Nadaraya-Watson kernel regression,¹⁴ or any of the classic nonparametric regression techniques (see, e.g., Masry 1996 for local polynomials and Song 2008 for series estimators), all the way to modern statistical learning prediction estimators (see Hastie, Tibshirani, and Friedman 2009 for several examples).

In the tail symmetric case (Assumption 1), $\mathbb{E}[X^*|X^* \leq 0, Z]$ may be estimated following (5). The conditional quantile $F_{X|Z=z}^{-1}(1 - F_{X|Z=z}(0))$ can be estimated with any of several existing techniques for nonparametric estimation of conditional quantiles (e.g., Samanta 1989 and Chaudhuri 1991) evaluated at the estimated probability $1 - \hat{F}_{X|Z=z}(0)$. The conditional trimmed mean $\mathbb{E}[X|X \geq F_{X|Z=z}^{-1}(1 - F_{X|Z=z}(0)), Z = z]$ can be estimated as a nonparametric regression of X onto Z at $Z = z$, using only observations such that $X \geq \hat{F}_{X|Z=z}^{-1}(1 - \hat{F}_{X|Z=z}(0))$. This can be implemented with any of the regression methods discussed in the estimation of $F_{X|Z=z}(\cdot)$. Note that the asymptotic results for quantile and trimmed mean estimation in the literature are developed for known probabilities and trimming points, respectively. Here, these quantities are evaluated at estimated points instead, which must be taken into account. See the discussion in Example 5.3.

In the location-scale case (Assumption 2), $\mathbb{E}[X^*|X^* \leq 0, Z = z]$ may be estimated following equation (6). This is a deterministic equation of $F_{X|Z=z}(0)$ and $\mathbb{E}[X|Z = z]$. $F_{X|Z=z}(0)$ may be estimated as discussed above, and $\mathbb{E}[X|Z = z]$ may be estimated similarly, using any of the nonparametric regression techniques discussed above. Alternatively, $\mathbb{E}[X|Z = z]$ may also be estimated via a varying coefficient maximum likelihood method (e.g., Cai, Fan, and Li 2000).

Remark 5.1 (Large-Dimension/Mixed Discrete-Continuous/Sparse Support Controls). There are well-known practical issues in implementing the approaches discussed above when Z has many dimensions, or when it is composed of multiple continuous and discrete variables, or when it is sparse in regions. Thus, in many applications, including ours, it may be necessary to implement a strategy that is adaptive for complex Z . Techniques aimed at nonparametric multivariate regression with complex regressors are a main concern of the statistical learning field. For a summary of this large literature, we refer the reader to Hastie, Tibshirani, and Friedman (2009), which examines many of the modern dimensionality reduction techniques for regressions, and thus covers the estimation of $F_{X|Z=z}(\cdot)$, $\mathbb{E}[X|Z = z]$, and $\mathbb{E}[X|X \geq F_{X|Z=z}^{-1}(1 - F_{X|Z=z}(0))]$. Analogous methods for conditional quantiles have also been studied extensively (e.g., White 1992; Chaudhuri and Loh 2002; Li and Racine 2008; Chen et al. 2019), which cover the estimation of $F_{X|Z=z}^{-1}(1 - F_{X|Z=z}(0))$.

In our application, we use a clustering strategy to reduce the dimension of Z . The idea is to group the values of Z into K clusters such that each cluster has the least amount of variation

possible in Z . We then substitute Z with \hat{C}_K , the vector of cluster indicators, in the estimation of the expectation. That is, if Z belongs to the k th cluster, then $\hat{E}[X^*|X^* \leq 0, Z] = \hat{E}[X^*|X^* \leq 0, \hat{C}_K = e_k(K)]$, where $e_k(K)$ is the k th canonical vector with K coordinates.

Clustering is a convenient strategy for dimensionality reduction because it discretizes the support of Z for the purposes of estimation: no matter the dimension of Z , \hat{C}_K can assume only K values: $e_1(K), \dots, e_K(K)$. This allows us to use the estimators developed for the discrete support case—see Examples 5.2 and 5.3, substituting Z with \hat{C}_K and z_l with $e_l(K)$. These estimators may be implemented conveniently using optimized, off-the-shelf packages in commonly used software. Details about the clustering strategy are found in Appendix D.

6. Testing the Identification Assumptions

In this section, we discuss how to test the two identification conditions: (3) and the distributional assumption, both separately and jointly. All tests we propose are simple modifications of existing, well-known tests. We implement the tests in our application, and the findings are discussed in Section 7 and Appendix E. A comprehensive study of the properties of these tests is beyond the scope of this article.

6.1. Testing Equation (3)

Let the null hypothesis be “ H_0 : (3) holds.” If H_0 is true, then (4) holds, and therefore $\mathbb{E}[Y|X, Z] = (\beta + \delta)X + Z'(\pi - \gamma\delta)$ for $X > 0$. That is, the conditional expectation for $X > 0$ must be linear in X and Z . Note that no assumption other than equation (3) was used to establish this fact. Therefore, we can test H_0 by applying any specification test to the regression of Y on X and Z for $X > 0$ (e.g., Ramsey’s 1969 RESET test). In Appendix E.2, we show plots that illustrate in our application how this testing idea may be visually displayed.

6.2. Testing the Distributional Assumption

As discussed in Section 3, identification of $\mathbb{E}[X^*|X^* \leq 0, Z]$ is often achieved by assuming that $X^*|Z = z$ belongs to some specific distribution family. Because the distribution of $X^*|Z$ is observed when $X^* > 0$, this assumption is often directly testable.

For example, the null hypothesis “ H_0 : $X^*|Z = z \sim \mathcal{N}(\mu_z, \sigma_z^2)$ ” (Assumption 2 with $H = \Phi$, see Example 3.1) can be tested by comparing the empirical distribution $\hat{F}_{X|Z=z}(x)$, and the assumed normal distribution $\Phi((x - \mu_z)/\sigma_z)$ for $x > 0$, where μ_z and σ_z can be estimated as described in Example 5.2. This test can be implemented with any distribution comparison test, such as the well-known two-sample Kolmogorov-Smirnov test, as well as various more powerful alternatives such as the test developed in Goldman and Kaplan (2018).¹⁵ Analogous tests can be built for most well-known parametric families.

¹⁴Let $K(u)$ be a kernel function ($\int_{-\infty}^{\infty} K(u) = 1$, and suppose if convenient that $K(u) \geq 0$ and $K(u)\mathbf{1}(|u| > 1) = 0$), let $k_n(Z_i - z) = K((Z_i - z)/h_n) / \sum_{i=1}^n K((Z_i - z)/h_n)$, for some sequence $h_n \rightarrow 0, nh_n \rightarrow \infty$. Then

$$\hat{F}_{X|Z=z}(\cdot) = \sum_{i=1}^n \mathbf{1}(X_i \leq \cdot) k_n(Z_i - z).$$

Conditions for uniform convergence of such estimators can be verified in the existing literature. See, for example, Andrews (1995) and Hansen (2008).

¹⁵This test is designed to have better power to detect deviations from the null at the extremities of distributions. Such deviations may be a particular concern in our setting, when including the upper tail in the comparison of the distributions is often necessary.

Tail symmetry (Assumption 1) is not directly testable. However, full symmetry around the mean implies tail symmetry and is testable if $F_{X|Z=z}(0) < 0.5$.¹⁶ This follows because, if $X^*|Z = z$ is symmetrically distributed around its mean, then $X|X \in I, Z = z$, where $I = (F_{X|Z=z}(0), 1 - F_{X|Z=z}(0))$, is also symmetrically distributed around its mean. Therefore, if $F_{X|Z=z}(0) < 0.5$, the null hypothesis “ $H_0: X^*|Z = z$ is symmetrically distributed around $\mathbb{E}[X^*|Z = z]$ ” can be tested by implementing any of a large number of existing tests of symmetry on $F_{X|X \in I, Z=z}$.¹⁷ In Appendix E.3, we show plots that illustrate in our application how this testing idea may be visually displayed.

6.3. Testing All Identifying Assumptions: Specification Tests

Suppose that the identified expectation is $\tilde{e}(Z)$, which may be different from the true expectation, $\mathbb{E}[X^*|X^* \leq 0, Z]$. Let the best linear predictor of Y given X, Z and $X + \tilde{e}(Z)\mathbf{1}(X = 0)$ be defined as

$$L(Y|X, Z, X + \tilde{e}(Z)\mathbf{1}(X = 0)) = \tilde{\beta}X + Z'\tilde{\theta}_Z + \tilde{\delta}(X + \tilde{e}(Z)\mathbf{1}(X = 0)).$$

If (3) holds and the expectation is correctly identified (i.e., $\tilde{e}(Z) = \mathbb{E}[X^*|X^* \leq 0, Z]$), then $L(Y|X, Z, X + \tilde{e}(Z)\mathbf{1}(X = 0)) = L(Y|X, Z, X + \mathbb{E}[X^*|X^* \leq 0, Z]\mathbf{1}(X = 0)) = \mathbb{E}[Y|X, Z]$, and $\tilde{\beta} = \beta$, $\tilde{\theta}_Z = (\gamma - \pi\delta)$, and $\tilde{\delta} = \delta$. This implies that it is possible to test the hypothesis “ H_0 : equation (3) holds, and $\mathbb{E}[X^*|X^* \leq 0, Z]$ is correctly identified” by testing whether $\mathbb{E}[Y|X, Z] = L(Y|X, Z, X + \tilde{e}(Z)\mathbf{1}(X = 0))$ using a specification test.¹⁸

Since the regression employs a generated regressor, $X + \hat{e}(Z)\mathbf{1}(X = 0)$, standard specification tests will have the wrong size. A correctly-sized specification test can be performed by simply adding functions of X and Z to the sets of controls used in the regression and testing if the coefficients of these terms are equal to zero using the correct covariance matrix estimator or the bootstrapped critical values discussed in Section 4 (see Remark C.3 in Appendix C.1).

We also propose a specification test of H_0 that is particularly suited to our setting, where the support of the treatment variable is bounded at one extreme with bunching at the boundary point. The test leverages sample truncation to detect nonlinearities caused by failures of H_0 . Specifically, we propose restricting the sample to $X \in [0, \alpha]$, for a given value of α , and testing whether the coefficient of X in the truncated regression is different from the coefficient of X in the full sample regression. This test can be implemented using off-the-shelf tests that compare coefficients of non-nested regressions. The rationale of this test and further details can be seen in Appendix E.4. There, we also present plots that illustrate in our application how this testing idea may be visually displayed.

¹⁶Precisely, $X^*|Z = z$ is symmetric around its mean if $F_{X^*|Z=z}(x) = 1 - F_{X^*|Z=z}(2\mathbb{E}[X^*|Z = z] - x)$ for all x .

¹⁷See, for example, Chen and Tripathi (2017), which is designed specifically for truncated variables as in our case, Niu et al. (2018) and the large list of citations therein, and Quesy (2021) for a more general test of copula homogeneity which can be used to test symmetry.

¹⁸Note that, in models without controls, this test has nontrivial power only against violations of (3), not against the misidentification of the expectation. This is because, if (2) and (3) hold without controls, $\mathbb{E}[Y|X] = L(Y|X, X + \tilde{e}\mathbf{1}(X = 0))$ holds for all \tilde{e} .

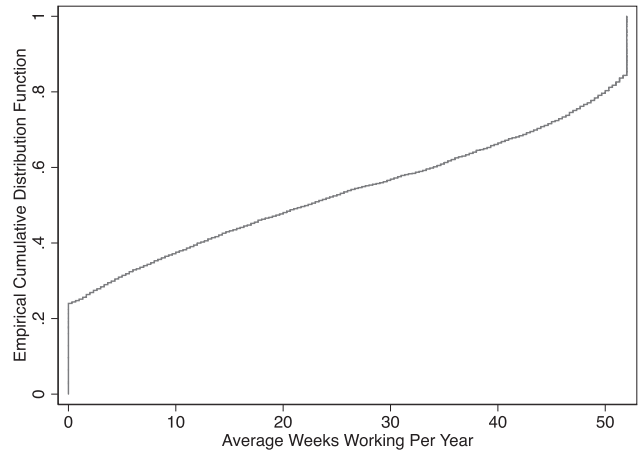


Figure 3. Evidence of bunching, maternal labor supply.

NOTE: This figure shows the empirical cdf for the average number of weeks per year in which the mother worked in the three years following her child's birth. Source: National Longitudinal Study of Youth, 1979 cohort, sample of mothers whose children were born from 1979 to 2002.

6.4. Testing All Identifying Assumptions: Additional Bunching Points

When there are multiple bunching points, including possibly some in the interior of the support of the treatment, one bunching point on the boundary of the support can be used to build the correction, while the remaining bunching points can be used to test the underlying assumptions of the correction method in the same regression. Specifically, suppose that a second bunching point at $X = \bar{x}$ exists, we propose testing the hypothesis “ H_0 : equation (3) holds and $\mathbb{E}[X^*|X^* \leq 0, Z]$ is correctly identified” by estimating the model

$$\mathbb{E}[Y|X, Z] = \beta X + Z'(\gamma - \pi\delta) + \delta(X + \mathbb{E}[X^*|X^* \leq 0, Z]\mathbf{1}(X = 0)) + \alpha_{\bar{x}}\mathbf{1}(X = \bar{x}),$$

and performing a t -test of whether $\alpha_{\bar{x}} = 0$. This is equivalent to performing the dummy test from Caetano et al. (2021) in this model. This test detects whether linearity holds and the correction has solved the endogeneity problem.¹⁹ The t -test must be implemented using the correct standard errors. The standard error estimator and the consistency of the bootstrap critical values are established in Appendix C.1 (see Remark C.3).

To illustrate the type of situation where this testing approach can be employed, Figure 3 shows the empirical cdf of maternal labor supply, measured as the average number of weeks per year the mother worked during the three years following the child's birth. The figure shows two bunching points, one at $X = 0$, and the other at $X = 52$. Maternal labor supply is the treatment variable of a large literature (see e.g., Caetano et al. 2022 and the references therein).

7. Application: The Effect of TV on Children's Skills

In this section, we apply our method to estimate the effect of time spent watching TV on children's skills using the 1997, 2002,

¹⁹When X has N bunching points, one can use $X = 0$ for the correction, add indicators of each of the remaining $N - 1$ bunching points, and do a joint test that all the coefficients of the dummies are equal to zero, as discussed in the multivariate version of the dummy test in Caetano et al. (2021).

Table 1. The effect of TV time on children's skills.

		(i) Uncorrected No controls	(ii) Uncorrected w/ controls	(iii) Hom. Tobit	(iv) Semipar. Tobit	(v) Tail symmetry
Cognitive	β	-0.57** (0.13)	-0.44** (0.10)	1.54 (0.98)	1.52 (0.98)	0.94 (0.68)
	δ			-1.92** (0.93)	-1.90** (0.93)	-1.31** (0.63)
Noncognitive	β	-0.28* (0.14)	-0.10 (0.14)	-3.05** (1.44)	-3.07** (1.42)	-2.11** (1.00)
	δ			2.85** (1.40)	2.87** (1.38)	1.91** (0.95)

NOTE: N=4330. Results are reported in terms of percentage points of the standard deviation of the outcome variable. For example, the results in the last column suggest that an increase of one hour per week watching TV leads to a reduction in noncognitive skills of 2.11 percentage points of one standard deviation. The list of observed controls in columns (ii)–(v) is described in Footnote 21. Bootstrapped standard errors in parentheses (1000 bootstrap samples). Columns (iii), (iv), and (v) show results using the Homoscedastic Tobit expectation estimator (Example 5.1), the Semiparametric Tobit expectation estimator (Example 5.2) and the Tail Symmetry expectation estimator (Example 5.3), respectively. Specifications in columns (iv) and (v) use 10 clusters (see Remark 5.1). See Figure 4 in Appendix D for a reproduction of the results in column (v) for different numbers of clusters. ** $p < 0.05$, * $p < 0.1$.

and 2007 Waves of the Child Development Supplement from the Panel Study of Income Dynamics (CDS-PSID).²⁰ More detailed applications of the method proposed in this article can be found in Caetano, Caetano, and Nielsen (2022), which studies the effect of enrichment activities on skills, and Caetano et al. (2022), which studies the effect of maternal labor supply on skills. As the application here uses the same sample as Caetano, Caetano, and Nielsen (2022), we refer the reader to that paper for details about the data and definition of skills.

We would like to estimate β in (3), where Y is either the cognitive or noncognitive skills of the child, X is the number of hours the child spent watching TV in a typical week, and Z is a vector of controls which includes a constant as well as characteristics of the child, family, and environment.²¹

Table 1 presents the estimates of β in (3) both with and without the endogeneity correction. Column (i) shows the simple regressions of TV time on skills without controls. Time spent watching TV is negatively correlated with both cognitive and noncognitive skills. Column (ii) adds observed controls, but does not include the correction control function. After controlling for observables, the estimates of β are closer to zero, but the cognitive estimate is still negative and significant.

In columns (iii)–(v), we show estimates of the corrected regression. Specifically, we follow (4), and estimate a regression of Y on X , Z and the control function $X + \hat{\mathbb{E}}[X^* | X^* \leq 0, Z] \mathbf{1}(X = 0)$. In column (iii), $\hat{\mathbb{E}}[X^* | X^* \leq 0, Z]$ uses the Homoscedastic Tobit estimator, described in Example 5.1. The estimates of $\mathbb{E}[X^* | X^* \leq 0, Z]$ in columns (iv) and (v) are obtained in two steps. First, we cluster the controls according to the method described in Remark 5.1. We obtain $K = 10$ clusters, and the cluster indicator vector \hat{C}_K . Second, for observation Z_i in

cluster l , we estimate $\hat{\mathbb{E}}[X^* | X^* \leq 0, Z = Z_i] = \hat{\mathbb{E}}[X^* | X^* \leq 0, \hat{C}_K = e_l(K)]$. In column (iv) we use the Semiparametric Tobit estimator described in Example 5.2, substituting Z with \hat{C}_K , and z_l with $e_l(K)$ in the formulas. In column (v) we use the tail symmetry estimator described in Example 5.3 with analogous substitutions.

The corrected β estimates are positive and insignificant for cognitive skills, but negative and significant for noncognitive skills. The Homoscedastic and Semiparametric Tobit estimates (columns (iii) and (iv)) are very similar. The tail symmetry estimates (column (v)) are closer to zero, yet the noncognitive estimate remains significant at 5%. Table 1 also displays the estimates of δ . All are significant at 5% and are negative for cognitive skills and positive for noncognitive skills, suggesting that selection on observables is strongly rejected in this context.

Note that the bootstrapped standard errors of $\hat{\beta}$ in the corrected models (columns (iii)–(v)) are much larger than the corresponding standard errors in the uncorrected model (column (ii)). The Eicker-White standard errors of the corrected model (Σ , see Theorem 4.1) gravitate around 95% of the bootstrapped standard errors for all specifications, so the penalty due to the estimation of the expectation turns out not to be important in explaining the larger standard errors. Rather, the standard errors are larger because much of the raw variation in X is contaminated by variation in the confounder. The uncorrected models attribute to X the confounding variation in Y , thus yielding artificially smaller standard errors. In this context, inference is affected twice by not appropriately controlling for unobservables: not only is there bias in the uncorrected estimator, but its standard error leads to the belief that effects might be estimated with a higher degree of precision than is achievable by a correct estimator.

Testing the Identification Assumptions

We mention here the key findings from the implementation of the tests discussed in Section 6 in our application. The precise null hypothesis, method used and details can be seen in Appendix E. In Appendix E.1, we show that there is strong evidence of endogeneity in the uncorrected model, which justifies our correction approach. In Appendix E.2, we test the hypothesis that (3) holds, and we do not reject it. In Appendix

²⁰See Zavodny (2006), Gentzkow and Shapiro (2008), and Munasib and Bhattacharya (2010) for recent empirical papers in this literature. These papers are well aware that watching television may be endogenous. Zavodny (2006) tackles endogeneity using fixed effects, Munasib and Bhattacharya (2010) uses IV, while Gentzkow and Shapiro (2008) uses the timing of the roll-out of children's programming to different local markets to obtain causal estimates.

²¹The variables included as controls are the child's age and squared age (in months), and indicators for: CDS wave (1997, 2002, and 2007), grade (thirteen variables, from kindergarten through grade 12), gender, ethnicity (black, Hispanic and other nonwhite ethnicity), whether the child has siblings, family income tercile, whether the mother is alive, and whether the father is alive.

E.3, we test the distributional assumption of $X^*|Z$ alone. The Homoscedastic Tobit estimator (column(iii)) is consistent under the homoscedastic Gaussian assumption, as discussed in [Example 5.1](#), and the Semiparametric Tobit estimator (column (iv)) is consistent under the semiparametric Gaussian assumption, as discussed in [Example 5.2](#). Our tests reject both assumptions. The tail symmetry estimator (column (v)) is consistent under the tail symmetry assumption, as discussed in [Example 5.3](#). Our tests do not reject the stronger condition of symmetry around the mean (see [Footnote 16](#)). Therefore, our preferred estimates are those in column (v) of [Table 1](#). It is worth noting that the estimates in columns (iii) and (iv) are similar to the results in column (v). This suggests that our tests are able to detect violations of the distributional assumptions even when the bias generated by these violations is likely small. This is consistent with what we find in our Monte Carlo study ([Appendix F](#)). Finally, in [Appendix E.4](#), we perform a joint test of whether (3) and the tail symmetry assumption hold and do not reject this joint hypothesis.

8. Concluding Remarks

This article shows how to leverage bunching at the lower (or upper) extremum of the distribution of the treatment variable to build a control function to correct for endogeneity. In linear models, the method consists of adding a generated regressor to the original regression. We also show that the method can be applied in many models used by practitioners, including some nonparametric and nonseparable models. The approach does not make exclusion restrictions. Instead, it makes testable functional and distributional shape assumptions. We study the asymptotic behavior of the estimated coefficients, prove the consistency of the bootstrap, and show that the method performs well in an empirical Monte Carlo study. Finally, we apply the method to estimate the effect of watching television on the skills of the child. A common finding in both the application and the Monte Carlo study is that, when violations of the identifying assumptions are large enough to generate a noticeable bias, these violations are easy to detect with the tests of the identification assumptions we develop.

The method proposed in this article opens up many paths for new research. Here we highlight a few: (a) discretizing the controls Z before estimating the expectation is convenient for the implementation of the method proposed in this article, as evidenced in our application, [Caetano, Caetano, and Nielsen \(2022\)](#), and [Caetano et al. \(2022\)](#). The advantages/drawbacks of the use of clusters in this context need to be investigated further. (b) The interaction of the correction with existing methods is promising. We mentioned the combination of this method with [Caetano et al.'s \(2021\)](#) test in [Section 6.4](#). The interaction of this approach with instrumental variables methods may also prove valuable.

Supplementary Materials

The appendix provides generalizations of the approach to more flexible functional forms, details referenced in the paper as well as further results on identification, estimation and testing, a Monte Carlo study, and proofs.

Acknowledgments

We thank [Marinho Bertanha](#), [Stéphane Bonhomme](#), [Juan Carlos Escanciano](#), [Bruno Ferman](#), [Simon Freyaldenhoven](#), [Dalia Ghanem](#), [Leonard Goff](#), [Guido Imbens](#), [Ingrid van Keilegom](#), [Salvador Navarro](#), [Rodrigo Pinto](#), [Vitor Possebom](#), [Pedro Sant'Anna](#), [Azeem Shaikh](#), [David Slichter](#), [Christopher Taber](#), [Aad van der Vaart](#), and seminar participants at various conferences and institutions for helpful discussions and comments. The analysis and conclusions set forth here are those of the authors and do not indicate concurrence by other members of the research staff, the Board of Governors, or the Federal Reserve System.

Disclosure Statement

The authors report there are no competing interests to declare.

References

- Andrews, D. W. (1994), "Asymptotics for Semiparametric Econometric Models via Stochastic Equicontinuity," *Econometrica: Journal of the Econometric Society*, 62, 43–72. [2]
- Andrews, D. W. (1995), "Nonparametric Kernel Estimation for Semiparametric Models," *Econometric Theory*, 11, 560–596. [8]
- Bertanha, M., McCallum, A. H., and Seegert, N. (2020), "Better Bunching, Nicer Notching," working paper. [1,2]
- Bertrand, M., Karlan, D., Mullainathan, S., Shafir, E., and Zinman, J. (2010), "What's Advertising Content Worth? Evidence from a Consumer Credit Marketing Field Experiment," *The Quarterly Journal of Economics*, 125, 263–306. [2]
- Bhutani, S., Klempel, M. C., Kroeger, C. M., Aggour, E., Calvo, Y., Trepanowski, J. F., Hoddy, K. K., and Varady, K. A. (2013), "Effect of Exercising While Fasting on Eating Behaviors and Food Intake," *Journal of the International Society of Sports Nutrition*, 10, 50. [2]
- Bischoff-Ferrari, H. A., Willett, W. C., Wong, J. B., Giovannucci, E., Dietrich, T., and Dawson-Hughes, B. (2005), "Fracture Prevention with Vitamin D Supplementation: A Meta-Analysis of Randomized Controlled Trials," *JAMA*, 293, 2257–2264. [2]
- Black, S. E., Devereux, P. J., and Salvanes, K. G. (2005), "The More the Merrier? The Effect of Family Size and Birth Order on Children's Education," *The Quarterly Journal of Economics*, 120, 669–700. [2]
- Bleemer, Z. (2018a), "The Effect of Selective Public Research University Enrollment: Evidence from California. Research & Occasional Paper Series: CSHE. 11.18," *Center for Studies in Higher Education*. [2]
- Bleemer, Z. (2018b), "Top Percent Policies and the Return to Postsecondary Selectivity," working paper. [2]
- Blomquist, N. S., Newey, W. K., Kumar, A., and Liang, C.-Y. (2019), "On Bunching and Identification of the Taxable Income Elasticity," CENMAP working paper. [1,2]
- Boserup, S. H., Kopczuk, W., and Kreiner, C. T. (2016), "The Role of Bequests in Shaping Wealth Inequality: Evidence from Danish Wealth Records," *American Economic Review*, 106, 656–661. [2]
- Boulianne, S. (2015), "Social Media Use and Participation: A Meta-Analysis of Current Research," *Information, Communication & Society*, 18, 524–538. [2]
- Brown, J. R., Coile, C. C., and Weisbenner, S. J. (2010), "The Effect of Inheritance Receipt on Retirement," *The Review of Economics and Statistics*, 92, 425–434. [2]
- Caetano, C. (2015), "A Test of Exogeneity Without Instrumental Variables in Models with Bunching," *Econometrica*, 83, 1581–1600. [1,2]
- Caetano, C., Caetano, G., Fe, H., and Nielsen, E. (2021), "A Dummy Test of Identification in Models with Bunching," working paper. [9,11]
- Caetano, C., Caetano, G., and Nielsen, E. (2022), "Should Children Do More Enrichment Activities? Leveraging Bunching to Correct for Endogeneity," working paper. [2,10,11]
- Caetano, C., Caetano, G., Nielsen, E., and Sanfelice, V. (2022), "The Effect of Maternal Labor Supply on Children's Skills," working paper. [2,9,10,11]
- Caetano, G., Kinsler, J., and Teng, H. (2019), "Towards Causal Estimates of Children's Time Allocation on Skill Development," *Journal of Applied Econometrics*, 34, 588–605. [2]

- Caetano, G., and Maheshri, V. (2018), "Identifying Dynamic Spillovers of Crime with a Causal Approach to Model Selection," *Quantitative Economics*, 9, 343–394. [2]
- Cai, Z., Fan, J., and Li, R. (2000), "Efficient Estimation and Inferences for Varying-Coefficient Models," *Journal of the American Statistical Association*, 95, 888–902. [8]
- Chaudhuri, P. (1991), "Global Nonparametric Estimation of Conditional Quantile Functions and their Derivatives," *Journal of Multivariate Analysis*, 39, 246–269. [8]
- Chaudhuri, P., and Loh, W.-Y. (2002), "Nonparametric Estimation of Conditional Quantiles Using Quantile Regression Trees," *Bernoulli*, 8, 561–576. [8]
- Chen, T., and Tripathi, G. (2017), "A Simple Consistent Test of Conditional Symmetry in Symmetrically Trimmed Tobit Models," *Journal of Econometrics*, 198, 29–40. [4,9]
- Chen, X., Li, D., Li, Q., and Li, Z. (2019), "Nonparametric Estimation of Conditional Quantile Functions in the Presence of Irrelevant Covariates," *Journal of Econometrics*, 212, 433–450. [8]
- Chen, X., Linton, O., and Van Keilegom, I. (2003), "Estimation of Semiparametric Models When the Criterion Function is not Smooth," *Econometrica*, 71, 1591–1608. [2,5,6]
- Cohen, M. A. (2008), "The Effect of Crime on Life Satisfaction," *The Journal of Legal Studies*, 37, S325–S353. [2]
- De Vito, A., Jacob, M., and Müller, M. A. (2019), "Avoiding Taxes to Fix the Tax Code," working paper. [2]
- D'Haultfœuille, X., Hoderlein, S., and Sasaki, Y. (2021), "Testing and Relaxing the Exclusion Restriction in the Control Function Approach," *Journal of Econometrics*. [1]
- Durrieu, G., Grama, I., Pham, Q.-K., and Tricot, J.-M. (2015), "Nonparametric Adaptive Estimation of Conditional Probabilities of Rare Events and Extreme Quantiles," *Extremes*, 18, 437–478. [7]
- Ekici, T., and Dunn, L. (2010), "Credit Card Debt and Consumption: Evidence from Household-Level Data," *Applied Economics*, 42, 455–462. [2]
- Eren, O., and Henderson, D. J. (2011), "Are We Wasting Our Children's Time by Giving Them More Homework?" *Economics of Education Review*, 30, 950–961. [2]
- Erhardt, E. C. (2017), "Microfinance Beyond Self-Employment: Evidence for Firms in Bulgaria," *Labour Economics*, 47, 75–95. [2]
- Erixson, O. (2017), "Health Responses to a Wealth Shock: Evidence from a Swedish Tax Reform," *The Journal of Population Economics*, 30, 1281–1336. [2]
- Fang, Z., and Santos, A. (2019), "Inference on Directionally Differentiable Functions," *The Review of Economic Studies*, 86, 377–412. [2]
- Fawzi, W. W., Chalmers, T. C., Herrera, M. G., and Mosteller, F. (1993), "Vitamin A Supplementation and Child Mortality: A Meta-Analysis," *JAMA*, 269, 898–903. [2]
- Ferreira, D., Ferreira, M. A., and Mariano, B. (2018), "Creditor Control Rights and Board Independence," *The Journal of Finance*, 73, 2385–2423. [2]
- Gentzkow, M., and Shapiro, J. M. (2008), "Preschool Television Viewing and Adolescent Test Scores: Historical Evidence from the Coleman Study," *The Quarterly Journal of Economics*, 123, 279–323. [10]
- Goff, L. (2020), "Treatment Effects in Bunching Designs: The Impact of the Federal Overtime Rule on Hours," Technical Report, working paper. [1]
- Goldman, M., and Kaplan, D. M. (2018), "Comparing Distributions by Multiple Testing Across Quantiles or CDF Values," *Journal of Econometrics*, 206, 143–166. [8]
- Hansen, B. E. (2008), "Uniform Convergence Rates for Kernel Estimation with Dependent Data," *Econometric Theory*, 24, 726–748. [8]
- Hastie, T., Tibshirani, R., and Friedman, J. (2009), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, New York: Springer. [8]
- Hazra, N. K., Kuiti, M. R., Finkelstein, M., and Nanda, A. K. (2017), "On Stochastic Comparisons of Maximum Order Statistics from the Location-Scale Family of Distributions," *Journal of Multivariate Analysis*, 160, 31–41. [5]
- Heckman, J. J. (1979), "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153–161. [1]
- Heckman, J. J., and Robb, R. (1985), "Alternative Methods for Evaluating the Impact of Interventions: An Overview," *Journal of Econometrics*, 30, 239–267. [1]
- Hernán, M. A., Takkouche, B., Caamaño-Isorna, F., and Gestal-Otero, J. J. (2002), "A Meta-Analysis of Coffee Drinking, Cigarette Smoking, and the Risk of Parkinson's Disease," *Annals of Neurology*, 52, 276–284. [2]
- Holt, K., Shehata, A., Strömbäck, J., and Ljungberg, E. (2013), "Age and the Effects of News Media Attention and Social Media Use on Political Interest and Participation: Do Social Media Function as Leveller?" *European Journal of Communication*, 28, 19–34. [2]
- James-Burdumy, S. (2005), "The Effect of Maternal Labor Force Participation on Child Development," *Journal of Labor Economics*, 23, 177–211. [2]
- Jenish, N., and Prucha, I. R. (2009), "Central Limit Theorems and Uniform Laws of Large Numbers for Arrays of Random Fields," *Journal of Econometrics*, 150, 86–98. [6]
- Klein, R., and Vella, F. (2010), "Estimating a Class of Triangular Simultaneous Equations Models Without Exclusion Restrictions," *Journal of Econometrics*, 154, 154–164. [1]
- Lavetti, K., and Schmutte, I. M. (2018), "Estimating Compensating Wage Differentials with Endogenous Job Mobility," working paper. [2]
- Lewbel, A. (2012), "Using Heteroscedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models," *Journal of Business & Economic Statistics*, 30, 67–80. [1]
- Li, Q., and Racine, J. S. (2008), "Nonparametric Estimation of Conditional CDF and Quantile Functions with Mixed Categorical and Continuous Data," *Journal of Business & Economic Statistics*, 26, 423–434. [8]
- Luoh, M.-C., and Herzog, A. R. (2002), "Individual Consequences of Volunteer and Paid Work in Old Age: Health and Mortality," *Journal of Health and Social Behavior*, 43, 490–509. [2]
- Masry, E. (1996), "Multivariate Local Polynomial Regression for Time Series: Uniform Strong Consistency and Rates," *Journal of Time Series Analysis*, 17, 571–599. [8]
- McDuffie, R. S., Beck, A., Bischoff, K., Cross, J., and Orleans, M. (1996), "Effect of Frequency of Prenatal Care Visits on Perinatal Outcome Among Low-Risk Women: A Randomized Controlled Trial," *JAMA*, 275, 847–851. [2]
- Melzer, B. T. (2011), "The Real Costs of Credit Access: Evidence from the Payday Lending Market," *The Quarterly Journal of Economics*, 126, 517–555. [2]
- Meyer, J. (1987), "Two-Moment Decision Models and Expected Utility Maximization," *The American Economic Review*, 77, 421–430. [5]
- Millimet, D. L., and Tchernis, R. (2013), "Estimation of Treatment Effects Without an Exclusion Restriction: With An Application to the Analysis of the School Breakfast Program," *Journal of Applied Econometrics*, 28, 982–1017. [1]
- Munasingh, A., and Bhattacharya, S. (2010), "Is the 'idiot's box' Raising Idiocy? Early and Middle Childhood Television Watching and Child Cognitive Outcome," *Economics of Education Review*, 29, 873–883. [10]
- Navarro, S. (2010), "Control Functions," in *Microeconometrics*, eds. S. N. Durlauf and L. E. Blume, pp. 20–28, London: Springer. [1]
- Niu, C., Guo, X., Li, Y., and Zhu, L. (2018), "Pairwise Distance-based Tests for Conditional Symmetry," *Computational Statistics & Data Analysis*, 128, 145–162. [9]
- Noordzij, M., Uiterwaal, C. S., Arends, L. R., Kok, F. J., Grobbee, D. E., and Geleijnse, J. M. (2005), "Blood Pressure Response to Chronic Intake of Coffee and Caffeine: A Meta-Analysis of Randomized Controlled Trials," *Journal of Hypertension*, 23, 921–928. [2]
- Oken, E., Levitan, E., and Gillman, M. (2008), "Maternal Smoking During Pregnancy and Child Overweight: Systematic Review and Meta-Analysis," *International Journal of Obesity*, 32, 201–210. [2]
- Pakes, A., and Pollard, D. (1989), "Simulation and the Asymptotics of Optimization Estimators," *Econometrica: Journal of the Econometric Society*, 57, 1027–1057. [2,6]
- Pang, J. (2017), "Do Subways Improve Labor Market Outcomes for Low-Skilled Workers," working paper, Syracuse University. [2]
- Peek, J., Rosengren, E. S., and Tootell, G. M. (2003), "Identifying the Macroeconomic Effect of Loan Supply Shocks," *Journal of Money, Credit and Banking*, 35, 931–946. [2]

- Pötscher, B. M., and Prucha, I. R. (1994), "Generic Uniform Convergence and Equicontinuity Concepts for Random Functions: An Exploration of the Basic Structure," *Journal of Econometrics*, 60, 23–63. [6]
- Powell, J. L. (1986), "Symmetrically Trimmed Least Squares Estimation for Tobit Models," *Econometrica: journal of the Econometric Society*, 54, 1435–1460. [4]
- Quesy, J.-F. (2021), "A Szekely–Rizzo Inequality for Testing General Copula Homogeneity Hypotheses," *Journal of Multivariate Analysis*, 186, 104815. [9]
- Ramsey, J. B. (1969), "Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis," *Journal of the Royal Statistical Society, Series B*, 31, 350–371. [8]
- Reynolds, K., Lewis, B., Nolen, J. D. L., Kinney, G. L., Sathya, B., and He, J. (2003), "Alcohol Consumption and Risk of Stroke: A Meta-Analysis," *JAMA*, 289, 579–588. [2]
- Richardson, T., Elliott, P., and Roberts, R. (2013), "The Relationship between Personal Unsecured Debt and Mental and Physical Health: A Systematic Review and Meta-Analysis," *Clinical Psychology Review*, 33, 1148–1162. [2]
- Rozenas, A., Schutte, S., and Zhukov, Y. (2017), "The Political Legacy of Violence: The Long-Term Impact of Stalin's Repression in Ukraine," *The Journal of Politics*, 79, 1147–1161. [2]
- Samanta, M. (1989), "Non-Parametric Estimation of Conditional Quantiles," *Statistics & Probability Letters*, 7, 407–412. [8]
- Song, K. (2008), "Uniform Convergence of Series Estimators Over Function Spaces," *Econometric Theory*, 24, 1463–1499. [8]
- Tobin, J. (1958), "Estimation of Relationships for Limited Dependent Variables," *Econometrica*, 26, 24–36. [7]
- van der Vaart, A. (1994), "Maximum Likelihood Estimation with Partially Censored Data," *The Annals of Statistics*, 22, 1896–1916. [7]
- Van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press. [6]
- White, H. (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817–838. [5,6]
- (1992), "Nonparametric Estimation of Conditional Quantiles Using Neural Networks," in *Computing Science and Statistics*, eds. C. Page, and R. LePage, pp. 190–199, New York: Springer. [8]
- Wong, W.-K., and Ma, C. (2008), "Preferences Over Location-Scale Family," *Economic Theory*, 37, 119–146. [5]
- Zavodny, M. (2006), "Does Watching Television Rot Your Mind? Estimates of the Effect on Test Scores," *Economics of Education Review*, 25, 565–573. [10]