Optimal sup-norm rates and uniform inference on nonlinear functionals of nonparametric IV regression

Xiaohong Chen
Cowles Foundation for Research in Economics, Yale University

Timothy M. Christensen
Department of Economics, New York University

This paper makes several important contributions to the literature about non-parametric instrumental variables (NPIV) estimation and inference on a structural function h_0 and functionals of h_0. First, we derive sup-norm convergence rates for computationally simple sieve NPIV (series two-stage least squares) estimators of h_0 and its derivatives. Second, we derive a lower bound that describes the best possible (minimax) sup-norm rates of estimating h_0 and its derivatives, and show that the sieve NPIV estimator can attain the minimax rates when h_0 is approximated via a spline or wavelet sieve. Our optimal sup-norm rates surprisingly coincide with the optimal root-mean-squared rates for severely ill-posed problems, and are only a logarithmic factor slower than the optimal root-mean-squared rates for mildly ill-posed problems. Third, we use our sup-norm rates to establish the uniform Gaussian process strong approximations and the score bootstrap uniform confidence bands (UCBs) for collections of nonlinear functionals of h_0 under primitive conditions, allowing for mildly and severely ill-posed problems. Fourth, as applications, we obtain the first asymptotic pointwise and uniform inference results for plug-in sieve t-statistics of exact consumer surplus (CS) and deadweight loss (DL) welfare functionals under low-level conditions when demand is estimated via sieve NPIV. Our real data application of UCBs for exact CS and DL functionals of gasoline demand reveals interesting patterns and is applicable to other goods markets.

Keywords. Series two-stage least squares, optimal sup-norm convergence rates, uniform Gaussian process strong approximation, score bootstrap uniform confi-

dence bands, nonlinear welfare functionals, nonparametric demand with endogeneity.

JEL classification. C13, C14, C36.

1. Introduction

Well-founded empirical evaluation of economic policy is often based on inference on nonlinear welfare functionals of nonparametric or semiparametric structural models. This paper makes several important contributions to estimation and inference on a flexible (i.e., nonparametric) structural function h_0 and nonlinear functionals of h_0 within the framework of a nonparametric instrumental variables (NPIV) model,

$$Y_i = h_0(X_i) + u_i, \qquad E[u_i|W_i] = 0, \tag{1}$$

where h_0 is an unknown function, X_i is a vector of continuous *endogenous* regressors, W_i is a vector of (conditional) instrumental variables, and the conditional distribution of X_i given W_i is unspecified.

Given a random sample $\{(Y_i, X_i, W_i)\}_{i=1}^n$ (of size n) from the NPIV model (1), our first two main theoretical results address how well one may estimate h_0 and its derivatives simultaneously in sup-norm loss, that is, we bound

$$\sup_x |\widehat{h}(x) - h_0(x)| \quad \text{and} \quad \sup_x |\partial^k \widehat{h}(x) - \partial^k h_0(x)|$$

for estimators $\widehat{h}$ of h_0, where $\partial^k h(x)$ denotes the kth partial derivatives of h with respect to components of x. We first provide upper bounds on sup-norm convergence rates for the computationally simple sieve NPIV (i.e., series two-stage least-squares (2SLS)) estimators (Newey and Powell (2003), Ai and Chen (2003), Blundell, Chen, and Kristensen (2007)). We then derive a lower bound that describes the best possible (i.e., minimax) sup-norm convergence rates among all estimators for h_0 and its derivatives, and show that the sieve NPIV estimator can attain the minimax lower bound when a spline or wavelet basis is used to approximate h_0.[1] Next, we apply our sup-norm rate results to establish the uniform Gaussian process strong approximation and the validity of score bootstrap uniform confidence bands (UCBs) for collections of possibly nonlinear functionals of h_0 under primitive conditions.[2] This includes valid score bootstrap UCBs for h_0 and its derivatives as special cases. Finally, as important applications, we establish first pointwise and uniform inference results for two leading nonlinear welfare functionals of a nonparametric demand function h_0 estimated via sieve NPIV, namely the exact consumer surplus (CS) and deadweight loss (DL) arising from price changes at different income levels when prices (and possibly income) are endogenous.[3] We present two

---

[1]The optimal sup-norm rates for estimating h_0 are in the first version (Chen and Christensen (2013)); the optimal sup-norm rates for estimating derivatives of h_0 are in the second version (Chen and Christensen (2015a)).

[2]The uniform strong approximation and the score bootstrap UCB results are in the second version (Chen and Christensen (2015a)); see Theorem B.1 and its proof in that version.

[3]The pointwise inference results on exact CS and DL are in the second version (Chen and Christensen (2015a)).

real data applications to illustrate the easy implementation and usefulness of the score bootstrap UCBs based on sieve NPIV estimators. The first application is to nonparametric exact CS and DL functionals of gasoline demand; the second is to nonparametric Engel curves and their derivatives. The UCBs reveal new interesting and sensible patterns in both data applications. We note that the score bootstrap UCBs for exact CS and DL nonlinear functionals are new to the literature even when the prices might be exogenous. Empiricists could jump to Section 2 to read the sieve score bootstrap UCBs procedure and these real data applications without reading the rest of the more theoretical sections.

Regardless of whether the regressor X_i is endogenous or not, sup-norm convergence rates provide sharper measures of how well h_0 and its derivatives can be estimated nonparametrically than the usual L^2-norm (i.e., root-mean-squared) rates. This is also why, in the existing literature on nonparametric models without endogeneity, consistent specification tests in sup-norm (i.e., Kolmogorov–Smirnov type statistics) are widely used. Further, sup-norm rates are particularly useful for controlling nonlinearity bias when conducting inference on *highly nonlinear* (i.e., beyond quadratic) functionals of h_0. In addition to being useful in constructing pointwise and uniform confidence bands for nonlinear functionals of h_0 via plug-in estimators, the sup-norm rates for estimating h_0 are also useful in semiparametric two-step procedures when h_0 enters the second-stage moment conditions (equalities or inequalities) nonlinearly.

Despite the usefulness of sup-norm convergence rates in nonparametric estimation and inference, as yet there are no published results on optimal sup-norm convergence rates for estimating h_0 or its derivatives in the NPIV model (1). This is because, unlike nonparametric least-squares (LS) regression (i.e. estimation of $h_0(x) = E[Y_i|X_i = x]$ when X_i is exogenous), estimation of h_0 in the NPIV model (1) is a difficult ill-posed inverse problem with an unknown operator (Newey and Powell (2003), Carrasco, Florens, and Renault (2007)). Intuitively, h_0 in model (1) is identified by the integral equation

$$E[Y_i|W_i = w] = Th_0(w) := \int h_0(x) f_{X|W}(x|w)\,\mathrm{d}x,$$

where T must be inverted to obtain h_0. Since integration smoothes out features of h_0, a small error in estimating $E[Y_i|W_i = w]$ using the data $\{(Y_i, X_i, W_i)\}_{i=1}^n$ may lead to a large error in estimating h_0. In addition, the conditional density $f_{X|W}$ and, hence, the operator T, are generally unknown, so T must be also estimated from the data. Due to the difficult ill-posed inverse nature, even the L^2-norm convergence rates for estimating h_0 in model (1) have not been established until recently.[4] In particular, Hall and Horowitz (2005) derived minimax L^2-norm convergence rates for mildly ill-posed NPIV models and showed that their estimators can attain the optimal L^2-norm rates for h_0. Chen and Reiss (2011) derived minimax L^2-norm convergence rates for mildly and severely ill-posed NPIV models and showed that sieve NPIV estimators can attain the optimal

---

[4]See, for example, Hall and Horowitz (2005), Blundell, Chen, and Kristensen (2007), Chen and Reiss (2011), Darolles, Fan, Florens, and Renault (2011), Horowitz (2011), Chen and Pouzo (2012), Gagliardini and Scaillet (2012), Florens and Simoni (2012), Kato (2013), and references therein.

rates.[5] Moreover, it is generally much harder to obtain optimal nonparametric convergence rates in sup-norm than in L^2-norm.[6]

In this paper, we derive the best possible (i.e., minimax) sup-norm convergence rates of any estimator of h_0 and its derivatives in mildly and severely ill-posed NPIV models. Surprisingly, the optimal sup-norm convergence rates for estimating h_0 and its derivatives coincide with the optimal L^2-norm rates for severely ill-posed problems and are only a power of $\log n$ slower than optimal L^2-norm rates for mildly ill-posed problems. We also obtain sup-norm convergence rates for sieve NPIV estimators of h_0 and its derivatives. We show that a sieve NPIV estimator using a spline or wavelet basis to approximate h_0 can attain the minimax sup-norm rates for estimating both h_0 and its derivatives. When specializing to series LS regression (without endogeneity), our results automatically imply that spline and wavelet series LS estimators will also achieve the optimal sup-norm rates of Stone (1982) for estimating the derivatives of a nonparametric LS regression function, which strengthen the recent sup-norm optimality results in Belloni et al. (2015) and Chen and Christensen (2015b) for estimating regression function h_0 itself. We focus on the sieve NPIV estimator because it has been used in empirical work, can be implemented as easily as 2SLS, and can reduce to simple series LS when the regressor X_i is exogenous. Moreover, both h_0 and its derivatives may be simultaneously estimated at their respective optimal convergence rates via a sieve NPIV estimator when the same sieve dimension is used to approximate h_0. This is a desirable property to practitioners. In addition, the sieve NPIV estimator for h_0 in model (1) and our proof of its sup-norm rates could be easily extended to estimating unknown functions in other semiparametric models with nonparametric endogeneity, such as a system of shape-invariant Engel curve instrumental variable (IV) regression models (Blundell, Chen, and Kristensen (2007)).

We provide two important applications of our results on sup-norm convergence rates in detail: both are about inferences on nonlinear functionals of h_0 based on plug-in sieve NPIV estimators; see Section 6 for discussions of additional applications. Inference on highly nonlinear (i.e., beyond quadratic) functionals of h_0 in a NPIV model is very difficult because of the combined effects of nonlinearity bias and the slow convergence rates (in sup-norm and L^2-norm) of any estimators of h_0. Indeed, our minimax rate results show that *any* estimator of h_0 in an ill-posed NPIV model must necessarily converge slower than its nonparametric LS counterpart. For example, the optimal sup- and L^2-norm rates for estimating h_0 in a severely ill-posed NPIV model are $(\log n)^{-\gamma}$ for some $\gamma > 0$. It is well known that a plug-in series LS estimate of a weighted quadratic functional could be root-n consistent. But a plug-in sieve NPIV estimate of a weighted quadratic functional of h_0 in a severely ill-posed NPIV model fails to be root-n consistent (Chen and Pouzo (2015)). In fact, we establish that the minimax convergence rate of any

---

[5]Appendix B extends the results in Chen and Reiss (2011) to L^2-norm optimality for estimating derivatives of h_0.

[6]Even for the simple nonparametric LS regression of h_0 (without endogeneity), the optimal sup-norm rates for series LS estimators of h_0 were not obtained until recently in Cattaneo and Farrell (2013) for locally partitioning series LS, Belloni, Chernozhukov, Chetverikov, and Kato (2015) for spline LS, and Chen and Christensen (2015b) for wavelet LS.

estimators of a simple weighted quadratic functional of h_0 in a severely ill-posed NPIV model is as slow as $(\log n)^{-a}$ for some $a > 0$ (see Appendix C).

In the first application, we extend the seminal work of Hausman and Newey (1995) about pointwise inference on *exact* CS and DL functionals of nonparametric demand without endogeneity to allow for prices, and possibly incomes, to be endogenous. According to Hausman (1981) and Hausman and Newey (1995, 2016, 2017), exact CS and DL functionals are the most widely used welfare and economic efficiency measures. Exact CS is a leading example of a complicated nonlinear functional of h_0, which is defined as the solution to a differential equation involving a demand function (Hausman (1981)). Hausman and Newey (1995) were the first to establish the pointwise asymptotic normality of plug-in kernel estimators of exact CS and DL functionals of a nonparametric demand without endogeneity. Vanhems (2010) was the first to estimate exact CS via the plug-in Hall and Horowitz (2005) kernel NPIV estimator of h_0 when price is endogenous, and she derived its convergence rate in L^2-norm for the mildly ill-posed case, but did not establish any inference results (such as the pointwise asymptotic normality). Our paper is the first to provide low-level sufficient conditions to establish inference results for plug-in (spline and wavelet) sieve NPIV estimators of exact CS and DL functionals, allowing for both mildly and severely ill-posed NPIV models. Precisely, we use our sup-norm convergence rates for sieve NPIV estimators of h_0 and its derivatives to locally linearize plug-in estimators of exact CS and DL, which then lead to asymptotic normality of sieve t-statistics for exact CS and DL under primitive sufficient conditions. We also establish the asymptotic normality of plug-in sieve NPIV t-statistic for an *approximate* CS functional, extending Newey (1997)'s result from nonparametric exogenous demand to endogenous demand. Recently, Chen and Pouzo (2015) presented a set of high-level conditions for the pointwise asymptotic normality of sieve t-statistics of possibly nonlinear functionals of h_0 in a general class of nonparametric conditional moment restriction models (including the NPIV model as a special case). They verified their high-level conditions for pointwise asymptotic normality of sieve t-statistics for linear and quadratic functionals. But without sup-norm convergence rate result, Chen and Pouzo (2015) were unable to provide low-level sufficient conditions for pointwise asymptotic normality of plug-in sieve NPIV estimators for complicated nonlinear (beyond quadratic) functionals such as the exact CS functional. This was actually the original motivation for us to derive sup-norm convergence rates for sieve NPIV estimators of h_0 and its derivatives.

In the second important application of our sup-norm rate results, we establish the uniform Gaussian process strong approximation and the validity of score bootstrap uniform confidence bands (UCBs) for collections of possibly nonlinear functionals of h_0, under primitive sufficient conditions that allow for mildly and severely ill-posed NPIV models. The low-level sufficient conditions for Gaussian process strong approximation and UCBs are applied to complicated nonlinear functionals such as collections of exact CS and DL functionals of nonparametric demand with endogenous price (and possibly income). When specializing to collections of linear functionals of the NPIV function h_0, our Gaussian process strong approximation and sieve score bootstrap UCBs for h_0 and its derivatives are valid under mild sufficient conditions. In particular, for a NPIV model

with a scalar endogenous regressor, our sufficient conditions are comparable to those in Horowitz and Lee (2012) for their notion of UCBs with a growing number of grid points by interpolation for h_0 estimated via the modified orthogonal series NPIV estimator of Horowitz (2011). When specialized to a nonparametric LS regression (with exogenous X_i), our results on the Gaussian strong approximation and score bootstrap UCBs for collections of nonlinear functionals of h_0, such as exact CS and DL functionals, are still new to the literature and complement the important results in Chernozhukov, Lee, and Rosen (2013) for h_0 and Belloni et al. (2015) for linear functionals of h_0 estimated via series LS.

Our sieve score bootstrap UCBs procedure is extremely easy to implement since it computes the sieve NPIV estimator only once using the data, and then perturbs the sieve score statistics by random weights that are mean zero and independent of the data. So it should be very useful to empirical researchers who conduct nonparametric estimation and inference on structural functions with endogeneity in diverse subfields of applied economics, such as consumer theory, industrial organization, labor economics, public finance, health economics, development, and trade, to name only a few. Two real data illustrations are presented in Section 2. In the first, we construct UCBs for exact CS and DL welfare functionals for a range of gasoline taxes at different income levels. For this illustration, we use the same data set as in Blundell, Horowitz, and Parey (2012, 2017) and estimate household gasoline demand via spline sieve NPIV (other data sets and other goods could be used). Despite the slow convergence rates of NPIV estimators, the UCBs for exact CS are particularly informative. In the second empirical illustration, we use the same data set as in Blundell, Chen, and Kristensen (2007) to estimate Engel curves for households with kids via a spline sieve NPIV and construct UCBs for Engel curves and their derivatives for various categories of household expenditure.

The rest of the paper is organized as follows. Section 2 presents the sieve NPIV estimator, the score bootstrap UCBs procedure, and two real-data applications. This section aims at empirical researchers. Section 3 establishes the minimax optimal sup-norm rates for estimating a NPIV function h_0 and its derivatives. Section 4 presents low-level sufficient conditions for the uniform Gaussian process strong approximation and sieve score bootstrap UCBs for collections of general nonlinear functionals of a NPIV function. Section 5 deals with pointwise and uniform inferences on exact CS and DL, and approximate CS functionals in nonparametric demand estimation with endogeneity. Section 6 concludes with discussions of additional applications of the sup-norm rates of sieve NPIV estimators. Appendix A contains additional results on sup-norm convergence rates. Appendix B presents optimal L^2-norm rates for estimating derivatives of a NPIV function under extremely weak conditions. Appendix C establishes the minimax lower bounds for estimating quadratic functionals of a NPIV function. The main supplemental appendix, available in a supplementary file on the journal website, http://qeconomics.org/supp/722/supplement.pdf, contains pointwise normality of sieve t-statistics for nonlinear functionals of NPIV under lower-level sufficient conditions than those in Chen and Pouzo (2015) (Appendix D), background material on B-spline and wavelet sieves (Appendix E), and useful lemmas on random matrices (Appendix F). The secondary supplemental appendix, available in a supplementary file on

the journal website, http://qeconomics.org/supp/722/code_and_data.zip, contains additional lemmas and all of the proofs (Appendix G).

2. Estimator and motivating applications to UCBs

This section describes the sieve NPIV estimator and a score bootstrap UCBs procedure for collections of functionals of the NPIV function. It mentions intuitively why sup-norm convergence rates of a sieve NPIV estimator are needed to formally justify the validity of the computationally simple score bootstrap UCBs procedure. It then present two real data applications of uniform inferences on functionals of a NPIV function: UCBs for exact CS and DL functionals of nonparametric demand with endogenous price, and UCBs for nonparametric Engel curves and their derivatives when the total expenditure is endogenous. This section is presented to practitioners.

Sieve NPIV estimators. Let $\{(Y_i, X_i, W_i)\}_{i=1}^n$ denote a random sample from the NPIV model (1). The sieve NPIV estimator $\widehat{h}$ of h_0 is simply the 2SLS estimator applied to some basis functions of X_i (the endogenous regressors) and W_i (the conditioning variables), namely

$$\widehat{h}(x) = \psi^J(x)'\widehat{c} \quad \text{with } \widehat{c} = \left[\Psi'B(B'B)^- B'\Psi\right]^- \Psi'B(B'B)^- B'Y, \tag{2}$$

where $Y = (Y_1, \ldots, Y_n)'$,

$$\psi^J(x) = \left(\psi_{J1}(x), \ldots, \psi_{JJ}(x)\right)', \qquad \Psi = \left(\psi^J(X_1), \ldots, \psi^J(X_n)\right)', \tag{3}$$

$$b^K(w) = \left(b_{K1}(w), \ldots, b_{KK}(w)\right)', \qquad B = \left(b^K(W_1), \ldots, b^K(W_n)\right)', \tag{4}$$

and $\{\psi_{J1}, \ldots, \psi_{JJ}\}$ and $\{b_{K1}, \ldots, b_{KK}\}$ are collections of basis functions of dimension J and K for approximating h_0 and the instrument space, respectively (Blundell, Chen, and Kristensen (2007), Chen and Pouzo (2012), Newey (2013)). The *regularization* parameter J is the dimension of the sieve for approximating h_0. The *smoothing* parameter K is the dimension of the instrument sieve. From the analogy with 2SLS, it is clear that we need $K \geq J$. Blundell, Chen, and Kristensen (2007), Chen and Reiss (2011), Chen and Pouzo (2012) have previously shown that $\lim_J(K/J) = c \in [1, \infty)$ can lead to an optimal L^2-norm convergence rate for the sieve NPIV estimator. Thus we assume that K grows to infinity at the same rate as that of J, say $J \leq K \leq cJ$ for some finite $c > 1$ for simplicity.[7] When $K = J$ and b^K and ψ^J are formed from the same orthogonal basis, the sieve NPIV estimator becomes Horowitz's (2011) modified orthogonal series NPIV estimator. Note that the sieve NPIV estimator (2) reduces to a series LS estimator $\widehat{h}(x) = \psi^J(x)'[\Psi'\Psi]^- \Psi'Y$ when $X_i = W_i$ is exogenous, $J = K$, and $\psi^J(x) = b^K(w)$ (Newey (1997), Huang (1998)).

---

[7]Monte Carlo evidence in (Blundell, Chen, and Kristensen (2007), Chen and Pouzo (2015)) and others suggest that sieve NPIV estimators often perform better with $K > J$ than with $K = J$, and that the regularization parameter J is important for finite sample performance while the parameter K is not as important as long as it is larger than J. See our second version (Chen and Christensen (2015a)) for the data-driven choice of J.

2.1 *Uniform confidence bands for nonlinear functionals*

One important motivating application is to *uniform inference* on a collection of nonlinear functionals $\{f_t(h_0) : t \in \mathcal{T}\}$, where $\mathcal{T}$ is an index set (e.g., an interval). Uniform inference may be performed via uniform confidence bands (UCBs) that contain the function $t \mapsto f_t(h_0)$ with prescribed coverage probability. UCBs for h_0 (or its derivatives) are obtained as a special case with $\mathcal{T} = \mathcal{X}$ (support of X_i) and $f_t(h_0) = h_0(t)$ (or $f_t(h_0) = \partial^k h_0(t)$ for the kth derivative). We present applications below to uniform inference on exact CS and DL functionals over a range of price changes as well as UCBs for Engel curves and their derivatives.

A $100(1 - \alpha)\%$ bootstrap-based UCB for $\{f_t(h_0) : t \in \mathcal{T}\}$ is constructed as

$$t \mapsto \left[f_t(\widehat{h}) - z^*_{1-\alpha} \frac{\widehat{\sigma}(f_t)}{\sqrt{n}}, f_t(\widehat{h}) + z^*_{1-\alpha} \frac{\widehat{\sigma}(f_t)}{\sqrt{n}} \right]. \tag{5}$$

In this display, $f_t(\widehat{h})$ is the plug-in sieve NPIV estimator of $f_t(h_0)$, $\widehat{\sigma}^2(f_t)$ is a *sieve variance estimator* for $f_t(\widehat{h})$, and $z^*_{1-\alpha}$ is a bootstrap-based critical value to be defined below.

To compute the sieve variance estimator for $f_t(\widehat{h})$ with $\widehat{h}(x) = \psi^J(x)'\widehat{c}$ given in (2), one would first compute the 2SLS covariance matrix estimator (but applied to basis functions) for $\widehat{c}$,

$$\widehat{\upsilon} = \left[\widehat{S}'\widehat{G}_b^{-1}\widehat{S} \right]^{-1} \widehat{S}'\widehat{G}_b^{-1}\widehat{\Omega}\widehat{G}_b^{-1}\widehat{S} \left[\widehat{S}'\widehat{G}_b^{-1}\widehat{S} \right]^{-1}, \tag{6}$$

where $\widehat{S} = B'\Psi/n$, $\widehat{G}_b = B'B/n$, $\widehat{\Omega} = n^{-1} \sum_{i=1}^n \widehat{u}_i^2 b^K(W_i) b^K(W_i)'$, and $\widehat{u}_i = Y_i - \widehat{h}(X_i)$. One then computes a "delta-method" correction term—a $J \times 1$ vector $Df_t(\widehat{h})[\psi^J] :=$ $(Df_t(\widehat{h})[\psi_{J1}], \ldots, Df_t(\widehat{h})[\psi_{JJ}])'$—by calculating $Df_t(\widehat{h})[v] = \lim_{\delta \to 0^+} [\delta^{-1}\{f_t(\widehat{h} + \delta v) - f_t(\widehat{h})\}]$, which is the (functional directional) derivative of f_t at $\widehat{h}$ in direction v for $v = \psi_{J1}, \ldots, \psi_{JJ}$. The sieve variance estimator for $f_t(\widehat{h})$ is then

$$\widehat{\sigma}^2(f_t) = \left(Df_t(\widehat{h})[\psi^J] \right)' \widehat{\upsilon} \left(Df_t(\widehat{h})[\psi^J] \right). \tag{7}$$

We use the following *sieve score bootstrap* procedure to calculate the critical value $z^*_{1-\alpha}$. Let $\varpi_1, \ldots, \varpi_n$ be independent and identically distributed (IID) random variables independent of the data with mean zero, unit variance, and finite third moment (e.g., $N(0, 1)$).[8] We define the bootstrap sieve t-statistic process $\{\mathbb{Z}_n^*(t) : t \in \mathcal{T}\}$ as

$$\mathbb{Z}_n^*(t) := \frac{(Df_t(\widehat{h})[\psi^J])' [\widehat{S}'\widehat{G}_b^{-1}\widehat{S}]^{-1} \widehat{S}'\widehat{G}_b^{-1}}{\widehat{\sigma}(f_t)} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n b^K(W_i)\widehat{u}_i\varpi_i \right) \quad \text{for each } t \in \mathcal{T}. \tag{8}$$

To compute $z^*_{1-\alpha}$, one would calculate $\sup_{t \in \mathcal{T}} |\mathbb{Z}_n^*(t)|$ for a large number of independent draws of $\varpi_1, \ldots, \varpi_n$. The critical value $z^*_{1-\alpha}$ is the $(1 - \alpha)$ quantile of $\sup_{t \in \mathcal{T}} |\mathbb{Z}_n^*(t)|$ over the draws. Note that this sieve score bootstrap procedure is different from the usual nonparametric bootstrap (based on resampling the data and then recomputing the estimator): here we only compute the estimator once and then perturb the sieve t-statistic process by the innovations $\varpi_1, \ldots, \varpi_n$.

---

[8]Other examples of distributions with these properties include the recentered exponential (i.e., $\varpi_i = \text{Exp}(1) - 1$), the Rademacher (i.e., ± 1 each with probability $\frac{1}{2}$), or the two-point distribution of Mammen (1993) (i.e., $(1 - \sqrt{5})/2$ with probability $(\sqrt{5} + 1)/(2\sqrt{5})$ and $(\sqrt{5} + 1)/\sqrt{2}$ with remaining probability).

An intuitive description of why sup-norm rates are very useful to justify this procedure is as follows. Under regularity conditions, the sieve t-statistic for an individual functional $f_t(h_0)$ admits an expansion

$$\frac{\sqrt{n}(f_t(\widehat{h}) - f_t(h_0))}{\widehat{\sigma}(f_t)} = \widehat{\mathbb{Z}}_n(t) + \text{nonlinear remainder term} \tag{9}$$

(see equation (18) for the definition of $\widehat{\mathbb{Z}}_n(t)$). The term $\widehat{\mathbb{Z}}_n(t)$ is a central limit theorem (CLT) term, that is, $\widehat{\mathbb{Z}}_n(t) \to_d N(0, 1)$ for each fixed $t \in \mathcal{T}$. Therefore, the sieve t-statistic for $f_t(h_0)$ also converges to a $N(0, 1)$ random variable provided that the nonlinear remainder term is asymptotically negligible (i.e., $o_p(1)$) (see Assumption 3.5 in Chen and Pouzo (2015)). Our sup-norm rates are very useful for providing weak regularity conditions under which the remainder is $o_p(1)$ for fixed t.[9] This justifies constructing confidence intervals for *individual* functionals $f_t(h_0)$ for any fixed $t \in \mathcal{T}$ by inverting the sieve t-statistic (on the left-hand side of display (9)) and using $N(0, 1)$ critical values. However, for *uniform* inference the usual $N(0, 1)$ critical values are no longer appropriate, as we need to consider the sampling error in estimating the whole process $t \mapsto f_t(h_0)$. For this purpose, display (9) is strengthened to be valid uniformly in $t \in \mathcal{T}$ (see Lemma 4.1). Under some regularity conditions, $\sup_{t \in \mathcal{T}} |\widehat{\mathbb{Z}}_n(t)|$ converges in distribution to the supremum of a (nonpivotal) Gaussian process. As its critical values are generally not available, we use the sieve score bootstrap procedure to estimate its critical values.

Section 4 formally justifies the use of this procedure for constructing UCBs for $\{f_t(h_0) : t \in \mathcal{T}\}$. The sup-norm rates are useful for controlling the nonlinear remainder terms for UCBs for collections of nonlinear functionals. Theorem 4.1 appears to be the first to establish the consistency of sieve score bootstrap UCBs for general nonlinear functionals of NPIV under low-level conditions, allowing for mildly and severely ill-posed problems. It includes as special cases the score bootstrap UCBs for nonlinear functionals of h_0 under exogeneity when h_0 is estimated via series LS and the score bootstrap UCBs for the NPIV function h_0 and its derivatives.[10] Theorem 4.1 is applied in Section 5 to formally justify the validity of score bootstrap UCBs for exact CS and DL functionals over a range of price changes when demand is estimated nonparametrically via sieve NPIV.

2.2 *Empirical application 1: UCBs for nonparametric exact CS and DL functionals*

Here we apply our methodology to study the effect of gasoline price changes on household welfare. We extend the important work by Hausman and Newey (1995) on pointwise confidence bands for exact CS and DL of demand without endogeneity to UCBs for exact CS and DL of demand with endogeneity.

---

[9]Chen and Pouzo (2015) verified their high-level Assumption 3.5 for a plug-in sieve estimator of a weighted quadratic functional example. Without sup-norm convergence rates, it is difficult to verify their Assumption 3.5 for nonlinear functionals (such as the exact CS) that are more complicated than quadratic functionals.

[10]One also needs to use sup-norm convergence rates of $\widehat{h}$ to h_0 to build a valid UCB for $\{h_0(t) : t \in \mathcal{X}\}$.

Let the demand of consumer i be

$$Q_i = h_0(P_i, Y_i) + u_i,$$

where Q_i is quantity, P_i is price, which may be endogenous, Y_i is income of consumer i, and u_i is an error term.[11] Hausman (1981) shows that the exact CS from a price change from p^0 to p^1 at income level y, denoted $S_y(p^0)$, solves

$$\frac{\partial S_y(p(u))}{\partial u} = -h_0\big(p(u), y - S_y(p(u))\big)\frac{dp(u)}{du}, \tag{10}$$

$$S_y(p(1)) = 0,$$

where $p : [0, 1] \to \mathbb{R}$ is a twice continuously differentiable path with $p(0) = p^0$ and $p(1) = p^1$. The corresponding DL functional $D_y(p^0)$ is

$$D_y(p^0) = S_y(p^0) - (p^1 - p^0)h_0(p^1, y). \tag{11}$$

As is evident from (10) and (11), exact CS and DL are (typically nonlinear) functionals of h_0. An exception is when demand is independent of income, in which case exact CS and DL are linear functionals of h_0. Let $t = (p^0, p^1, y)$ index the initial price, final price, and income level, and let $\mathcal{T} \subseteq [\underline{p}^0, \overline{p}^0] \times [\underline{p}^1, \overline{p}^1] \times [\underline{y}, \overline{y}]$ denote a range of price changes and/or incomes over which inference is to be performed. To denote dependence on h_0, we use the notation

$$f_{\mathrm{CS},t}(h) = \text{solution to (10) with } h \text{ in place of } h_0, \tag{12}$$

$$f_{\mathrm{DL},t}(h) = f_{\mathrm{CS},t}(h) - (p^1 - p^0)h(p^1, y), \tag{13}$$

so $S_y(p^0) = f_{\mathrm{CS},t}(h_0)$ and $D_y(p^0) = f_{\mathrm{DL},t}(h_0)$.

We estimate exact CS and DL using the plug-in estimators $f_{\mathrm{CS},t}(\widehat{h})$ and $f_{\mathrm{DL},t}(\widehat{h})$. The sieve variance estimators $\widehat{\sigma}^2(f_{\mathrm{CS},t})$ and $\widehat{\sigma}^2(f_{\mathrm{DL},t})$ are as described in (7) with the delta-method correction terms

$$Df_{\mathrm{CS},t}(\widehat{h})[\psi^J] = \int_0^1 \psi^J\big(p(u), y - \widehat{S}_y(p(u))\big)e^{-\int_0^u \partial_2\widehat{h}(p(v), y - \widehat{S}_y(p(v)))p'(v)\,dv}p'(u)\,du, \tag{14}$$

$$Df_{\mathrm{DL},t}(\widehat{h})[\psi^J] = Df_{\mathrm{CS},t}(\widehat{h})[\psi^J] - (p^1 - p^0)\psi^J(p^1, y), \tag{15}$$

where $p'(u) = \frac{dp(u)}{du}$, $\partial_2 h$ denotes the partial derivative of h with respect to its second argument, and $\widehat{S}_y(p(u))$ denotes the solution to (10) with $\widehat{h}$ in place of h_0.

We use the 2001 National Household Travel Survey gasoline demand data from Blundell, Horowitz, and Parey (2012, 2017).[12] The main variables are annual household

---

[11]Endogeneity may also be an issue in the estimation of static models of labor supply, in which Q_i represents hours worked, P_i is the wage, and Y_i is other income. In this setting it is reasonable to allow for endogeneity of both P_i and Y_i (see Blundell, Duncan, and Meghir (1998), Blundell, MaCurdy, and Meghir (2007), and references therein).

[12]We are grateful to Matthias Parey for sharing the data set with us. We refer the reader to Section 3 of Blundell, Horowitz, and Parey (2012) for a detailed description of the data.

TABLE 1. Summary statistics for gasoline demand data.

	Quantity (gal)	Price ($/gal)	Income ($)
Mean	1455	1.33	58,307
25th %	871	1.28	42,500
Median	1269	1.32	57,500
75th %	1813	1.40	72,500
Std dev	894	0.07	19,584

gasoline consumption (in gallons), average price (in dollars per gallon) in the county in which the household is located, household income, and distance from the Gulf coast to the capital of the state in which the household is located. Due to censoring, we consider the subset of households with incomes less than $100,000 per year. To keep households somewhat homogeneous, we select households with incomes above $25,000 per year (the 8th percentile) that have at most six inhabitants and one or two drivers. The resulting sample has size $n = 2753$.[13] Table 1 presents summary statistics.

We estimate the household gasoline demand function in levels via sieve NPIV using distance as an instrument for price. To implement the estimator, we form Ψ_J by taking a tensor product of quartic B-spline bases of dimension 5 for both price and income (so $J = 25$) and form B_K by taking a tensor product of quartic B-spline bases of dimension 8 for distance and 5 for income (so $K = 40$) with interior knots spaced evenly at quantiles.

We consider exact CS and DL resulting from price increases from $p^0 \in$ [$1.20, $1.40] to $p^1 =$ $1.40 at income levels of $y =$ $42,500 (low) and $y =$ $72,500 (high). We estimate exact CS at each initial price level by solving the ordinary differential equation (ODE (10)) by backward differences. We construct UCBs for exact CS as described above by setting $\mathcal{T} =$ [$1.20, $1.40] $\times$ {$1.40} $\times$ {$42,500} for the low-income group and $\mathcal{T} =$ [$1.20, $1.40] $\times$ {$1.40} $\times$ {$72,500} for the high-income group, $f_t(h) = f_{\mathrm{CS},t}(h)$ from display (12), and $Df_t(\widehat{h})[\psi^J] = Df_{\mathrm{CS},t}(\widehat{h})[\psi^J]$ from display (14). The ODE (10) is solved numerically by backward differences and the integrals in (14) are computed numerically. UCBs for DL are formed similarly, $f_t(h) = f_{\mathrm{DL},t}(h)$ from display (13), and $Df_t(\widehat{h})[\psi^J] = Df_{\mathrm{DL},t}(\widehat{h})[\psi^J]$ from display (15). We draw the bootstrap innovations ϖ_i from Mammen's two-point distribution with 1000 bootstrap replications.

The exact CS and DL estimates are presented in Figure 1 together with their UCBs. It is clear that exact CS is much more precisely estimated than DL. This is to be expected, since exact CS is computed by essentially integrating over one argument of the estimated demand function and is therefore smoother than the DL functional, which depends on h_0 estimated at the point (p^1, y). In fact, even though the sieve NPIV $\widehat{h}$ itself converges slowly, the UCBs for exact CS are still quite informative. At their widest point (with initial price $1.20), the 95% UCBs for exact CS for low-income households are [$259, $314]. In terms of comparison across high- and low-income households, the exact CS estimates

---

[13]We also exclude one household that reports 14,635 gallons; the next largest is 8089 gallons. Similar results are obtained using the full set of $n = 4811$ observations.
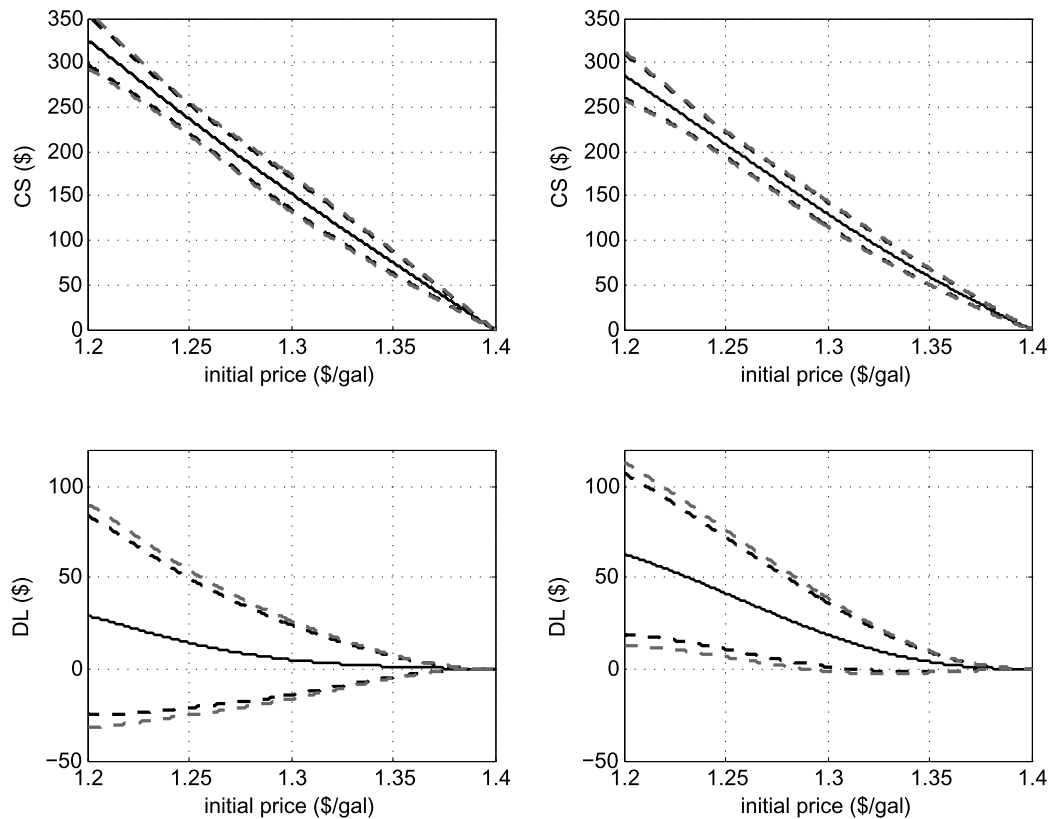
FIGURE 1. Estimated CS and DL from a price increase to $1.40/gal (solid black line) and their bootstrap UCBs (dashed black lines are 90%; dashed grey lines are 95%) when demand is estimated via sieve NPIV. Left panels are for household income of $72,500; right panels are for household income of $42,500.

are higher for the high-income households, whereas DL estimates are higher for the low-income households.

Figure 2 displays estimates obtained when we treat price as exogenous and estimate demand (h_0) by series LS regression. This is a special case of the preceding analysis with $X_i = W_i = (P_i, Y_i)'$, $K = J$, and $\psi^J = b^K$. These estimates display several notable features. First, the exact CS estimates are very similar whether demand is estimated via series LS or via sieve NPIV. Second, the UCBs for exact CS estimates are of a similar width to those obtained when demand was estimated via sieve NPIV, even though NPIV is an ill-posed inverse problem, whereas nonparametric LS regression is not. Third, the UCBs for DL are noticeably narrower when demand is estimated via series LS than when demand is estimated via sieve NPIV. Fourth, the DL estimates for LS and sieve NPIV are similar for high-income households, but quite different for low-income households. This is consistent with Blundell, Horowitz, and Parey (2017), who find some evidence of endogeneity in gasoline prices for low-income groups.
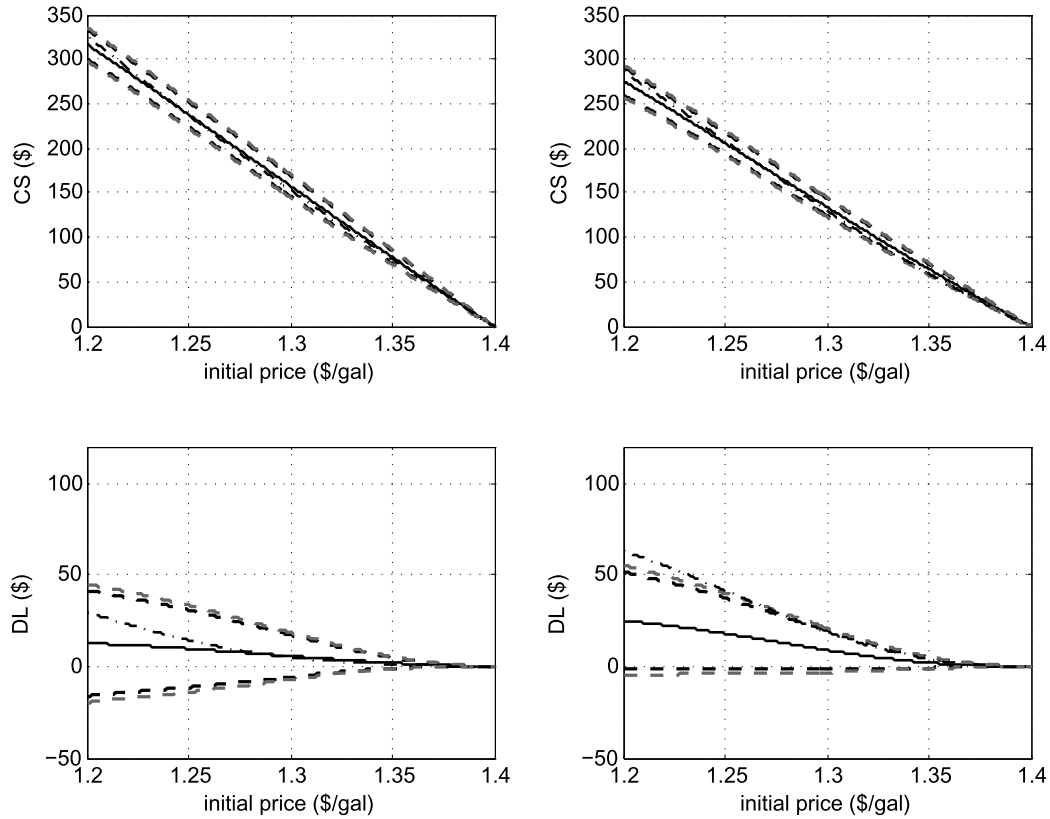
Figure 2. Estimated CS and DL from a price increase to $1.40/gal (solid black lines) and their bootstrap UCBs (dashed black lines are 90%; dashed grey lines are 95%) when demand is estimated via series LS. CS and DL when demand is estimated via NPIV are also shown (black dash–dot lines). Left panels are for household income of $72,500; right panels are for household income of $42,500.

2.3 *Empirical application 2: UCBs for Engel curves and their derivatives*

Engel curves describe the household budget share for expenditure categories as a function of total household expenditure. Following Blundell, Chen, and Kristensen (2007), we use sieve NPIV to estimate Engel curves, taking log total household income as an instrument for log total household expenditure. We use data from the 1995 British Family Expenditure Survey, focusing on the subset of married or cohabitating couples with one or two children, with the head of household aged between 20 and 55 and in work. This leaves a sample of size $n = 1027$. We consider six categories of nondurables and services expenditure: food in, food out, alcohol, fuel, travel, and leisure.

We construct UCBs for Engel curves as described above by setting $\mathcal{T} = [4.75, 6.25]$ (approximately the 5th–95th percentile of log expenditure), $f_t(h) = h(t)$, and $Df_t(\widehat{h})[\psi^J] = \psi^J(t)$. We also construct UCBs for derivatives of the Engel curves by setting $\mathcal{T} = [4.75, 6.25]$, $f_t(h)$ to be the derivative of h evaluated at t, and $Df_t(\widehat{h})[\psi^J]$ to be

Figure 3. Estimated Engel curves (black line) with bootstrap uniform confidence bands (dashed black lines are 90%; dashed grey lines are 95%). The *x* axis is log total household expenditure; the *y* axis is household budget share.

the vector formed by taking derivatives of $\psi_{J1}, \ldots, \psi_{JJ}$ evaluated at t. For both constructions, we use a quartic B-spline basis of dimension $J = 5$ for Ψ_J and a quartic B-spline basis of dimension $K = 9$ for B_K, with interior knots evenly spaced at quantiles (an important feature of sieve estimators is that the same sieve dimension can be used for optimal estimation of the function and its derivatives; this is not the case for kernel-based estimators). We draw the bootstrap innovations ϖ_i from Mammen's two-point distribution with 1000 bootstrap replications.

The Engel curves presented in Figure 3 and their derivatives presented in Figure 4 exhibit several interesting features. The curves for food in and fuel (necessary goods) are both downward sloping, with the curve for fuel exhibiting a pronounced downward slope at lower-income levels. The derivative of the curve for fuel is negative, though the UCBs are positive at the extremities. In contrast, the curve for leisure expenditure (luxury good) is strongly upward sloping and its derivative is positive except at low-income levels. Remaining curves for food out, alcohol, and travel appear to be non-monotonic.
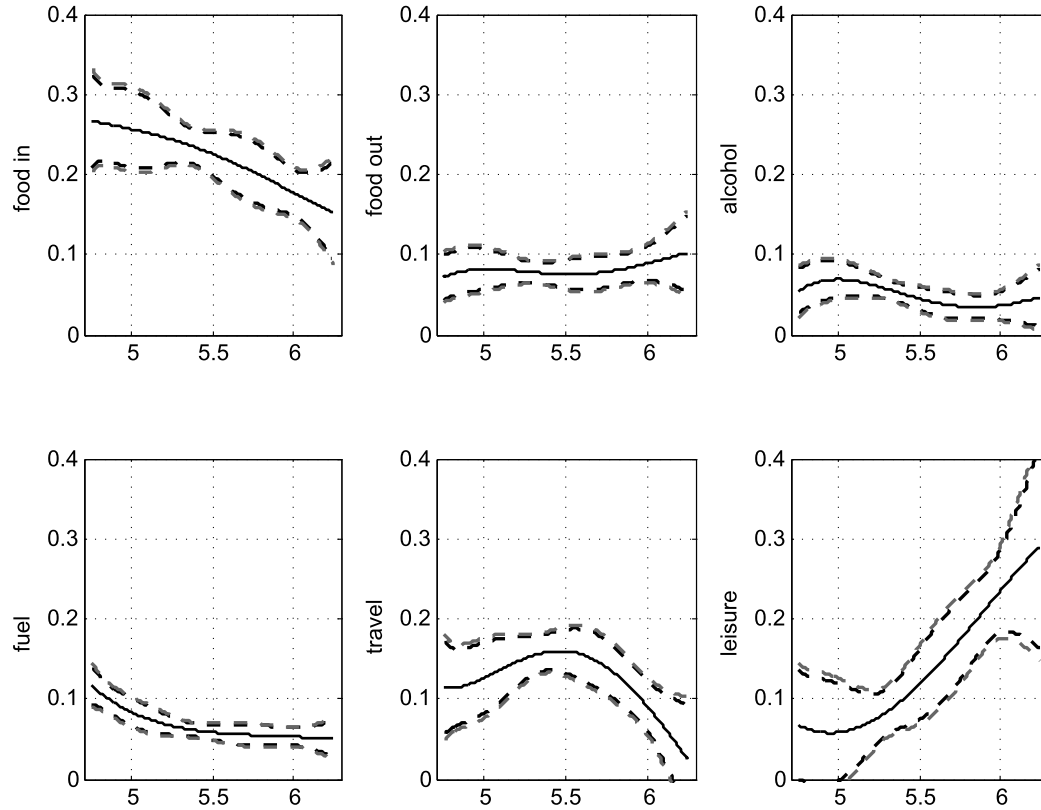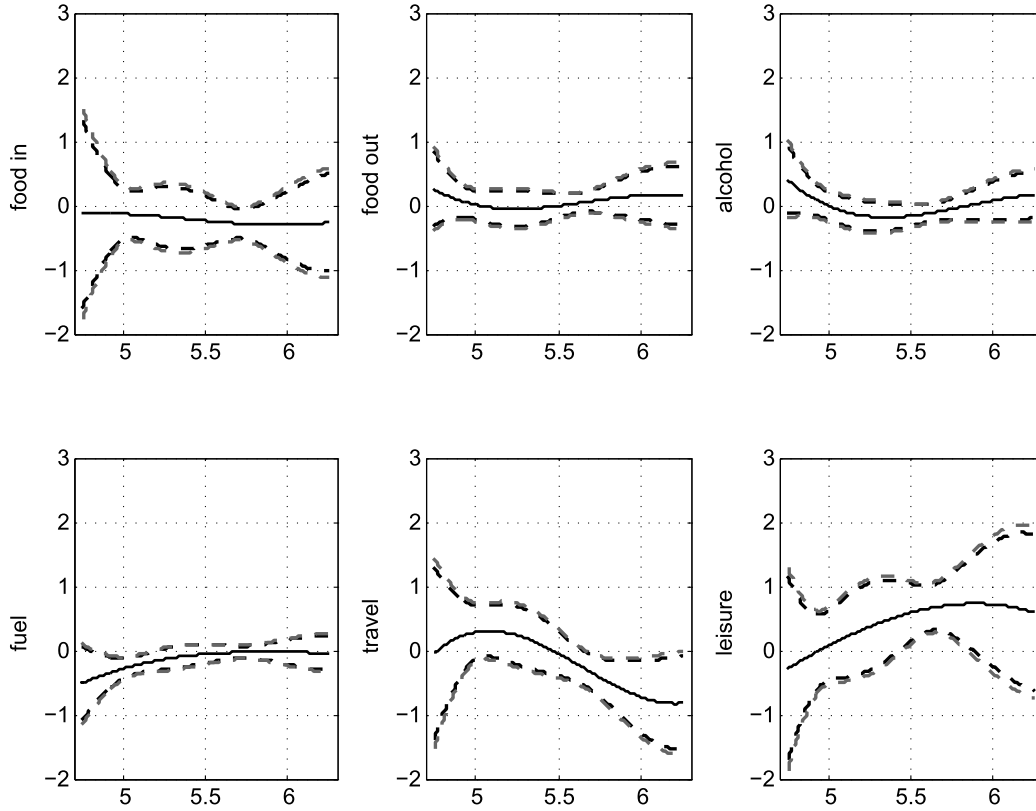
FIGURE 4. Estimated Engel curve derivatives (black line) with bootstrap uniform confidence bands (dashed black lines are 90%; dashed grey lines are 95%).

3. OPTIMAL SUP-NORM CONVERGENCE RATES

This section presents several results on sup-norm convergence rates. Section 3.1 presents upper bounds on sup-norm convergence rates of NPIV estimators of h_0 and its derivatives. Section 3.2 presents (minimax) lower bounds. Section 3.3 considers NPIV models with endogenous and exogenous regressors that are useful in empirical studies.

Notation. We work on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The term $\mathcal{A}^c$ denotes the complement of an even $\mathcal{A} \in \mathcal{F}$. We abbreviate "with probability approaching 1" to wpa1, and say that a sequence of events $\{\mathcal{A}_n\} \subset \mathcal{F}$ holds wpa1 if $\mathbb{P}(\mathcal{A}_n^c) = o(1)$. For a random variable X, we define the space $L^q(X)$ as the equivalence class of all measurable functions of X with finite qth moment if $1 \leq q < \infty$; when $q = \infty$, we denote $L^\infty(X)$ as the set of all bounded measurable functions $g : \mathcal{X} \to \mathbb{R}$ endowed with the sup-norm $\|g\|_\infty = \sup_x |g(x)|$. Let $\langle \cdot, \cdot \rangle_X$ denote the inner product on $L^2(X)$. For matrix and vector norms, $\|\cdot\|_{\ell^q}$ denotes the vector ℓ^q-norm when applied to vectors and the operator norm induced by the vector ℓ^q-norm when applied to matrices. If a and b are scalars, we let $a \vee b := \max\{a, b\}$ and $a \wedge b := \min\{a, b\}$. Minimum and maximum eigenvalues are denoted by $\lambda_{\min}$ and $\lambda_{\max}$. If $\{a_n\}$ and $\{b_n\}$ are sequences of positive numbers, we say that $a_n \lesssim b_n$ if $\limsup_{n \to \infty} a_n/b_n < \infty$ and we say that $a_n \asymp b_n$ if $a_n \lesssim b_n$ and $b_n \lesssim a_n$.

Sieve measure of ill-posedness. For a NPIV model (1), an important quantity is the *measure of ill-posedness*, which, roughly speaking, measures how much the conditional expectation $h \mapsto E[h(X_i)|W_i = w]$ smoothes out h. Let $T : L^2(X) \to L^2(W)$ denote the conditional expectation operator given by

$$Th(w) = E[h(X_i)|W_i = w].$$

Let $\Psi_J = \text{clsp}\{\psi_{J1}, \ldots, \psi_{JJ}\} \subset L^2(X)$ and $B_K = \text{clsp}\{b_{K1}, \ldots, b_{KK}\} \subset L^2(W)$ denote the sieve spaces for the endogenous variables and instrumental variables, respectively. Let $\Psi_{J,1} = \{h \in \Psi_J : \|h\|_{L^2(X)} = 1\}$. The *sieve L^2 measure of ill-posedness* is

$$\tau_J = \sup_{h \in \Psi_J : h \neq 0} \frac{\|h\|_{L^2(X)}}{\|Th\|_{L^2(W)}} = \frac{1}{\inf_{h \in \Psi_{J,1}} \|Th\|_{L^2(W)}}.$$

Following Blundell, Chen, and Kristensen (2007), we call a NPIV model (1) with X_i being a d-dimensional random vector

  (i) *mildly ill-posed* if $\tau_J = O(J^{\varsigma/d})$ for some $\varsigma > 0$

 (ii) *severely ill-posed* if $\tau_J = O(\exp(\frac{1}{2}J^{\varsigma/d}))$ for some $\varsigma > 0$.

See our second version (Chen and Christensen (2015a)) for simple consistent estimation of the sieve measure of ill-posedness τ_J.

3.1 *Sup-norm convergence rates*

We first introduce some basic conditions on the basic NPIV model (1) and the sieve spaces.

ASSUMPTION 1. (i) *The variable X_i has compact rectangular support $\mathcal{X} \subset \mathbb{R}^d$ with nonempty interior and the density of X_i is uniformly bounded away from 0 and ∞ on $\mathcal{X}$; (ii) W_i has compact rectangular support $\mathcal{W} \subset \mathbb{R}^{d_w}$ and the density of W_i is uniformly bounded away from 0 and ∞ on $\mathcal{W}$; (iii) $T : L^2(X) \to L^2(W)$ is injective; (iv) $h_0 \in \mathcal{H} \subset L^\infty(X)$, and $\cup_J \Psi_J$ is dense in $(\mathcal{H}, \|\cdot\|_{L^\infty(X)})$.*

ASSUMPTION 2. *We have* (i) $\sup_{w \in \mathcal{W}} E[u_i^2|W_i = w] \leq \overline{\sigma}^2 < \infty$ *and* (ii) $E[|u_i|^{2+\delta}] < \infty$ *for some $\delta > 0$.*

The following assumptions concern the basis functions. Define

$$G_\psi = G_{\psi,J} = E[\psi^J(X_i)\psi^J(X_i)'] = E[\Psi'\Psi/n],$$

$$G_b = G_{b,K} = E[b^K(W_i)b^K(W_i)'] = E[B'B/n],$$

$$S = S_{KJ} = E[b^K(W_i)\psi^J(X_i)'] = E[B'\Psi/n].$$

We assume throughout that the basis functions are not linearly dependent, that is, S has full column rank J, and $G_{\psi,J}$ and $G_{b,K}$ are positive definite for each J and K, that is,

$e_J = \lambda_{\min}(G_{\psi,J}) > 0$ and $e_{b,K} = \lambda_{\min}(G_{b,K}) > 0$, although e_J and $e_{b,K}$ could go to zero as $K \geq J$ goes to infinity. Let

$$\zeta_\psi = \zeta_{\psi,J} = \sup_x \left\| G_\psi^{-1/2} \psi^J(x) \right\|_{\ell^2}, \qquad \zeta_b = \zeta_{b,K} = \sup_w \left\| G_b^{-1/2} b^K(w) \right\|_{\ell^2},$$

$$\xi_\psi = \xi_{\psi,J} = \sup_x \left\| \psi^J(x) \right\|_{\ell^1}$$

for each J and K, and define $\zeta = \zeta_J = \zeta_{b,K} \vee \zeta_{\psi,J}$. Note that $\zeta_{\psi,J}$ has some useful properties: $\|h\|_\infty \leq \zeta_{\psi,J} \|h\|_{L^2(X)}$ for all $h \in \Psi_J$ and $\sqrt{J} = (E[\|G_\psi^{-1/2} \psi^J(X)\|_{\ell^2}^2])^{1/2} \leq \zeta_{\psi,J} \leq \xi_{\psi,J}/\sqrt{e_J}$; clearly $\zeta_{b,K}$ has similar properties.

We say that the sieve basis for Ψ_J is Hölder continuous if there exist finite constants $\omega \geq 0$, $\omega' > 0$ such that $\|G_{\psi,J}^{-1/2}\{\psi^J(x) - \psi^J(x')\}\|_{\ell^2} \lesssim J^\omega \|x - x'\|_{\ell^2}^{\omega'}$ for all $x, x' \in \mathcal{X}$.

ASSUMPTION 3. (i) *The basis spanning* Ψ_J *is Hölder continuous,* (ii) $\tau_J \zeta^2/\sqrt{n} = O(1)$, *and* (iii) $\zeta^{(2+\delta)/\delta}\sqrt{(\log n)/n} = o(1)$.

Let $\Pi_J : L^2(X) \to \Psi_J$ denote the $L^2(X)$ orthogonal (i.e., least-squares) projection onto Ψ_J, namely $\Pi_J h_0 = \arg\min_{h \in \Psi_J} \|h_0 - h\|_{L^2(X)}$, and let $\Pi_K : L^2(W) \to B_K$ denote the $L^2(W)$ orthogonal (i.e., least-squares) projection onto B_K. Let $Q_J h_0 = \arg\min_{h \in \Psi_J} \|\Pi_K T(h_0 - h)\|_{L^2(W)}$ denote the sieve 2SLS projection of h_0 onto Ψ_J. We may write $Q_J h_0 = \psi^J(\cdot)' c_{0,J}$, where

$$c_{0,J} = [S' G_b^{-1} S]^{-1} S' G_b^{-1} E[b^K(W_i) h_0(X_i)].$$

ASSUMPTION 4. *We have* (i) $\sup_{h \in \Psi_{J,1}} \|(\Pi_K T - T)h\|_{L^2(W)} = o(\tau_J^{-1})$, (ii) $\tau_J \times \|T(h_0 - \Pi_J h_0)\|_{L^2(W)} \leq \text{const} \times \|h_0 - \Pi_J h_0\|_{L^2(X)}$, *and* (iii) $\|Q_J(h_0 - \Pi_J h_0)\|_\infty \leq O(1) \times \|h_0 - \Pi_J h_0\|_\infty$.

Discussion of Assumptions. Assumption 1 is standard. Assumption 1(iii) is stronger than needed for convergence rates in sup-norm only. We impose it as a common sufficient condition for convergence rates in both sup-norm and L^2-norm (Appendix B). For the sup-norm convergence rate only, Assumption 1(iii) could be replaced by the following alternative weaker identification condition:

ASSUMPTION 1. (iii-sup) *We have* $h_0 \in \mathcal{H} \subset L^\infty(X)$, *and* $T[h - h_0] = 0 \in L^2(W)$ *for any* $h \in \mathcal{H}$ *implies that* $\|h - h_0\|_\infty = 0$.

This in turn is implied by the injectivity of $T : L^\infty(X) \to L^2(W)$ (or the bounded completeness), which is weaker than the injectivity of $T : L^2(X) \to L^2(W)$ (i.e., the L^2 completeness). Bounded completeness or L^2-completeness condition is often assumed in models with endogeneity (e.g., Newey and Powell (2003), Carrasco, Florens, and Renault (2007), Blundell, Chen, and Kristensen (2007), Andrews (2011), Chen, Chernozhukov, Lee, and Newey (2014)) and is generically satisfied according to Andrews (2011). The parameter space $\mathcal{H}$ for h_0 is typically taken to be a Hölder or Sobolev class of smooth functions. Assumption 1(i) could be relaxed to unbounded support, and the

proofs need to be modified slightly using wavelet basis and weighted compact embedding results; see, for example, Blundell, Chen, and Kristensen (2007), Chen and Pouzo (2012), Triebel (2006), and references therein. To present the sup-norm rate results in a clean way we stick to the simplest Assumption 1. Assumption 2 is also imposed for sup-norm convergence rates for series LS regression under exogeneity (e.g., Chen and Christensen (2015b)). Assumption 3(i) is satisfied by many commonly used sieve bases, such as splines, wavelets, and cosine bases. Assumption 3(ii) and (iii) restrict the rate at which J can grow with n. Upper bounds for $\zeta_{\psi,J}$ and $\zeta_{b,K}$ are known for commonly used bases, for instance, under Assumption 1(i) and (ii), $\zeta_{b,K} = O(\sqrt{K})$ and $\zeta_{\psi,J} = O(\sqrt{J})$ for (tensor-product) polynomial spline, wavelet and cosine bases, and $\zeta_{b,K} = O(K)$ and $\zeta_{\psi,J} = O(J)$ for (tensor-product) orthogonal polynomial bases; see, for example, Newey (1997), Huang (1998) and Appendix E. Assumption 4(i) is a mild condition on the approximation properties of the basis used for the instrument space and is similar to the first part of Assumption 5(iv) of Horowitz (2014). In fact, $\|(\Pi_K T - T)h\|_{L^2(W)} = 0$ for all $h \in \Psi_J$ when the basis functions for B_K and Ψ_J form either a Riesz basis or an eigenfunction basis for the conditional expectation operator. Assumption 4(ii) is the usual L^2 "stability condition" imposed in the NPIV literature (cf. Assumption 6 in Blundell, Chen, and Kristensen (2007) and Assumption 5.2(ii) in Chen and Pouzo (2012)). Assumption 4(iii) is a new L^∞ stability condition to control the sup-norm bias. It turns out that Assumption 4(ii) and (iii) are also automatically satisfied by Riesz bases; see Appendix A for further discussions and sufficient conditions.

To derive the sup-norm (uniform) convergence rate, we split $\|\widehat{h} - h_0\|_\infty$ into "bias" and "standard deviation" terms and derive sup-norm convergence rates for the two terms. Specifically, let

$$\widetilde{h}(x) = \psi^J(x)'\widetilde{c} \quad \text{with } \widetilde{c} = \big[\Psi'B(B'B)^- B'\Psi\big]^- \Psi'B(B'B)^- B'H_0,$$

where $H_0 = (h_0(X_1), \ldots, h_0(X_n))'$. We refer loosely to $\|\widetilde{h} - h_0\|_\infty$ as the bias term and to $\|\widehat{h} - \widetilde{h}\|_\infty$ as the standard deviation (or sometimes "variance") term. Both are random quantities. We first bound the sup-norm standard deviation term in the following lemma.

LEMMA 3.1. *Let Assumptions* 1(i) *and* (iii), 2(i) *and* (ii), 3(ii) *and* (iii), *and* 4(i) *hold.*

   (i) *Then* $\|\widehat{h} - \widetilde{h}\|_\infty = O_p(\tau_J \xi_{\psi,J}\sqrt{(\log J)/(ne_J)})$.
   (ii) *If Assumption* 3(i) *also holds, then* $\|\widehat{h} - \widetilde{h}\|_\infty = O_p(\tau_J \zeta_{\psi,J}\sqrt{(\log n)/n})$.

Recall that $\sqrt{J} \le \zeta_{\psi,J} \le \xi_{\psi,J}/\sqrt{e_J}$. Result (ii) of Lemma 3.1 provides a slightly tighter upper bound on the variance term than result (i) does, while result (i) allows for slightly more general basis to approximate h_0. For splines and wavelets, we show in Appendix E that $\xi_{\psi,J}/\sqrt{e_J} \lesssim \sqrt{J}$, so results (i) and (ii) produce the same tight upper bound $\|\widehat{h} - \widetilde{h}\|_\infty = O_p(\tau_J\sqrt{(J\log n)/n})$ when $J \asymp n^r$ for some constant $r > 0$.

Before we present an upper bound on the bias term in Theorem 3.1(i) below, we mention one more property of the sieve space Ψ_J that is crucial for sharp bounds on the sup-norm bias term. Let $h_{0,J} \in \Psi_J$ denote the best approximation to h_0 in sup-norm,

that is, $h_{0,J}$ solves $\inf_{h \in \Psi_J} \|h_0 - h\|_\infty$. Then by Lebesgue's lemma (DeVore and Lorentz (1993, p. 30))

$$\|h_0 - \Pi_J h_0\|_\infty \leq \big(1 + \|\Pi_J\|_\infty\big) \times \|h_0 - h_{0,J}\|_\infty,$$

where $\|\Pi_J\|_\infty$ is the Lebesgue constant for the sieve Ψ_J. Recently it has been established that $\|\Pi_J\|_\infty \lesssim 1$ when Ψ_J is spanned by a tensor-product B-spline basis (Huang (2003)) or a tensor-product Cohen–Daubechies–Vial (CDV) wavelet basis (Chen and Christensen (2015b)).[14] Boundedness of the Lebesgue constant is crucial for attaining optimal sup-norm rates.

THEOREM 3.1.

(i) *Let Assumptions* 1(iii), 3(ii), *and* 4 *hold. Then*

$$\|\widetilde{h} - h_0\|_\infty = O_p\big(\|h_0 - \Pi_J h_0\|_\infty\big).$$

(ii) *Let Assumptions* 1(i), (iii), *and* (iv), 2(i) *and* (ii), 3(ii), *and* (iii), *and* 4 *hold. Then*

$$\|\widehat{h} - h_0\|_\infty = O_p\big(\|h_0 - \Pi_J h_0\|_\infty + \tau_J \xi_{\psi,J}\sqrt{(\log J)/(ne_J)}\big).$$

(iii) *Further, if the linear sieve Ψ_J satisfies $\|\Pi_J\|_\infty \lesssim 1$ and $\xi_{\psi,J}/\sqrt{e_J} \lesssim \sqrt{J}$, then*

$$\|\widehat{h} - h_0\|_\infty = O_p\big(\|h_0 - h_{0,J}\|_\infty + \tau_J\sqrt{(J \log J)/n}\big).$$

Theorem 3.1(ii) and (iii) follows directly from part (i) (for bias) and Lemma 3.1(i) (for standard deviation). See Appendix A for additional details about bounds on sup-norm bias.

The following corollary provides concrete sup-norm convergence rates of $\widehat{h}$ and its derivatives. To introduce the result, let $B^p_{\infty,\infty}$ denote the Hölder space of smoothness $p > 0$ and let $\|\cdot\|_{B^p_{\infty,\infty}}$ denote its norm (see Section 1.11.10 of Triebel (2006)). Let $B_\infty(p, L) = \{h \in B^p_{\infty,\infty} : \|h\|_{B^p_{\infty,\infty}} \leq L\}$ denote a Hölder ball of smoothness $p > 0$ and radius $L \in (0, \infty)$. Let $\alpha_1, \ldots, \alpha_d$ be nonnegative integers, let $|\alpha| = \alpha_1 + \cdots + \alpha_d$, and define

$$\partial^\alpha h(x) := \frac{\partial^{|\alpha|} h}{\partial^{\alpha_1} x_1 \cdots \partial^{\alpha_d} x_d} h(x).$$

Of course, if $|\alpha| = 0$, then $\partial^\alpha h = h$.[15]

COROLLARY 3.1. *Let Assumptions* 1(i), (ii), *and* (iii) *and* 4 *hold. Let $h_0 \in B_\infty(p, L)$, Ψ_J be spanned by a B-spline basis of order $\gamma > p$ or a CDV wavelet basis of regularity $\gamma > p$, and let B_K be spanned by a cosine, spline, or wavelet basis.*

---

[14]See DeVore and Lorentz (1993) and Belloni et al. (2015) for examples of other bases with bounded Lebesgue constant or with Lebesgue constant diverging slowly with the sieve dimension.

[15]If $|\alpha| > 0$, then we assume h and its derivatives can be continuously extended to an open set containing $\mathcal{X}$.

(i) *If Assumption* 3(ii) *holds, then*

$$\left\|\partial^\alpha \widetilde{h} - \partial^\alpha h_0\right\|_\infty = O_p\big(J^{-(p-|\alpha|)/d}\big) \quad \text{for all } 0 \le |\alpha| < p.$$

(ii) *If Assumptions* 2(i) *and* (ii) *and* 3(ii) *and* (iii) *hold, then*

$$\left\|\partial^\alpha \widehat{h} - \partial^\alpha h_0\right\|_\infty = O_p\big(J^{-(p-|\alpha|)/d} + \tau_J J^{|\alpha|/d}\sqrt{(J\log J)/n}\big) \quad \text{for all } 0 \le |\alpha| < p.$$

(ii)(a) *Mildly ill-posed case. With* $p \ge d/2$ *and* $\delta \ge d/(p+\varsigma)$, *choosing* $J \asymp (n/\log n)^{d/(2(p+\varsigma)+d)}$ *implies that Assumption* 3(ii) *and* (iii) *hold and*

$$\left\|\partial^\alpha \widehat{h} - \partial^\alpha h_0\right\|_\infty = O_p\big((n/\log n)^{-(p-|\alpha|)/(2(p+\varsigma)+d)}\big).$$

(ii)(b) *Severely ill-posed case. Choosing* $J = (c_0 \log n)^{d/\varsigma}$ *with* $c_0 \in (0,1)$ *implies that Assumption* 3(ii) *and* (iii) *hold and*

$$\left\|\partial^\alpha \widehat{h} - \partial^\alpha h_0\right\|_\infty = O_p\big((\log n)^{-(p-|\alpha|)/\varsigma}\big).$$

Corollary 3.1 shows that, for sieve NPIV estimators, taking derivatives has the same impact on the bias and standard deviation terms in terms of the order of convergence, and that the same choice of sieve dimension J can lead to optimal sup-norm convergence rates for estimating h_0 and its derivatives simultaneously (since they match the lower bounds in Theorem 3.2 below). When specializing to series LS regression (without endogeneity, i.e., $\tau_J = 1$), Corollary 3.1(ii)(a) with $\varsigma = 0$ automatically implies that spline and wavelet series LS estimators will also achieve the optimal sup-norm rates of Stone (1982) for estimating the derivatives of a nonparametric LS regression function. This strengthens the recent results in Belloni et al. (2015) and Chen and Christensen (2015b) for sup-norm rate optimality of spline and wavelet LS estimators of the regression function h_0 itself. This is in contrast to kernel-based LS regression estimators where different choices of bandwidth are needed for the optimal rates of estimating h_0 and its derivatives.

Corollary 3.1 is useful for estimating functions with certain shape properties. For instance, if $h_0 : [a, b] \to \mathbb{R}$ is strictly monotone and/or strictly concave/convex, then knowing that $\partial\widehat{h}(x)$ and/or $\partial^2\widehat{h}(x)$ converge uniformly to $\partial h_0(x)$ and/or $\partial^2 h_0(x)$ implies that $\widehat{h}$ will also be strictly monotone and/or strictly concave/convex wpa1. In this paper, we shall illustrate the usefulness of Corollary 3.1 in controlling the nonlinear remainder terms for pointwise and uniform inferences on highly nonlinear (i.e., beyond quadratic) functionals of h_0; see Sections 4 and 5 for details.

3.2 *Lower bounds*

We now establish that the sup-norm rates obtained in Corollary 3.1 are the best possible (i.e., minimax) sup-norm convergence rates for estimating h_0 and its derivatives.

To establish a lower bound, we require a *link condition* that relates smoothness of T to the parameter space for h_0. Let $\widetilde{\psi}_{j,k,G}$ denote a tensor-product CDV wavelet basis for $[0,1]^d$ of regularity $\gamma > p$. Appendix E provides details on the construction and properties of this basis.

CONDITION LB. (i) Assumption 1(i)–(iii) hold, (ii) $E[u_i^2|W_i = w] \geq \underline{\sigma}^2 > 0$ uniformly for $w \in \mathcal{W}$, and (iii) there is a positive decreasing function ν such that $\|Th\|_{L^2(W)}^2 \lesssim \sum_{j,G,k} [\nu(2^j)]^2 \langle h, \widetilde{\psi}_{j,k,G} \rangle_X^2$ holds for all $h \in B_\infty(p,L)$.

Condition LB is standard in the optimal rate literature (see Hall and Horowitz (2005) and Chen and Reiss (2011)). The mildly ill-posed case corresponds to choosing $\nu(t) = t^{-\varsigma}$, and says roughly that the conditional expectation operator T makes p-smooth functions of X into $(\varsigma + p)$-smooth functions of W. The severely ill-posed case, which corresponds to choosing $\nu(t) = \exp(-\frac{1}{2}t^\varsigma)$ and says roughly that T maps smooth functions of X into "supersmooth" functions of W.

THEOREM 3.2. *Let Condition* LB *hold for the NPIV model with a random sample* $\{(X_i, Y_i, W_i)\}_{i=1}^n$. *Then, for any* $0 \leq |\alpha| < p$,

$$\liminf_{n \to \infty} \inf_{\widehat{g}_n} \sup_{h \in B_\infty(p,L)} \mathbb{P}_h\big(\|\widehat{g}_n - \partial^\alpha h\|_\infty \geq cr_n\big) \geq c' > 0,$$

where

$$r_n = \begin{bmatrix} (n/\log n)^{-(p-|\alpha|)/(2(p+\varsigma)+d)} & \text{in the mildly ill-posed case,} \\ (\log n)^{-(p-|\alpha|)/\varsigma} & \text{in the severely ill-posed case,} \end{bmatrix}$$

$\inf_{\widehat{g}_n}$ *denotes the infimum over all estimators of* $\partial^\alpha h$ *based on the sample of size* n, $\sup_{h \in B_\infty(p,L)} \mathbb{P}_h$ *denotes the sup over* $h \in B_\infty(p, L)$, *and distributions of* (X_i, W_i, u_i) *that satisfy Condition* LB *with fixed* ν, *and the finite positive constants* c *and* c' *do not depend on* n.

According to Theorem 3.2 and Theorem B.2 (in Appendix B), the minimax lower bounds in sup-norm for estimating h_0 and its derivatives coincide with those in L^2 for severely ill-posed NPIV problems, and are only a factor of $[\log(n)]^\epsilon$ (with $\epsilon = \frac{p-|\alpha|}{2(p+\varsigma)+d} < \frac{p}{2p+d} < \frac{1}{2}$) worse than those in L^2 for mildly ill-posed problems. Our proof of sup-norm lower bound for NPIV models is similar to that of Chen and Reiss (2011) for L^2-norm lower bound. Similar sup-norm lower bounds for density deconvolution were recently obtained by Lounici and Nickl (2011).

3.3 *Models with endogenous and exogenous regressors*

In many empirical studies, some regressors might be endogenous while others are exogenous. Consider the model

$$Y_i = h_0(X_{1i}, Z_i) + u_i, \tag{16}$$

where X_{1i} is a vector of endogenous regressors and Z_i is a vector of exogenous regressors. Let $X_i = (X_{1i}', Z_i')'$. Here the vector of instrumental variables W_i is of the form $W_i = (W_{1i}', Z_i')'$, where W_{1i} are instruments for X_{1i}. We refer to this as the *partially endogenous case*. The sieve NPIV estimator is implemented in exactly the same way as the *fully*

endogenous setting in which X_i consists only of endogenous variables, just like 2SLS with endogenous and exogenous regressors.[16] Our convergence rates presented in Section 3.1 and Appendix B apply equally to the partially endogenous model (16) under the stated regularity conditions: all that differs between the two cases is the interpretation of the sieve measure of ill-posedness.

Consider first the fully endogenous case where $T : L^2(X) \to L^2(W)$ is compact under mild conditions on the conditional density of X given W (see, e.g., Newey and Powell (2003), Blundell, Chen, and Kristensen (2007), Darolles et al. (2011), Andrews (2011)). Then T admits a singular value decomposition (SVD) $\{\phi_{0j}, \phi_{1j}, \mu_j\}_{j=1}^\infty$, where $(T^*T)^{1/2}\phi_{0j} = \mu_j\phi_{0j}$, $\mu_j \geq \mu_{j+1}$ for each j, and $\{\phi_{0j}\}_{j=1}^\infty$ and $\{\phi_{1j}\}_{j=1}^\infty$ are orthonormal bases for $L^2(X)$ and $L^2(W)$, respectively. Suppose that Ψ_J spans $\phi_{0j}, \ldots, \phi_{0J}$. Then the sieve measure of ill-posedness is $\tau_J = \mu_J^{-1}$.

Now consider the partially endogenous case. Similar to Horowitz (2011), we suppose that for each value of z the conditional expectation operator $T_z : L^2(X_1|Z = z) \to L^2(W_1|Z = z)$ given by $(T_z h)(w_1) = E[h(X_1)|W_{1i} = w_1, Z_i = z]$ is compact. Then each T_z admits a SVD $\{\phi_{0j,z}, \phi_{1j,z}, \mu_{j,z}\}_{j=1}^\infty$, where $T_z\phi_{0j,z} = \mu_{j,z}\phi_{1j,z}$, $(T_z^*T_z)^{1/2}\phi_{0j,z} = \mu_{j,z}\phi_{0j,z}$, $(T_zT_z^*)^{1/2}\phi_{1j,z} = \mu_{j,z}\phi_{1j,z}$, $\mu_{j,z} \geq \mu_{j+1,z}$ for each j and z, and $\{\phi_{0j,z}\}_{j=1}^\infty$ and $\{\phi_{1j,z}\}_{j=1}^\infty$ are orthonormal bases for $L^2(X_1|Z = z)$ and $L^2(W_1|Z = z)$, respectively, for each z. The following result adapts Lemma 1 of Blundell, Chen, and Kristensen (2007) to the partially endogenous setting.

LEMMA 3.2. *Let T_z be compact with SVD $\{\phi_{0j,z}, \phi_{1j,z}, \mu_{j,z}\}_{j=1}^\infty$ for each z. Let $\mu_j^2 = E[\mu_{j,Z_i}^2]$ and $\phi_{0j}(\cdot, z) = \phi_{0j,z}(\cdot)$ for each z and j. Then*

(i) $\tau_J \geq \mu_J^{-1}$.

(ii) *If, in addition, $\phi_{01}, \ldots, \phi_{0J} \in \Psi_J$, then $\tau_J \leq \mu_J^{-1}$.*

Consider the following partially endogenous stylized example from Hoderlein and Holzmann (2011). Let X_{1i}, W_{1i}, and Z_i be scalar random variables with

$$\begin{pmatrix} X_{1i} \\ W_{1i} \\ Z_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{xw} & \rho_{xz} \\ \rho_{xw} & 1 & \rho_{wz} \\ \rho_{xz} & \rho_{wz} & 1 \end{pmatrix}\right).$$

Then

$$\left(\begin{array}{c|c} \dfrac{X_{1i} - \rho_{xz}z}{\sqrt{1-\rho_{xz}^2}} & \\ \dfrac{W_{1i} - \rho_{wz}z}{\sqrt{1-\rho_{wz}^2}} & Z_i = z \end{array}\right) \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{xw|z} \\ \rho_{xw|z} & 1 \end{pmatrix}\right), \tag{17}$$

---

[16]All that changes here is that J may grow more quickly as the degree of ill-posedness will be smaller. In contrast, other NPIV estimators based on estimating the conditional densities of the regressors and instrumental variables must be implemented separately for each value of z (Hall and Horowitz (2005), Horowitz (2011), Gagliardini and Scaillet (2012)).

where

$$\rho_{xw|z} = \frac{\rho_{xw} - \rho_{xz}\rho_{wz}}{\sqrt{(1-\rho_{xz}^2)(1-\rho_{wz}^2)}}$$

is the partial correlation between X_{1i} and W_{1i} given Z_i. For each $j \geq 1$, let H_j denote the jth Hermite polynomial (the Hermite polynomials form an orthonormal basis with respect to Gaussian density). Since $T_z : L^2(X_1|Z = z) \to L^2(W_1|Z = z)$ is compact for each z, it follows from Mehler's formula that T_z has a SVD $\{\phi_{0j,z}, \phi_{1j,z}, \mu_{j,z}\}_{j=1}^{\infty}$ with

$$\phi_{0j,z}(x_1) = H_{j-1}\left(\frac{x_1 - \rho_{xz}z}{\sqrt{1-\rho_{xz}^2}}\right), \qquad \phi_{1j,z}(w_1) = H_{j-1}\left(\frac{w_1 - \rho_{wz}z}{\sqrt{1-\rho_{wz}^2}}\right), \qquad \mu_{j,z} = |\rho_{xw|Z}|^{j-1}$$

for each z. Since $\mu_{J,z} = |\rho_{xw|z}|^{J-1}$ for each z, we have $\mu_J = |\rho_{xw|z}|^{J-1} \asymp |\rho_{xw|z}|^J$. If X_{1i} and W_{1i} are uncorrelated with Z_i, then $\mu_J = |\rho|^{J-1}$, where $\rho = \rho_{xw}$.

In contrast, consider the following fully endogenous model in which X_i and W_i are bivariate with

$$\begin{pmatrix} X_{1i} \\ X_{2i} \\ W_{1i} \\ W_{2i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & \rho_1 & 0 \\ 0 & 1 & 0 & \rho_2 \\ \rho_1 & 0 & 1 & 0 \\ 0 & \rho_2 & 0 & 1 \end{pmatrix}\right),$$

where ρ_1 and ρ_2 are such that the covariance matrix is invertible. It is straightforward to verify that T has singular value decomposition with

$$\phi_{0j}(x) = H_{j-1}(x_1)H_{j-1}(x_2), \qquad \phi_{1j}(w) = H_{j-1}(w_1)H_{j-2}(w_2), \qquad \mu_j = |\rho_1\rho_2|^{j-1},$$

and $\mu_J = \rho^{2(J-1)}$ if $\rho_1 = \rho_2 = \rho$. Thus, the measure of ill-posedness diverges faster in the fully endogenous case ($\mu_J = \rho^{2(J-1)}$) than that in the partially endogenous case ($\mu_J = |\rho|^{J-1}$).

4. Uniform inference on collections of nonlinear functionals

In this section, we apply our sup-norm rate results and tight bounds on random matrices (in Appendix F) to establish uniform Gaussian process strong approximation and the consistency of the score bootstrap UCBs defined in (5) for collections of (possibly) nonlinear functionals $\{f_t(\cdot) : t \in \mathcal{T}\}$ of a NPIV function h_0. See Section 6 for discussions of other applications.

We consider functionals $f_t : \mathcal{H} \subset L^{\infty}(X) \to \mathbb{R}$ for each $t \in \mathcal{T}$ for which $Df_t(h)[v] = \lim_{\delta \to 0^+}[\delta^{-1}\{f_t(h+\delta v) - f_t(h)\}]$ exists for all $v \in \mathcal{H} - \{h_0\}$ for all h in a small neighborhood of h_0 (where the neighborhood is independent of t). This is trivially true for, say, $f_t(h) = h(t)$ with $\mathcal{T} \subseteq \mathcal{X}$ for UCBs for h_0. Let $\Omega = E[u_i^2 b^K(W_i)b^K(W_i)']$. Then the 2SLS covariance matrix for $\widehat{c}$ (given in (2)) is

$$\mho = [S'G_b^{-1}S]^{-1}S'G_b^{-1}\Omega G_b^{-1}S[S'G_b^{-1}S]^{-1},$$

and the sieve variance for $f_t(\widehat{h})$ is

$$\big[\sigma_n(f_t)\big]^2 = \big(Df_t(h_0)\big[\psi^J\big]\big)'\mho\big(Df_t(h_0)\big[\psi^J\big]\big).$$

Assumption 2 (continued). (iii) *We have* $E[u_i^2|W_i = w] \geq \underline{\sigma}^2 > 0$ *uniformly for all* $w \in \mathcal{W}$ *and* (iv) $\sup_w E[|u_i|^3|W_i = w] < \infty$.

Assumption 2(iii) and (iv) are reasonably mild conditions used to derive the uniform limit theory. Define

$$v_n(f_t)(x) = \psi^J(x)'\big[S'G_b^{-1}S\big]^{-1}Df_t(h_0)\big[\psi^J\big], \qquad \widehat{v}_n(f_t)(x) = \psi^J(x)'\big[S'G_b^{-1}S\big]^{-1}Df_t(\widehat{h})\big[\psi^J\big],$$

where, for each fixed t, $v_n(f_t)$ could be viewed as a "sieve 2SLS Riesz representer." Note that $v_n(f_t) = \widehat{v}_n(f_t)$ whenever f_t is linear. Under Assumption 2(i) and (iii), we have that

$$\big[\sigma_n(f_t)\big]^2 \asymp Df_t(h_0)\big[\psi^J\big]'\big[S'G_b^{-1}S\big]^{-1}Df_t(h_0)\big[\psi^J\big] = \big\|\Pi_K Tv_n(f_t)\big\|_{L^2(W)}^2 \quad \text{uniformly in } t.$$

Following Chen and Pouzo (2015), we call $f_t(\cdot)$ an irregular functional of h_0 (i.e., slower than $\sqrt{n}$ estimable) if $\sigma_n(f_t) \nearrow +\infty$ as $n \to \infty$. This includes the evaluation functionals $h_0(t)$ and $\partial^\alpha h_0(t)$ as well as $f_{\mathrm{CS},t}(h_0)$ and $f_{\mathrm{DL},t}(h_0)$. In this paper, we shall focus on applications of sup-norm rate results to inference on irregular functionals.

Assumption 5. *Let* η_n *and* η'_n *be sequences of nonnegative numbers such that* $\eta_n = o(1)$ *and* $\eta'_n = o(1)$. *Let* $\sigma_n(f_t) \nearrow +\infty$ *as* $n \to \infty$ *for each* $t \in \mathcal{T}$. *Either* (a) *or* (b) *of the following options holds*:

(a) *The functional* f_t *is a linear functional for each* $t \in \mathcal{T}$ *and* $\sup_{t \in \mathcal{T}} \sqrt{n}(\sigma_n(f_t))^{-1} \times |f_t(\widetilde{h}) - f_t(h_0)| = O_p(\eta_n)$.

(b) *The functional* (i) $v \mapsto Df_t(h_0)[v]$ *is a linear functional for each* $t \in \mathcal{T}$; (ii)

$$\sup_{t \in \mathcal{T}}\left|\sqrt{n}\frac{f_t(\widehat{h}) - f_t(h_0)}{\sigma_n(f_t)} - \sqrt{n}\frac{Df_t(h_0)[\widehat{h} - \widetilde{h}]}{\sigma_n(f_t)}\right| = O_p(\eta_n);$$

(iii) $\sup_{t \in \mathcal{T}} \frac{\|\Pi_K T(\widehat{v}_n(f_t) - v_n(f_t))\|_{L^2(W)}}{\sigma_n(f_t)} = O_p(\eta'_n)$.

Assumption 5(a) and (b)(i) and (ii) are similar to uniform-in-t versions of Assumption 3.5 of Chen and Pouzo (2015). Assumption 5(b)(iii) controls any additional error arising in the estimation of $\sigma_n(f_t)$ by $\widehat{\sigma}(f_t)$ (given in equation (7)) due to nonlinearity of $f_t(\cdot)$, and is automatically satisfied with $\eta'_n = 0$ when $f_t(\cdot)$ is a linear functional.

The next remark presents a set of sufficient conditions for Assumption 5 when $\{f_t : t \in \mathcal{T}\}$ are irregular functionals of h_0. Since the functionals are irregular, the quantity $\underline{\sigma}_n := \inf_{t \in \mathcal{T}} \sigma_n(f_t)$ will typically satisfy $\underline{\sigma}_n \nearrow +\infty$ as $n \to \infty$. Our sup-norm rates for $\widehat{h}$ and $\widetilde{h}$, together with divergence of $\underline{\sigma}_n$, help to control the nonlinearity bias terms.

Remark 4.1. Let $\mathcal{H}_n \subseteq \mathcal{H}$ be a sequence of neighborhoods of h_0 with $\widehat{h}, \widetilde{h} \in \mathcal{H}_n$ wpa1 and assume $\underline{\sigma}_n := \inf_{t \in \mathcal{T}} \sigma_n(f_t) > 0$ for each n. Then Assumption 5(a) is implied by (a') and Assumption 5(b) is implied by (b'), where the following alternative statements hold:

(a′) (i) The functional f_t is a linear functional for each $t \in \mathcal{T}$ and there exists α with $|\alpha| \geq 0$ such that $\sup_t |f_t(h - h_0)| \lesssim \|\partial^\alpha h - \partial^\alpha h_0\|_\infty$ for all $h \in \mathcal{H}_n$, and (ii) $n^{1/2}\underline{\sigma}_n^{-1}\|\partial^\alpha \widetilde{h} - \partial^\alpha h_0\|_\infty = O_p(\eta_n)$.

(b′) (i) The functional $v \mapsto Df_t(h_0)[v]$ is a linear functional for each $t \in \mathcal{T}$ and there exists α with $|\alpha| \geq 0$ such that $\sup_t |Df_t(h_0)[h - h_0]| \lesssim \|\partial^\alpha h - \partial^\alpha h_0\|_\infty$ for all $h \in \mathcal{H}_n$.

(ii) There are α_1 and α_2 with $|\alpha_1|, |\alpha_2| \geq 0$ such that

(ii.1) $\quad \sup_t |f_t(\widehat{h}) - f_t(h_0) - Df_t(h_0)[\widehat{h} - h_0]| \lesssim \|\partial^{\alpha_1}\widehat{h} - \partial^{\alpha_1}h_0\|_\infty \|\partial^{\alpha_2}\widehat{h} - \partial^{\alpha_2}h_0\|_\infty,$

(ii.2) $\quad n^{1/2}\underline{\sigma}_n^{-1}(\|\partial^{\alpha_1}\widehat{h} - \partial^{\alpha_1}h_0\|_\infty \|\partial^{\alpha_2}\widehat{h} - \partial^{\alpha_2}h_0\|_\infty + \|\partial^\alpha \widetilde{h} - \partial^\alpha h_0\|_\infty) = O_p(\eta_n).$

(iii) Additionally, $\sup_{t \in \mathcal{T}} \dfrac{(\tau_J)\sqrt{\sum_{j=1}^J (Df_t(\widehat{h})[(G_\psi^{-1/2}\psi^J)_j] - Df_t(h_0)[(G_\psi^{-1/2}\psi^J)_j])^2}}{\sigma_n(f_t)} = O_p(\eta_n').$

Condition (a′)(i) is automatically satisfied by functionals of the form $f_t(h) = \partial^\alpha h(t)$ with $\mathcal{T} \subseteq \mathcal{X}$ and $\mathcal{H}_n = \mathcal{H}$. Conditions (a′)(i) and (b′)(i) and (ii) are sufficient conditions that are formulated to take advantage of the sup-norm rate results in Section 3. For example, conditions (b′)(i) and (ii.1) are easily satisfied by exact CS and DL functionals (Lemma A.1 of Hausman and Newey (1995)). Condition (b′)(ii.2) is simply satisfied by applying our sup-norm rate results. Condition (b′)(iii) is a sufficient condition for Assumption 5(b)(iii) and is needed for uniform-in-t consistent estimation of $\sigma_n(f_t)$ by $\widehat{\sigma}(f_t)$ only, and is automatically satisfied with $\eta_n' = 0$ when $f_t(\cdot)$ is a linear functional.

The next assumption concerns the set of normalized sieve 2SLS Riesz representers, given by

$$u_n(f_t)(x) = v_n(f_t)(x)/\sigma_n(f_t).$$

Let d_n denote the semimetric on $\mathcal{T}$ given by $d_n(t_1, t_2)^2 = E[(u_n(f_{t_1})(X_i) - u_n(f_{t_2})(X_i))^2]$ and let $N(\mathcal{T}, d_n, \epsilon)$ be the ϵ-covering number of $\mathcal{T}$ with respect to d_n. Let η_n and η_n' be from Assumption 5, and let $\delta_{h,n}$ be a sequence of positive constants such that $\|\widehat{h} - h_0\|_\infty = O_p(\delta_{h,n}) = o_p(1)$. Denote $\delta_{V,n} \equiv [\zeta_{b,K}^{(2+\delta)/\delta}\sqrt{(\log K)/n}]^{\delta/(1+\delta)} + \tau_J\zeta\sqrt{(\log J)/n} + \delta_{h,n}$.

ASSUMPTION 6. (i) *There is a sequence of finite constants $c_n \gtrsim 1$ that could grow to infinity such that*

$$1 + \int_0^\infty \sqrt{\log N(\mathcal{T}, d_n, \epsilon)}\, d\epsilon = O(c_n)$$

and (ii) *there is a sequence of constants $r_n > 0$ decreasing to zero slowly such that*

(ii)(a) $\ r_n c_n \lesssim 1$ *and* $\frac{\zeta_{b,K}J^2}{r_n^3\sqrt{n}} = o(1)$,

(ii)(b) $\ \tau_J\zeta\sqrt{(J\log J)/n} + \eta_n + (\delta_{V,n} + \eta_n') \times c_n = o(r_n)$, *with $\eta_n' \equiv 0$ when $f_t(\cdot)$ is linear.*

Assumption 6(i) is a mild regularity condition requiring that the class $\{u_n(f_t) : t \in \mathcal{T}\}$ not be too complex; see Remark 4.2 below for sufficient conditions to bound c_n. Assumption 6(ii) strengthens conditions on the growth rate of J. Condition $\frac{\zeta_{b,K}J^2}{r_n^3\sqrt{n}} = o(1)$ of

Assumption 6(ii)(a) is used to apply Yurinskii's coupling Chernozhukov, Lee, and Rosen (2013), Pollard (2002, Theorem 10, p. 244) to derive a uniform Gaussian process strong approximation to the linearized sieve process $\{\widehat{\mathbb{Z}}_n(t) : t \in \mathcal{T}\}$ (defined in equation (18)). This condition could be improved if other types of strong approximation probability tools are used. Assumption 6(ii)(b) ensures that both the *nonlinear* remainder terms and the error in estimating $\sigma_n(f_t)$ by $\widehat{\sigma}(f_t)$ vanish sufficiently fast. While the consistency of $\widehat{\sigma}(f)$ is enough for the pointwise asymptotic normality of the plug-in sieve t-statistic for $f(h_0)$ (see Theorem D.1 in Appendix D), we need the rate of convergence for uniform inference

$$\sup_{t \in \mathcal{T}} \left| \frac{\sigma_n(f_t)}{\widehat{\sigma}(f_t)} - 1 \right| = O_p\big(\delta_{V,n} + \eta'_n\big),$$

which is established using our results on sup-norm convergence rates of sieve NPIV; see Lemma G.4 in Appendix G.

REMARK 4.2. Let Assumptions 1(iii) and 4(i) hold. Let $\mathcal{T}$ be a compact subset in $\mathbb{R}^{d_T}$, and let there exist positive sequences Γ_n and γ_n such that for any $t_1, t_2 \in \mathcal{T}$,

$$\sup_{h \in \Psi_J : \|h\|_{L^2(X)} = 1} \left| \big(Df_{t_1}(h_0)[h] - Df_{t_2}(h_0)[h]\big) \right| \leq \Gamma_n \|t_1 - t_2\|_{\ell^2}^{\gamma_n}.$$

Then Assumption 6(i) holds with $c_n = 1 + \int_0^\infty \sqrt{\{(d_T/\gamma_n) \log(\Gamma_n \tau_J/(\epsilon \underline{\sigma}_n))\} \vee 0} \, d\epsilon$.

The next lemma is about uniform Bahadur representation and uniform Gaussian process strong approximation for the sieve t-statistic process for (possibly) nonlinear functionals of NPIV. Define

$$\widehat{\mathbb{Z}}_n(t) = \frac{\big(Df_t(h_0)[\psi^J]\big)' [S'G_b^{-1}S]^{-1} S' G_b^{-1/2}}{\sigma_n(f_t)} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n G_b^{-1/2} b^K(W_i) u_i \right), \qquad (18)$$

$$\mathbb{Z}_n(t) = \frac{\big(Df_t(h_0)[\psi^J]\big)' [S'G_b^{-1}S]^{-1} S' G_b^{-1/2}}{\sigma_n(f_t)} \mathcal{Z}_n$$

with $\mathcal{Z}_n \sim N(0, G_b^{-1/2} \Omega G_b^{-1/2})$. Note that $\mathbb{Z}_n(t)$ is a Gaussian process indexed by $t \in \mathcal{T}$.

LEMMA 4.1. *Let Assumptions* 1(iii), 2, 3(ii) *and* (iii), 4(i), 5, *and* 6 *hold. Then*

$$\sup_{t \in \mathcal{T}} \left| \frac{\sqrt{n}\big(f_t(\widehat{h}) - f_t(h_0)\big)}{\widehat{\sigma}(f_t)} - \mathbb{Z}_n(t) \right| = \sup_{t \in \mathcal{T}} \left| \frac{\sqrt{n}\big(f_t(\widehat{h}) - f_t(h_0)\big)}{\widehat{\sigma}(f_t)} - \widehat{\mathbb{Z}}_n(t) \right| + o_p(r_n)$$

$$= o_p(r_n). \qquad (19)$$

Lemma 4.1 is used in this paper to establish the consistency of the sieve score bootstrap for estimating the critical values of the uniform sieve t-statistic process, $\sup_{t \in \mathcal{T}} |\frac{\sqrt{n}(f_t(\widehat{h}) - f_t(h_0))}{\widehat{\sigma}(f_t)}|$, for a NPIV model. The strong approximation result, however, is also useful for various applications to testing equality and/or inequality (such as shape) constraints on $f_t(h_0)$, and is therefore of independent interest.

In what follows, $\mathbb{P}^*(\cdot)$ denotes a probability measure conditional on the data $Z^n := \{(X_i, Y_i, W_i)\}_{i=1}^n$. Recall that $\mathbb{Z}_n^*(t)$ is defined in equation (8).

THEOREM 4.1. *Let conditions of Lemma* 4.1 *hold. Let* $\eta_n' \sqrt{J} = o(r_n)$ *for nonlinear* $f_t()$. *Let the bootstrap weights* $\{\varpi_i\}_{i=1}^n$ *be IID with zero mean, unit variance, and finite third moment, and independent of the data. Then*

$$\sup_{s \in \mathbb{R}} \left| \mathbb{P}\left(\sup_{t \in \mathcal{T}} \left| \frac{\sqrt{n}(f_t(\widehat{h}) - f_t(h_0))}{\widehat{\sigma}(f_t)} \right| \le s \right) - \mathbb{P}^*\left(\sup_{t \in \mathcal{T}} \left| \mathbb{Z}_n^*(t) \right| \le s \right) \right| = o_p(1). \tag{20}$$

Theorem 4.1 appears to be the first to establish consistency of a sieve score bootstrap for uniform inference on general nonlinear functionals of NPIV under low-level conditions. When specializing to collections of linear functionals, Lemma 4.1, Theorem 4.1, and Corollary 3.1 immediately imply the following result.

COROLLARY 4.1. *Consider a collection of linear functionals* $\{f_t(h_0) = \partial^\alpha h_0(t) : t \in \mathcal{T}\}$ *of the NPIV function* h_0, *with* $\mathcal{T}$ *a compact convex subset of* $\mathcal{X}$. *Let Assumptions* 1(i), (ii), *and* (iii) *and* 2 (*with* $\delta \ge 1$) *hold, let* $h_0 \in B_\infty(p, L)$, *let* Ψ_J *be formed from a B-spline basis of regularity* $\gamma > (p \vee 2 + |\alpha|)$, *let* B_K *be a B-spline, wavelet, or cosine basis, and let* $\sigma_n(f_t) \asymp \tau_J J^a$ *uniformly in* t *with* $a = \frac{1}{2} + \frac{|\alpha|}{d}$. *For* $\kappa \in [1/2, 1]$ *we set* $J^5 (\log n)^{6\kappa}/n = o(1)$, $\tau_J J (\log J)^{\kappa + 0.5}/\sqrt{n} = o(1)$, *and* $J^{-p/d} = o([\log J]^{-\kappa} \tau_J \sqrt{J/n})$. *Then results* (19) (*with* $r_n = (\log J)^{-\kappa}$) *and* (20) *hold for* $f_t(h_0) = \partial^\alpha h_0(t)$.

Recently Horowitz and Lee (2012) developed a notion of UCBs for a NPIV function h_0 of a scalar endogenous regressor $X_i \in [0, 1]$ based on interpolation over a growing number of uniformly generated random grid points on $[0, 1]$, with h_0 estimated via the modified orthogonal series NPIV estimator of Horowitz (2011).[17] When specializing Corollary 4.1 to a NPIV function of a scalar regressor (i.e., $d = 1$ and $|\alpha| = 0$), our sufficient conditions are comparable to theirs (see their Theorem 4.1). Our score bootstrap UCBs would be computationally much simpler for a NPIV function of a multivariate endogenous regressor X_i, however.

When X_i is exogenous, the sieve NPIV estimator $\widehat{h}$ reduces to the series LS estimator of a nonparametric regression $h_0(x) = E[Y_i|W_i = x]$ with $X_i = W_i$, $K = J$, and $b^K = \psi^J$ with $\tau_J = 1$. Lemma 4.1 and Theorem 4.1 immediately imply the validity of Gaussian strong approximation and sieve score bootstrap UCBs for collections of general nonlinear functionals of a nonparametric LS regression. We note that the regularity conditions in Lemma 4.1 and Theorem 4.1 are much weaker for models with exogenous regressors. For instance, when specializing Corollary 4.1 to a nonparametric LS regression with exogenous regressor X_i, the conditions on J simplify to $J^5(\log n)^{6\kappa}/n = o(1)$ and $J^{-p/d} = o([\log J]^{-\kappa}\sqrt{J/n})$ for $\kappa \in [1/2, 1]$, and results (19) (with $r_n = [\log J]^{-\kappa}$) and (20) both hold for linear functionals $\{f_t(h_0) = \partial^\alpha h(t_0) : t \in \mathcal{T}\}$ of $h_0(\cdot) = E[Y_i|X_i = \cdot]$. These

---

[17] Remark 4 in Horowitz and Lee (2012) mentioned that their notion of UCB is different from the standard UCBs. They also proved the consistency of their bootstrap confidence bands over a fixed finite number of grid points.

conditions on J are the same as those in Chernozhukov, Lee, and Rosen (2013) for h_0 (see their Theorem 7) and Belloni et al. (2015) for linear functionals of h_0 (see their Theorem 5.5 with $r_n = [\log J]^{-1/2}$) estimated via series LS.

To the best of our knowledge, there is no published work on uniform Gaussian process strong approximation and sieve score bootstrap for general *nonlinear functionals* of sieve NPIV or series LS regression. The results in this section are thus presented as nontrivial applications of our sup-norm rate results for sieve NPIV, and are not aimed at weakest sufficient conditions.

4.1 *Monte Carlo*

We now evaluate the finite sample performance of our sieve score bootstrap UCBs for h_0 in NPIV model (1). We use the experimental design of Newey and Powell (2003), in which IID draws are generated from

$$\begin{pmatrix} u_i \\ V_i^* \\ W_i^* \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 & 0 \\ 0.5 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right)$$

from which we then set $X_i^* = W_i^* + V_i^*$. To ensure compact support of the regressor and instrument, we rescale X_i^* and W_i^* by defining $X_i = \Phi(X_i^*/\sqrt{2})$ and $W_i = \Phi(W_i^*)$, where Φ is the Gaussian cumulative distribution function (cdf). We use $h_0(x) = 4x - 2$ for our *linear* design and $h_0(x) = \log(|16x - 8| + 1)\,\text{sgn}(x - \frac{1}{2})$ for our *nonlinear* design (our nonlinear h_0 is a rescaled version of the h_0 used in Newey and Powell (2003)). Note that p for the nonlinear h_0 is between 1 and 2, so h_0 is not particularly smooth ($h_0'(x)$ has a kink at $x = \frac{1}{2}$).

We generate 1000 samples of length 1000 and implement our procedure using a B-spline basis for B_K and Ψ_J. For each simulation, we calculate the 90%, 95%, and 99% uniform confidence bands for h_0 over the support $[0.05, 0.95]$ with 1000 bootstrap replications for each simulation. We draw the bootstrap innovations ϖ_i from the two-point distribution of Mammen (1993). We then calculate the Monte Carlo (MC) coverage probabilities of our uniform confidence bands.

Figure 5 displays the estimated structural function $\widehat{h}$ and confidence bands together with a scatter plot of the sample (X_i, Y_i) data for the nonlinear design. The true function h_0 is seen to lie inside the UCBs. The results of this MC experiment are presented in Table 2. By comparing the MC coverage probabilities with their nominal values, it is clear that the uniform confidence bands for the linear design are slightly too conservative. However, the uniform confidence bands for the nonlinear design using cubic B-splines to approximate h_0 have MC converge much closer to the nominal coverage probabilities.

5. Pointwise and uniform inference on nonparametric welfare functionals

We now apply our sup-norm rate results to study pointwise and uniform inference on nonlinear welfare functionals in nonparametric demand estimation with endogeneity.
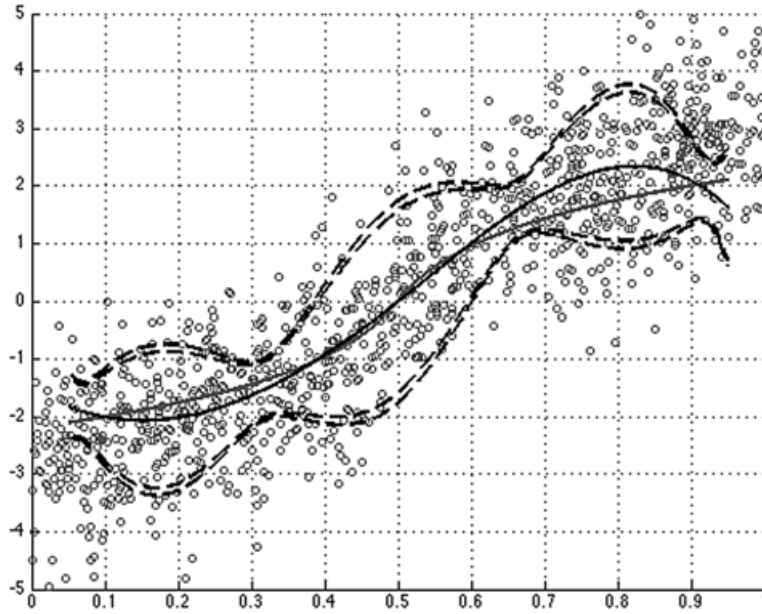
FIGURE 5. The 90% and 95% uniform confidence bands for h_0 (dashed lines; innermost are 90%), NPIV estimate $\widehat{h}$ (solid black line), and true structural function h_0 (solid grey line) for the nonlinear design.

TABLE 2. MC coverage probabilities of uniform confidence bands for h_0. Results are presented for cubic (C) and quartic (Q) B-spline bases for Ψ_J and B_K. Confidence interval is abbreviated CI.

Ψ_J	B_K	J	K	Design 1: Linear h_0			Design 2: Nonlinear h_0		
				90% CI	95% CI	99% CI	90% CI	95% CI	99% CI
C	C	5	5	0.962	0.983	0.996	0.896	0.942	0.987
C	C	5	6	0.957	0.983	0.996	0.845	0.924	0.981
C	Q	5	5	0.961	0.982	0.996	0.884	0.939	0.985
C	Q	5	6	0.958	0.983	0.997	0.846	0.921	0.981
Q	Q	5	5	0.964	0.984	0.997	0.913	0.948	0.989
Q	Q	5	6	0.961	0.985	0.996	0.886	0.937	0.983

First, we provide mild sufficient conditions under which plug-in sieve t-statistics for exact CS and DL and approximate CS functionals are asymptotically $N(0, 1)$, allowing for mildly and severely ill-posed NPIV models (Sections 5.1 and 5.2). Second, under stronger sufficient conditions but still allowing for severely ill-posed NPIV models, the validity of uniform Gaussian process strong approximations and sieve score bootstrap UCBs for exact CS and DL over a range of taxes and/or incomes (Section 5.3) are presented. When specialized to inference on exact CS and DL and approximate CS functionals of nonparametric demand estimation without endogeneity, our pointwise asymptotic normality

results are valid under sufficient conditions weaker than those in the existing literature, while our uniform inference results appear to be new (Section 5.4).

Previously, Hausman and Newey (1995) and Newey (1997) provided sufficient conditions for pointwise asymptotic normality for plug-in nonparametric LS estimators of exact CS and DL functionals and of approximate CS functionals, respectively, when prices and incomes are exogenous. Vanhems (2010) studied consistency and convergence rates of kernel-based plug-in estimators of CS functionals allowing for mildly ill-posed NPIV models. Blundell, Horowitz, and Parey (2012) and Hausman and Newey (2016) estimated CS and DL of nonparametric gasoline demand allowing for prices to be endogenous, but did not provide theoretical justification for their inference approach under endogeneity. Therefore, although presented as applications of our sup-norm rate results, our inference results contribute nicely to the literature on nonparametric welfare analysis.

5.1 *Pointwise inference on exact CS and DL with endogeneity*

Here we present primitive regularity conditions for pointwise asymptotic normality of the sieve t-statistics for exact CS and DL. We suppress dependence of the functionals on $t = (\mathsf{p}^0, \mathsf{p}^1, \mathsf{y})$.

Let $\mathsf{X}_i = (\mathsf{P}_i, \mathsf{Y}_i)$. We assume in what follows that the support of both P_i and Y_i is bounded away from zero. If both P_i and Y_i are endogenous, let W_i be a 2×1 vector of instruments. Let $T : L^2(\mathsf{X}) \to L^2(\mathsf{W})$ be compact and injective with singular value decomposition (SVD) $\{\phi_{0j}, \phi_{1j}, \mu_j\}_{j=1}^\infty$, where

$$T\phi_{0j} = \mu_j \phi_{1j}, \qquad (T^*T)^{1/2}\phi_{0j} = \mu_j \phi_{0j}, \qquad (TT^*)^{1/2}\phi_{1j} = \mu_j \phi_{1j},$$

and $\{\phi_{0j}\}_{j=1}^\infty$ and $\{\phi_{0j}\}_{j=1}^\infty$ are orthonormal bases for $L^2(\mathsf{X})$ and $L^2(\mathsf{W})$, respectively. If P_i is endogenous but Y_i is exogenous, we take $\mathsf{W}_i = (\mathsf{W}_{1i}, \mathsf{Y}_i)'$ with W_{1i} an instrument for P_i. Let $T_\mathsf{y} : L^2(\mathsf{P}|\mathsf{Y} = \mathsf{y}) \to L^2(\mathsf{W}_1|\mathsf{Y} = \mathsf{y})$ be compact and injective with SVD $\{\phi_{0j,\mathsf{y}}, \phi_{1j,\mathsf{y}}, \mu_{j,\mathsf{y}}\}_{j=1}^\infty$ for each y, where

$$T_\mathsf{y}\phi_{0j,\mathsf{y}} = \mu_{j,\mathsf{y}}\phi_{1j,\mathsf{y}}, \qquad (T_\mathsf{y}^*T_\mathsf{y})^{1/2}\phi_{0j,\mathsf{y}} = \mu_{j,\mathsf{y}}\phi_{0j,\mathsf{y}}, \qquad (T_\mathsf{y}T_\mathsf{y}^*)^{1/2}\phi_{1j,\mathsf{y}} = \mu_{j,\mathsf{y}}\phi_{1j,\mathsf{y}},$$

and $\{\phi_{0j,\mathsf{y}}\}_{j=1}^\infty$ and $\{\phi_{0j,\mathsf{y}}\}_{j=1}^\infty$ are orthonormal bases for $L^2(\mathsf{P}|\mathsf{Y} = \mathsf{y})$ and $L^2(\mathsf{W}_1|\mathsf{Y} = \mathsf{y})$, respectively. In this case, we define $\phi_{0j}(\mathsf{p}, \mathsf{y}) = \phi_{0j,\mathsf{y}}(\mathsf{p})$, $\phi_{1j}(\mathsf{w}_1, \mathsf{y}) = \phi_{1j,\mathsf{y}}(\mathsf{w}_1)$, and $\mu_j^2 = E[\mu_{j,\mathsf{Y}_i}^2]$ (see Section 3.3 for further details).

In both cases, we follow Chen and Pouzo (2015) and assume that Ψ_J and B_K are Riesz bases in that they span $\phi_{01}, \ldots, \phi_{0J}$ and $\phi_{11}, \ldots, \phi_{1K}$, respectively. This implies that $\tau_J \asymp \mu_J^{-1}$. For fixed $\mathsf{p}^0, \mathsf{p}^1$, and y we define

$$a_j = a_j(\mathsf{p}^0, \mathsf{p}^1, \mathsf{y}) = \int_0^1 \big(\phi_{0j}\big(\mathsf{p}(u), \mathsf{y} - \mathsf{S}_\mathsf{y}\big(\mathsf{p}(u)\big)\big)e^{-\int_0^u \partial_2 h_0(\mathsf{p}(v), \mathsf{y} - \mathsf{S}_\mathsf{y}(\mathsf{p}(v)))\mathsf{p}'(v)\,\mathrm{d}v}\mathsf{p}'(u)\big)\,\mathrm{d}u$$

for the exact CS functional.

ASSUMPTION CS. (i) *The random vectors* X_i *and* W_i *both have compact rectangular support and densities bounded away from* 0 *and* ∞; (ii) $h_0 \in B_\infty(p, L)$ *with* $p > 2$ *and*

$0 < L < \infty$; (iii) $E[\mathsf{u}_i^2|\mathsf{W}_i = w]$ *is uniformly bounded away from* 0 *and* ∞, $E[|\mathsf{u}_i|^{2+\delta}]$ *is finite for some* $\delta > 0$, *and* $\sup_w E[u_i^2\{|u_i| > \ell(n)\}|\mathsf{W}_i = w] = o(1)$ *for any positive sequence with* $\ell(n) \nearrow \infty$; (iv) Ψ_J *is spanned by a* (*tensor-product*) *B-spline basis of order* $\gamma > p$ *or continuously differentiable wavelet basis of regularity* $\gamma > p$ *and* B_K *is spanned by a* (*tensor-product*) *B-spline, wavelet, or cosine basis;* (v) $J^{(2+\delta)/(2\delta)}\sqrt{(\log n)/n} = o(1)$ *and*

$$\frac{\sqrt{n}}{\left(\sum_{j=1}^{J}(a_j/\mu_j)^2\right)^{1/2}} \times \left(J^{-p/2} + \mu_J^{-2}\frac{J^2\sqrt{\log J}}{n}\right) = o(1).$$

Assumption CS(i)–(iv) are standard even for series LS regression without endogeneity. Let $[\sigma_n(f_{\mathrm{CS}})]^2 = (Df_{\mathrm{CS}}(h_0)[\psi^J])'\mho(Df_{\mathrm{CS}}(h_0)[\psi^J])$ be the sieve variance of the plug-in sieve NPIV estimator $f_{\mathrm{CS}}(\widehat{h}_0)$. Then these assumptions imply that $[\sigma_n(f_{\mathrm{CS}})]^2 \asymp \sum_{j=1}^{J}(a_j/\mu_j)^2 \lesssim J\mu_J^{-2}$. Assumption CS(v) is sufficient for Remark 4.1(b′) for a fixed t.

Our first result is for exact CS functionals, established by applying Theorem D.1 in Appendix D. Let

$$\widehat{\sigma}^2(f_{\mathrm{CS}}) = Df_{\mathrm{CS}}(\widehat{h})[\psi^J]'\widehat{\mho}Df_{\mathrm{CS}}(\widehat{h})[\psi^J]$$

with

$$Df_{\mathrm{CS}}(\widehat{h})[\psi^J] = \int_0^1 \psi^J\big(\mathsf{p}(u), \mathsf{y} - \widehat{\mathsf{S}}_\mathsf{y}(\mathsf{p}(u))\big)e^{-\int_0^u \partial_2\widehat{h}(\mathsf{p}(v), \mathsf{y} - \widehat{\mathsf{S}}_\mathsf{y}(\mathsf{p}(v)))\mathsf{p}'(v)\,\mathrm{d}v}\mathsf{p}'(u)\,\mathrm{d}u.$$

THEOREM 5.1. *Let Assumption CS hold. Then the sieve* t-*statistic for* $f_{\mathrm{CS}}(h_0)$ *is asymptotically* $N(0,1)$, *that is,*

$$\sqrt{n}\frac{f_{\mathrm{CS}}(\widehat{h}) - f_{\mathrm{CS}}(h_0)}{\widehat{\sigma}(f_{\mathrm{CS}})} \to_d N(0,1).$$

Since $\mu_j > 0$ decreases as j increases, we could use the relation

$$\mu_J^{-2}J \gtrsim \mu_J^{-2}\sum_{j=1}^{J}a_j^2 \geq \sum_{j=1}^{J}(a_j/\mu_j)^2 \tag{21}$$

$$\geq \max\left(\left(\min_{1\leq j\leq J}a_j^2\right)\sum_{j=1}^{J}\mu_j^{-2}, \max_{1\leq j\leq J}(a_j^2\mu_j^{-2}), \mu_1^{-2}\sum_{j=1}^{J}a_j^2\right)$$

to provide simpler sufficient conditions for Assumption CS(v) that could be satisfied by both mildly and severely ill-posed NPIV models. Corollary 5.1 provides one set of concrete sufficient conditions for Assumption CS(v).

COROLLARY 5.1. *Let Assumption* CS(i)–(iv) *hold and let* $a_j^2 \asymp j^a$ *for* $a \leq 0$. *Then* $[\sigma_n(f_{\mathrm{CS}})]^2 \asymp \sum_{j=1}^{J}(j^a\mu_j^{-2})$.

(i) *Mildly ill-posed case Let $\mu_j \asymp j^{-\varsigma/2}$ for $\varsigma \geq 0$, $a + \varsigma > -1$. Then*

$$\left[\sigma_n(f_{\mathrm{CS}})\right]^2 \asymp J^{(a+\varsigma)+1}.$$

Further, if $\delta \geq 2/(2 + \varsigma - a)$, $nJ^{-(p+a+\varsigma+1)} = o(1)$, and $J^{3+\varsigma-a}(\log n)/n = o(1)$, then Assumption CS(v) *is satisfied, and the sieve t-statistic for $f_{\mathrm{CS}}(h_0)$ is asymptotically $N(0, 1)$.*

(ii) *Severely ill-posed case. Let $\mu_j \asymp \exp(-\frac{1}{2}j^{\varsigma/2})$, $\varsigma > 0$ and $J = (\log(n/(\log n)^\varrho))^{2/\varsigma}$ for $\varrho > 0$. Then*

$$\left[\sigma_n(f_{\mathrm{CS}})\right]^2 \gtrsim \frac{n}{(\log n)^\varrho} \times \left(\log(n/(\log n)^\varrho)\right)^{2a/\varsigma}.$$

Further, if $\varrho > 0$ is chosen such that $2p > \varrho\varsigma - 2a$ and $\varrho\varsigma > 8 - 2a$, then Assumption CS(v) *is satisfied and the sieve t-statistic for $f_{\mathrm{CS}}(h_0)$ is asymptotically $N(0, 1)$.*

Note that in Corollary 5.1, J may be chosen to satisfy the stated conditions in the mildly ill-posed case whenever $p > 2 - 2a$ and in the severely ill-posed case whenever $p > 4 - 2a$.

Our next result is for DL functionals. Note that DL is the sum of CS and a tax receipts functional, namely $(\mathsf{p}^1 - \mathsf{p}^0)h_0(\mathsf{p}^1, \mathsf{y})$. Note that the tax receipts functional is typically less smooth and hence converges slower than the CS functional. Therefore, $[\sigma_n(f_{\mathrm{DL}})]^2 = (Df_{\mathrm{DL}}(h_0)[\psi^J])'\mho(Df_{\mathrm{DL}}(h_0)[\psi^J])$ will typically grow on the order of $(\tau_J\sqrt{J})^2$, which is the growth order of the sieve variance term for estimating the unknown NPIV function h_0 at a fixed point. For this reason we do not derive the joint asymptotic distribution of $f_{\mathrm{CS}}(\widehat{h})$ and $f_{\mathrm{DL}}(\widehat{h})$. The next result adapts Theorem 5.1 to derive asymptotic normality of plug-in sieve t-statistics for DL functionals. Let

$$\widehat{\sigma}^2(f_{\mathrm{DL}}) = Df_{\mathrm{DL}}(\widehat{h})[\psi^J]'\widehat{\mho}Df_{\mathrm{DL}}(\widehat{h})[\psi^J]$$

with

$$Df_{\mathrm{DL}}(\widehat{h})[\psi^J] = Df_{\mathrm{CS}}(\widehat{h})[\psi^J] - (\mathsf{p}^1 - \mathsf{p}^0)\psi^J(\mathsf{p}^1, \mathsf{y}).$$

THEOREM 5.2. *Let Assumption* CS(i)–(iv) *hold. Let $\sigma_n(f_{\mathrm{DL}}) \asymp \mu_J^{-1}\sqrt{J}$, $\sqrt{n}\mu_J J^{-(p+1)/2} = o(1)$ and $(J^{(2+\delta)/(2\delta)}\sqrt{\log n} \vee \mu_J^{-1}J^{3/2}\sqrt{\log J})/\sqrt{n} = o(1)$. Then*

$$\sqrt{n}\frac{f_{\mathrm{DL}}(\widehat{h}) - f_{\mathrm{DL}}(h_0)}{\widehat{\sigma}(f_{\mathrm{DL}})} \to_d N(0, 1).$$

5.2 Pointwise inference on approximate CS with endogeneity

Suppose instead that demand of consumer i for some good is estimated in logs, that is,

$$\log \mathsf{Q}_i = h_0(\log \mathsf{P}_i, \log \mathsf{Y}_i) + u_i. \tag{22}$$

As h_0 is the log-demand function, any linear functional of demand is a nonlinear functional of h_0. One such example is the weighted average demand functional of the form

$$f_A(h) = \int w(\mathsf{p})e^{h(\log \mathsf{p}, \log \mathsf{y})}\, d\mathsf{p},$$

where $w(\mathsf{p})$ is a nonnegative weighting function and y is fixed. With $w(\mathsf{p}) = \mathbb{1}\{\underline{\mathsf{p}} \leq \mathsf{p} \leq \overline{\mathsf{p}}\}$, the functional $f(h)$ may be interpreted as the approximate CS. The functional is defined for fixed y, so it will typically be an irregular functional of h_0.

The setup is similar to the previous subsection. Let $\mathsf{X}_i = (\log \mathsf{P}_i, \log \mathsf{Y}_i)$. If both P_i and Y_i are endogenous, we let W_i be a 2×1 vector of instruments and let $T : L^2(\mathsf{X}) \to L^2(\mathsf{W})$ be compact with SVD $\{\phi_{0j}, \phi_{1j}, \mu_j\}_{j=1}^{\infty}$. If P_i is endogenous but Y_i is exogenous, we let $\mathsf{W}_i = (\mathsf{W}_{1i}, \log \mathsf{Y}_i)'$ with W_{1i} an instrument for P_i, and let $T_{\mathsf{y}} : L^2(\log \mathsf{P} | \log \mathsf{Y} = \log \mathsf{y}) \to L^2(\mathsf{W}_1 | \log \mathsf{Y} = \log \mathsf{y})$ be compact with SVD $\{\phi_{0j,\mathsf{y}}, \phi_{1j,\mathsf{y}}, \mu_{j,\mathsf{y}}\}_{j=1}^{\infty}$ for each y. In this case, we define $\phi_{0j}(\log \mathsf{p}, \log \mathsf{y}) = \phi_{0j,\mathsf{y}}(\log \mathsf{p})$, $\phi_{1j}(\mathsf{w}_1, \log \mathsf{y}) = \phi_{1j,\mathsf{y}}(\mathsf{w}_1)$, and $\mu_j^2 = E[\mu_{j,\mathsf{Y}_i}^2]$. We again assume that Ψ_J and B_K are Riesz bases. For each $j \geq 1$, define

$$a_j = a_j(\mathsf{y}) = \int w(\mathsf{p}) e^{h_0(\log \mathsf{p}, \log \mathsf{y})} \phi_{0j}(\log \mathsf{p}, \log \mathsf{y}) \, d\mathsf{p}.$$

The next result follows from Theorem D.1 (in Appendix D). Let

$$\widehat{\sigma}^2(f_A) = Df_A(\widehat{h})[\psi^J]' \widehat{\mho} Df_A(\widehat{h})[\psi^J]$$

with

$$Df_A(\widehat{h})[\psi^J] = \int w(\mathsf{p}) e^{\widehat{h}(\log \mathsf{p}, \log \mathsf{y})} \psi^J(\log \mathsf{p}, \log \mathsf{y}) \, d\mathsf{p}.$$

THEOREM 5.3. *Let Assumption* CS(i)–(iv) *hold for the log-demand model* (22) *with* $p > 0$, *and let* $J^{(2+\delta)/(2\delta)} \sqrt{(\log n)/n} = o(1)$ *and*

$$\frac{\sqrt{n}}{\left(\sum_{j=1}^{J} (a_j/\mu_j)^2\right)^{1/2}} \times \left(J^{-p/2} + \mu_J^{-2} \frac{J^{3/2} \sqrt{\log J}}{n}\right) = o(1).$$

Then

$$\frac{\sqrt{n}(f_A(\widehat{h}) - f_A(h_0))}{\widehat{\sigma}(f_A)} \to_d N(0, 1).$$

5.3 *Uniform inference on collections of exact CS and DL functionals with endogeneity*

Here we apply Lemma 4.1 and Theorem 4.1 to present sufficient conditions for uniform Gaussian process strong approximations and bootstrap UCBs for exact CS and DL under endogeneity. We maintain the setup described at the beginning of Section 5.1. We take $t = (\mathsf{p}^0, \mathsf{p}^1, \mathsf{y}) \in \mathcal{T} = [\underline{\mathsf{p}}^0, \overline{\mathsf{p}}^0] \times [\underline{\mathsf{p}}^1, \overline{\mathsf{p}}^1] \times [\underline{\mathsf{y}}, \overline{\mathsf{y}}]$, where the intervals $[\underline{\mathsf{p}}^0, \overline{\mathsf{p}}^0]$ and $[\underline{\mathsf{p}}^1, \overline{\mathsf{p}}^1]$ are in the interior of the support of P_i and $[\underline{\mathsf{y}}, \overline{\mathsf{y}}]$ is in the interior of the support of Y_i. For each $t \in \mathcal{T}$ we let

$$
\begin{aligned}
a_{j,t} &= a_{j,t}(\mathsf{p}^0, \mathsf{p}^1, \mathsf{y}) \\
&= \int_0^1 \left(\phi_{0j}(\mathsf{p}(u), \mathsf{y} - \mathsf{S}_{\mathsf{y}}(\mathsf{p}(u))) e^{-\int_0^u \partial_2 h_0(\mathsf{p}(v), \mathsf{y} - \mathsf{S}_{\mathsf{y}}(\mathsf{p}(v))) \mathsf{p}'(v) \, dv} \mathsf{p}'(u)\right) du
\end{aligned}
\tag{23}
$$

for each $j \geq 1$ (where $\mathsf{p}(u)$ is a smooth price path from $\mathsf{p}^0 = \mathsf{p}(0)$ to $\mathsf{p}^1 = \mathsf{p}(1)$). Also define $\underline{\sigma}_n = \inf_{t \in \mathcal{T}} ((\sum_{j=1}^{J} (a_{j,t}/\mu_j)^2)^{1/2}$.

ASSUMPTION U-CS. (i) *The function $E[\mathsf{u}_i^2 | W_i = w]$ is uniformly bounded away from 0, $E[|\mathsf{u}_i|^{2+\delta}]$ is finite with $\delta \geq 1$, and $\sup_w E[|\mathsf{u}_i|^3 | W_i = w]$ is finite;* (ii) *the Hölder condition in Remark 4.2 holds with $\gamma_n = \gamma$ and $\Gamma_n \lesssim J^c$ for some finite positive constants γ and c;* (iii) *$J^5(\log n)^3/n = o(1)$, $\frac{\sqrt{n(\log J)}}{\underline{\sigma}_n} J^{-p/2} = o(1)$;* (iv) *letting $\eta_n' = \frac{J^{3/2}\mu_J^{-1}}{\underline{\sigma}_n}(J^{-p/2} + \mu_J^{-1}\sqrt{J(\log J)/n})$, either* (iv)(a) *$\eta_n'(\log J) = o(1)$ or* (iv)(b) *$\eta_n'\sqrt{J(\log J)} = o(1)$.*

Assumption U-CS(i) is slightly stronger than Assumption CS(iii) (since $\delta = 1$ in Assumption U-CS(i) is enough). Assumption U-CS(ii) is made for simplicity to verify Assumption 6(i); other sufficient conditions could also be used. Assumption U-CS(iii) and (iv)(a) strengthen Assumption CS(v) to ensure uniform Gaussian process strong approximation with an error rate of $r_n = (\log J)^{-1/2}$. Again, one could use bounds on $\underline{\sigma}_n$ that are analogous to relation (21) to provide sufficient conditions for Assumption U-CS(iii) and (iv) that could be satisfied by mildly and severely ill-posed NPIV models. See Remark 5.1 below for one concrete set of such sufficient conditions.

REMARK 5.1. Let $\underline{\sigma}_n^2 \gtrsim \sum_{j=1}^{J}(j^a \mu_j^{-2})$ for $a \leq 0$.

(i) Mildly ill-posed case. Let $\mu_j \asymp j^{-\varsigma/2}$ for $\varsigma \geq 0$ and $a + \varsigma > -1$. Let $J^{5\vee(4+\varsigma-a)}(\log n)^3/n = o(1)$ and $nJ^{-(p+a+\varsigma+1)}(\log J) = o(1)$. Then Assumption U-CS(iii) and (iv) hold.

(ii) Severely ill-posed case. Let $\mu_J \asymp \exp(-\frac{1}{2}j^{\varsigma/2})$, $\varsigma > 0$. Let $J = (\log(n/(\log n)^\varrho))^{2/\varsigma}$ with $\varrho > 0$ chosen such that $2p > \varrho\varsigma - 2a$ and $\varrho\varsigma > 10 - 2a$. Then Assumption U-CS(iii) and (iv) hold.

The next results are about the uniform Gaussian process strong approximation and validity of score bootstrap UCBs for exact CS and DL functionals.

THEOREM 5.4. *Let Assumptions CS(i), (ii), and (iv) and U-CS(i), (ii), and (iii) hold.*

(i) *If Assumption U-CS(iv)(a) holds, then result (19) (with $r_n = (\log J)^{-1/2}$) holds for $f_t = f_{\mathrm{CS},t}$.*

(ii) *If Assumption U-CS(iv)(b) holds, then result (20) also holds for $f_t = f_{\mathrm{CS},t}$.*

In the next theorem the condition $\underline{\sigma}_n \asymp \mu_J^{-1}\sqrt{J}$ is implied by the assumption that $\sigma_n(f_{\mathrm{DL},t}) \asymp \mu_J^{-1}\sqrt{J}$ uniformly for $t \in \mathcal{T}$, which is reasonable for the DL functional.

THEOREM 5.5. *Let Assumptions CS(i), (ii), and (iv) and U-CS(i), (ii), and (iii) hold with $\underline{\sigma}_n \asymp \mu_J^{-1}\sqrt{J}$.*

(i) *If Assumption U-CS(iv)(a) holds, then result (19) (with $r_n = (\log J)^{-1/2}$) holds for $f_t = f_{\mathrm{DL},t}$.*

(ii) *If Assumption U-CS(iv)(b) holds, then result (20) also holds for $f_t = f_{\mathrm{DL},t}$.*

5.4 *Inference on welfare functionals without endogeneity*

This subsection specializes the pointwise and uniform inference results for welfare functionals from the preceding subsections to nonparametric demand estimation with exogenous price and income. Precisely, we let $X_i = W_i$, $J = K$, $b^K = \psi^J$, and $\mu_J \asymp 1$, $\tau_J \asymp 1$, and so the sieve NPIV estimator reduces to the usual series LS estimator of $h_0(x) = E[Y_i | W_i = x]$.

The next two corollaries are direct consequences of our Theorems 5.1, 5.2, and 5.3 for pointwise asymptotic normality of sieve t-statistics for exact CS and DL and approximate CS functionals under exogeneity; hence, the proofs are omitted.

COROLLARY 5.2. *Let Assumption* CS(i)–(iv) *hold with* $X_i = W_i$, $J = K$, $b^K = \psi^J$, *and* $\mu_J \asymp 1$, *and let* $\sum_{j=1}^J a_j^2 \gtrsim J^{a+1}$ *with* $0 \geq a \geq -1$.

(i) *Let* $nJ^{-(p+a+1)} = o(1)$, $J^{3-a}(\log J)/n = o(1)$, *and* $\delta \geq 2/(2-a)$. *Then the sieve t-statistic for* $f_{\mathrm{CS}}(h_0)$ *is asymptotically* $N(0,1)$.

(ii) *Let* $nJ^{-(p+1)} = o(1)$, $J^3(\log J)/n = o(1)$, *and* $a = 0$, $\delta \geq 1$. *Then the sieve t-statistic for* $f_{\mathrm{DL}}(h_0)$ *is asymptotically* $N(0,1)$.

Previously Hausman and Newey (1995) established the pointwise asymptotic normality of t-statistics for exact CS and DL based on plug-in kernel LS estimators of demand without endogeneity. They also established root-n asymptotic normality of t-statistics for *averaged* exact CS and DL (i.e., CS/DL averaged over a range of incomes) based on plug-in power series LS estimator of demand without endogeneity, under some regularity conditions including that $\sup_x E[|u_i|^4 | X_i = x] < \infty$ (which, in our notation, implies $\delta = 2$), $p = \infty$ (i.e., h_0 is infinitely times differentia,ble) and $J^{22}/n = o(1)$. Corollary 5.2 complements their work by providing conditions for the pointwise asymptotic normality of exact CS and DL functionals based on spline and wavelet LS estimators of demand.

COROLLARY 5.3. *Let Assumption* CS(i)–(iv) *hold for the log-demand model* (22) *with* $X_i = W_i$, $J = K$, $b^K = \psi^J$, $\mu_J \asymp 1$, *and* $p > 0$, *and let* $\sum_{j=1}^J a_j^2 \gtrsim J^{c+1}$ *with* $0 \geq c \geq -1$. *Let* $nJ^{-(p+c+1)} = o(1)$, $J^{2-c}(\log J)/n = o(1)$, *and* $\delta \geq 2/(1-c)$. *Then the sieve t-statistic for* $f_A(h_0)$ *is asymptotically* $N(0,1)$.

Previously Newey (1997) established the pointwise asymptotic normality of t-statistics for *approximate* CS functionals based on plug-in series LS estimators of exogenous demand under some regularity conditions including that $\sup_x E[|u_i|^4 | X_i = x] < \infty$ (which implies $\delta = 2$), $nJ^{-p} = o(1)$, and either $J^6/n = o(1)$ for power series or $J^4/n = o(1)$ for splines.

The final corollary is a direct consequence of our Theorems 5.4 and 5.5 and Remark 5.1 for uniform inferences based on sieve t processes for exact CS and DL nonlinear functionals under exogeneity; hence, its proof is omitted.

Corollary 5.4. *Let Assumptions* CS(i), (ii), *and* (iv) *and* U-CS(i) *and* (ii) *hold with* $\mathsf{X}_i = \mathsf{W}_i$, $J = K$, $b^K = \psi^J$, *and* $\mu_J \asymp 1$. *Let* $\underline{\sigma}_n^2 \gtrsim J^{a+1}$ *with* $0 \geq a \geq -1$. *Let* $J^5(\log n)^3/n = o(1)$ *and* $nJ^{-(p+a+1)}(\log J) = o(1)$. *Then results* (19) (*with* $r_n = (\log J)^{-1/2}$) *and* (20) *hold for* $f_t = f_{\mathrm{CS},t}, f_{\mathrm{DL},t}$.

We note that $\underline{\sigma}_n^2 \asymp J$ (or $a = 0$) for $f_t = f_{\mathrm{DL},t}$. Corollary 5.4 appears to be a new addition to the existing literature. The sufficient conditions for uniform inference for collections of nonlinear exact CS and DL functionals of nonparametric demand estimation under exogeneity are mild and simple.

6. Conclusion

This paper makes several important contributions to inference on nonparametric models with endogeneity. We derive the minimax sup-norm convergence rates for estimating the structural NPIV function h_0 and its derivatives. We also provide upper bounds for sup-norm convergence rates of computationally simple sieve NPIV (series 2SLS) estimators using any sieve basis to approximate unknown h_0, and show that the sieve NPIV estimator using a spline or wavelet basis can attain the minimax sup-norm rates. These rate results are particularly useful for establishing the validity of pointwise and uniform inference procedures for nonlinear functionals of h_0. In particular, we use our sup-norm rates to establish the uniform Gaussian process strong approximation and the validity of score bootstrap-based UCBs for collections of nonlinear functionals of h_0 under primitive conditions, allowing for mildly and severely ill-posed problems. We illustrate the usefulness of our UCBs procedure with two real data applications to nonparametric demand analysis with endogeneity. We establish the pointwise and uniform limit theories for sieve t-statistics for exact (and approximate) CS and DL nonlinear functionals under low-level conditions when the demand function is estimated via sieve NPIV. Our theoretical and empirical results for CS and DL are new additions to the literature on nonparametric welfare analysis.

We conclude the paper by mentioning some further extensions and applications of sup-norm convergence rates of sieve NPIV estimators.

Extensions to semiparametric IV models

Although our rate results are presented for purely nonparametric IV models, the results may be adapted easily to some semiparametric models with nonparametric endogeneity, such as partially linear IV regression (Ai and Chen (2003), Florens, Johannes, and Van Bellegem (2012)), shape-invariant Engel curve IV regression (Blundell, Chen, and Kristensen (2007)), and single index IV regression (Chen et al. (2014)), to list a few. For example, consider the partially linear NPIV model

$$Y_i = X'_{1i}\beta_0 + h_0(X_{2i}) + u_i, \qquad E[u_i|W_{1i}, W_{2i}] = 0,$$

where X_{1i} and X_{2i} are of dimensions d_1 and d_2 and do not contain elements in common, and $W_i = (W_{1i}, W_{2i})$ is the (conditional) IV. See Florens, Johannes, and Van Bellegem (2012), Chen et al. (2014) for identification of (β_0, h_0) in this model. We can still

estimate (β_0, h_0) via sieve NPIV or series 2SLS as before, replacing Ψ and B in equations (3 and (4) by

$$\psi^J(x) = \left(x_1', \psi_2^J(x_2)'\right)', \qquad \psi_2^J(x) = \left(\psi_{J1}(x_2), \ldots, \psi_{JJ}(x_2)\right)',$$

$$b^K(w) = \left(w_1', b_2^K(w_2)'\right)', \qquad b_2^K(w) = \left(b_{K1}(w_2), \ldots, b_{KK}(w_2)\right)',$$

where $x = (x_1', x_2')'$, $w = (w_1', w_2')'$, $\psi_{J1}, \ldots, \psi_{JJ}$ denotes a sieve of dimension J for approximating $h_0(x_2)$, and $b_{K1}, \ldots, b_{KK}$ denotes a sieve of dimension K for the instrument space for W_2. We then partition $\widehat{c}$ in (2) into $\widehat{c} = (\widehat{\beta}', \widetilde{c}_2')'$ and set $\widehat{h}(x) = \psi_2^J(x)'\widehat{c}_2$. Note that $\widehat{\beta}$ is root-n consistent and asymptotically normal for β_0 under mild conditions (see Ai and Chen (2003), Chen and Pouzo (2009)), and, hence, would not affect the optimal convergence rate of $\widehat{h}$ to h_0. Our rate results may be slightly altered to derive sup-norm convergence rates for $\widehat{h}$ and its derivatives.

Nonparametric specification testing in NPIV models

Structural models may specify a parametric form $m_{\theta_0}(x)$, where $\theta_0 \in \Theta \subseteq \mathbb{R}^{d_\theta}$ for the unknown structural function $h_0(x)$ in NPIV model (1). We may be interested in testing the parametric model $\{m_\theta : \theta \in \Theta\}$ against a nonparametric alternative that only assumes some smoothness on h_0. Specification tests for nonparametric regression without endogeneity have typically been performed via either a quadratic-form-based statistic or a Kolmogorov–Smirnov (KS) type sup statistic.[18] However, specification tests for NPIV models have so far only been performed via quadratic-form-based statistics; see, for example, Horowitz (2006, 2011, 2012), Blundell and Horowitz (2007), Breunig (2015). Equipped with our sup-norm rate and UCBs results for the NPIV function and its derivatives, one could also perform specification tests in NPIV models using KS type statistics of the form

$$T_n = \sup_x \frac{\left|\widehat{h}(x) - \widehat{m}(x, \widehat{\theta})\right|}{s_n(x)},$$

where $\widehat{\theta}$ is a first-stage estimator of θ_0, and $\widehat{m}(x, \widehat{\theta})$ is obtained from series 2SLS regression of $m(X_1, \widehat{\theta}), \ldots, m(X_n, \widehat{\theta})$ on the same basis functions as in $\widehat{h}$, and $s_n(x)$ is a normalization factor. Alternatively, one could consider a KS statistic formed in terms of the projection of $[\widehat{h}(x) - \widehat{m}(x, \widehat{\theta})]$ onto the instrument space. Sup-norm convergence rates and uniform limit theory derived in this paper would be useful in deriving the large-sample distribution of these KS type statistics. Further, based on our rate results (in sup- and L^2-norm) for estimating derivatives of h_0 in a NPIV model, one could also perform nonparametric tests of significance by testing whether partial derivatives of the NPIV function h_0 are identically zero, via KS or quadratic-form-based test statistics.

If one is interested in specifications or inferences on functionals directly, then one might consider KS type sup statistics for (possibly nonlinear) functionals directly. For

---

[18]See, for example, Bierens (1982), Hardle and Mammen (1993), Hong and White (1995), Fan and Li (1996), Lavergne and Vuong (1996), Stinchcombe and White (1998), and Horowitz and Spokoiny (2001) to list a few.

example, if one is interested in the exact CS functional of a demand and concerns about the potential endogeneity of price, then one could estimate the exact CS functional using a series LS estimated demand (under exogeneity) and series 2SLS estimated demand (under endogeneity), and then compare the two estimated exact CS functionals via a KS type or a quadratic-form-based test. In fact, the score bootstrap-based UCBs reported in Figure 2 indicate that such a test based on the exact CS functional directly could be quite informative.

Semiparametric two-step procedures with NPIV first stage

Many semiparametric two-step or multi-step estimation and inference procedures involve a nonparametric first stage. There are many theoretical results when the first stage is a purely nonparametric LS regression (without endogeneity) and its sup-norm convergence rate is used to assist subsequent analysis. For structural estimation and inference, it is natural to allow for the presence of nonparametric endogeneity in the first stage as well. For instance, if there is endogeneity present in the conditional moment inequality application of the famous intersection bound paper of Chernozhukov, Lee, and Rosen (2013), one could simply use our sup-norm rate and UCBs results for sieve NPIV instead of their series LS regression in the first stage. As another example, consider semiparametric two-step generalized method of moments (GMM) models $E[g(Z_i, \theta_0, h_0(X_i))] = 0$, where h_0 is the NPIV function in model (1), g is a $\mathbb{R}^{d_g}$-valued vector of moment functions with $d_g \geq d_\theta$, and $\theta_0 \in \mathbb{R}^{d_\theta}$ is a finite-dimensional parameter of interest, such as the average exact CS parameter of a nonparametric demand function with endogeneity. A popular estimator $\widehat{\theta}$ of θ_0 is a solution to the semiparametric two-step GMM with a weighting matrix $\widehat{W}$,

$$\min_{\theta} \left(\frac{1}{n} \sum_{i=1}^{n} g(Z_i, \theta, \widehat{h}(X_i)) \right)' \widehat{W} \left(\frac{1}{n} \sum_{i=1}^{n} g(Z_i, \theta, \widehat{h}(X_i)) \right),$$

where $\widehat{h}$ is a sieve NPIV estimator of h_0. When h_0 enters the moment function $g(\cdot)$ nonlinearly, sup-norm convergence rates of $\widehat{h}$ to h_0 are useful in deriving the asymptotic properties of $\widehat{\theta}$.

APPENDIX A: ADDITIONAL LEMMAS FOR SUP-NORM RATES

Let $s_{\min}(A)$ denote the minimum singular value of a rectangular matrix A. For a positive-definite symmetric matrix A, we let $A^{1/2}$ be its positive definite square root. We define $s_{JK} = s_{\min}(G_b^{-1/2} S G_\psi^{-1/2})$, which satisfies

$$s_{JK}^{-1} = \sup_{h \in \Psi_J : h \neq 0} \frac{\|h\|_{L^2(X)}}{\|\Pi_K T h\|_{L^2(W)}} \geq \tau_J$$

for all $K \geq J > 0$. The following lemma is used throughout the paper.

LEMMA A.1. *Let Assumptions* 1(iii) *and* 4(i) *hold. Then* $(1 - o(1))s_{JK}^{-1} \leq \tau_J \leq s_{JK}^{-1}$ *as* $J \to \infty$.

Before we provide a bound on the sup-norm bias term, we present some sufficient conditions for Assumption 4(iii). This involves three projections of h_0 onto the sieve approximating space Ψ_J. These projections imply different, but closely related, approximation biases for h_0. Recall that $\Pi_J : L^2(X) \to \Psi_J$ is the $L^2(X)$ orthogonal (i.e., least squares) projection onto Ψ_J, namely $\Pi_J h_0 = \arg\min_{h \in \Psi_J} \|h_0 - h\|_{L^2(X)}$, and $Q_J h_0 = \arg\min_{h \in \Psi_J} \|\Pi_K T(h_0 - h)\|_{L^2(W)}$ is the sieve 2SLS projection of h_0 onto Ψ_J. Let $\pi_J h_0 = \arg\min_{h \in \Psi_J} \|T(h_0 - h)\|_{L^2(W)}$ denote the IV projection of h_0 onto Ψ_J. Note that each of these projections is nonrandom.

Instead of Assumption 4(iii), we could impose the following version.

ASSUMPTION 4. (iii′) *We have* $(\zeta_{\psi,J}\tau_J) \times \|(\Pi_K T - T)(Q_J h_0 - \pi_J h_0)\|_{L^2(W)} \leq \text{const} \times \|Q_J h_0 - \pi_J h_0\|_{L^2(X)}$.

Assumption 4(iii′) seems mild and is automatically satisfied by Riesz basis. This is because $\|(\Pi_K T - T)h\|_{L^2(W)} = 0$ for all $h \in \Psi_J$ when the basis functions for B_K and Ψ_J form either a Riesz basis or an eigenfunction basis for the conditional expectation operator. The following lemma collects some useful facts about the approximation properties of $\pi_J h_0$.

LEMMA A.2. *Let Assumptions* 1(iii) *and* 4(ii) *hold.*

(i) *Then we have* $\|h_0 - \pi_J h_0\|_{L^2(X)} \asymp \|h_0 - \Pi_J h_0\|_{L^2(X)}$.

(ii) *If Assumption* 4(i) *also holds, then* $\|Q_J h_0 - \pi_J h_0\|_{L^2(X)} \leq o(1) \times \|h_0 - \pi_J h_0\|_{L^2(X)}$.

(iii) *Further, if Assumption* 4(iii′) *and*

$$\|\Pi_J h_0 - \pi_J h_0\|_\infty \leq \text{const} \times \|h_0 - \Pi_J h_0\|_\infty \tag{24}$$

hold, then Assumption 4(iii) *is satisfied.*

In light of Lemma A.2 parts (i) and (ii), condition (24) seems mild. In fact, condition (24) is trivially satisfied when the basis for Ψ_J is a Riesz basis because then $\pi_J h_0 = \Pi_J h_0$ (see Section 6 in Chen and Pouzo (2015)). See Lemma G.1 in Appendix G for more detailed relations among $\Pi_J h_0$, $\pi_J h_0$, and $Q_J h_0$.

The next lemma provides a bound on the sup-norm bias term.

LEMMA A.3. *Let Assumptions* 1(iii), 3(ii), *and* 4 *hold. Then*

(i) $\|\widetilde{h} - \Pi_J h_0\|_\infty \leq O_p(1) \times \|h_0 - \Pi_J h_0\|_\infty$

(ii) $\|\widetilde{h} - h_0\|_\infty \leq O_p(1 + \|\Pi_J\|_\infty) \times \|h_0 - h_{0,J}\|_\infty$.

APPENDIX B: OPTIMAL L^2-NORM RATES FOR DERIVATIVES

Here we show that the sieve NPIV estimator can attain the optimal L^2-norm convergence rates for estimating h_0 and its derivatives under much weaker conditions. The

optimal L^2-norm rates for sieve NPIV derivative estimation presented in this section are new, and should be very useful for inference on some nonlinear functionals involving derivatives such as $f(h) = \|\partial^\alpha h\|^2_{L^2(X)}$.

Instead of Assumption 1(iii), we impose the following condition for identification in $(\mathcal{H}, \|\cdot\|_{L^2(X)})$.

ASSUMPTION 1. (iii′) *We have $h_0 \in \mathcal{H} \subset L^2(X)$, and $T[h - h_0] = 0 \in L^2(W)$ for any $h \in \mathcal{H}$ implies that $\|h - h_0\|_{L^2(X)} = 0$.*

THEOREM B.1. *Let Assumptions 1(iii′) and 4(i) and (ii) hold, and let $\tau_J \zeta \sqrt{(\log J)/n} = o(1)$.*

(i) *Then we have $\|\widetilde{h} - h_0\|_{L^2(X)} \leq O_p(1) \times \|h_0 - \Pi_J h_0\|_{L^2(X)}$.*

(ii) *Further, if Assumption 2(i) holds, then*

$$\|\widehat{h} - h_0\|_{L^2(X)} = O_p\big(\|h_0 - \Pi_J h_0\|_{L^2(X)} + \tau_J \sqrt{J/n}\big).$$

The following corollary provides concrete L^2-norm convergence rates of $\widehat{h}$ and its derivatives. Let $B^p_{2,2}$ denote the Sobolev space of smoothness $p > 0$, let $\|\cdot\|_{B^p_{2,2}}$ denote a Sobolev norm of smoothness p, and let $B_2(p, L) = \{h \in B^p_{2,2} : \|h\|_{B^p_{2,2}} \leq L\}$, where radius $0 < L < \infty$ (Triebel (2006, Section 1.11)).

COROLLARY B.1. *Let Assumptions 1(i), (ii), and (iii′) and 4(i) and (ii) hold. Let $h_0 \in B_2(p, L)$, let Ψ_J be spanned by a cosine basis, B-spline basis of order $\gamma > p$, or CDV wavelet basis of regularity $\gamma > p$, and let B_K be spanned by a cosine, spline, or wavelet basis. Let $\tau_J \sqrt{(J \log J)/n} = o(1)$ hold.*

(i) *Then $\|\partial^\alpha \widetilde{h} - \partial^\alpha h_0\|_{L^2(X)} = O_p(J^{-(p-|\alpha|)/d})$ for all $0 \leq |\alpha| < p$.*

(ii) *Further if Assumption 2(i) holds, then*

$$\big\|\partial^\alpha \widehat{h} - \partial^\alpha h_0\big\|_{L^2(X)} = O_p\big(J^{-(p-|\alpha|)/d} + \tau_J J^{|\alpha|/d}\sqrt{J/n}\big) \quad \text{for all } 0 \leq |\alpha| < p.$$

(ii)(a) *Mildly ill-posed case. Choosing $J \asymp n^{d/(2(p+\varsigma)+d)}$ yields $\tau_J \sqrt{(J \log J)/n} = o(1)$ and*

$$\big\|\partial^\alpha \widehat{h} - \partial^\alpha h_0\big\|_{L^2(X)} = O_p\big(n^{-(p-|\alpha|)/(2(p+\varsigma)+d)}\big).$$

(ii)(b) *Severely ill-posed case. Choosing $J = (c_0 \log n)^{d/\varsigma}$ for any $c_0 \in (0, 1)$ yields $\tau_J \sqrt{(J \log J)/n} = o(1)$ and*

$$\big\|\partial^\alpha \widehat{h} - \partial^\alpha h_0\big\|_{L^2(X)} = O_p\big((\log n)^{-(p-|\alpha|)/\varsigma}\big).$$

The conclusions of Corollary B.1 remain true for any basis B_K under the condition $\tau_J \zeta_b \sqrt{(\log J)/n} = o(1)$. Previously, assuming some rates on estimating the unknown operator T, Johannes, van Bellegem, and Vanhems (2011) obtained similar L^2-norm rates for derivatives of iteratively Tikhonov-regularized estimators in a NPIV model with scalar regressor X_i and scalar instrument W_i.

Our next theorem shows that the rates obtained in Corollary B.1 are optimal. It extends the earlier work by Chen and Reiss (2011) on L^2-norm lower bounds for h_0 to lower bounds for derivative estimation.

THEOREM B.2. *Let Condition LB hold with $B_2(p, L)$ in place of $B_\infty(p, L)$ for the NPIV model with a random sample $\{(X_i, Y_i, W_i)\}_{i=1}^n$. Then, for any $0 \leq |\alpha| < p$,*

$$\liminf_{n \to \infty} \inf_{\widehat{g}_n} \sup_{h \in B_2(p,L)} \mathbb{P}_h\big(\|\widehat{g}_n - \partial^\alpha h\|_{L^2(X)} \geq cn^{-(p-|\alpha|)/(2(p+\varsigma)+d)}\big) \geq c' > 0$$

in the mildly ill-posed case, and

$$\liminf_{n \to \infty} \inf_{\widehat{g}_n} \sup_{h \in B_2(p,L)} \mathbb{P}_h\big(\|\widehat{g}_n - \partial^\alpha h\|_{L^2(X)} \geq c(\log n)^{-(p-|\alpha|)/\varsigma}\big) \geq c' > 0$$

in the severely ill-posed case, where $\inf_{\widehat{g}_n}$ denotes the infimum over all estimators of $\partial^\alpha h$ based on the sample of size n, $\sup_{h \in B_2(p,L)} \mathbb{P}_h$ denotes the sup over $h \in B_2(p, L)$ and distributions of (X_i, W_i, u_i) that satisfy Condition LB with ν fixed, and the finite positive constants c and c' do not depend on n.

APPENDIX C: LOWER BOUNDS FOR QUADRATIC FUNCTIONALS

In this section, we study quadratic functionals of the form

$$f(h) = \int \big(\partial^\alpha h(x)\big)^2 \mu(x)\, \mathrm{d}x,$$

where $\mu(x) \geq \underline{\mu} > 0$ is a positive weighting function. These functionals are very important for nonparametric specification and goodness-of-fit testing, as outlined in the conclusion section. We derive lower bounds on convergence rates of estimators of the functional $f(h_0)$.

THEOREM C.1. *Let Condition LB hold with $B_2(p, L)$ in place of $B_\infty(p, L)$ for the NPIV model with a random sample $\{(X_i, Y_i, W_i)\}_{i=1}^n$. Then, for any $0 \leq |\alpha| < p$,*

$$\liminf_{n \to \infty} \inf_{\widehat{g}_n} \sup_{h \in B_2(p,L)} \mathbb{P}_h\big(|\widehat{g}_n - f(h)| > cr_n\big) \geq c' > 0,$$

where

$$r_n = \begin{bmatrix} n^{-1/2} & \text{in the mildly ill-posed case when } p \geq \varsigma + 2|\alpha| + d/4, \\ n^{-4(p-|\alpha|)/(4(p+\varsigma)+d)} & \text{in the mildly ill-posed case when } \varsigma < p < \varsigma + 2|\alpha| + d/4, \\ (\log n)^{-2(p-|\alpha|)/\varsigma} & \text{in the severely ill-posed case,} \end{bmatrix}$$

$\inf_{\widehat{g}_n}$ *denotes the infimum over all estimators of $f(h)$ based on the sample of size n,* $\sup_{h \in B_2(p,L)} \mathbb{P}_h$ *denotes the sup over $h \in B_2(p, L)$ and distributions (X_i, W_i, u_i) that satisfy Condition LB with ν fixed, and the finite positive constants c and c' do not depend on n.*

In the mildly ill-posed case, Theorem C.1 shows that the rate exhibits a so-called elbow phenomenon, in which $f(h_0)$ is $\sqrt{n}$-estimable when $p \geq \varsigma + 2|\alpha| + d/4$ and is irregular otherwise. Moreover, $f(h_0)$ is always irregular in the severely ill-posed case.

Consider estimation using the plug-in estimator $f(\widehat{h})$. Expanding the quadratic, we see that

$$f(\widehat{h}) - f(h_0) = \int \partial^\alpha h_0(x)\big(\partial^\alpha \widehat{h}(x) - \partial^\alpha h_0(x)\big)\mu(x)\,\mathrm{d}x + \big\|\partial^\alpha \widehat{h} - \partial^\alpha h_0\big\|^2_{L^2(\mu)}.$$

Under appropriate normalization, the first term on the right-hand side will be the CLT term. Consider the quadratic remainder term. Since μ is bounded away from zero and the density of X_i is bounded away from zero and infinity, the quadratic remainder term behaves like $\|\partial^\alpha \widehat{h} - \partial^\alpha h_0\|^2_{L^2(X)}$. In the mildly ill-posed case, the optimal convergence rate of this term has been shown to be $O_p(n^{-2(p-|\alpha|)/(2(p+\varsigma)+d)})$ (see Appendix B). This term vanishes faster than $n^{-1/2}$ provided that $p > \varsigma + 2|\alpha| + d/2$, which is a stronger condition than is required for $f(h_0)$ to be $\sqrt{n}$-estimable. Therefore, when $\varsigma + 2|\alpha| + d/4 < p < \varsigma + 2|\alpha| + d/2$, the weighted quadratic functional $f(h_0)$ is $\sqrt{n}$-estimable but its simple plug-in estimator $f(\widehat{h})$ fails to attain the optimal rate.

References

Ai, C. and X. Chen (2003), "Efficient estimation of models with conditional moment restrictions containing unknown functions." *Econometrica*, 71 (6), 1795–1843. [40, 74, 75]

Andrews, D. W. K. (2011), "Examples of l2-complete and boundedly-complete distributions." Cowles Foundation Discussion Paper No. 1801. [55, 60]

Belloni, A., V. Chernozhukov, D. Chetverikov, and K. Kato (2015), "Some new asymptotic theory for least squares series: Pointwise and uniform results." *Journal of Econometrics*, 186 (2), 345–366. [42, 44, 57, 58, 66]

Bierens, H. (1982), "Consistent model specification tests." *Journal of Econometrics*, 20, 105–134. [75]

Blundell, R., X. Chen, and D. Kristensen (2007), "Semi-nonparametric iv estimation of shape-invariant engel curves." *Econometrica*, 75 (6), 1613–1669. [40, 41, 42, 44, 45, 51, 54, 55, 56, 60, 74]

Blundell, R., A. Duncan, and C. Meghir (1998), "Estimating labor supply responses using tax reforms." *Econometrica*, 66 (4), 827–861. [48]

Blundell, R. and J. L. Horowitz (2007), "A non-parametric test of exogeneity." *The Review of Economic Studies*, 74 (4), 1035–1058. [75]

Blundell, R., J. L. Horowitz, and M. Parey (2012), "Measuring the price responsiveness of gasoline demand: Economic shape restrictions and nonparametric demand estimation." *Quantitative Economics*, 3 (1), 29–51. [44, 48, 68]

Blundell, R., J. L. Horowitz, and M. Parey (2017), "Nonparametric estimation of a non-separable demand function under the slutsky inequality restriction." *The Review of Economics and Statistics*, 99 (2), 291–304. [44, 48, 50]

Blundell, R., T. MaCurdy, and C. Meghir (2007), "Labor supply models: Unobserved heterogeneity, nonparticipation and dynamics." In *Handbook of Econometrics*, Vol. 6, Part A, 4667–4775, Chapter 69, Elsevier. [48]

Breunig, C. (2015), "Goodness-of-fit tests based on series estimators in nonparametric instrumental regression." *Journal of Econometrics*, 184 (2), 328–346. [75]

Carrasco, M., J.-P. Florens, and E. Renault (2007), "Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization." In *Handbook of Econometrics*, Vol. 6, Part B (J. J. Heckman and E. E. Leamer, eds.), 5633–5751, Chapter 77, Elsevier. [41, 55]

Cattaneo, M. D. and M. H. Farrell (2013), "Optimal convergence rates, bahadur representation, and asymptotic normality of partitioning estimators." *Journal of Econometrics*, 174 (2), 127–143. [42]

Chen, X., V. Chernozhukov, S. Lee, and W. K. Newey (2014), "Local identification of nonparametric and semiparametric models." *Econometrica*, 82 (2), 785–809. [55, 74]

Chen, X. and T. M. Christensen (2013), "Optimal uniform convergence rates for sieve nonparametric instrumental variables regression." arXiv preprint arXiv:1311.0412, Cowles Foundation Discussion Paper no. 1923 and Cemmap working paper CWP56/13. [40]

Chen, X. and T. M. Christensen (2015a), "Optimal sup-norm rates, adaptivity and inference in nonparametric instrumental variables estimation." arXiv preprint arXiv:1508.03365, Cowles Foundation Discussion Paper no. 1923r. [40, 45, 54]

Chen, X. and T. M. Christensen (2015b), "Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions." *Journal of Econometrics*, 188 (2), 447–465. [42, 56, 57, 58]

Chen, X. and D. Pouzo (2009), "Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals." *Journal of Econometrics*, 152 (1), 46–60. [75]

Chen, X. and D. Pouzo (2012), "Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals." *Econometrica*, 80 (1), 277–321. [41, 45, 56]

Chen, X. and D. Pouzo (2015), "Sieve Wald and QLR inferences on semi/nonparametric conditional moment models." *Econometrica*, 83 (3), 1013–1079. [42, 43, 44, 45, 47, 62, 68, 77]

Chen, X. and M. Reiss (2011), "On rate optimality for ill-posed inverse problems in econometrics." *Econometric Theory*, 27 (3), 497–521. [41, 42, 45, 59, 79]

Chernozhukov, V., S. Lee, and A. M. Rosen (2013), "Intersection bounds: Estimation and inference." *Econometrica*, 81 (2), 667–737. [44, 64, 66, 76]

Darolles, S., Y. Fan, J.-P. Florens, and E. Renault (2011), "Nonparametric instrumental regression." *Econometrica*, 79 (5), 1541–1565. [41, 60]

DeVore, R. A. and G. G. Lorentz (1993), *Constructive Approximation*. Grundlehren der mathematischen Wissenschaften, Vol. 303. Springer-Verlag, Berlin. [57]

Fan, Y. and Q. Li (1996), "Consistent model specification tests: Omitted variables and semiparametric functional forms." *Econometrica*, 64 (4), 865–890. [75]

Florens, J.-P., J. Johannes, and S. Van Bellegem (2012), "Instrumental regression in partially linear models." *The Econometrics Journal*, 15, 304–324. [74]

Florens, J.-P. and A. Simoni (2012), "Nonparametric estimation of an instrumental variables regression: A quasi-Bayesian approach based on regularized posterior." *Journal of Econometrics*, 170, 458–475. [41]

Gagliardini, P. and O. Scaillet (2012), "Tikhonov regularization for nonparametric instrumental variable estimators." *Journal of Econometrics*, 167 (1), 61–75. [41, 60]

Hall, P. and J. L. Horowitz (2005), "Nonparametric methods for inference in the presence of instrumental variables." *The Annals of Statistics*, 33 (6), 2904–2929. [41, 43, 59, 60]

Hardle, W. and E. Mammen (1993), "Comparing nonparametric versus parametric regression fits." *The Annals of Statistics*, 21 (4), 1926–1947. [75]

Hausman, J. A. (1981), "Exact consumer's surplus and deadweight loss." *The American Economic Review*, 71 (4), 662–676. [43, 48]

Hausman, J. A. and W. K. Newey (1995), "Nonparametric estimation of exact consumers surplus and deadweight loss." *Econometrica*, 63 (6), 1445–1476. [43, 47, 63, 68, 73]

Hausman, J. A. and W. K. Newey (2016), "Individual heterogeneity and average welfare." *Econometrica*, 84 (3), 1225–1248. [43, 68]

Hausman, J. A. and W. K. Newey (2017), "Nonparametric welfare analysis." *Annu. Rev. Econ.*, 9 (1), 521–546. [43]

Hoderlein, S. and H. Holzmann (2011), "Demand analysis as an ill-posed inverse problem with semiparametric specification." *Econometric Theory*, 27 (3), 609–638. [60]

Hong, Y. and H. White (1995), "Consistent specification testing via nonparametric series regression." *Econometrica*, 63 (5), 1133–1159. [75]

Horowitz, J. L. (2006), "Testing a parametric model against a nonparametric alternative with identification through instrumental variables." *Econometrica*, 74 (2), 521–538. [75]

Horowitz, J. L. (2011), "Applied nonparametric instrumental variables estimation." *Econometrica*, 79 (2), 347–394. [41, 44, 45, 60, 65, 75]

Horowitz, J. L. (2012), "Specification testing in nonparametric instrumental variable estimation." *Journal of Econometrics*, 167 (2), 383–396. Fourth Symposium on Econometric Theory and Applications (SETA). [75]

Horowitz, J. L. (2014), "Adaptive nonparametric instrumental variables estimation: Empirical choice of the regularization parameter." *Journal of Econometrics*, 180, 158–173. [56]

Horowitz, J. L. and S. Lee (2012), "Uniform confidence bands for functions estimated nonparametrically with instrumental variables." *Journal of Econometrics*, 168, 175–188. [44, 65]

Horowitz, J. L. and V. G. Spokoiny (2001), "An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative." *Econometrica*, 69 (3), 599–631. [75]

Huang, J. Z. (1998), "Projection estimation in multiple regression with application to functional ANOVA models." *The Annals of Statistics*, 26 (1), 242–272. [45, 56]

Huang, J. Z. (2003), "Local asymptotics for polynomial spline regression." *The Annals of Statistics*, 31 (5), 1600–1635. [57]

Johannes, J., S. van Bellegem, and A. Vanhems (2011), "Convergence rates for ill-posed inverse problems with an unknown operator." *Econometric Theory*, 27 (3), 522–545. [78]

Kato, K. (2013), "Quasi-Bayesian analysis of nonparametric instrumental variables models." *The Annals of Statistics*, 41 (5), 2359–2390. [41]

Lavergne, P. and Q. H. Vuong (1996), "Nonparametric selection of regressors: The nonnested case." *Econometrica*, 64 (1), 207–219. [75]

Lounici, K. and R. Nickl (2011), "Global uniform risk bounds for wavelet deconvolution estimators." *The Annals of Statistics*, 39 (1), 201–231. [59]

Mammen, E. (1993), "Bootstrap and wild bootstrap for high dimensional linear models." *The Annals of Statistics*, 21 (1), 255–285. [46, 66]

Newey, W. K. (1997), "Convergence rates and asymptotic normality for series estimators." *Journal of Econometrics*, 79 (1), 147–168. [43, 45, 56, 68, 73]

Newey, W. K. (2013), "Nonparametric instrumental variables estimation." *American Economic Review: Papers and Proceedings*, 103 (3), 550–556. [45]

Newey, W. K. and J. L. Powell (2003), "Instrumental variable estimation of nonparametric models." *Econometrica*, 71 (5), 1565–1578. [40, 41, 55, 60, 66]

Pollard, D. (2002), *A User's Guide to Measure Theoretic Probability*. Cambridge University Press, Cambridge. [64]

Stinchcombe, M. B. and H. White (1998), "Consistent specification testing with nuisance parameters present only under the alternative." *Econometric Theory*, 14 (3), 295–325. [75]

Stone, C. J. (1982), "Optimal global rates of convergence for nonparametric regression." *The Annals of Statistics*, 10 (4), 1040–1053. [42, 58]

Triebel, H. (2006), *Theory of Function Spaces III*. Birkhäuser, Basel. [56, 57, 78]

Vanhems, A. (2010), "Non-parametric estimation of exact consumer surplus with endogeneity in price." *Econometrics Journal*, 13 (3), S80–S98. [43, 68]