

Nonparametric Specification Testing in Random Parameter Models*

Christoph Breunig[†] Stefan Hoderlein[‡]

Humboldt-Universität zu Berlin

Boston College

February 7, 2016

Abstract

In this paper, we suggest and analyze a new class of specification tests for random coefficient models. These tests allow to assess the validity of central structural features of the model, in particular linearity in coefficients and generalizations of this notion like a known nonlinear functional relationship. They also allow to test for degeneracy of the distribution of a random coefficient, i.e., whether a coefficient is fixed or random, including whether an associated variable can be omitted altogether. Our tests are nonparametric in nature, and use sieve estimators of the characteristic function. We analyze their power against both global and local alternatives in large samples and through a Monte Carlo simulation study. Finally, we apply our framework to analyze the specification in a heterogeneous random coefficients consumer demand model.

Keywords: Nonparametric specification testing, random coefficients, unobserved heterogeneity, sieve method, characteristic function, consumer demand.

JEL classification: C12, C14

*We thank seminar participants at Northwestern and WIAS Berlin. Christoph Breunig's research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 Economic Risk.

[†]Christoph Breunig, Humboldt-Universität zu Berlin, Spandauer Straße 1, 10178 Berlin, Germany, e-mail: christoph.breunig@hu-berlin.de.

[‡]Stefan Hoderlein, Boston College, Department of Economics, 140 Commonwealth Ave, Chestnut Hill, MA 02467, USA, email: stefan.hoderlein@yahoo.com.

1 Introduction

Heterogeneity of individual agents is now widely believed to be an important - if not the most important - source of unobserved variation in a typical microeconomic application. Increasingly, the focus of econometrics shifts towards explicitly modeling this central feature of the model through random parameters, as opposed to searching for fixed parameters that summarize only, say, the mean effect. However, as always when additional features are being introduced, this step increases the risk of model misspecification and therefore introducing bias. This suggests to use all the information available in the data to assess the validity of the chosen specification through a test before performing the main analysis. A second important feature of a specification test is that we may be able to find a restricted model that is easier to implement than the unrestricted one. This feature is particularly important in models of complex heterogeneity, which are generically only weakly identified and therefore estimable only under great difficulties.

This paper proposes a family of nonparametric specification tests in models with complex heterogeneity. We focus on the important class of random coefficient models, i.e., models in which there is a finite (d_b dimensional) vector of continuously distributed and heterogeneous parameters $B \in \mathbb{R}^{d_b}$, and a known structural function g which relates these coefficients and a d_x dimensional vector of observable explanatory variables X to a continuous dependent variable Y , i.e.,

$$Y = g(X, B). \quad (1.1)$$

Throughout this paper, we assume that X is independent of B (however, as we discuss below, this does not preclude extensions where some variables in the system are endogenous). The leading example in this class of models is the linear random coefficient model, where $g(X, B) = X'B$, but we also propose specification tests in models where g is nonlinear. Indeed, in extensions we also consider the case where Y is binary, and/or where Y is a vector.

The simple linear model with independent random coefficients is well suited to illustrate our contribution and to explain the most important features of such a nonparametric specification test. This model is known to be exactly point identified in the sense that there is a one-to-one mapping from the conditional probability density function of the observable variables, $f_{Y|X}$ to the density of random coefficients f_B such that the true density of random coefficients is associated with exactly one density of observables (see, e.g., Beran et al. [1996] and Hoderlein et al. [2010]). However, despite the one-to-one mapping between population density of the data and density of random coefficients, the model imposes structure that can be used to assess the validity of the model. For instance, in the very same model, the conditional expectation is linear, i.e., $E[Y|X] = b_0 + b_1X_1 + \dots + b_kX_k$, where $b_j = E[B_j]$. This means that a standard linear

model specification test for quadratic terms in X , or, somewhat more elaborate, nonparametric specification tests involving a nonparametric regression as alternative could be used to test the specification. Similarly, in this model the conditional skedastic function is at most quadratic in X , so any evidence of higher order terms can be taken as rejection of this linear random coefficients specification, too. However, both of these tests do not use the entire distribution of the data, and hence do not allow us to discern between the truth and certain alternatives.

In contrast, our test will be based on the characteristic function of the data, i.e., we use the entire distribution of the data to assess the validity of the specification. In the example of the linear model, we compare the distance between a series least squares estimator of the unrestricted characteristic function $E[\exp(itY)|X]$, and an estimator of the restricted one, which is $E[\exp(it(X'B))|X] = \int \exp(it(X'b))f_B(b)db$, where the probability density function f_B of the random coefficients B is replaced by a sieve minimum distance estimator under the hypothesis of linearity. More specifically, using the notation $\varepsilon(X, t) = E[\exp(itY) - \exp(it(X'B))|X]$, our test is based on the observation that under the null hypothesis of linearity, $\varepsilon(x, t) = 0$ holds, for all (x, t) , or equivalently,

$$\int E[|\varepsilon(X, t)|^2] \varpi(t) dt = 0,$$

for any strictly positive weighting function ϖ . Our test statistic is then given by the sample counterpart

$$S_n \equiv n^{-1} \sum_{j=1}^n \int |\hat{\varepsilon}_n(X_j, t)|^2 \varpi(t) dt,$$

where $\hat{\varepsilon}_n$ denotes an estimator of ε as described above. We reject the null hypothesis of linearity if the statistic S_n becomes too large.

This test uses evidently the entire distribution of the data to assess the validity of the specification. It therefore implicitly uses all available comparisons between the restricted and the unrestricted model, not just the ones between, say a linear conditional mean and a nonparametric conditional mean. Moreover, it does not even require that these conditional means (or higher order moments) exist. To see that our test uses the information contained in the conditional moments, consider again the linear random coefficients model. Using a series expansion of the exponential function, $\varepsilon(X, t) = 0$ is equivalent to

$$\sum_{l=0}^{\infty} (it)^l \{E[Y^l|X] - E[(X'B)^l|X]\} / (l!) = 0,$$

provided all moments exist. This equation holds true, if and only if, for every coefficient $l \geq 1$:

$$E[Y^l|X] = E[(X'B)^l|X],$$

i.e., there is equality of all of these conditional moments. This implies, in particular, the first and second conditional moment equation $E[Y|X] = X'E[B]$ and $E[Y^2|X] = X'E[BB']X$. As

such, our test exploits potential discrepancies in any of the conditional moments, and works even if some or all of them do not exist.

Our test is consistent against a misspecification of model (1.1) in the sense that, under the alternative, there exists no vector of random coefficients B satisfying the model equation (1.1) for a known function g . Indeed, such a misspecification leads to a deviation of the unrestricted from the restricted conditional characteristic function. Moreover, our test is also consistent against certain specific other alternatives, e.g., if the null is the linear random coefficient model and the alternative is a higher order polynomial with random parameters.

However, we can also use the same testing principle to analyze whether or not a parameter is nonrandom, which usually allows for a $\sqrt{n}$ consistent estimator for this parameter, and whether it has in addition mean zero which implies that we may omit the respective variable altogether. This is important, because from a nonparametric identification perspective random coefficient models are weakly identified (i.e., stem from the resolution of an ill posed inverse problem), a feature that substantially complicates nonparametric estimation¹. If we think of a parametric model as an approximation to a more complex nonparametric model, this is likely also going to affect the finite sample behavior of any parametric estimator. As such it is desirable to reduce the number of dimensions of random parameters as much as possible, and our test may serve as guidance in this process.

Finally, it is important to note that our method also applies to other point identified random coefficient models such as models that are linear in parameters, but where X is replaced by a elementwise transformation of the covariates (i.e., X_j is replaced by $h_j(X_j)$ with unknown h_j . See Gautier and Hoderlein [2015] for the formal argument that establishes identification). The reason is that the mean regression in these models is still of an additive structure, i.e., in particular it does preclude interaction terms among the variables that feature across all moments.

Extensions. While setting up the basic framework is a contribution in itself, a key insight in this paper is that testing is possible even if the density of random coefficients is not point identified under the null hypothesis. This is important, because many structural models are not linear in an index. As such, it is either clear that they are not point identified in general and at best set identified (see Hoderlein et al. [2014], for such an example), or identification is unknown. To give an example of such a model, consider a stylized version of the workhorse QUAIDS model of consumer demand (Banks et al. [1997]), where demand for a good Y is defined through:

$$Y = B_0 + B_1X + B_2X^2,$$

¹In a nonparametric sense, there is a stronger curse of dimensionality associated with random coefficient models than with nonparametric density estimation problems (see, e.g., Hoderlein et al. [2010]).

where B_j denotes parameters, and X log total expenditure. For reasons outlined in Masten [2015], the joint density of random parameters B_0, B_1, B_2 is not point identified in general. Our strategy is now to solve a functional minimization problem that minimizes a similar distance as outlined above between restricted and unrestricted model, and allows us to obtain one element in this set as minimizer. If the distance between the restricted model and the unrestricted model is larger than zero, we conclude that we can reject the null that the model is, in our example, a heterogeneous QUAIDS. However, if the distance is not significantly different from zero, there still may be other non-QUAIDS models which achieve zero distance, and which we therefore cannot distinguish from the heterogeneous QUAIDS model. As such, in the partially identified case we do not have power against all possible alternatives, and our test becomes conservative.

Interestingly, even if our model is not identified under the null hypothesis, such as in the case of the random coefficients QUAIDS model, our test still has power against certain alternatives, e.g., any higher polynomial random coefficient model. Again, since our test compares all conditional moments, $\varepsilon(X, t) = 0$ for all t implies that the cubic model $Y = \tilde{B}_0 + X\tilde{B}_1 + X^2\tilde{B}_2 + X^3\tilde{B}_3$ with random coefficients $(\tilde{B}_0, \tilde{B}_1, \tilde{B}_2, \tilde{B}_3)$ or any other higher polynomial model is misspecified. In this sense, our test has power even in situations where neither the null nor the alternative models are identified.

The second extension is that our testing principle extends to systems of equations, i.e., situations in which the endogenous variable is not a scalar, but a vector, by replacing the scalar conditional characteristic function with a vector valued one, i.e., $E[\exp(it'Y)|X = x]$. For instance, one may reformulate the triangular random coefficients model of Hoderlein et al. [2014], where $Y_1 = A_0 + A_1Y_2$, $Y_2 = B_2 + B_3X$ as

$$\begin{aligned} Y_1 &= B_0 + B_1X, \\ Y_2 &= B_2 + B_3X, \end{aligned}$$

with $B = (B_0, B_1, B_2, B_3) \perp X$, and then either use the minimum distance principle outlined above, or, under the additional assumptions in Hoderlein et al. [2014], their estimator for the restricted model.

Finally, we may extend the approach outlined in this paper to binary or discrete dependent variables, provided we have a special regressor Z , as in Lewbel [2000]. In this case, we replace the density of the data with the marginal probability with respect to the special regressor; otherwise, most of the above reasoning remains virtually unchanged.

Related Literature. As already mentioned, this paper draws upon several literatures. The first is nonparametric random coefficients models, a recently quite active line of work, including work on the linear model (Beran and Hall [1992], Beran et al. [1996], and Hoderlein et al. [2010]),

the binary choice model (Ichimura and Thompson [1998] and Gautier and Kitamura [2013]), and the treatment effects model (Gautier and Hoderlein [2015]). Related is also the wider class of models analyzed in Fox and Gandhi [2009] and Lewbel and Pendakur [2013], who both analyze nonlinear random coefficient models, Masten [2015] and Matzkin [2012], who both discuss identification of random coefficients in a simultaneous equation model, Hoderlein et al. [2014] who analyze a triangular random coefficients model, and Dunker et al. [2013] and Fox and Lazzati [2012] who analyze games.

As far as we know, the general type of specification tests we propose in this paper is new to the literature. In linear semiparametric random coefficient models, Beran [1993] proposes a minimum distance estimator for the unknown distributional parameter of the random coefficient distribution. Within this framework of a parametric joint random coefficients' distribution, Beran also proposes goodness of fit testing procedures. Also in a parametric setup where the unknown random coefficient distribution follows a parametric model, Swamy [1970] establishes a test for equivalence of random coefficient across individuals, i.e., a test for degeneracy of the random coefficient vector. We emphasize that with our testing methodology, despite less restrictive distributional assumptions, we are able to test degeneracy of a subvector of B while others are kept as random. Another test in linear parametric random coefficient models was proposed by Andrews [2001], namely a test for degeneracy of some random coefficients. In contrast, our nonparametric testing procedure is based on detecting differences in conditional characteristic function representation and, as we illustrate below, we do not obtain boundary problems as in Andrews [2001].

In this paper, we use sieve estimators for the unknown distributional elements. In the econometrics literature, sieve methodology was recently used to construct Wald statistics (see Chen and Pouzo [2015]) or nonparametric specification tests (see Breunig [2015b]), and, in nonparametric instrumental regression, tests based on series estimators have been proposed by Horowitz [2012] and Breunig [2015a]. Moreover, in the nonparametric IV model, tests for parametric specification have been proposed by Horowitz [2006] and Horowitz and Lee [2009], while Blundell and Horowitz [2007] proposes a test of exogeneity. Santos [2012] develops hypothesis tests which are robust to a failure of identification. More generally, there is a large literature on model specification tests based on nonparametric regression estimators in L^2 distance starting with Härdle and Mammen [1993]. Specification tests in nonseparable were proposed by Hoderlein et al. [2011] and Lewbel et al. [2015]. None of these tests is applicable to specification testing in random coefficient models.

Finally, our motivation is partly driven by consumer demand, where heterogeneity plays an important role. Other than the large body of work reviewed above we would like to mention the recent work by Hausman and Newey [2013], Blundell et al. [2010], see Lewbel [1999] for a

review of earlier work.

Overview of Paper. In the second section, we introduce our test formally, and discuss its large sample properties in the baseline scenario. We distinguish between general specification tests, and subcases where we can separate additively a part of the model which contains only covariates and fixed coefficients from the remainder. In the third section, we focus on the extensions discussed above. The finite sample behavior is investigated through a Monte Carlo study in the fourth section. Finally, we apply all concepts to analyze the validity of a heterogeneous QAIDS (Banks et al. [1997]) model which is the leading parametric specification in consumer demand.

2 The Test Statistic and its Asymptotic Properties

2.1 Examples of Testable Hypotheses

In the wider class of models encompassed by (1.1), we consider two different types of hypotheses tests. First, we provide a general test for the hypothesis that the structural relation of the covariates and random coefficients to the outcome coincides with a known function g . We thus consider the hypothesis ²

$$H_{\text{mod}} : Y = g(X, B) \quad \text{for some random parameters } B.$$

The alternative hypothesis is given by $P(Y \neq g(X, B) \quad \text{for all random parameters } B) > 0$. An important example is testing the hypothesis of linearity; that is, whether with probability one

$$H_{\text{lin}} : Y = X'B,$$

in which case the distribution of B is identified. Another example is a quadratic form of the function g in each component of the vector of covariates X , i.e., we want to assess the null hypothesis $H_{\text{quad}} : Y = B_0 + X'B_1 + (X^2)'B_2$ for some $B = (B_0, B_1, B_2)$, where a squared vector is understood element-wise. Note that in the former example f_B is point identified, while in the latter example it is only partially identified. This fact will generally result in a lack of power against certain alternatives.

The second type of hypotheses our test allows to consider is whether a subvector of B , say, B_2 , is deterministic (or, equivalently, has a degenerate distribution). I.e., we want to consider the following hypothesis

$$H_{\text{deg}} : B_2 = b_2 \quad \text{for all } B = (B_0, B_1, B_2) \text{ satisfying (1.1).}$$

²Equalities involving random variables are understood as equalities with probability one, even if we do not say so explicitly.

While this type of hypothesis could be considered in the most general model, motivated by the linear (or polynomial) model we will confine ourselves to functions G that have a partially linear structure, such that

$$H_{\text{part-lin}} : Y = B_0 + X_1' B_1 + g_2(X, B_2), \quad \text{for some random parameters } B = (B_0, B_1, B_2),$$

for a known function g_2 holds true. In the two main cases outlined above, this covers the following examples of hypotheses: First, in a linear model, i.e., $Y = B_0 + X_1' B_1 + X_2' B_2$, it allows to test whether the coefficient on X_2 is deterministic. Put reversely, we may test the null

$$H_{\text{deg-lin}} : Y = B_0 + X_1' B_1 + X_2' b_2,$$

against the alternative that B_2 is random.

A second example arises if, in the quadratic model, we want to test a specification with deterministic second order terms, i.e.

$$H_{\text{deg-quad}} : Y = B_0 + X_1' B_1 + (X_2^2)' b_2,$$

against the alternative that B_2 is random. In either case, the alternative is given by $P(B_2 \neq b_2 \text{ for some } B \text{ satisfying (1.1)}) > 0$.

2.2 The Test Statistic

Our test statistic is based on the L^2 distance between an unrestricted conditional characteristic function and a restricted one. We show below that each null hypothesis is then equivalent to

$$\varepsilon(X, t) = 0 \text{ for all } t, \tag{2.1}$$

where $\varepsilon : \mathbb{R}^{d_x+1} \rightarrow \mathbb{C}$ is a complex valued, measurable function. Our testing procedure is hence based on the L^2 distance of ε to zero. Equation (2.1) is equivalent to

$$\int E[|\varepsilon(X, t)|^2] \varpi(t) dt = 0,$$

for some strictly positive weighting function ϖ with $\int \varpi(t) dt < \infty$. In the following examples, we provide explicit forms for the function ε . The analysis is based on the assumption of independence of covariates X and random coefficients B . See also the discussion after Assumption 1 below.

Example 1 (Testing functional form restrictions). *The null hypothesis H_{mod} is equivalent to the following equation involving conditional characteristic functions*

$$E[\exp(itY)|X] = E[\exp(itg(X, B))|X],$$

for each $t \in \mathbb{R}$, a known function g , and some random parameters B . Hence, equation (2.1) holds true with

$$\varepsilon(X, t) = E[\exp(itY) - \exp(itg(X, B))|X]. \quad (2.2)$$

As already mentioned, if the function g is nonlinear the probability density function (p.d.f.) of the random coefficients B , denoted by f_B , does not need to be point identified. On the other hand, if g is the inner product of its entries, then (2.1) holds true with

$$\varepsilon(X, t) = E[\exp(itY) - \exp(itX'B)|X],$$

and in this case the distribution of B is identified (see, e.g., Hoderlein et al. [2010]). While our test, based on the function ε , is in general consistent against a failure of the null hypothesis H_{mod} , it is also consistent against certain alternative models such as higher order polynomials. To illustrate this, we consider the random coefficient QUAIDS model, which we also study in our application. Under the maintained hypothesis we have $Y = \tilde{B}_0 + \tilde{B}_1X + \tilde{B}_2X^2$ for random coefficients $\tilde{B}_0$, $\tilde{B}_1$, and $\tilde{B}_2$ (also independent of X). In this case, the conditional first and second moment equation implied by equation (2.2) yield $E[\tilde{B}_2] = 0$ and $Var(\tilde{B}_2) = 0$, respectively. We conclude $\tilde{B}_2 = 0$.

Example 2 (Testing degeneracy). Under a partially linear structure $H_{part-lin}$, the null hypothesis H_{deg} implies the equality of conditional characteristic functions, i.e.,

$$E[\exp(itY)|X] = E[\exp(it(B_0 + X_1'B_1))|X_1] \exp(itg_2(X, b_2)), \quad (2.3)$$

for each $t \in \mathbb{R}$. Therefore, equation (2.1) holds with

$$\varepsilon(X, t) = E[\exp(itY)|X] - E[\exp(it(B_0 + X_1'B_1))|X_1] \exp(itg_2(X, b_2)).$$

We assume throughout the paper that the parameter b_2 is identified through equation (2.3). Our test, based on the function ε , has power against a failure of H_{deg} if the function g_2 is an elementwise transformation of each component of the vector X_2 (see Gautier and Hoderlein [2015]). Moreover, we also see that our test has power in the random coefficient QUAIDS model. Under the maintained hypothesis we have $Y = \tilde{B}_0 + \tilde{B}_1X + \tilde{B}_2X^2$ for random coefficients $\tilde{B}_0$, $\tilde{B}_1$, and $\tilde{B}_2$ (also independent of X). In this case, the conditional first and second moment equation implied by equation (2.3) yield $E[\tilde{B}_2] = b_2$ and $E[\tilde{B}_2^2] = b_2^2$, respectively. We conclude that $\tilde{B}_2$ has to be degenerate with $\tilde{B}_2 = b_2$. Similarly, our test is able to reject any higher order polynomials with random parameters in the nonlinear part.

As already mentioned, we use the fact that equation (2.1) is equivalent to

$$\int E[|\varepsilon(X, t)|^2] \varpi(t) dt = 0,$$

for some strictly positive weighting function ϖ . Our test statistic is given by the sample counterpart to this expression, which is

$$S_n \equiv n^{-1} \sum_{j=1}^n \int |\widehat{\varepsilon}_n(X_j, t)|^2 \varpi(t) dt,$$

where $\widehat{\varepsilon}_n$ is a consistent estimator of ε . Below, we show that the statistic S_n is asymptotically normally distributed after standardization. As the test is one sided, we reject the null hypothesis at level α when the standardized version of S_n is larger than the $(1 - \alpha)$ -quantile of $\mathcal{N}(0, 1)$.

We consider a series estimator for the conditional characteristic function of Y given X , i.e., $\varphi(x, t) \equiv E[\exp(itY)|X = x]$. To do so, let us introduce a vector of basis functions denoted by $p_m(\cdot) = (p_1(\cdot), \dots, p_m(\cdot))'$ for some integer $m \geq 1$. Further, let $\mathbf{X}_m \equiv (p_m(X_1), \dots, p_m(X_n))'$ and $\mathbf{Y}_n(t) = (\exp(itY_1), \dots, \exp(itY_n))$. We replace φ by the series least squares estimator

$$\widehat{\varphi}_n(x, t) \equiv p_{m_n}(x) (\mathbf{X}_{m_n}' \mathbf{X}_{m_n})^{-1} \mathbf{X}_{m_n}' \mathbf{Y}_n(t),$$

where the integer m_n increases with sample size n . We compare this unrestricted conditional expectation estimator to a restricted one which depends on the hypothesis under consideration.

Example 3 (Testing functional form restrictions). *Let us introduce the integral transform $(\mathcal{F}_g f)(X, t) \equiv \int \exp(itg(X, b))f(b)db$, which coincides with the Fourier transform evaluated at tX , if g is linear.³ If g is nonlinear, then the random coefficient's p.d.f. f_B does not need to be identified through $\varphi = \mathcal{F}_g f$. We estimate the function ε by*

$$\widehat{\varepsilon}_n(X_j, t) = \widehat{\varphi}_n(X_j, t) - (\mathcal{F}_g \widehat{f}_{Bn})(X_j, t),$$

where the estimator $\widehat{f}_{Bn}$ is a sieve minimum distance estimator given by

$$\widehat{f}_{Bn} \in \arg \min_{f \in \mathcal{B}_n} \left\{ \sum_{j=1}^n \int |\widehat{\varphi}_n(X_j, t) - (\mathcal{F}_g f)(X_j, t)|^2 \varpi(t) dt \right\} \quad (2.4)$$

and $\mathcal{B}_n = \{\phi(\cdot) = \sum_{l=1}^{k_n} \beta_l q_l(\cdot)\}$ is a linear sieve space of dimension $k_n < \infty$ with basis functions $\{q_l\}_{l \geq 1}$. Here, k_n and m_n increase with sample size n . As we see below, we require that m_n increases faster than k_n . Next, using the notation $\mathbf{F}_n(t) = ((\mathcal{F}_g q_{k_n})(X_1, t), \dots, (\mathcal{F}_g q_{k_n})(X_n, t))'$, the minimum norm estimator of f_B given in (2.4) coincides with $\widehat{f}_{Bn}(\cdot) = q_{k_n}(\cdot)' \widehat{\beta}_n$ where

$$\widehat{\beta}_n = \left(\int \mathbf{F}_n(t)' \mathbf{F}_n(t) \varpi(t) dt \right)^- \int \mathbf{F}_n(t)' \Phi_n(t) \varpi(t) dt$$

and $\Phi_n(t) = (\widehat{\varphi}_n(X_1, t), \dots, \widehat{\varphi}_n(X_n, t))'$. The exponent $-$ denotes the Moore–Penrose generalized inverse. As a byproduct, we thus extend the minimum distance estimation principle of Beran and Millar [1994] to nonlinear random coefficient models and the sieve methodology.

³The Fourier transform is given by $(\mathcal{F}\phi)(t) \equiv \int \exp(itz)\phi(z)dz$ for a function $\phi \in L^1(\mathbb{R}^d)$ while its inverse is $(\mathcal{F}^{-1}\widetilde{\phi})(z) \equiv (2\pi)^{-d} \int \exp(-itz)\widetilde{\phi}(t)dt$.

Example 4 (Testing degeneracy). Under the partially linear structure $H_{part-lin}$ and hypothesis H_{deg} , Assumption 1 implies

$$E[Y|X] = b_0 + X_1' b_1 + g_2(X, b_2), \quad (2.5)$$

where $b_0 \equiv E[B_0]$ and $b_1 \equiv E[B_1]$. Let $\widehat{b}_{2n}$ denote the nonlinear least squares estimator of b_2 . We denote $p_{\underline{k}_n}(\cdot) = (p_1(\cdot), \dots, p_{k_n}(\cdot))'$ and $\mathbf{X}_{1n} \equiv (p_{\underline{k}_n}(X_{11}), \dots, p_{\underline{k}_n}(X_{1n}))'$ which is a $n \times k_n$ matrix. Consequently, we estimate the function ε by

$$\widehat{\varepsilon}_n(X_j, t) = \widehat{\varphi}_n(X_j, t) - p_{\underline{k}_n}(X_{1j})' (\mathbf{X}_{1n}' \mathbf{X}_{1n})^{-1} \mathbf{X}_{1n}' \mathbf{U}_n \exp(itg_2(X_j, \widehat{b}_{2n})),$$

where $\mathbf{U}_n = (\exp(it(Y_1 - g_2(X_1, \widehat{b}_{2n}))), \dots, \exp(it(Y_n - g_2(X_n, \widehat{b}_{2n}))))'$.

2.3 The Asymptotic Distribution of the Statistic under the Null Hypothesis

As a consequence of the previous considerations, we distinguish between two main hypotheses, i.e., functional form restrictions and degeneracy of some random coefficients. Both types of tests require certain common assumptions, and we start out this section with a subsection where we discuss the assumptions we require in both cases. Thereafter, we analyze each of the two types of tests in a separate subsection, and provide additional assumptions to obtain the test's asymptotic distribution under each null hypothesis.

2.3.1 General Assumptions for Inference

Assumption 1. *The random vector X is independent of B .*

Assumption 1 is crucial for the construction of our test statistic. Full independence is commonly assumed in the random coefficients literature (see, for instance, Beran [1993], Beran et al. [1996], Hoderlein et al. [2010], or any of the random coefficient references mentioned in the introduction). It is worth noting that this assumption can be relaxed by assuming independence of X and B conditional on additional variables that are available to the econometrician, allowing for instance for a control function solution to endogeneity as in Hoderlein and Sherman [2015], or simply controlling for observables in the spirit of the unconfoundedness assumption in the treatment effects literature. Further, $\mathcal{X}$ denotes the support of X .

Assumption 2. (i) *We observe a sample $((Y_1, X_1), \dots, (Y_n, X_n))$ of independent and identically distributed (i.i.d.) copies of (Y, X) . (ii) *There exists a strictly positive and nonincreasing sequence $(\lambda_n)_{n \geq 1}$ such that, uniformly in n , the smallest eigenvalue of $\lambda_n^{-1} E[p_{\underline{m}_n}(X)p_{\underline{m}_n}(X)']$ is bounded away from zero. (iii) *There exists a constant $C \geq 1$ and a sequence of positive integers $(m_n)_{n \geq 1}$ satisfying $\sup_{x \in \mathcal{X}} \|p_{\underline{m}_n}(x)\|^2 \leq C m_n$ with $m_n^2 \log n = o(n\lambda_n)$.***

Assumption 2 (ii) – (iii) restricts the magnitude of the approximating functions $\{p_l\}_{l \geq 1}$ and imposes nonsingularity of their second moment matrix. Assumption 2 (iii) holds, for instance, for polynomial splines, Fourier series and wavelet bases. Moreover, this assumption ensures that the smallest eigenvalue of $E[p_{\underline{m}_n}(X)p_{\underline{m}_n}(X)']$ is not too small relative to the dimension m_n . In Assumption 2 (ii), we assume that the eigenvalues of the matrix $E[p_{\underline{m}_n}(X)p_{\underline{m}_n}(X)']$ may tend to zero at the rate λ_n which was recently also assumed by Chen and Christensen [2015]. On the other hand, the sequence $(\lambda_n)_{n \geq 1}$ is bounded away from zero if $\{p_l\}_{l \geq 1}$ forms an orthonormal basis on the compact support of X and the p.d.f. of X is bounded away from zero (cf. Proposition 2.1 of Belloni et al. [2015]). The next result provides sufficient condition for Assumption 2 (ii) to hold even if the sequence of eigenvalues $(\lambda_n)_{n \geq 1}$ tends to zero.

Proposition 5. *Assume that $\{p_l\}_{l \geq 1}$ forms an orthonormal basis on $\mathcal{X}$ with respect to a measure ν . Let $(\lambda_n)_{n \geq 1}$ be a sequence that tends to zero. Suppose that, for some constant $0 < c < 1$, for all $n \geq 1$ and any vector $a_n \in \mathbb{R}^{m_n}$ the inequality*

$$\int (a_n' p_{\underline{m}_n}(x))^2 \mathbb{1}\{f(x) < \lambda_n\} \nu(dx) \leq c \int (a_n' p_{\underline{m}_n}(x))^2 \nu(dx) \quad (2.6)$$

holds, where $f = dF_X/d\nu$. Then, Assumption 2 (ii) is satisfied.

Condition (2.6) is violated, for instance, if $dF_X/d\nu$ vanishes on some subset $\mathcal{A}$ of the support of ν with $\nu(\mathcal{A}) > 0$. Estimation of conditional expectations with respect to X is more difficult when the marginal p.d.f. f_X is close to zero on the support $\mathcal{X}$. In this case, the rate of convergence will slow down relative to λ_n (see Lemma 2.4 in Chen and Christensen [2015] in case of series estimation). As we see from Proposition 2.6, λ_n plays the role of a truncation parameter used in kernel estimation of conditional densities to ensure that the denominator is bounded away from zero.

To derive our test's asymptotic distribution, we standardize S_n by subtracting the mean and dividing through a variance which we introduce in the following. Let $V \equiv (Y, X)$, and denote by δ a complex valued function which is the difference of $\exp(itY)$ and the restricted conditional characteristic function, i.e., $\delta(V, t) = \exp(itY) - (\mathcal{F}_g f_B)(X, t)$ in case of H_{mod} , and $\delta(V, t) = \exp(itY) - E[\exp(it(B_0 + X_1' B_1)) | X_1] \exp(itg_2(X, b_2))$ in case of H_{deg} . Moreover, note that $\int E[\delta(V, t) | X] \varpi(t) dt = 0$ holds.

Definition 6. *Denote by $P_n = E[p_{\underline{m}_n}(X)p_{\underline{m}_n}(X)']$, and define*

$$\begin{aligned} \mu_{m_n} &\equiv \int E[|\delta(V, t)|^2 p_{\underline{m}_n}(X)' P_n^{-1} p_{\underline{m}_n}(X)] \varpi(t) dt \quad \text{and} \\ \varsigma_{m_n} &\equiv \left(\int \int \left\| P_n^{-1/2} E[\delta(V, s) \overline{\delta(V, t)} p_{\underline{m}_n}(X) p_{\underline{m}_n}(X)'] P_n^{-1/2} \right\|_F^2 \varpi(s) \varpi(t) ds dt \right)^{1/2}. \end{aligned}$$

Here, we use the notation $\bar{\phi}$ for the complex conjugate of a function ϕ , and $\|\cdot\|_F$ to denote the Frobenius norm.

Assumption 3. *There exists some constant $C > 0$ such that $E[|\int \delta(V, t)\varpi(t)dt|^2|X] \geq C$.*

Assumption 3 ensures that the conditional variance of $\int \delta(V, t)\varpi(t)dt$ is uniformly bounded away from zero. Assumptions of this type are commonly required to obtain asymptotic normality of series estimators (see Assumption 4 of Newey [1997] or Theorem 4.2 of Belloni et al. [2015]). As we show in the appendix, Assumption 3 implies $\varsigma_{m_n} \geq C\sqrt{m_n}$.

2.3.2 Testing Functional Form Restrictions

We now present conditions that are sufficient to provide the test's asymptotic distribution under the null hypothesis H_{mod} . To do so, let us introduce the norm $\|\phi\|_{\varpi} = (\int E|\phi(X, t)|^2\varpi(dt))^{1/2}$ and the linear sieve space $\Phi_n \equiv \{\phi : \phi(\cdot) = \sum_{j=1}^{m_n} \beta_j p_l(\cdot)\}$. Moreover, $\|\cdot\|$ and $\|\cdot\|_{\infty}$, respectively, denote the Euclidean norm and the supremum norm. Let us introduce $A_n = \int E[(\mathcal{F}_g q_{k_n})(X, t)(\mathcal{F}_g q_{k_n})(X, t)']\varpi(t)dt$ and its empirical analog $\hat{A}_n = (n^{-1} \int \mathbf{F}_n(t)' \mathbf{F}_n(t)\varpi(t)dt)^{-}$ (see also Example 3).

Assumption 4. (i) *For any p.d.f. f_B satisfying $\varphi = \mathcal{F}_g f_B$ there exists $\Pi_{k_n} f_B \in \mathcal{B}_n$ such that $n\|\mathcal{F}_g(\Pi_{k_n} f_B - f_B)\|_{\varpi}^2 = o(\sqrt{m_n})$.* (ii) *There exists $\Pi_{m_n} \varphi \in \Phi_n$ such that $n\|\Pi_{m_n} \varphi - \varphi\|_{\varpi}^2 = o(\sqrt{m_n})$ and $\|\Pi_{m_n} \varphi - \varphi\|_{\infty} = O(1)$.* (iii) *It holds $k_n = o(\sqrt{m_n})$.* (iv) *It holds $A_n^- = O(1)$ and $P(\text{rank}(A_n) = \text{rank}(\hat{A}_n)) = 1 + o(1)$.* (v) *There exists a constant $C > 0$ such that $\sum_{l \geq 1} (\int_{\mathbb{R}^{d_b}} \phi(b) q_l(b) db)^2 \leq C \int_{\mathbb{R}^{d_b}} \phi^2(b) db$ for all square integrable functions ϕ .*

Assumption 4 (i) is a requirement on the sieve approximation error for all functions f_B that belong to the identified set $\mathcal{I}_g \equiv \{f : f \text{ is a p.d.f. with } \varphi = \mathcal{F}_g f\}$. This condition ensures that the bias for estimating any f_B in the identified set $\mathcal{I}_g$ is asymptotically negligible. Assumption 4 (ii) determines the sieve approximation error for the function φ . Consider the linear case and let $\|\mathcal{F}(\Pi_{k_n} f_B - f_B)\|_{\varpi} = O(k_n^{-s/d_x})$ for some constant $s > 0$, then Assumptions 4 (i) and (iii) are satisfied if $m_n \sim n^{\zeta}$ and $k_n \sim n^{\kappa}$ where $d_x(1 - \zeta/2)/(2s) < \kappa < \zeta/2$.⁴ We thus require $\zeta > 2d_x/(2s + d_x)$, so s has to increase with dimension d_x , which reflects a curse of dimensionality. In this case, Assumption 4 (ii) automatically holds if $\|\Pi_{m_n} \varphi - \varphi\|_{\varpi} = O(m_n^{-s/d_x})$ and we may choose κ to balance variance and bias, i.e., $\kappa = d_x/(2s + d_x)$.⁵ For further discussion and examples of sieve bases, we refer to Chen [2007]. Assumption 4 (iv) ensures that the sequence of generalized inverse matrices is bounded and imposes a rank condition. This condition is

⁴We use the notation $a_n \sim b_n$ for $cb_n \leq a_n \leq Cb_n$ given two constant $c, C > 0$ and all $n \geq 1$.

⁵This choice of k_n corresponds indeed to the optimal smoothing parameter choice in nonparametric random coefficient model if $s = r + (d_x - 1)/2$ where r corresponds to the smoothness of f_B (see Hoderlein et al. [2010] in case of kernel density estimation).

sufficient and necessary for convergence in probability of generalized inverses of random matrices with fixed dimension (see Andrews [1987] for generalized Wald tests). Assumption 4 (v) is satisfied if $\{q_l\}_{l \geq 1}$ forms a Riesz basis in $L^2(\mathbb{R}^{d_b}) \equiv \{\phi : \int_{\mathbb{R}^{d_b}} \phi^2(s) ds < \infty\}$. The following result establishes asymptotic normality of our standardized test statistic.

Theorem 7. *Let Assumptions 1–4 hold with $\delta(V, t) = \exp(itY) - (\mathcal{F}_g f_B)(X, t)$. Then, under H_{mod} we obtain*

$$(\sqrt{2}\varsigma_{m_n})^{-1}(n S_n - \mu_{m_n}) \xrightarrow{d} \mathcal{N}(0, 1).$$

Remark 8 (Estimation of Critical Values). *The asymptotic results of the previous theorem depends on unknown population quantities. As we see in the following, the critical values can be easily estimated. We define $\delta_n(V, t) = \exp(itY) - (\mathcal{F}_g \hat{f}_{B_n})(X, t)$, and*

$$\sigma_n(s, t) = \left(\delta_n(V_1, s) \overline{\delta_n(V_1, t)}, \dots, \delta_n(V_n, s) \overline{\delta_n(V_n, t)} \right)'.$$

We replace μ_{m_n} and ς_{m_n} , respectively, by the estimators

$$\hat{\mu}_{m_n} = \int \text{tr} \left((\mathbf{X}'_n \mathbf{X}_n)^{-1/2} \mathbf{X}'_n \text{diag}(\sigma_n(t, t)) \mathbf{X}_n (\mathbf{X}'_n \mathbf{X}_n)^{-1/2} \right) \varpi(t) dt$$

and

$$\hat{\varsigma}_{m_n} = \left(\int \int \left\| (\mathbf{X}'_n \mathbf{X}_n)^{-1/2} \mathbf{X}'_n \text{diag}(\sigma_n(s, t)) \mathbf{X}_n (\mathbf{X}'_n \mathbf{X}_n)^{-1/2} \right\|_F^2 \varpi(s) \varpi(t) ds dt \right)^{1/2}.$$

Proposition 9. *Under the conditions of Theorem 7, we obtain*

$$\varsigma_{m_n} \hat{\varsigma}_{m_n}^{-1} = 1 + o_p(1) \quad \text{and} \quad \hat{\mu}_{m_n} = \mu_{m_n} + o_p(\sqrt{m_n}).$$

The asymptotic distribution of our standardized test statistic remains unchanged if we replace μ_{m_n} and ς_{m_n} by estimators introduced in the last remark. This is summarized in following corollary, which follows immediately from Theorem 7 and Proposition 9.

Corollary 10. *Under the conditions of Theorem 7, we obtain*

$$(\sqrt{2}\hat{\varsigma}_{m_n})^{-1}(n S_n - \hat{\mu}_{m_n}) \xrightarrow{d} \mathcal{N}(0, 1).$$

An alternative way to obtain critical values is the bootstrap which, for testing nonlinear functionals in nonparametric instrumental regression, was considered by Chen and Pouzo [2015]. In our situation, the critical values can be easily estimated and the finite sample properties of our testing procedure are promising, thus we do not elaborate bootstrap procedures here. In the following example, we illustrate our sieve minimum distance approach for estimating f_B in the case of linearity of g .

Example 11 (Linear Case). *Let g be linear and recall that in this case the integral transform $\mathcal{F}_g$ coincides with the Fourier transform $\mathcal{F}$. For the sieve space $\mathcal{B}_n$, we consider as basis functions Hermite functions given by*

$$q_l(x) = \frac{(-1)^l}{\sqrt{2^l l!} \sqrt{\pi}} \exp(x^2/2) \frac{d^l}{dx^l} \exp(-x^2).$$

These functions form an orthonormal basis of $L^2(\mathbb{R})$. Hermite functions are also eigenfunctions of the Fourier transform with

$$(\mathcal{F}q_l)(\cdot) = \sqrt{2\pi} (-i)^{-l} q_l(\cdot).$$

Let us introduce the notation $\tilde{q}_l(\cdot) \equiv (-i)^{-l} q_l(\cdot)$ and $\mathbf{X}_n(t) = (\tilde{q}_{k_n}(tX_1)', \dots, \tilde{q}_{k_n}(tX_n'))'$. Thus, the estimator of f_B given in (2.4) simplifies to $\hat{f}_{Bn}(\cdot) = \underline{q}_{k_n}(\cdot)' \hat{\beta}_n$ where

$$\hat{\beta}_n = \min_{\beta \in \mathbb{R}^{k_n}} \sum_{j=1}^n \int |\hat{\varphi}_n(X_j, t) - \tilde{q}_{k_n}(tX_j)' \beta|^2 \varpi(t) dt. \quad (2.7)$$

An explicit solution of (2.7) is given by

$$\hat{\beta}_n = \left(\int \mathbf{X}_n(t)' \mathbf{X}_n(t) \varpi(t) dt \right)^{-} \int \mathbf{X}_n(t)' \Phi_n(t) \varpi(t) dt$$

where $\Phi_n(t) = (\hat{\varphi}_n(X_1, t), \dots, \hat{\varphi}_n(X_n, t))'$. We emphasize that under the previous assumptions, the matrix $\int \mathbf{X}_n(t)' \mathbf{X}_n(t) \varpi(t) dt$ will be nonsingular with probability approaching one.

2.3.3 Testing Degeneracy

In the following, d_{x_1} and d_{x_2} denote the dimensions of X_1 and X_2 , respectively. We introduce the function $h(\cdot, t) = E[\exp(it(Y - g_2(X, b_2)) | X_1 = \cdot)]$ and a linear sieve space $\mathcal{H}_n \equiv \{\phi : \phi(x_1) = \sum_{j=1}^{k_n} \beta_j p_j(x_1) \text{ for } x_1 \in \mathbb{R}^{d_{x_1}}\}$. The series least squares estimator of h is denoted by $\hat{h}_n(\cdot) = \underline{p}_{k_n}(\cdot)' (\mathbf{X}_{1n}' \mathbf{X}_{1n})^{-1} \mathbf{X}_{1n}' \mathbf{U}_n$ where $\mathbf{U}_n = (\exp(it(Y_1 - g_2(X_1, \hat{b}_{2n}))), \dots, \exp(it(Y_n - g_2(X_n, \hat{b}_{2n}))))'$. Further, let $\tilde{g}(x, t, b) \equiv \exp(itg_2(x, b))$. Below we denote the vector of partial derivatives of $\tilde{g}$ with respect to b by $\tilde{g}_b$.

Assumption 5. (i) *The hypothesis $H_{\text{part-lin}}$ holds.* (ii) *There exists $\Pi_{k_n} h \in \mathcal{H}_n$ such that $n \|\Pi_{k_n} h - h\|_{\varpi}^2 = o(\sqrt{m_n})$.* (iii) *The parameter b_2 is identified in equation (2.5) and belongs to the interior of a compact parameter space $\mathcal{B} \subset \mathbb{R}^{d_{b_2}}$.* (iv) *The function $\tilde{g}$ is partially differentiable with respect to b and $\int E \sup_{b_2 \in \mathcal{B}} \|\tilde{g}_b(X, t, b_2)\|^2 \varpi(t) dt < \infty$.* (v) *It holds $k_n = o(\sqrt{m_n})$.*

Assumption 5 (ii) determines the required asymptotic behavior of the sieve approximation bias for estimating h . This condition ensures that the bias for estimating the function h is asymptotically negligible but does not require undersmoothing of the estimator $\hat{h}_n$. To see

this, let $\|\Pi_{k_n} h - h\|_{\varpi} = O(k_n^{-s/d_{x_1}})$ for some constant $s > 0$. Assumptions 5 (ii) and (v) are satisfied if $m_n \sim n^{\zeta}$ and $k_n \sim n^{\kappa}$ where $d_{x_1}(1 - \zeta/2)/(2s) < \kappa < \zeta/2$. We thus require $\zeta > 2d_{x_1}/(2s + d_{x_1})$ and we may choose κ to balance variance and bias, i.e., $\kappa = d_{x_1}/(2s + d_{x_1})$. In this case, Assumption 4 (ii) automatically holds if $\|\Pi_{m_n} \varphi - \varphi\|_{\varpi} = O(m_n^{-s/d_x})$ and $2d_{x_1} \geq d_x$. If g_2 is linear, Assumption 5 (iv) holds true if $E\|X\|^2 < \infty$ and $\int t^2 \varpi(t) dt < \infty$.

Theorem 12. *Let Assumptions 1–3, 4 (ii), and 5 hold, with $\delta(V, t) = \exp(itY) - h(X_1, t)\tilde{g}(X, t, b_2)$. Then, under H_{deg} we obtain*

$$(\sqrt{2}\varsigma_{m_n})^{-1}(n S_n - \mu_{m_n}) \xrightarrow{d} \mathcal{N}(0, 1).$$

The critical values can be estimated as in Remark 8 but where now $\delta_n(V, t) = \exp(itY) - \hat{h}_n(X_1, t) \exp(itg_2(X, \hat{b}_{2n}))$. The following result shows that, by doing so, the asymptotic distribution of our standardized test statistic remains unchanged. This corollary follows directly from Theorem 12 and the proof of Proposition 9; hence we omit its proof.

Corollary 13. *Under the conditions of Theorem 12 it holds*

$$(\sqrt{2}\hat{\varsigma}_{m_n})^{-1}(n S_n - \hat{\mu}_{m_n}) \xrightarrow{d} \mathcal{N}(0, 1).$$

Remark 14 (Comparison to Andrews [2001]). *It is instructive to compare our setup and results to Andrews [2001], who considers the random coefficient model:*

$$Y = B_0 + B_1 X_1 + (b_2 + \tau \tilde{B}_2) X_2,$$

where $E[B_0 \cdot B_1 | X] = 0$, B_1 is independent of $\tilde{B}_2$, and $E[B_1 | X] = E[\tilde{B}_2 | X] = 0$. In this model, degeneracy of the second random coefficient is equivalent to $\tau = 0$ and degeneracy fails if $\tau > 0$. So under H_{deg} the parameter τ is on the boundary of the maintained hypothesis with $\tau \in [0, \infty)$.

In contrast, we rely in this paper on independence of B to X under the maintained hypothesis. In this case, the hypothesis of degeneracy is equivalent to a conditional characteristic function equation as explained in Example 2 and which is not possible given the assumptions of Andrews [2001]. This is why in our framework we automatically avoid the boundary problem that is apparent in Andrews [2001].

2.4 Consistency against a fixed alternative

In the following, we establish consistency of our test when the difference of restricted and unrestricted conditional characteristic functions does not vanish, i.e., when

$$P\left(\int |\varepsilon(X, t)|^2 \varpi(t) dt \neq 0\right) > 0. \quad (2.8)$$

In case of testing functional form restrictions, this is equivalent to a failure of the null hypothesis H_{mod} . A deviation of conditional characteristic functions can be also caused by alternative models with a different structural function (see Example 1). We only discuss the global power for testing functional form restrictions here, but the results for testing degeneracy follow analogously (of course, in this case we have to be more restrictive about the shape of g_2 as discussed in Example 2). The next proposition shows that our test for testing functional form restrictions has the ability to reject a deviation of restricted and unrestricted conditional characteristic functions with probability one as the sample size grows to infinity.

Proposition 15. *Let Assumptions 1–4 be satisfied. Consider a sequence $(\gamma_n)_{n \geq 1}$ satisfying $\gamma_n = o(n\varsigma_{m_n}^{-1})$. Then, under (2.8) we have*

$$P\left((\sqrt{2}\widehat{\varsigma}_{m_n})^{-1}(n S_n - \widehat{\mu}_{m_n}) > \gamma_n\right) = 1 + o(1).$$

2.5 Asymptotic distribution under local alternatives

We now study the power of our testing procedure against a sequence of linear local alternatives that tends to zero as the sample size tends to infinity. First, we consider deviations from the hypothesis of known functional form restriction. Under H_{mod} , the identified set in the nonseparable model (1.1) is given by $\mathcal{I}_g = \{f : f \text{ is a p.d.f. with } \varphi = \mathcal{F}_g f\}$. We consider the following sequence of local alternatives

$$\varphi = \mathcal{F}_g(f_B + \Delta\sqrt{\varsigma_{m_n}/n}), \quad (2.9)$$

for some function $\Delta \in L^1(\mathbb{R}^{d_b}) \cap L^2(\mathbb{R}^{d_b})$. Here, while $f_B \in \mathcal{I}$, we assume that $f_B + \Delta\sqrt{\varsigma_{m_n}/n}$ does not belong to the identified set $\mathcal{I}$. Equation (2.9) can be written equivalently as

$$E[\exp(itY)|X] = E[\exp(itg(X, B))|X] + \int \exp(itg(X, s))\Delta(s)ds\sqrt{\varsigma_{m_n}/n}.$$

The next result establishes asymptotic normality under (2.9) of the standardized test statistic S_n for testing functional form restrictions.

Proposition 16. *Let the assumptions of Theorem 7 be satisfied. Then, under (2.9) we obtain*

$$(\sqrt{2}\widehat{\varsigma}_{m_n})^{-1}(n S_n - \widehat{\mu}_{m_n}) \xrightarrow{d} \mathcal{N}\left(2^{-1/2}\|\mathcal{F}_g\Delta\|_{\varpi}^2, 1\right).$$

As we see from Proposition 16, our test can detect linear alternatives at the rate $\sqrt{\varsigma_{m_n}/n}$. We now study deviations from the hypothesis of degeneracy under the maintained hypothesis H_{lin} . Under the maintained hypothesis of linearity, any deviance in conditional characteristic functions is equivalent to a failure of a degeneracy of the random coefficients B_2 . Let us denote

$B_{deg} \equiv (B_1, b_2)$ with associated p.d.f. $f_{B_{deg}}$. We consider the following sequence of linear local alternatives

$$f_B(\cdot) = f_{B_{deg}}(\cdot) + \Delta(\cdot)\sqrt{\varsigma_{m_n}/n}, \quad (2.10)$$

for some density function $\Delta \in L^1(\mathbb{R}^{d_b}) \cap L^2(\mathbb{R}^{d_b})$ which is not degenerate at b_2 . Applying the Fourier transform to equation (2.10) yields

$$E[\exp(itX'B)|X] = E[\exp(it(B_0 + X_1'B_1))|X] \exp(itX_2'b_2) + \int \exp(itX's)\Delta(s)ds\sqrt{\varsigma_{m_n}/n}.$$

The next result establishes asymptotic normality under (2.10) for the standardized test statistic S_n for testing degeneracy. This corollary follows by similar arguments used to establish Proposition 16 and hence we omit the proof.

Corollary 17. *Let the assumptions of Theorem 12 be satisfied. Then, under (2.10) we obtain*

$$(\sqrt{2}\widehat{\varsigma}_{m_n})^{-1}(nS_n - \widehat{\mu}_{m_n}) \xrightarrow{d} \mathcal{N}\left(2^{-1/2}\|\mathcal{F}\Delta\|_{\varpi}^2, 1\right).$$

3 Extensions

In this section, we show that our testing procedures can be extended to two different models. First, we consider the class of heterogeneous binary response models. Second, we discuss an extension of linear random coefficient models to system of equations. In both cases, we again discuss testing functional form restrictions and testing degeneracy of some random coefficients separately.

3.1 Binary Response Models

We consider the binary response model

$$Y = \mathbb{1}\{g(X, B) < Z\}, \quad (3.1)$$

where, besides the dependent variable Y and covariates X , a *special regressor* Z is observed as well. In the following, we assume that (X, Z) is independent of B . In contrast to the previous section, the test in the binary response model is based on the difference of a partial derivative of the conditional success probability $P(Y = 1|X, Z)$ and a restricted transformation of the p.d.f. f_B .

Testing functional form restrictions. In the binary response model (3.1), observe that

$$\begin{aligned} P[Y = 1|X = x, Z = z] &= \int \mathbb{1}\{z > g(x, b)\} f_B(b)db \\ &= \int_{-\infty}^z \int_{P_{x,s}} f_B(b)d\nu(b)ds, \end{aligned}$$

where ν is the Lebesgue measure on the lower dimensional hyperplane $P_{x,s} = \{b : g(x, b) = s\}$. Consequently, it holds

$$\psi(x, z) \equiv \partial_z P[Y = 1 | X = x, Z = z] = \int_{P_{x,z}} f_B(b) d\nu(b).$$

Again by considering conditional characteristic functions, the null hypothesis H_{mod} is equivalent to $E[\exp(itY)|X] = E[\exp(itg(X, B))|X]$ for some random coefficient B . By using the above integral representation of ψ we equivalently obtain $(\mathcal{F}\psi(X, \cdot))(t) = (\mathcal{F}_g f_B)(t, X)$ (recall the definition of the integral transform $(\mathcal{F}_g f)(X, t) \equiv \int \exp(itg(X, b))f(b)db$). Due to technical reason, we invert the Fourier transform and conclude that equation (2.1) holds true with

$$\varepsilon(X, z) = \psi(X, z) - (\mathcal{F}^{-1}[(\mathcal{F}_g f_B)(X, \cdot)])(z).$$

In the case of a linear g , the random coefficient density f_B is thus identified through the Radon transform, see also Gautier and Hoderlein [2015].

To estimate the function ε , we replace ψ by a series least squares estimator. Let us introduce the matrix $\mathbf{W}_n = (p_{\underline{m}_n}(X_1, Z_1), \dots, p_{\underline{m}_n}(X_n, Z_n))'$ where the basis function p_l , $l \geq 1$, are assumed to be differentiable with respect to the $(d_x + 1)$ -th entry. We estimate ψ by

$$\hat{\psi}_n(x, z) = \partial_z p_{\underline{m}_n}(x, z)' (\mathbf{W}_n' \mathbf{W}_n)^{-1} \mathbf{Y}_n,$$

where $\mathbf{Y}_n = (Y_1, \dots, Y_n)'$. Consequently, we replace the function ε by

$$\hat{\varepsilon}_n(X_j, z) = \hat{\psi}_n(X_j, z) - (\mathcal{F}^{-1}[(\mathcal{F}_g \hat{f}_{Bn})(X_j, \cdot)])(z),$$

where $\hat{f}_{Bn}$ is the sieve minimum distance estimator given by

$$\hat{f}_{Bn} \in \arg \min_{f \in \mathcal{B}_n} \left\{ \sum_{j=1}^n \int |\hat{\psi}_n(X_j, z) - (\mathcal{F}^{-1}[(\mathcal{F}_g f)(X_j, \cdot)])(z)|^2 \varpi(z) dz \right\}$$

and $\mathcal{B}_n = \{\phi(\cdot) = \sum_{l=1}^{k_n} \beta_l q_l(\cdot)\}$. Our test statistic is $S_n = n^{-1} \sum_{j=1}^n \int |\hat{\varepsilon}_n(X_j, z)|^2 \varpi(z) dz$ where, in this section, ϖ is an integrable weighting function on the support of Z .

We introduce an m_n dimensional linear sieve space $\Psi_n \equiv \{\phi : \phi(x, z) = \sum_{j=1}^{m_n} \beta_j p_l(x, z)\}$. Let $p_{\underline{m}_n}(X, Z)$ be a tensor-product of vectors of basis functions $p_{\underline{m}_{n_1}}(X)$ and $p_{\underline{m}_{n_2}}(Z)$ for integers m_{n_1} and m_{n_2} with $m_n = m_{n_1} \cdot m_{n_2}$. We assume that $\partial_z p_{\underline{m}_{n_2}}(z) = (p_0(z), 2p_1(z), \dots, m_{n_2} p_{m_{n_2}-1}(z))'$. Further, let τ_l denote the squared integer that is associated with $\partial_z p_l$. In Definition 6, $p_l(X)$ has to be replaced by $\tau_l p_l(X, Z)$. Let $B_n = \int E[(\mathcal{F}^{-1}[(\mathcal{F}_g q_{\underline{k}_n})(X, \cdot)])(z) (\mathcal{F}^{-1}[(\mathcal{F}_g q_{\underline{k}_n})(X, \cdot)](z))'] \varpi(z) dz$, which is denoted by $\hat{B}_n$ when the expectation is replaced by the sample mean.

Assumption 6. (i) The random vector (X, Z) is independent of B . (ii) For any p.d.f. f_B satisfying $\mathcal{F}\psi = \mathcal{F}_g f_B$ there exists $\Pi_{k_n} f_B \in \Psi_n$ such that $n \|\psi - \mathcal{F}^{-1} \mathcal{F}_g \Pi_{k_n} f_B\|_{\varpi}^2 = o(\sqrt{m_n})$. (iii) There exists $\Pi_{m_n} \psi \in \Psi_n$ such that $n \|\Pi_{m_n} \psi - \psi\|_{\varpi}^2 = o(\sqrt{m_n})$. (iv) It holds $B_n^- = O(1)$ and $P(\text{rank}(B_n) = \text{rank}(\hat{B}_n)) = 1 + o(1)$. (v) It holds $k_n = o(\sqrt{m_n})$ and $m_n^2 (\log n) \sum_{l=1}^{m_n} \tau_l = o(n \lambda_n)$.

Assumption 6 is similar to Assumption 4. Note that due to the partial derivatives of the basis functions we need to be more restrictive about the dimension parameter m_n , which is captured in Assumption 6 (iv). The following result establishes the asymptotic distribution of our test statistic under H_{mod} in the binary response model (3.1).

Proposition 18. *Let Assumptions 2, 3, and 6 hold with $\delta(Y, X, Z) = Y - \int \mathbb{1}\{Z > X'b\} f_B(b) db$. Then, under H_{mod} we have*

$$(\sqrt{2\varsigma_{m_n}})^{-1}(n S_n - \mu_{m_n}) \xrightarrow{d} \mathcal{N}(0, 1).$$

The critical values can be estimated as in Remark 8 but where now $\delta_n(Y, X, Z) = Y - \int \mathbb{1}\{Z \geq X'b\} \hat{f}_{Bn}(b) db$.

Testing degeneracy. To keep the presentation simple, we only consider the linear case in the following. Under H_{lin} , the binary response model (3.1) simplifies to

$$Y = \mathbb{1}\{X'B < Z\}. \quad (3.2)$$

The null hypothesis H_{deg} can thus be written as

$$\int \exp(itz) \psi(X, z) dz = \int \exp(itz) \psi(X_1, z - X_2'b_2) dz.$$

By nonsingularity of the Fourier transform, we conclude that H_{deg} is equivalent to equation (2.1) where

$$\varepsilon(X, z) = \psi(X, z) - \psi(X_1, z - X_2'b_2).$$

If ψ only depends on X_1 , we consider the estimator $\hat{\psi}_{1n}(x_1, z) = \partial_z p_{\underline{k}_n}(x_1, z)' (\mathbf{W}'_{n1} \mathbf{W}_{n1})^{-1} \mathbf{Y}_n$, where $\mathbf{W}_{n1} = (p_{\underline{k}_n}(X_{11}, Z_1), \dots, p_{\underline{k}_n}(X_{1n}, Z_n))'$. We propose a minimum distance estimator of b_2 given by

$$\hat{b}_{2n} = \operatorname{argmin}_{\beta \in \mathcal{B}} \sum_{j=1}^n \int \left| \hat{\psi}_n(X_j, t) - \hat{\psi}_{1n}(X_{1j}, t - \beta' X_2) \right|^2 \varpi(t) dt. \quad (3.3)$$

Consequently, we estimate the function ε by $\hat{\varepsilon}_n(X_j, z) = \hat{\psi}_n(X_j, z) - \hat{\psi}_{1n}(X_{1j}, z - \hat{b}_{2n} X_{2j})$.

Proposition 19. *Let Assumptions 2, 3, 6 (i), (iii), (v) with $\delta(Y, X, Z) = Y - P(Y = 1|X_1, Z - X_2'b_2)$ hold true. Assume that $n \int E|(\Pi_{\underline{k}_n} \psi)(X_1, z) - \psi(X_1, z)|^2 \varpi(z) dz = o(\sqrt{m_n})$. Then, under H_{deg} we have*

$$(\sqrt{2\varsigma_{m_n}})^{-1}(n S_n - \mu_{m_n}) \xrightarrow{d} \mathcal{N}(0, 1).$$

The critical values can be estimated as in Remark 8 by replacing $P(Y = 1|X_1, Z - X_2'b_2)$ by a series least squares estimator.

3.2 Application to Systems of Equations

In this subsection, we apply our testing procedure to systems of equations, i.e., situations in which the endogenous variable is not a scalar, but a vector. For simplicity, we consider in the following only the bivariate case. Formally, we consider the model

$$Y = g(X, B), \quad (3.4)$$

for some function g and $Y \in \mathbb{R}^2$. Again the vector of random coefficients $B = (B_0, B_1, B_2, B_3)$ is assumed to be independent of the covariates X .

Testing functional form restrictions. Null hypothesis H_{mod} is equivalent to equation (2.1) with

$$\varepsilon(X, t) = E[\exp(it'Y) - \exp(it'g(X, B))|X]$$

for some $t \in \mathbb{R}^2$. Our test of H_{mod} is now based on $S_n \equiv n^{-1} \sum_{j=1}^n \int |\hat{\varepsilon}_n(X_j, t)|^2 \varpi(t) dt$ where $\hat{\varepsilon}_n$ is the estimator of ε introduced in Example 3 but with a multivariate index t and ϖ being a weighting function on $\mathbb{R}^2$. Under a slight modification of assumptions required for Theorem 7, asymptotic normality of the standardized test statistic S_n follows under H_{mod} .

Testing degeneracy. In the partially linear case (i.e., $H_{\text{part-lin}}$ holds), the random coefficient model (3.4) simplifies to

$$\begin{aligned} Y_1 &= B_0 + B'_{11}X_1 + B'_{12}X_2 \\ Y_2 &= B_2 + B'_{31}X_1 + g_2(X_2, B_{32}). \end{aligned}$$

This model is identified if B_{32} is degenerate (see Hoderlein et al. [2014]). A test for degeneracy of $H_{\text{deg}} : B_{32} = b$, for some non-stochastic vector b , uses only the second equation, i.e.,

$$E[\exp(itY_2)|X] = E[\exp(it(B_2 + B'_{31}X_1))|X_1] \exp(itg(X_{32}, b)).$$

We can consequently use the testing methodology developed in Section 2.3.3.

4 Monte Carlo Experiments

In this section, we study the finite-sample performance of our test by presenting the results of a Monte Carlo simulation. The experiments use a sample size of 500 and there are 1000 Monte Carlo replications in each experiment. As throughout the paper, we structure this section again in a part related to testing functional form restrictions, and a part related to testing degeneracy.

4.1 Testing Functional Form Restrictions

In each experiment, we generate realizations of regressors X from $X \sim \mathcal{N}(0, 2)$ and random coefficients $B = (B_1, B_2)'$ from $B \sim \mathcal{N}(0, A)$ where

$$A = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}.$$

We simulate a random intercept $B_0 \perp (B_1, B_2)$ according to the standard normal distribution. Realizations of the dependent variable Y are generated either by the linear model

$$Y = \eta B_0 + X B_1, \quad (4.1)$$

the quadratic model

$$Y = \eta B_0 + X B_1 + X^2 B_2, \quad (4.2)$$

or the nonlinear model

$$Y = \eta B_0 + X B_1 + \sqrt{|X|} B_2, \quad (4.3)$$

where the constant η is either $\sqrt{0.5}$ or 1. Note that the random coefficient density f_B is neither point identified in model (4.2) nor in model (4.3). However, recall that even if the model is not point identified under the maintained hypothesis, our testing procedure may still be able to detect certain failures of the null hypothesis, in particular if they arise from differences in conditional moments. Consider, for example, a test of linearity of the heterogeneous QUAIDS model (4.2), where the first two conditional moments yield $E[B_2] = 0$ and $2 \text{Cov}(B_0, B_2) + 2x \text{Cov}(B_1, B_2) + x^2 \text{Var}(B_2) = 0$, for all x . Consequently, $P(\int |\varepsilon(X, t)|^2 \varpi(t) dt \neq 0) > 0$ if and only if $P(B_2 \neq 0) > 0$. We also observe in the finite sample experiment that our testing procedure is able to detect such deviations.

The test is implemented using Hermite functions, and uses the standardization described in Remark 8. When (4.1) is the true model, we estimate the random coefficient density as described in Example 11, where we make use of the fact that the hermite functions are eigenfunctions of the Fourier transform. If (4.2) is the true model, the integral transform $\mathcal{F}_g$ is computed using numerical integration. In both cases, the weighting function ϖ coincides with the standard normal p.d.f.. If (4.1) is the correct model, we use $k_n = 4 (= 3 + 2)$ Hermite functions to estimate the density of the bivariate random coefficients (B_0, B_1) and let $m_n = 9$. If (4.2) is the correct model, we have an additional dimension which accounts for the nonlinear part. Here, the choice of Hermite basis functions is $k_n = 7 (= 3 + 2 \cdot 2)$ with $m_n = 12$ if $\eta = \sqrt{0.5}$ and $k_n = 9 (= 3 + 2 \cdot 3)$ with $m_n = 16$ if $\eta = 1$. We thus increase the dimension parameters k_n and m_n as the noise level η becomes larger, i.e., the model becomes more complex. Note that m_n could be any integer larger than $\text{const.} \times k_n^2$ that is smaller than $n^{1/2}$ (up to logs).

<i>Null Model</i>	<i>Alt. Model</i>	η	<i>Empirical Rejection probabilities at level</i>		
H_{mod}	<i>True DGP</i>		<i>0.010</i>	<i>0.050</i>	<i>0.100</i>
(4.1)		0.5	0.003	0.027	0.076
(4.2)			0.008	0.034	0.072
(4.1)	(4.2)		0.698	0.911	0.958
(4.1)	(4.3)		0.178	0.491	0.683
(4.2)	(4.1)		0.714	0.928	0.980
(4.2)	(4.3)		0.864	0.982	0.994
(4.1)		1	0.007	0.047	0.127
(4.2)			0.007	0.058	0.154
(4.1)	(4.2)		0.491	0.804	0.907
(4.1)	(4.3)		0.102	0.368	0.615
(4.2)	(4.1)		0.557	0.875	0.970
(4.2)	(4.3)		0.558	0.905	0.978

Table 1: Rows 1,2,7,8 depict the empirical rejection probabilities if H_{mod} holds true, the rows 3–6 and 9–12 show the finite sample power of our tests against various alternatives. The first column states the null model while the second shows the alternative model and is left empty if the null model is the correct model. Column 3 specifies the noise level of the data generating process. Columns 4–6 depict the empirical rejection probabilities for different nominal levels.

In practice, we let k_n such that it minimizes the value of the test statistic. I.e., if $s(k_n, m_n)$ denotes the value of the test statistic, a guideline for parameter choice in practice is given by the minimum-maximum principle $\min_{1 \leq k_n < n^{1/4}} \max_{k_n^2 < m_n < n^{1/2}} \{s(k_n, m_n)\}$.

The empirical rejection probabilities of our tests are shown in Table 1 at nominal levels 0.010, 0.050, and 0.100. We also note that the models are normalized and hence, the null and alternative have the same variance. The differences between the nominal and empirical rejection probabilities are small under the correct functional form restrictions, as is obvious from rows 1, 2, 7, and 8. Comparing the empirical rejections probabilities in rows 3–6 and 9–12, we see our tests become less powerful as the parameter η increases, as was to be expected. On the other hand, we observe from this table that our tests have power to detect nonlinear alternatives even in cases where the model under the maintained hypothesis is not identified. This is in line with our observation that these alternatives imply deviations between the restricted and unrestricted characteristic functions. Comparing rows 3, 9 with 4, 10 in Table 1, we observe that our test rejects the quadratic model (4.2) more often than the nonlinear model (4.3). From

rows 5, 11 and 6, 12 we see that our test rejects the nonlinear model (4.3) slightly more often than the linear model (4.1).

We have also tried different data generating processes, such as a cubic polynomial with random coefficients. In this case, our test of linearity led to empirical rejection probabilities which were close to one for all nominal levels considered and hence these results are not reported here. Regarding consistency, we also conducted experiments with larger sample sizes. In particular, we saw that the slight tendency of our test statistic to under-reject for small η , see in Table 1 in rows 1 and 2, diminishes as we increase the sample size to $n = 1000$. Not surprisingly, when $n = 1000$ also the empirical rejection probabilities in rows 3–6 and 9–12 increase.

4.2 Testing Degeneracy

In each experiment, we generate realizations of X from $X \sim \mathcal{N}(0, A)$ and random coefficients $B = (B_1, B_2)'$ from $B \sim \mathcal{N}(0, A_\rho)$, where

$$A = \begin{pmatrix} 1 & .5 \\ .5 & 1 \end{pmatrix} \quad \text{and} \quad A_\rho = \begin{pmatrix} 2 & \rho \\ \rho & 2 \end{pmatrix},$$

for some constant $\rho > 0$, which varies in the experiments. Further, we generate the dependent variable Y as

$$Y = B_0 + B_1 X_1 + X_2,$$

if the null hypothesis H_{deg} holds. For the alternative, we generate the dependent variable Y using

$$Y = B_0 + B_1 X_1 + \eta B_2 X_2,$$

for some constant $\eta > 0$, which varies in the simulations below.

The test is implemented as described in Example 4 with B-splines, and uses the standardization described in Remark 8 with $\delta_n(V, t) = \exp(itY) - \widehat{h}_n(X_1, t) \exp(itg_2(X, \widehat{b}_{2n}))$. To estimate the restricted conditional characteristic function, we use B-splines of order 2 with one knot (hence, $k_n = 4$), and for the unrestricted one a tensor-product of this B-spline basis functions and a quadratic polynomial (hence, $m_n = 12$). In practice, we may employ the minimum-maximum principle for parameter choice, as described in the previous subsection.

The empirical rejection probabilities for testing degeneracy are shown in Table 2 at nominal levels 0.010, 0.050, and 0.100. Again we normalize the models to ensure that the null and alternative have the same variance. The differences between the nominal and empirical rejection probabilities are small under fixed coefficient for X_2 , as is obvious from rows 1 and 5. From Table 2, we also see from rows 2–4 or 6–8 that our test rejects the alternative model more

<i>Alt. Model</i>		<i>Empirical Rejection probabilities at level</i>		
η	ρ	<i>0.010</i>	<i>0.050</i>	<i>0.100</i>
	1	0.003	0.048	0.110
0.75		0.033	0.177	0.352
1		0.142	0.479	0.837
1.25		0.378	0.738	0.876
	1.5	0.003	0.050	0.105
0.75		0.118	0.396	0.643
1		0.367	0.741	0.837
1.25		0.627	0.904	0.967

Table 2: Rows 1 and 5 depict the empirical rejection probabilities under degeneracy of the coefficient of X_2 , the rows 2–4 and 6–8 show the finite sample power of our tests against various alternatives. Column 1 depicts the value of η in the correct model and is empty if the null model is correct. Column 2 specifies covariance of B_1 and B_2 . Columns 4–6 depict the empirical rejection probabilities for different nominal levels.

often for a larger variance of B_2 , as we expect. Moreover, the empirical rejection probabilities increase as the covariance of B_1 and B_2 becomes larger, as we see by comparing rows 2–4 with 6–8.

5 Application

5.1 Motivation: Consumer Demand

Heterogeneity plays an important role in classical consumer demand. The most popular class of parametric demand systems is the almost ideal (AI) class, pioneered by Deaton and Muellbauer [1980]. In the AI model, instead of quantities budget shares are being considered and they are being explained by log prices and log total expenditure⁶. The model is linear in log prices and a term that involves log total expenditure linearly but which is divided by a price index that depends on parameters of the utility function. In applications, one frequent shortcut is that the price index is replaced by an actual price index, another is that homogeneity of degree zero is

⁶The use of total expenditure as wealth concept is standard practice in the demand literature and, assuming the existence of preferences, is satisfied under an assumption of separability of the labor supply from the consumer demand decision, see Lewbel [1999].

imposed, which means that all prices and total expenditure are relative to a price index. This step has the beneficial side effect that it removes general inflation as well.

A popular extension in this model allows for quadratic terms in total expenditure (QUAIDS, Banks et al. [1997]). Since we focus in this paper on the budget share for food at home (BSF), which, due at least in parts to satiation effects, is often documented to decline steadily across the total expenditure range, we want to assess whether quadratic terms are really necessary. Note that prices enter the quadratic term in a nonlinear fashion, however, due to the fact that we have very limited price variation, we can treat the nonlinear expression involving prices as a constant. This justifies the use of real total expenditure as regressor, even in the quadratic term. In other words, we thus consider an Engel curve QUAIDS model. Moreover, we want to allow for preference heterogeneity, and hence consider a heterogeneous population model:

$$BSF_i = B_{0i} + B_{1i} \log(TotExp_i) + B_{2i} (\log(TotExp_i))^2 + b_4 W_{1i} + b_5 W_{2i}. \quad (5.1)$$

Unobserved heterogeneity is reflected in the three random coefficients B_{0i} , B_{1i} and B_{2i} . This additive specification can be thought of as letting the mean of the random intercept B_{0i} depend on covariates. To account for observed heterogeneity in preferences, we include in addition household covariates as regressors. Specifically, we use principal components to reduce the vector of remaining household characteristics to a few orthogonal, approximately continuous components. We only use two principal components, denoted W_{1i} and W_{2i} . While including these additional controls in this form is arguably ad hoc, we perform some robustness checks like alternating the component or adding several others, and the results do not change appreciably.

We implement the test statistics as described in the Monte Carlo section. For testing degeneracy, we estimate the conditional characteristic functions as described in Example 4. For testing functional form restrictions, our test is implemented as described in Example 3, where in the linear case we employ the estimation procedure in Example 11. In both cases, we choose the dimension parameters k_n and m_n by the minimum-maximum principle explained in the Monte Carlo section.

5.2 The Data: The British Family Expenditure Survey

The FES reports a yearly cross section of labor income, expenditures, demographic composition, and other characteristics of about 7,000 households. We use years 2008 and 2009. As is standard in the demand system literature, we focus on the subpopulation of two person households where both are adults, at least one is working, and the head of household is a white collar worker. This is to reduce the impact of measurement error; see Lewbel [1999] for a discussion. We thus have a sample of size 543, which is similar to the one considered in the Monte Carlo section.

We form several expenditure categories, but focus on the food at home category. This category contains all food expenditure spent for consumption at home; it is broad since more detailed accounts suffer from infrequent purchases (the recording period is 14 days) and are thus often underreported. Food consumption accounts for roughly 20% of total expenditure. Results actually displayed were generated by considering consumption of food versus nonfood items. We removed outliers by excluding the upper and lower 2.5% of the population in the three groups. We form food budget shares by dividing the expenditures for all food items by total expenditures, as is standard in consumer demand. The following table provides summary statistics of the economically important variables:

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	St. Dev.
Food share	0.008	0.137	0.178	0.188	0.232	0.591	0.075
log(TotExp)	4.207	5.534	5.788	5.782	6.066	6.927	0.448

5.3 Results

For testing degeneracy of the coefficient B_2 , we estimate the coefficient under H_{deg} , i.e., we assume that this coefficient is fixed. The ordinary least squares estimate is -0.009 with standard error 0.008 . This significant role of nonlinearity is also picked up by our procedure. Table 3 shows the different values of the test statistics and p-values at nominal level 0.05 . As we see from Table 3, our test fails to reject the model (5.1) with degenerate B_{2i} but rejects the linear random coefficient model where $B_{2i} = 0$. Unsurprisingly, we also fail to reject the random coefficient QUAIDS model. The dimension parameters k_n and m_n are chosen via the proposed minimum-maximum principle. It is interesting to note that we chose higher order basis functions to account for the random coefficient of the quadratic term. This also supports the hypothesis that the marginal p.d.f. B_2 is close to the Dirac measure.

Null Hypothesis	linear RC	quadratic RC	RC with fixed coeff. on quadratic term in TotExp
value of test	2.1289	1.4200	0.6551
p-values	0.0166	0.0778	0.2562

Table 3: *Values of the tests with p-values when null hypothesis is either a linear random coefficient model (i.e., $B_{2i} = 0$ in (5.1)), a quadratic random coefficient model (i.e., random B_{2i} in (5.1)), or a random coefficient model with degenerate coefficient on the quadratic term (i.e., $B_{2i} = b_2$ in (5.1) for some fixed b_2).*

The analysis thus far assumes that total expenditure is exogenous. However, in consumer demand it is commonly thought that log total expenditure is endogenous and is hence instrumented for, typically by labor income, say Z , see Lewbel [1999]. One might thus argue that we reject our hypotheses not due to a failure to functional form restrictions, but because of a violation of exogeneity of total expenditure. Therefore, we follow Imbens and Newey [2009], and model the endogeneity through a structural heterogeneous equation that relates total expenditure X to the instrument labor income Z , i.e.,

$$X = \psi(Z, U),$$

where U denotes a scalar unobservable. Following Imbens and Newey [2009], we assume that the instrument Z is exogenous, i.e., we assume $Z \perp (B, U)$, implying $X \perp B|U$, and we assume that the function ψ is strictly monotonic in U . Finally, we employ the common normalization that $U|Z$ is uniformly distributed on the unit interval $[0, 1]$. Then, the disturbance U is identified through the conditional cumulative distribution function of X given Z , i.e.,

$$U = F_{X|Z}(X|Z).$$

Since $X \perp B|U$, we then simply modify our testing procedure by additionally conditioning on controls U . In the consumer demand literature, this control function approach was also considered by Hoderlein [2011]. However, since the theory is outside of the scope of this paper, we do not adjust for estimation error in this variable depending on the smoothness assumptions one is willing to impose in this step, which may lead to a higher variance.

The results of this modification are summarized in Table 4. As we see from this table, the

Null Hypothesis	linear RC	quadratic RC	RC with fixed coeff. on quadratic term in TotExp
value of test	2.0661	1.3978	0.2304
p-values	0.0194	0.0810	0.4089

Table 4: *Values of the test statistics with p-values, when additionally corrected for endogeneity.*

value of the modified test statistics are smaller, once we introduce the instrument Z in a control function approach. This possibly indicates that there is some endogeneity bias in the first case; however, our main conclusions remain unchanged: We soundly reject the linear RC model, and fail to reject H_{deg} and H_{mod} .

6 Conclusion

This paper develops nonparametric specification testing for random coefficient models. We employ a sieve strategy to obtain tests for both the functional form of the structural equation,

e.g., for linearity in random parameters, as well as for the important question of whether or not a parameter can be omitted. While the former can be used to distinguish between various models, including such models where the density of random coefficients is not necessarily point identified, the latter types of test reduce the dimensionality of the random coefficients density. From a nonparametric perspective, this is an important task, because random coefficient models are known to suffer from very slow rates of convergence, see Hoderlein et al. [2010]. We establish the large sample behavior of our test statistics, and show that our tests work well in a finite sample experiment, and allow to obtain reasonable results in a consumer demand application.

Mathematical Appendix

Throughout the proofs, we will use $C > 0$ to denote a generic finite constant that may be different in different uses. We use the notation $a_n \lesssim b_n$ to denote $a_n \leq Cb_n$ for all $n \geq 1$. Further, for ease of notation we write $\sum_j$ for $\sum_{j=1}^n$. Recall that $\|\cdot\|$ denotes the usual Euclidean norm, while for a matrix A , $\|A\|$ is the operator norm. Further, $\|\phi\|_X \equiv \sqrt{E(\phi^2(X))}$ and $\langle \phi, \psi \rangle_X \equiv E[\phi(X)\psi(X)]$. Recall the notation $P_n = E[p_{\underline{m}_n}(X)p_{\underline{m}_n}(X)']$.

Proofs of Section 2.

PROOF OF PROPOSITION 5. Let us denote $f = \frac{dF_X}{d\nu}$. For some constant $0 < c < 1$, for all $n \geq 1$, and any $a_n \in \mathbb{R}^{m_n}$ we have

$$\begin{aligned} \|a_n\|^2 &= \int (a'_n p_{\underline{m}_n}(x))^2 \mathbb{1}\{f(x) \geq \lambda_n\} \nu(dx) + \int (a'_n p_{\underline{m}_n}(x))^2 \mathbb{1}\{f(x) < \lambda_n\} \nu(dx) \\ &\leq \lambda_n^{-1} \int (a'_n p_{\underline{m}_n}(x))^2 f(x) \nu(dx) + c \int (a'_n p_{\underline{m}_n}(x))^2 \nu(dx). \end{aligned}$$

Consequently,

$$C\lambda_n I_{m_n} \leq P_n$$

where I_{m_n} denotes the $m_n \times m_n$ dimensional identity matrix. □

By Assumption 2, the eigenvalues of $\lambda_n^{-1}P_n$ are bounded away from zero and hence, it may be assumed that $P_n = \lambda_n I_{m_n}$. Otherwise, consider a linear transformation of $p_{\underline{m}_n}$ of the form $\tilde{p}_{\underline{m}_n} \equiv (P_n/\lambda_n)^{-1/2} p_{\underline{m}_n}$ where $\sup_{x \in \mathcal{X}} \|\tilde{p}_{\underline{m}_n}(x)\| \leq Cm_n$ using that the smallest eigenvalue of $(P_n/\lambda_n)^{-1/2}$ is bounded away from zero uniformly in n .

Lemma 6.1. *It holds $\varsigma_{m_n} \geq C\sqrt{m_n}$.*

Proof. Without loss of generality it may be assumed that $\int \varpi(t)dt = 1$. By the definition of ς_{m_n} we conclude

$$\begin{aligned}
\varsigma_{m_n}^2 &\geq \lambda_n^{-2} \sum_{l=1}^{m_n} \int \int \left| E[\delta(V, s) \overline{\delta(V, t)} p_l^2(X)] \right|^2 \varpi(s) \varpi(t) ds dt \\
&\geq \lambda_n^{-2} \sum_{l=1}^{m_n} \left(E[|\int \delta(V, t) \varpi(t) dt|^2 p_l^2(X)] \right)^2 && \text{(by Jensen's inequality)} \\
&\geq C \lambda_n^{-2} \sum_{l=1}^{m_n} (E[p_l^2(X)])^2 && \text{(by Assumption 3)} \\
&= C m_n.
\end{aligned}$$

□

In the following, we make use of the notations $\widehat{P}_n = n^{-1} \sum_j p_{\underline{m}_n}(X_j) p_{\underline{m}_n}(X_j)'$ and $\widehat{\gamma}_n(t) \equiv (n \widehat{P}_n)^{-1} \sum_j \exp(it Y_j) p_{\underline{m}_n}(X_j)$. Let $\widehat{A}_n = n^{-1} \sum_j \int (\mathcal{F}_g q_{\underline{k}_n})(X_j, t) (\mathcal{F}_g q'_{\underline{k}_n})(X_j, t) \varpi(t) dt$ and $A_n = E[\int (\mathcal{F}_g q_{\underline{k}_n})(X, t) (\mathcal{F}_g q'_{\underline{k}_n})(X, t) \varpi(t) dt]$. Recall $\widehat{\beta}_n = (n \widehat{A}_n)^{-1} \sum_j \int (\mathcal{F}_g q_{\underline{k}_n})(X_j, t) \widehat{\varphi}_n(X_j, t) \varpi(t) dt$ and let $\beta_n = A_n^{-1} \int E[(\mathcal{F}_g q_{\underline{k}_n})(X, t) \varphi(X, t)] \varpi(t) dt$.

PROOF OF THEOREM 7. We make use of the decomposition

$$\begin{aligned}
n S_n &= \sum_j \int |\widehat{\varepsilon}_n(X_j, t)|^2 \varpi(t) dt \\
&= \sum_j \int |p_{\underline{m}_n}(X_j)' \widehat{\gamma}_n(t) - \Pi_{m_n} \varphi(X_j, t)|^2 \varpi(t) dt \\
&\quad + 2 \sum_j \int (p_{\underline{m}_n}(X_j)' \widehat{\gamma}_n(t) - \Pi_{m_n} \varphi(X_j, t)) (\Pi_{m_n} \varphi(X_j, t) - (\mathcal{F}_g \widehat{f}_{B_n})(X_j, t)) \varpi(t) dt \\
&\quad + \sum_j \int |\Pi_{m_n} \varphi(X_j, t) - (\mathcal{F}_g \widehat{f}_{B_n})(X_j, t)|^2 \varpi(t) dt \\
&= I_n + 2 II_n + III_n \quad (\text{say}).
\end{aligned}$$

Consider I_n . We conclude

$$\begin{aligned}
I_n &= n \int \left(\widehat{\gamma}_n(t) - \langle \varphi(\cdot, t), p_{\underline{m}_n} \rangle_X \right)' \widehat{P}_n \left(\widehat{\gamma}_n(t) - \langle \varphi(\cdot, t), p_{\underline{m}_n} \rangle_X \right) \varpi(t) dt \\
&= n^{-1} \int \left(\sum_j \left(\exp(itY_j) - \Pi_{m_n} \varphi(X_j, t) \right) p_{\underline{m}_n}(X_j) \right)' \widehat{P}_n^{-1} \\
&\quad \times \left(\sum_j \left(\exp(itY_j) - \Pi_{m_n} \varphi(X_j, t) \right) p_{\underline{m}_n}(X_j) \right) \varpi(t) dt \\
&= \lambda_n^{-1} \int \left\| n^{-1/2} \sum_j \left(\exp(itY_j) - \Pi_{m_n} \varphi(X_j, t) \right) p_{\underline{m}_n}(X_j) \right\|^2 \varpi(t) dt \\
&\quad + n^{-1} \int \left(\sum_j \left(\exp(itY_j) - \Pi_{m_n} \varphi(X_j, t) \right) p_{\underline{m}_n}(X_j) \right)' (\widehat{P}_n^{-1} - \lambda_n^{-1} I_n) \\
&\quad \times \left(\sum_j \left(\exp(itY_j) - \Pi_{m_n} \varphi(X_j, t) \right) p_{\underline{m}_n}(X_j) \right) \varpi(t) dt \\
&= B_{1n} + B_{2n} \quad (\text{say}).
\end{aligned}$$

Since $(\Pi_{m_n} \varphi(X, t) - \varphi(X, t)) p_{\underline{m}_n}(X)$ is a centered random variable for all t it is easily seen that $B_{1n} = \lambda_n^{-1} \int \left\| n^{-1/2} \sum_j \left(\exp(itY_j) - \varphi(X_j, t) \right) p_{\underline{m}_n}(X_j) \right\|^2 \varpi(t) dt + o_p(1)$. Thus, Lemma 6.2 yields $(\sqrt{2}\zeta_{m_n})^{-1} (B_{1n} - \mu_{m_n}) \xrightarrow{d} \mathcal{N}(0, 1)$. To show that $B_{2n} = o_p(\sqrt{m_n})$ note that

$$\|\widehat{P}_n^{-1} - \lambda_n^{-1} I_n\| \leq \lambda_n^{-1} \|(\widehat{P}_n / \lambda_n)^{-1}\| \|I_n - \widehat{P}_n / \lambda_n\| = O_p\left(\sqrt{(m_n \log n) / (n \lambda_n^2)}\right)$$

by Lemma 6.2 of Belloni et al. [2015]. Further, from $E[(\exp(itY) - \Pi_{m_n} \varphi(X, t)) p_l(X)] = 0$, $1 \leq l \leq m_n$, we deduce

$$\begin{aligned}
&n^{-1} \int E \left\| \sum_j \left(\exp(itY_j) - \Pi_{m_n} \varphi(X_j, t) \right) p_{\underline{m}_n}(X_j) \right\|^2 \varpi(t) dt \\
&\lesssim \int \varpi(t) dt E \|p_{\underline{m}_n}(X)\|^2 + \sup_{x \in \mathcal{X}} \|p_{\underline{m}_n}(x)\|^2 \sum_{l=1}^{m_n} \int \langle \varphi(\cdot, t), p_l \rangle_X^2 \varpi(t) dt E[p_l^2(X)] \\
&\lesssim m_n \lambda_n.
\end{aligned} \tag{6.1}$$

The result follows due to condition $m_n^2 \log n = o(n \lambda_n)$. Thereby, it is sufficient to prove $II_n + III_n = o_p(\sqrt{m_n})$. Consider III_n . We observe

$$III_n \lesssim \sum_j \int |\mathcal{F}_g(\widehat{f}_{Bn} - \Pi_{k_n} f_B)(X_j, t)|^2 \varpi(t) dt + \sum_j \int |(\mathcal{F}_g \Pi_{k_n} f_B)(X_j, t) - \Pi_{m_n} \varphi(X_j, t)|^2 \varpi(t) dt,$$

where $\sum_j \int |(\mathcal{F}_g \Pi_{k_n} f_B)(X_j, t) - \Pi_{m_n} \varphi(X_j, t)|^2 \varpi(t) dt = o_p(\sqrt{m_n})$ and

$$\begin{aligned}
\sum_j \int |\mathcal{F}_g(\widehat{f}_{Bn} - \Pi_{k_n} f_B)(X_j, t)|^2 \varpi(t) dt &= (\widehat{\beta}_n - \beta_n)' \sum_j \int (\mathcal{F}_g q_{\underline{k}_n})(X_j, t) (\mathcal{F}_g q_{\underline{k}_n})(X_j, t)' \varpi(t) dt (\widehat{\beta}_n - \beta_n) \\
&= n(\widehat{\beta}_n - \beta_n)' \widehat{A}_n (\widehat{\beta}_n - \beta_n).
\end{aligned}$$

Let us introduce the vector $\tilde{\beta}_n = (n\hat{A}_n^-) \sum_j \int (\mathcal{F}_g q_{k_n})(X_j, t) \varphi(X_j, t) \varpi(t) dt$. Using the property of Moore-Penrose inverses that $\hat{A}_n = \hat{A}_n \hat{A}_n^- \hat{A}_n$, we conclude

$$\begin{aligned} n(\hat{\beta}_n - \beta_n)' \hat{A}_n (\hat{\beta}_n - \beta_n) &\lesssim n(\hat{\beta}_n - \tilde{\beta}_n)' \hat{A}_n (\hat{\beta}_n - \tilde{\beta}_n) + n(\tilde{\beta}_n - \beta_n)' \hat{A}_n (\tilde{\beta}_n - \beta_n) \\ &\lesssim \left\| n^{-1/2} \sum_j \int (\mathcal{F}_g q_{k_n})(X_j, t) (\hat{\varphi}_n(X_j, t) - \varphi(X_j, t)) \varpi(t) dt \right\|^2 \|\hat{A}_n^-\| \\ &\quad + \left\| n^{-1/2} \sum_j \int (\mathcal{F}_g q_{k_n})(X_j, t) \varphi(X_j, t) \varpi(t) dt \right\|^2 \|\hat{A}_n^- - A_n^-\|^2 \|\hat{A}_n\| \\ &\quad + \left\| n^{-1/2} \sum_j \int \left((\mathcal{F}_g q_{k_n})(X_j, t) \varphi(X_j, t) - E[(\mathcal{F}_g q_{k_n})(X, t) \varphi(X, t)] \right) \varpi(t) dt \right\|^2 \|A_n^-\|^2 \|\hat{A}_n\| \end{aligned}$$

From Lemma 6.3 we have $\|\hat{A}_n^- - A_n^-\| = O_p(\sqrt{(\log n)k_n/n})$. By Assumption 4 (v) it holds $\|A_n^-\| = O(1)$ and thus, $\|\hat{A}_n^-\| \leq \|\hat{A}_n^- - A_n^-\| + \|A_n^-\| = O_p(1)$. Thereby, it is sufficient to consider

$$\begin{aligned} &\left\| n^{-1/2} \sum_j \int (\mathcal{F}_g q_{k_n})(X_j, t) (\hat{\varphi}_n(X_j, t) - \varphi(X_j, t)) \varpi(t) dt \right\|^2 \\ &\quad \lesssim \left\| n^{-1/2} \sum_j \int (\mathcal{F}_g q_{k_n})(X_j, t) p_{\underline{m}_n}(X_j)' (\hat{\gamma}_n(t) - \langle \varphi(\cdot, t), p_{\underline{m}_n} \rangle_X) \varpi(t) dt \right\|^2 \\ &\quad \quad + \left\| n^{-1/2} \sum_j \int (\mathcal{F}_g q_{k_n})(X_j, t) (\Pi_{m_n} \varphi(X_j, t) - \varphi(X_j, t)) \varpi(t) dt \right\|^2 \\ &\quad \lesssim n \left\| \int E[(\mathcal{F}_g q_{k_n})(X, t) p_{\underline{m}_n}(X)] (\hat{\gamma}_n(t) - \langle \varphi(\cdot, t), p_{\underline{m}_n} \rangle_X) \varpi(t) dt \right\|^2 \\ &\quad \quad + n \left\| \int E[(\mathcal{F}_g q_{k_n})(X, t) (\Pi_{m_n} \varphi(X, t) - \varphi(X, t))] \varpi(t) dt \right\|^2 + O_p(k_n) \\ &\quad = O_p(k_n + n \|\Pi_{m_n} \varphi - \varphi\|_{\varpi}^2) \end{aligned}$$

which can be seen as follows. Let $\langle \cdot, \cdot \rangle_{\varpi}$ denote the inner product induced by the norm $\|\cdot\|_{\varpi}$.

We calculate

$$\begin{aligned} \left\| \int E[(\mathcal{F}_g q_{k_n})(X, t) (\Pi_{m_n} \varphi(X, t) - \varphi(X, t))] \varpi(t) dt \right\|^2 &= \sum_{l=1}^{k_n} \langle \mathcal{F}_g q_l, \Pi_{m_n} \varphi - \varphi \rangle_{\varpi}^2 \\ &= \sum_{l=1}^{k_n} \left(\int q_l(b) E[(\mathcal{F}_g^* (\Pi_{m_n} \varphi - \varphi))(X, b)] db \right)^2 \\ &\lesssim \int \left(E[(\mathcal{F}_g^* (\Pi_{m_n} \varphi - \varphi))(X, b)] \right)^2 db \\ &\lesssim \|\Pi_{m_n} \varphi - \varphi\|_{\varpi}^2 \end{aligned}$$

where $\mathcal{F}_g^*$ is the adjoint operator of $\mathcal{F}_g$ given by $(\mathcal{F}_g^* \phi)(b) = \int E[\exp(itg(X, b)) \phi(X, t)] \varpi(t) dt$. Consequently, we have $n(\hat{\beta}_n - \beta_n)' \hat{A}_n (\hat{\beta}_n - \beta_n) = O_p(k_n + n \|\Pi_{m_n} \varphi - \varphi\|_{\varpi}^2) = o_p(\sqrt{m_n})$ and,

in particular, $III_n = o_p(\sqrt{m_n})$. Consider II_n . From above we infer $n\|\hat{\beta}_n - \beta_n\|^2 = O_p(k_n + n\|II_{m_n}\varphi - \varphi\|_\varpi^2)$ by employing that $\hat{A}_n^-$ is stochastically bounded. Thereby, we obtain

$$\begin{aligned}
|II_n|^2 &\lesssim \left| \int \sum_j (\exp(itY_j) - II_{m_n}\varphi(X_j, t)) p_{\underline{m_n}}(X_j)' \langle \varphi(\cdot, t) - (\mathcal{F}_g II_{k_n} f_B)(\cdot, t), p_{\underline{m_n}} \rangle_X \varpi(t) dt \right|^2 \\
&\quad + \left| \int \sum_j (\exp(itY_j) - II_{m_n}\varphi(X_j, t)) p_{\underline{m_n}}(X_j)' \langle (\mathcal{F}_g(II_{k_n} f_B - \hat{f}_{Bn}))(\cdot, t), p_{\underline{m_n}} \rangle_X \varpi(t) dt \right|^2 + o_p(m_n) \\
&\lesssim n \int E \left| II_{m_n}\varphi(X, t) - (II_{m_n} \mathcal{F}_g II_{k_n} f_B)(X, t) \right|^2 \varpi(t) dt \\
&\quad + n \int E \left\| (\exp(itY) - II_{m_n}\varphi(X, t)) (II_{m_n} \mathcal{F}_g q_{\underline{k_n}})(X, t) \right\|^2 \varpi(t) dt \|\hat{\beta}_n - \beta_n\|^2 + o_p(m_n) \\
&= O_p\left(n\|\mathcal{F}_g(II_{k_n} f_B - f_B)\|_\varpi^2 + k_n(k_n + n\|II_{m_n}\varphi - \varphi\|_\varpi^2)\right) + o_p(m_n) \\
&= o_p(m_n)
\end{aligned}$$

where we used that $\|II_{m_n}\varphi - \varphi\|_\infty = O(1)$ and $\sum_{l=1}^{k_n} \|II_{m_n} \mathcal{F}_g q_l\|_\varpi^2 = O(k_n)$, which completes the proof. $\square$

We require the following notation. Let us introduce the covariance matrix estimator $\hat{\Sigma}_{m_n}(s, t) = n^{-1} \sum_j p_{\underline{m_n}}(X_j) p_{\underline{m_n}}(X_j)' \delta_n(V_j, s) \delta_n(V_j, t)$ where $\delta_n(V_j, s) = \exp(itY) - (\mathcal{F}_g \hat{f}_{Bn})(X, t)$. Further, we define $\tilde{\delta}_n(V, t) = \exp(itY) - (\mathcal{F}_g II_{k_n} f_B)(X, t)$ and introduce the matrix $\tilde{\Sigma}_{m_n}(s, t) = n^{-1} \sum_j p_{\underline{m_n}}(X_j) p_{\underline{m_n}}(X_j)' \tilde{\delta}_n(V_j, s) \tilde{\delta}_n(V_j, t)$.

PROOF OF PROPOSITION 9. To keep the presentation of the proof simple, we do not consider estimation of P_n in $\hat{\zeta}_{m_n}$ and $\hat{\mu}_{m_n}$. We make use of the relationship

$$\begin{aligned}
\delta_n(\cdot, s) \delta_n(\cdot, t) - \tilde{\delta}_n(\cdot, s) \tilde{\delta}_n(\cdot, t) &= \tilde{\delta}_n(\cdot, s) ((\mathcal{F}_g \hat{f}_{Bn})(\cdot, t) - (\mathcal{F}_g II_{k_n} f_B)(\cdot, t)) \\
&\quad + \delta_n(\cdot, t) ((\mathcal{F}_g \hat{f}_{Bn})(\cdot, s) - (\mathcal{F}_g II_{k_n} f_B)(\cdot, s)).
\end{aligned}$$

Observe

$$\begin{aligned}
&\int \int \|\hat{\Sigma}_{m_n}(s, t) - \tilde{\Sigma}_{m_n}(s, t)\|_F^2 \varpi(s) ds \varpi(t) dt \\
&\lesssim \int \int \left\| n^{-1} \sum_j p_{\underline{m_n}}(X_j) p_{\underline{m_n}}(X_j)' \tilde{\delta}_n(V_j, s) ((\mathcal{F}_g \hat{f}_{Bn})(X_j, t) - (\mathcal{F}_g II_{k_n} f_B)(X_j, t)) \right\|_F^2 \varpi(s) ds \varpi(t) dt \\
&\quad + \int \int \left\| n^{-1} \sum_j p_{\underline{m_n}}(X_j) p_{\underline{m_n}}(X_j)' \delta_n(V_j, t) ((\mathcal{F}_g \hat{f}_{Bn})(X_j, s) - (\mathcal{F}_g II_{k_n} f_B)(X_j, s)) \right\|_F^2 \varpi(s) ds \varpi(t) dt \\
&= I_n + II_n \quad (\text{say}).
\end{aligned}$$

We conclude

$$\begin{aligned}
I_n &\leq \int \int \left\| \frac{1}{n} \sum_j \tilde{\delta}_n(V_j, s) p_{\underline{m}_n}(X_j) p_{\underline{m}_n}(X_j)' (\mathcal{F}_g p_{\underline{k}_n})(X_j, t)' (\hat{\beta}_n - \beta_n) \right\|_F^2 \varpi(s) ds \varpi(t) dt \\
&\leq \int \int \left\| E[\tilde{\delta}_n(V, s) p_{\underline{m}_n}(X) p_{\underline{m}_n}(X)' (\mathcal{F}_g q_{\underline{k}_n})(X, t)] (\hat{\beta}_n - \beta_n) \right\|_F^2 \varpi(s) ds \varpi(t) dt + o_p(1) \\
&\leq \|\hat{\beta}_n - \beta_n\|^2 \\
&\quad \times O_p \left(\sum_{j,l=1}^{m_n} \sum_{l'=1}^{k_n} \int \int \left(E[(\varphi(X, s) - (\mathcal{F}_g \Pi_{k_n} f_B)(X, s)) (\mathcal{F}_g q_{l'})(X, t) p_j(X) p_l(X)] \right)^2 \varpi(s) ds \varpi(t) dt \right) \\
&\leq \|\hat{\beta}_n - \beta_n\|^2 \\
&\quad \times O_p \left(\sum_{l=1}^{m_n} \sum_{l'=1}^{k_n} \int \int E |(\varphi(X, s) - (\mathcal{F}_g \Pi_{k_n} f_B)(X, s)) (\mathcal{F}_g q_{l'})(X, t) p_l(X)|^2 \varpi(s) ds \varpi(t) dt \right) \\
&= O_p \left(m_n k_n^2 / n + m_n k_n \|\Pi_{m_n} \varphi - \varphi\|_{\varpi}^2 \right) = o_p(1).
\end{aligned}$$

Here, we used $\|\hat{\beta}_n - \beta_n\|^2 = O_p(k_n/n + \|\Pi_{m_n} \varphi - \varphi\|_{\varpi}^2)$ which can be seen as in the proof of Theorem 7. Since $I_n = o_p(1)$ we conclude

$$\begin{aligned}
II_n &\lesssim \int \int \left\| (\hat{\beta}_n - \beta_n)' E[(\mathcal{F}_g q_{\underline{k}_n})(X, s) p_{\underline{m}_n}(X) p_{\underline{m}_n}(X)' (\mathcal{F}_g q_{\underline{k}_n})(X, t)] (\hat{\beta}_n - \beta_n) \right\|_F^2 \varpi(s) ds \varpi(t) dt + o_p(1) \\
&\leq \|\hat{\beta}_n - \beta_n\|^4 \sum_{j,l=1}^{m_n} \int \int E \left[\|(\mathcal{F}_g q_{\underline{k}_n})(X, s)\| \|(\mathcal{F}_g q_{\underline{k}_n})(X, t)\| p_j(X) p_l(X) \right]^2 \varpi(s) ds \varpi(t) dt + o_p(1) \\
&\leq C m_n^2 \|\hat{\beta}_n - \beta_n\|^4 \left(\int E \|(\mathcal{F}_g q_{\underline{k}_n})(X, t)\|^2 \varpi(t) dt \right)^2 + o_p(1) \\
&= O_p \left(m_n^2 k_n / n^2 + m_n^2 \|\Pi_{m_n} \varphi - \varphi\|_{\varpi}^4 \right) = o_p(1).
\end{aligned}$$

by using $k_n = o(\sqrt{m_n})$. Finally, it is to see that $\varsigma_{m_n}^2 - \int \int \|\tilde{\Sigma}_{m_n}(s, t)\|^2 \varpi(s) ds \varpi(t) dt = o_p(1)$, which proves $\varsigma_{m_n} \hat{\varsigma}_{m_n}^{-1} = 1 + o_p(1)$. In particular, convergence of the trace of $\hat{\Sigma}_{m_n}(t, t)$ to the trace of $\Sigma_{m_n}(t, t)$ follows by using $|\hat{\mu}_{m_n} - \mu_{m_n}|^2 \leq m_n \int \|\hat{\Sigma}_{m_n}(t, t) - \Sigma_{m_n}(t, t)\|_F^2 \varpi(t) dt = o_p(m_n)$. $\square$

PROOF OF THEOREM 12. We make use of the decomposition

$$\begin{aligned}
nS_n &= \sum_j \int |p_{\underline{m}_n}(X_j)' (\hat{\gamma}_n(t) - \langle \varphi(\cdot, t), p_{\underline{m}_n} \rangle_X)|^2 \varpi(t) dt \\
&\quad + 2 \sum_j \int \left(p_{\underline{m}_n}(X_j)' (\hat{\gamma}_n(t) - \langle \varphi(\cdot, t), p_{\underline{m}_n} \rangle_X) \right) \\
&\quad \times \left(\Pi_{m_n} \varphi(X_j, t) - \hat{h}_n(X_{1j}, t) \tilde{g}(X_j, t, \hat{b}_{2n}) \right) \varpi(t) dt \\
&\quad + \sum_j \int |\Pi_{m_n} \varphi(X_j, t) - \hat{h}_n(X_{1j}, t) \tilde{g}(X_j, t, \hat{b}_{2n})|^2 \varpi(t) dt \\
&= I_n + 2 II_n + III_n \quad (\text{say})
\end{aligned}$$

where we used $\langle h(\cdot, t)\tilde{g}(\cdot, t, b_2), p_{\underline{m}_n} \rangle_X = \langle \varphi(\cdot, t), p_{\underline{m}_n} \rangle_X$. Consider I_n . As in the proof of Theorem 7 we have

$$I_n = n\lambda_n^{-1} \int \left\| n^{-1/2} \sum_j \left(\exp(itY_j) - h(X_{1j}, t)\tilde{g}(X_j, t, b_2) \right) p_{\underline{m}_n}(X_j) \right\|^2 \varpi(t) dt + o_p(\sqrt{m_n}).$$

Thus, Lemma 6.2 yields $(\sqrt{2}\zeta_{m_n})^{-1}(I_n - \mu_{m_n}) \xrightarrow{d} \mathcal{N}(0, 1)$. Consider III_n . Since $|\tilde{g}(X_j, t, b)| \leq 1$ for all b we evaluate

$$\begin{aligned} III_n &\lesssim \sum_j \int |\Pi_{m_n} \varphi(X_j, t) - \varphi(X_j, t)|^2 \varpi(t) dt \\ &\quad + \sum_j \int |h(X_{1j}, t) - \hat{h}_n(X_{1j}, t)|^2 \varpi(t) dt \\ &\quad + \sum_j \int |\hat{h}_n(X_{1j}, t)|^2 |\tilde{g}(X_j, t, b_2) - \tilde{g}(X_j, t, \hat{b}_{2n})|^2 \varpi(t) dt. \end{aligned}$$

It holds $\int \|\hat{h}_n(\cdot, t) - \Pi_{k_n} h(\cdot, t)\|_{X_1}^2 \varpi(t) dt = O_p(k_n/n)$ as we see in the following. We have

$$\begin{aligned} &\lambda_n \int \|\hat{h}_n(\cdot, t) - \Pi_{k_n} h(\cdot, t)\|_{X_1}^2 \varpi(t) dt \\ &\leq \lambda_n \left\| \left(\sum_j p_{\underline{k}_n}(X_j) p_{\underline{k}_n}(X_j)' \right)^{-1} \right\| \int \left\| \sum_j \left(\Pi_{k_n} h(X_{1j}, t) - \exp(it(Y_j - g_2(X_j, \hat{b}_{2n}))) \right) p_{\underline{k}_n}(X_{1j}) \right\|^2 \varpi(t) dt \\ &\lesssim \int \left\| n^{-1} \sum_j \left(\Pi_{k_n} h(X_{1j}, t) - \exp(it(Y_j - g_2(X_j, b_2))) \right) p_{\underline{k}_n}(X_{1j}) \right\|^2 \varpi(t) dt \\ &\quad + \|\hat{b}_{2n} - b_2\|^2 \sum_{l=1}^{k_n} \int \left\| n^{-1} \sum_j \exp(itY_j) \tilde{g}_b(X_j, t, \tilde{b}_{2n}) p_l(X_{1j}) \right\|^2 \varpi(t) dt + o_p(1), \end{aligned}$$

by a Taylor series expansion, where $\tilde{b}_{2n}$ is between $\hat{b}_{2n}$ and b_2 . As in relation (6.1), from $E[(\Pi_{k_n} h(X, t) - \exp(it(Y - g_2(X, b_2)))) p_{\underline{k}_n}(X)] = 0$ we deduce

$$\int E \left\| n^{-1} \sum_j \left(\Pi_{k_n} h(X_{1j}, t) - \exp(it(Y_j - g_2(X_j, b_2))) \right) p_{\underline{k}_n}(X_j) \right\|^2 \varpi(t) dt = O(n^{-1} k_n \lambda_n).$$

Further, since $\int E \sup_{b_2} \|\tilde{g}_b(X, t, b_2)\|^2 \varpi(t) dt \leq C$ we have

$$\begin{aligned} &E \left(\sum_{l=1}^{k_n} \int \left\| n^{-1} \sum_j \exp(itY_j) \tilde{g}_b(X_j, t, \tilde{b}_{2n}) p_l(X_j) \right\|^2 \varpi(t) dt \right)^{1/2} \\ &\leq E \left[\|p_{\underline{k}_n}(X)\| \left(\int \|\tilde{g}_b(X, t, \tilde{b}_{2n})\|^2 \varpi(t) dt \right)^{1/2} \right] \\ &\leq (E \|p_{\underline{k}_n}(X)\|^2)^{1/2} \left(\int E \sup_{b_2 \in \mathcal{B}} \|\tilde{g}_b(X, t, b_2)\|^2 \varpi(t) dt \right)^{1/2} \\ &= O(\sqrt{\lambda_n k_n}). \end{aligned}$$

This establishes the rate for the estimator $\hat{h}_n$. In light of condition $n\|II_{k_n}h - h\|_{\varpi}^2 = o(\sqrt{m_n})$ and since $n\|b_2 - \hat{b}_{2n}\|^2 = O_p(1)$ and $k_n = o(\sqrt{m_n})$ we conclude $III_n = o_p(\sqrt{m_n})$. It remains to show $II_n = o_p(\sqrt{m_n})$, which follows by

$$\begin{aligned}
|II_n| &\lesssim \left| \int \sum_j (\exp(itY_j) - II_{m_n}\varphi(X_j, t)) p_{\underline{m_n}}(X_j)' \langle II_{m_n}\varphi(\cdot, t) - II_{k_n}h(\cdot, t) \tilde{g}(\cdot, t, b_2), p_{\underline{m_n}} \rangle_X \varpi(t) dt \right| \\
&\quad + \left| \int \sum_j (\exp(itY_j) - II_{m_n}\varphi(X_j, t)) p_{\underline{m_n}}(X_j)' \langle II_{k_n}h(\cdot, t) \tilde{g}(\cdot, t, b_2) - \hat{h}_n(\cdot, t) \tilde{g}(\cdot, t, \hat{b}_{2n}), p_{\underline{m_n}} \rangle_X \varpi(t) dt \right| \\
&\quad + o_p(\sqrt{m_n}) \\
&= O_p(\sqrt{n}\|II_{k_n}h - h\|_{\varpi}) + o_p(\sqrt{m_n}) \\
&\quad + O_p\left(\left(k_n \int E \sup_{b_2 \in \mathcal{B}} \left\| \sum_{l=1}^{m_n} p_l(X) \langle \tilde{g}_b(\cdot, t, b_2) p'_{\underline{k_n}}, p_l \rangle_X \right\|^2 \varpi(t) dt\right)^{1/2}\right) \\
&= o_p(\sqrt{m_n}),
\end{aligned}$$

using that $\int E \sup_{b_2 \in \mathcal{B}} \left\| \sum_{l=1}^{m_n} p_l(X) \langle \tilde{g}_b(\cdot, t, b_2) p'_{\underline{k_n}}, p_l \rangle_X \right\|^2 \varpi(t) dt \leq \sum_{l=1}^{k_n} E[p_l^2(X)] = O(k_n)$, which proves the result. $\square$

PROOF OF PROPOSITION 15. For the proof it is sufficient to show $S_n \geq C\|\varepsilon\|_{\varpi}^2 + o_p(1)$. The proof of Theorem 7 together with the basic inequality $(a - b)^2 \geq a^2 - b^2$ implies that

$$\begin{aligned}
S_n &= \lambda_n^{-1} \sum_{l=1}^{m_n} \int \left| n^{-1} \sum_j \delta(V_j, t) p_l(X_j) \right|^2 \varpi(t) dt + o_p(1) \\
&\gtrsim \sum_{l=1}^{m_n} \int |E[\delta(V, t) p_l(X)]|^2 \varpi(t) dt + o_p(1) \\
&\gtrsim C(\|\varepsilon\|_{\varpi}^2/2 - \|II_{m_n}\varepsilon - \varepsilon\|_{\varpi}^2) + o_p(1) \\
&\gtrsim \|\varepsilon\|_{\varpi}^2 + o_p(1),
\end{aligned}$$

by using that $(\lambda_n)_{n \geq 1}$ is a nonincreasing sequence. $\square$

PROOF OF PROPOSITION 16. Following the proof Theorem 7, it is easily seen that

$$\begin{aligned}
nS_n &= \lambda_n^{-1} \sum_{l=1}^{m_n} \int \left| n^{-1/2} \sum_j \delta(V_j, t) p_l(X_j) \right|^2 \varpi(t) dt \\
&\quad + \sum_j \int |(\mathcal{F}_g II_{k_n} f_B)(X_j, t) - II_{m_n}\varphi(X_j, t)|^2 \varpi(t) dt + o_p(\sqrt{m_n}).
\end{aligned}$$

Further, under the sequence of local alternatives (2.9), we calculate

$$\begin{aligned}
\sum_j \int |(\mathcal{F}_g II_{k_n} f_B)(X_j, t) - II_{m_n}\varphi(X_j, t)|^2 \varpi(t) dt &= n\|\mathcal{F}_g f_B - \varphi\|_{\varpi}^2 + o_p(\sqrt{m_n}) \\
&= \varsigma_{m_n}^{-1} \|\mathcal{F}_g \Delta\|_{\varpi}^2 + o_p(\sqrt{m_n}),
\end{aligned}$$

which proves the result. $\square$

Proofs of Section 3.

In the following, we make use of the notation $\hat{\alpha}_n \equiv (n\hat{R}_n)^{-1} \sum_j Y_j p_{\underline{m}_n}(X_j, Z_j)$ where $\hat{R}_n = n^{-1} \sum_j p_{\underline{m}_n}(X_j, Z_j) p_{\underline{m}_n}(X_j, Z_j)'$. The Kronecker product for matrices is denoted by $\otimes$.

PROOF OF PROPOSITION 18. We make use of the decomposition

$$\begin{aligned}
nS_n &= \sum_j \int \left| \partial_z p_{\underline{m}_n}(X_j, z)' (\hat{\alpha}_n - E[\mathbb{1}\{Z > g(X, B)\} p_{\underline{m}_n}(X, Z)] \right|^2 \varpi(z) dz \\
&\quad + 2 \sum_j \int \partial_z p_{\underline{m}_n}(X_j, z)' \left(\hat{\alpha}_n - E[\mathbb{1}\{Z > g(X, B)\} p_{\underline{m}_n}(X, Z)] \right) \\
&\quad \times \left(\partial_z p_{\underline{m}_n}(X_j, z)' E[\mathbb{1}\{Z > g(X, B)\} p_{\underline{m}_n}(X, Z)] - (\mathcal{F}^{-1}[(\mathcal{F}_g \hat{f}_{Bn})(X_j, \cdot)])(z) \right) \varpi(z) dz \\
&\quad + \sum_j \int \left| \partial_z p_{\underline{m}_n}(X_j, z)' E[\mathbb{1}\{Z > g(X, B)\} p_{\underline{m}_n}(X, Z)] - (\mathcal{F}^{-1}[(\mathcal{F}_g \hat{f}_{Bn})(X_j, \cdot)])(z) \right|^2 \varpi(z) dz \\
&= I_n + 2II_n + III_n \quad (\text{say}).
\end{aligned}$$

Consider I_n . For all $l \geq 1$, the derivative of a basis function p_l is given by lp_{l-1} . Since p_l forms an orthonormal basis in $L^2_{\varpi}(\mathbb{R})$ is holds

$$\begin{aligned}
I_n &= \frac{n}{\lambda_n} (\hat{\beta}_{m_n} - E[\mathbb{1}\{Z > g(X, B)\} p_{\underline{m}_n}(X, Z)])' (I_{m_{1n}} \otimes T_n) (\hat{\beta}_{m_n} - E[\mathbb{1}\{Z > g(X, B)\} p_{\underline{m}_n}(X, Z)]) \\
&\quad + o_p(\sqrt{m_n})
\end{aligned}$$

where T_n is a $m_{2n} \times m_{2n}$ diagonal matrix with l -th diagonal element is given by $(l-1)^2$. It holds

$$I_n = \lambda_n^{-1} \sum_{l=1}^{m_n} \tau_l \left| n^{-1/2} \sum_j \left(Y_j - \int \mathbb{1}\{Z_j > g_2(X_j, b)\} f_B(b) db \right) p_l(X_j, Z_j) \right|^2 + o_p(1).$$

Thus, Lemma 6.2 yields $(\sqrt{2}\varsigma_{m_n})^{-1} (I_n - \mu_{m_n}) \xrightarrow{d} \mathcal{N}(0, 1)$. Consider III_n . We have

$$\begin{aligned}
III_n &\lesssim \sum_j \int \left| (\Pi_{m_n} \psi)(X_j, z) - \psi(X_j, z) \right|^2 \varpi(z) dz + \sum_j \int \left| (\mathcal{F}^{-1}[(\mathcal{F}_g(\hat{f}_{Bn} - f_B))(X_j, \cdot)])(z) \right|^2 \varpi(z) dz \\
&= A_{n1} + A_{n2}.
\end{aligned}$$

We have $A_{n1} = O_p(n \|\Pi_{m_n} \psi - \psi\|_{\varpi}^2) = o_p(\sqrt{m_n})$ and

$$\begin{aligned}
A_{n2} &\lesssim \sum_j \int \left| (\mathcal{F}^{-1}[(\mathcal{F}_g(\hat{f}_{Bn} - \Pi_{k_n} f_B))(X_j, \cdot)])(z) \right|^2 \varpi(z) dz \\
&\quad + \sum_j \int \left| (\mathcal{F}^{-1}[(\mathcal{F}_g(\Pi_{k_n} f_B - f_B))(X_j, \cdot)])(z) \right|^2 \varpi(z) dz,
\end{aligned}$$

where the second summand on the right hand is of the order $o_p(\sqrt{m_n})$. Further,

$$\begin{aligned} & \sum_j \int \left| (\mathcal{F}^{-1}[(\mathcal{F}_g(\widehat{f}_{Bn} - \Pi_{k_n} f_B))(X_j, \cdot)])(z) \right|^2 \varpi(z) dz \\ &= (\widehat{\beta}_n - \beta_n)' \sum_j \int (\mathcal{F}^{-1}[(\mathcal{F}_g q_{\underline{k}_n})(X_j, \cdot)])(z) (\mathcal{F}^{-1}[(\mathcal{F}_g q_{\underline{k}_n})(X_j, \cdot)])(z)' \varpi(z) dz (\widehat{\beta}_n - \beta_n) \end{aligned}$$

and thus, following the proof of Theorem 7 we obtain $A_{n2} = o_p(\sqrt{m_n})$. Similarly as in the proof of Theorem 7 it can be seen that $II_n = o_p(\sqrt{m_n})$, which completes the proof. $\square$

PROOF OF PROPOSITION 19. We decompose our test statistic as

$$\begin{aligned} nS_n &= \sum_j \int \left| \partial_z p_{\underline{m}_n}(X_j, z) \left(\widehat{\alpha}_n - E[Y p_{\underline{m}_n}(X, Z)] \right) \right|^2 \varpi(z) dz \\ &+ 2 \sum_j \int \left(\partial_z p_{\underline{m}_n}(X_j, z) \widehat{\alpha}_n - (\Pi_{m_n} \psi)(X_j, z) \right) \left((\Pi_{m_n} \psi)(X_j, z) - \widehat{\psi}_{1n}(X_{1j}, z - X'_{2j} \widehat{b}_{2n}) \right) \varpi(z) dz \\ &+ \sum_j \int \left| (\Pi_{m_n} \psi)(X_j, z) - \widehat{\psi}_{1n}(X_{1j}, z - X'_{2j} \widehat{b}_{2n}) \right|^2 \varpi(z) dz \\ &= I_n + 2II_n + III_n \quad (\text{say}). \end{aligned}$$

Consider I_n . As in the proof of Proposition 18 we obtain

$$I_n = \lambda_n^{-1} \sum_{l=1}^{m_n} \tau_l \left| n^{-1/2} \sum_j \left(Y_j - \int \mathbb{1}\{Z_j \geq X'_1 b_1 + X'_2 b_2\} f_{B_1}(b_1) db_1 \right) p_l(X_j, Z_j) \right|^2 + o_p(1),$$

and thus, Lemma 6.2 yields $(\sqrt{2} \varsigma_{m_n})^{-1} (I_n - \mu_{m_n}) \xrightarrow{d} \mathcal{N}(0, 1)$. We evaluate for III_n as follows

$$\begin{aligned} III_n &\lesssim \sum_j \int \left| (\Pi_{m_n} \psi)(X_j, z) - \psi(X_j, z) \right|^2 \varpi(z) dz \\ &+ \sum_j \int \left| \psi(X_{1j}, z - X'_{2j} b_2) - \widehat{\psi}_n(X_j, z) \right|^2 \varpi(z) dz \\ &+ \sum_j \int \left| \widehat{\psi}_n(X_j, z) - \widehat{\psi}_{1n}(X_{1j}, z - X'_{2j} \widehat{b}_{2n}) \right|^2 \varpi(z) dz \end{aligned}$$

The definition of the estimator $\widehat{b}_2$ in (3.3) yields

$$\begin{aligned} & \sum_j \int \left| \widehat{\psi}_n(X_j, z) - \widehat{\psi}_{1n}(X_{1j}, z - X'_{2j} \widehat{b}_{2n}) \right|^2 \varpi(z) dz \\ & \leq \sum_j \int \left| \widehat{\psi}_n(X_j, z) - \widehat{\psi}_{1n}(X_{1j}, z - X'_{2j} b_2) \right|^2 \varpi(z) dz \\ & \lesssim \sum_j \int \left| \widehat{\psi}_n(X_j, z) - \psi(X_j, z) \right|^2 \varpi(z) dz \\ & + \sum_j \int \left| \widehat{\psi}_{1n}(X_{1j}, z - X'_{2j} b_2) - \psi(X_{1j}, z - X'_{2j} b_2) \right|^2 \varpi(z) dz \\ & = o_p(\sqrt{m_n}). \end{aligned}$$

It thus follows $III_n = o_p(\sqrt{m_n})$. Similarly as in the proof of Theorem 12 it can be shown that $II_n = o_p(\sqrt{m_n})$. $\square$

Technical Appendix.

Lemma 6.2. *Let Assumptions 1–3 hold true. Then*

$$(\sqrt{2}\varsigma_{m_n})^{-1} \left(\lambda_n^{-1} \sum_{l=1}^{m_n} \int \left| n^{-1/2} \sum_j \delta(V_j, t) p_l(X_j) \right|^2 \varpi(t) dt - \mu_{m_n} \right) \xrightarrow{d} \mathcal{N}(0, 1).$$

Proof. Let us denote the real and imaginary parts of $\delta(V, t) p_l(X)$ by $\delta_l^R(V, t) = \text{Re}(\delta(V, t)) p_l(X)$ and $\delta_l^I(V, t) = \text{Im}(\delta(V, t)) p_l(X)$, respectively. We have

$$\begin{aligned} & \sum_{l=1}^{m_n} \int \left| (\lambda_n n)^{-1/2} \sum_j \delta(V_j, t) p_l(X_j) \right|^2 \varpi(t) dt \\ &= \sum_{l=1}^{m_n} \int \left\| (\lambda_n n)^{-1/2} \sum_j \left(\delta_l^R(V_j, t), \delta_l^I(V_j, t) \right)' \right\|^2 \varpi(t) dt \\ &= (\lambda_n n)^{-1} \sum_{l=1}^{m_n} \sum_j \int \left\| \left(\delta_l^R(V_j, t), \delta_l^I(V_j, t) \right)' \right\|^2 \varpi(t) dt \\ &\quad + (\lambda_n n)^{-1} \sum_{l=1}^{m_n} \sum_{j \neq j'} \int \left(\delta_l^R(V_j, t) \delta_l^R(V_{j'}, t) + \delta_l^I(V_j, t) \delta_l^I(V_{j'}, t) \right) \varpi(t) dt \\ &= I_n + II_n \quad (\text{say}). \end{aligned}$$

We observe

$$\begin{aligned} E|I_n - \mu_{m_n}|^2 &= \text{Var} \left((\lambda_n n)^{-1} \sum_{l=1}^{m_n} \sum_j \int \left| \delta(V_j, t) p_l(X_j) \right|^2 \varpi(t) dt \right) \\ &\leq \lambda_n^{-2} n^{-1} E \left[\int |\delta(V, t)|^4 \varpi(t) dt \left(\sum_{l=1}^{m_n} p_l^2(X) \right)^2 \right] \\ &\leq C \sup_{x \in \mathcal{X}} \|p_{\underline{m}_n}(x)\|^2 \lambda_n^{-2} n^{-1} \sum_{l=1}^{m_n} E[p_l^2(X)] = O(m_n^2 n^{-1} \lambda_n^{-1}) = o(1) \end{aligned}$$

using that $\int \sup_v |\delta(v, t)|^4 \varpi(t) dt$ is bounded. Consider II_n . Let us introduce the Martingale difference array $Q_{nj} = \sqrt{2}(\varsigma_{m_n} n)^{-1} \sum_{l=1}^{m_n} \sum_{j'=1}^{j-1} \int \left(\delta_l^R(V_j, t) \delta_l^R(V_{j'}, t) + \delta_l^I(V_j, t) \delta_l^I(V_{j'}, t) \right) \varpi(t) dt$ for $j = 2, \dots, n$, and zero otherwise. Further,

$$(\sqrt{2}\varsigma_{m_n})^{-1} II_n = \sqrt{2}(\varsigma_{m_n} n)^{-1} \sum_{j < j'} \sum_{l=1}^{m_n} \int \left(\delta_l^R(V_j, t) \delta_l^R(V_{j'}, t) + \delta_l^I(V_j, t) \delta_l^I(V_{j'}, t) \right) \varpi(t) dt = \sum_j Q_{nj}.$$

It remains to show that $\sum_j Q_{nj} \xrightarrow{d} \mathcal{N}(0, 1)$, which follows by Lemma A.3 of Breunig [2015b] by using the following computations. To show $\sum_{j=1}^{\infty} E|Q_{nj}|^2 \leq 1$ observe that

$$\sum_{j \neq j'} \int \left(\delta_l^I(V_j, t) \delta_l^R(V_{j'}, t) - \delta_l^R(V_j, t) \delta_l^I(V_{j'}, t) \right) \varpi(t) dt = 0$$

and $E[X_{1j}X_{1j'}] = 0$ for $j \neq j'$. Thus, for $j = 2, \dots, n$ we have

$$\begin{aligned}
E|Q_{nj}|^2 &= \frac{2(j-1)}{n^2\varsigma_{m_n}^2} E \left| \sum_{l=1}^{m_n} \int \delta_l(V_1, t) \overline{\delta_l(V_2, t)} \varpi(t) dt \right|^2 \\
&= \frac{2(j-1)}{n^2\varsigma_{m_n}^2} \sum_{l, l'=1}^{m_n} \int \int E \left[\delta_l(V, s) \overline{\delta_{l'}(V, t)} \right] E \left[\overline{\delta_l(V, s)} \delta_{l'}(V, t) \right] \varpi(s) ds \varpi(t) dt \\
&= \frac{2(j-1)}{n^2\varsigma_{m_n}^2} \sum_{l, l'=1}^{m_n} \int \int \left| E \left[\delta_l(V, s) \overline{\delta_{l'}(V, t)} \right] \right|^2 \varpi(s) ds \varpi(t) dt \\
&= \frac{2(j-1)}{n^2}
\end{aligned}$$

by the definition of ς_{m_n} and thus $\sum_j E|Q_{nj}|^2 = 1 - 1/n$. $\square$

Recall $\widehat{A}_n = n^{-1} \int \mathbf{F}_n(t)' \mathbf{F}_n(t) \varpi(t) dt$ and $A_n = \int E[(\mathcal{F}_g p_{\underline{k}_n})(X, t) (\mathcal{F}_g p_{\underline{k}_n})(X, t)'] \varpi(t) dt$.

Lemma 6.3. *Under the conditions of Theorem 7 it holds*

$$\widehat{A}_n^- = A_n^- + O_p\left(\sqrt{(\log n)k_n/n}\right).$$

Proof. On the set $\Omega \equiv \left\{ \|\widehat{A}_n^-\| \|\widehat{A}_n - A_n\| < 1/4, \text{ rank}(A_n) = \text{rank}(\widehat{A}_n) \right\}$, it holds $R(\widehat{A}_n) \cap R(A_n)^\perp = \{0\}$ by Corollary 3.1 of Chen et al. [1996], where R denotes the range of a mapping. Consequently, by using properties of the Moore-Penrose pseudoinverse it holds on the set Ω :

$$\begin{aligned}
\widehat{A}_n^- - A_n^- &= -\widehat{A}_n^-(\widehat{A}_n - A_n)A_n^- + \widehat{A}_n^-(\widehat{A}_n^-)'(\widehat{A}_n - A_n)'(I_{k_n} - A_n A_n^-) \\
&\quad + (I_{k_n} - \widehat{A}_n \widehat{A}_n^-)(\widehat{A}_n - A_n)'(A_n^-)'A_n^-,
\end{aligned}$$

see derivation of equation (3.19) in Theorem 3.10 on page 345 of Nashed [2014]. Applying the operator norm and using the fact that $I_{k_n} - A_n A_n^-$ and $I_{k_n} - \widehat{A}_n \widehat{A}_n^-$ as projections have operator norm bounded by one, we obtain

$$\begin{aligned}
\|\widehat{A}_n^- - A_n^-\| \mathbb{1}_\Omega &= (\|\widehat{A}_n^-\| \|\widehat{A}_n - A_n\| \|A_n^-\| + \|\widehat{A}_n^-\|^2 \|\widehat{A}_n - A_n\| + \|A_n^-\|^2 \|\widehat{A}_n - A_n\|) \mathbb{1}_\Omega \\
&\leq 3 \|\widehat{A}_n - A_n\| \max \left\{ \|A_n^-\|^2, \|\widehat{A}_n^-\|^2 \right\} \mathbb{1}_\Omega.
\end{aligned}$$

By Theorem 3.2 of Chen et al. [1996] it holds $\|\widehat{A}_n^-\| \mathbb{1}_\Omega \leq 3\|A_n^-\| = O(1)$. Consequently, Lemma 6.2 of Belloni et al. [2015] yields $\|\widehat{A}_n^- - A_n^-\| \mathbb{1}_\Omega = O_p(\sqrt{k_n(\log n)/n})$. The assertion follows by $\mathbb{1}_\Omega = 1$ with probability approaching one. $\square$

References

D. W. Andrews. Asymptotic results for generalized wald tests. *Econometric Theory*, 3(03): 348–358, 1987.

- D. W. Andrews. Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica*, pages 683–734, 2001.
- J. Banks, R. Blundell, and A. Lewbel. Quadratic engel curves and consumer demand. *Review of Economics and Statistics*, 79(4):527–539, 1997.
- A. Belloni, V. Chernozhukov, D. Chetverikov, and K. Kato. Some new asymptotic theory for least squares series: Pointwise and uniform results. *Journal of Econometrics*, 186(2):345–366, 2015.
- R. Beran. Semiparametric random coefficient regression models. *Annals of the Institute of Statistical Mathematics*, 45(4):639–654, 1993.
- R. Beran and P. Hall. Estimating coefficient distributions in random coefficient regressions. *The Annals of Statistics*, pages 1970–1984, 1992.
- R. Beran and P. W. Millar. Minimum distance estimation in random coefficient regression models. *The Annals of Statistics*, 22(4):1976–1992, 1994.
- R. Beran, A. Feuerverger, P. Hall, et al. On nonparametric estimation of intercept and slope distributions in random coefficient regression. *The Annals of Statistics*, 24(6):2569–2592, 1996.
- R. Blundell and J. Horowitz. A nonparametric test of exogeneity. *Review of Economic Studies*, 74(4):1035–1058, Oct 2007.
- R. Blundell, D. Kristensen, and R. Matzkin. Stochastic demand and revealed preference. Technical report, working paper, 2010.
- C. Breunig. Goodness-of-fit tests based on series estimators in nonparametric instrumental regression. *Journal of Econometrics*, 184(2):328–346, 2015a.
- C. Breunig. Specification testing in nonparametric instrumental quantile regression. Technical report, working paper, 2015b.
- G. Chen, M. Wei, and Y. Xue. Perturbation analysis of the least squares solution in hilbert spaces. *Linear algebra and its applications*, 244:69–80, 1996.
- X. Chen. Large sample sieve estimation of semi-nonparametric models. *Handbook of Econometrics*, 2007.

- X. Chen and T. M. Christensen. Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions. *Journal of Econometrics*, 2015.
- X. Chen and D. Pouzo. Sieve quasi likelihood ratio inference on semi/nonparametric conditional moment models. *Econometrica*, 83(3):1013–1079, 2015.
- A. Deaton and J. Muellbauer. An almost ideal demand system. *American Economic Review*, pages 312–326, 1980.
- F. Dunker, S. Hoderlein, and H. Kaido. Random coefficients in static games of complete information. Technical report, cemmap working paper, Centre for Microdata Methods and Practice, 2013.
- J. T. Fox and A. Gandhi. Identifying heterogeneity in economic choice and selection models using mixtures. In *James Heckman and B. Citeseer*, 2009.
- J. T. Fox and N. Lazzati. Identification of potential games and demand models for bundles. Technical report, National Bureau of Economic Research, 2012.
- E. Gautier and S. Hoderlein. A triangular treatment effect model with random coefficients in the selection equation. Technical report, Toulouse School of Economics, 2015.
- E. Gautier and Y. Kitamura. Nonparametric estimation in random coefficients binary choice models. *Econometrica*, 81(2):581–607, 2013.
- W. Härdle and E. Mammen. Comparing nonparametric versus parametric regression fits. *The Annals of Statistics*, pages 1926–1947, 1993.
- J. Hausman and W. Newey. Individual heterogeneity and average welfare. Technical report, Centre for Microdata Methods and Practice, 2013.
- S. Hoderlein. How many consumers are rational? *Journal of Econometrics*, 164(2):294–309, 2011.
- S. Hoderlein and R. Sherman. Identification and estimation in a correlated random coefficients binary response model. *Journal of Econometrics*, 2015.
- S. Hoderlein, J. Klemelä, and E. Mammen. Analyzing the random coefficient model nonparametrically. *Econometric Theory*, 26(03):804–837, 2010.

- S. Hoderlein, L. Su, and H. White. Specification testing for nonparametric structural models with monotonicity in unobservables. *V UCSD Department of Economics Working Paper*, 2011.
- S. Hoderlein, H. Holzmann, and A. Meister. The triangular model with random coefficients. Technical report, Boston College, 2014.
- J. L. Horowitz. Testing a parametric model against a nonparametric alternative with identification through instrumental variables. *Econometrica*, 74(2):521–538, 2006.
- J. L. Horowitz. Specification testing in nonparametric instrumental variables estimation. *Journal of Econometrics*, 167:383–396, 2012.
- J. L. Horowitz and S. Lee. Testing a parametric quantile-regression model with an endogenous explanatory variable against a nonparametric alternative. *Journal of Econometrics*, 152(2):141–152, 2009.
- H. Ichimura and T. S. Thompson. Maximum likelihood estimation of a binary choice model with random coefficients of unknown distribution. *Journal of Econometrics*, 86(2):269–295, 1998.
- G. W. Imbens and W. K. Newey. Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica*, 77(5):1481–1512, 2009.
- A. Lewbel. Consumer demand systems and household expenditure. *Handbook of Applied Econometrics, Blackwell Handbooks in economics*, 1999.
- A. Lewbel. Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables. *Journal of Econometrics*, 97(1):145–177, 2000.
- A. Lewbel and K. Pendakur. Unobserved preference heterogeneity in demand using generalized random coefficients. Technical report, Boston College Department of Economics, 2013.
- A. Lewbel, X. Lu, and L. Su. Specification testing for transformation models with an application to generalized accelerated failure-time models. *Journal of Econometrics*, 184(1):81–96, 2015.
- M. Masten. Random coefficients on endogenous variables in simultaneous equations models. Technical report, cemmap working paper, Centre for Microdata Methods and Practice, 2015.
- R. L. Matzkin. Identification in nonparametric limited dependent variable models with simultaneity and unobserved heterogeneity. *Journal of Econometrics*, 166(1):106–115, 2012.

- M. Z. Nashed. *Generalized Inverses and Applications: Proceedings of an Advanced Seminar Sponsored by the Mathematics Research Center, the University of WisconsinMadison, October 8-10, 1973*. Number 32. Elsevier, 2014.
- W. K. Newey. Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics*, 79(1):147–168, 1997.
- A. Santos. Inference in nonparametric instrumental variables with partial identification. *Econometrica*, 80(1):213–275, 2012.
- P. A. Swamy. Efficient inference in a random coefficient regression model. *Econometrica*, pages 311–323, 1970.