Specification Testing in Random Coefficient Models

Christoph Breunig (Humboldt University Berlin)
Stefan Hoderlein (Boston College)

Discussion Paper No. 77

March 5, 2018

Collaborative Research Center Transregio 190 | www.rationality-and-competition.de
Ludwig-Maximilians-Universität München | Humboldt-Universität zu Berlin
Spokesperson: Prof. Dr. Klaus M. Schmidt, University of Munich, 80539 Munich, Germany
+49 (89) 2180 3405 | info@rationality-and-competition.de

Specification Testing in Random Coefficient Models*

Christoph Breunig † Stefan Hoderlein‡

Humboldt-Universität zu Berlin Boston College

March 2, 2018

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, generalizations of this notion like a known nonlinear functional relationship, or 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 provide formal power analysis against global as well as against local alternatives. Moreover, we perform a Monte Carlo simulation study, and apply the tests to analyze the degree of nonlinearity in a heterogeneous random coefficients demand model. While we find some evidence against the popular QUAIDS specification with random coefficients, it is not strong enough to reject the specification at the conventional significance level.

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

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 microeconometric application.

*We thank seminar participants at Northwestern and WIAS Berlin. Financial support by Deutsche Forschungsgemeinschaft through CRC TRR 190 is gratefully acknowledged.

†Christoph Breunig, Department of Economics, 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.

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[1], 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 as well as a d_x dimensional vector of observable explanatory variables X to a continuous dependent variable Y, i.e.,

$$Y = g(X, B). \tag{1.1}$$

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. Despite the fact that in this model 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]), the model imposes structure that can be used to assess the validity of the specification. For instance, in the very same model, the conditional expectation is linear, i.e., $E[Y|X] = b_0 + b_1X_1 + ... + b_kX_k$, 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

[1] We refer to models with several unobservables, e.g., random coefficients models, nonseparable models, treatment effects etc, as (models with) "complex heterogeneity".

can again be taken as rejection of this linear random coefficients specification. 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, or equivalently,

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

for any strictly positive integrable weighting function ϖ, which is not required to be a pdf and whose choice is discussed in the simulation section.[2] Our test statistic is then given by the sample counterpart

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

where $\widehat{\varepsilon}_n$ 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

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

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

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

[2]This type of weighting is standard in the literature, see the weighted L^2 test statistic by Su and White [2007], or the empirical likelihood test proposed by Chen et al. [2013]. For a complex number $z \in \mathbb{C}$ the absolute value is given by $|z| = \sqrt{z\bar{z}}$.

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 estimation[3].

Another 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 that we will pursue in the application, consider a single cross section of the workhorse QUAIDS model of consumer demand (Banks et al. [1997]). Note that in a cross section prices often do not vary (or only very minimally, see, e.g., the commonly used British FES data), and the demand model for a good Y, in our example food at home, is therefore defined through:

$$Y = B_0 + B_1 X + B_2 X^2,$$

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

[3] In 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]).

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. In contrast, our test has power against certain alternatives even if our model is not identified under the null hypothesis. As an example, in the application we consider testing the random coefficients QUAIDS model against higher order polynomials; in this case, $\varepsilon(X, t) = 0$ for all t implies that, e.g., the cubic model $Y = \widetilde{B}_0 + X\widetilde{B}_1 + X^2\widetilde{B}_2 + X^3\widetilde{B}_3$ with random coefficients $(\widetilde{B}_0, \widetilde{B}_1, \widetilde{B}_2, \widetilde{B}_3)$ is misspecified[4].

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], and to systems of equations, see section 3 as well as an additional online appendix, see Breunig and Hoderlein [2017].

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

[4]In addition, 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 element-wise 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).

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].

While our test is the first that uses characteristic functions to test hypotheses about random coefficients, in other econometric setups tests based on comparing characteristic functions have been proposed. For instance, Su and White [2007] considered a test of conditional independence, Chen and Hong [2010] proposed a goodness-of-fit test for multifactor continuous-time Markov models, and Chen et al. [2013] considered an empirical likelihood test for correct specification for Markov processes.

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] and Chen and Pouzo [2012] for sieve minimum distance estimation) 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 models 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. Moreover, in contrast to nonparametric specification tests in instrumental variable models in Horowitz [2012] and Breunig [2015a] who assume bounded support, we explicitly allow for regressors with large support which is required to ensure identification of random coefficient models in general. This results in a very different setup as densities have to be allowed to be close to zero, which leads to slower rates of convergence and rules out the approach of density weighting considered in Horowitz [2012].

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 additively separate 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 QUAIDS (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. First, we provide a general test for the hypothesis that the structural relation between the covariates, the random coefficients, and the outcome variable coincides with a known function g. We thus consider the hypothesis [5]

$$H_{\text{mod}} : \text{there exist some distributions of random parameters } B \text{ such that } Y = g(X, B).$$

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

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

in which case the distribution of B is point 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 the square of the vector X is understood element-wise. Note that in the latter example the distribution of the random parameters B is only partially identified. As already discussed above, 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). More specifically, we want to consider the following hypothesis

$$H_{\text{deg}} : B_2 = b_2 \quad \text{for some distributions of random parameters satisfying (1.1).}$$

The alternative is $P\big(B_2 \neq b_2 \text{ for all distributions of random parameters } B \text{ satisfying (1.1)}\big) > 0$.

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

While the hypothesis H_{deg} could be considered in more general models, motivated by the linear (or polynomial) model we will confine ourselves to functions g that are additively separable in the sense that

$$H_{\text{add}}: \; Y = g_1(X, B_{-2}) + g_2(X, B_2), \tag{2.1}$$

where g_1 and g_2 denote two known functions, and we use the notation $B_{-2} = (B_0, B_1')'$. The leading example for this type of hypothesis is of course when g_1 is a linear function of a subvector X_1 of covariates X, in which case we obtain a partially linear structure, i.e.,

$$H_{\text{part-lin}}: \; Y = B_0 + X_1' B_1 + g_2(X, B_2), \tag{2.2}$$

where g_2 is a known function. This covers the following examples of hypotheses already outlined in the introduction: 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, i.e., 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. Obviously, in this case b_2 is identified by standard linear mean regression identification conditions. A second example arises if, in the quadratic model, we want to test a specification with deterministic second order terms, i.e.

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

against the alternative that B_2 is random. Note that in the latter two hypotheses, identification of b_2 follows as in parametric mean regression and, in equation (2.2), point identification under the null holds for instance if $g_2(X, b_2) = h(X_2)'b_2$ for some vector valued function h such that the associated rank condition is satisfied. In the Monte Carlo study and the application, we will only consider the case where b_2 is point identified, which we consider to be the leading case. However, we would like to point out that the test applies also more generally to situations where b_2 does not need to be point identified, as in the most general case defined by hypothesis H_{add}, albeit with a loss of power against some alternatives.

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

$$\varepsilon(X, t) = 0 \text{ for all } t, \tag{2.3}$$

where $\varepsilon : \mathbb{R}^{d_x+1} \to \mathbb{C}$ is a complex valued, measurable function. Our testing procedure is based on the L^2 distance of ε to zero. Equation (2.3) is equivalent to

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

for some strictly positive weighting function ϖ with $\int \varpi(t)dt < \infty$. Our test statistic is given by the sample counterpart to this expression, which is

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

where $\widehat{\varepsilon}_n$ is a consistent estimator of ε. Below, we show that the statistic S_n is (after standardization) asymptotically standard normally distributed. 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 $p_{\underline{m}}(\cdot) = (p_1(\cdot),\ldots,p_m(\cdot))'$ for some integer $m \geq 1$. Further, let $\mathbf{X}_m \equiv \big(p_{\underline{m}}(X_1),\ldots,p_{\underline{m}}(X_n)\big)'$ and $\mathbf{Y}_n(t) = \big(\exp(itY_1),\ldots,\exp(itY_n)\big)$. We replace φ by the series least squares estimator

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

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.

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*

$$E[\exp(itY)|X] = \int \exp(itg(X,b))f_B(b)db,$$

for each $t \in \mathbb{R}$, a known function g, and some random parameters B, with probability density function (p.d.f.) f_B. Hence, equation (2.3) holds true with

$$\varepsilon(X,t) = E[\exp(itY)|X] - \int \exp(itg(X,b))f_B(b)db. \tag{2.4}$$

As already mentioned, if the function g is nonlinear the p.d.f. of random coefficients B is not *necessarily point identified. On the other side, if g is the inner product of its entries, then (2.3) holds true with*

$$\varepsilon(X,t) = E[\exp(itY)|X] - \int \exp(itX'b) f_B(b)db,$$

and in this case the distribution of B is point 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 which are not point identified. To illustrate this, consider testing linearity of the random coefficient QUAIDS model which is given by $Y = \widetilde{B}_0 + \widetilde{B}_1 X + \widetilde{B}_2 X^2$ for random coefficients $\widetilde{B}_0$, $\widetilde{B}_1$, and $\widetilde{B}_2$ (also independent of X). In this case, the conditional first and second moment equation implied by equation (2.3) yield $E[\widetilde{B}_2] = 0$ and $Var(\widetilde{B}_2) = 0$, respectively. We thus conclude that $\widetilde{B}_2 = 0$ with probability one.

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.[6] *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

$$\widehat{f}_{Bn} \in \arg\min_{f\in\mathcal{B}_n} \Big\{ \sum_{j=1}^n \int |\widehat{\varphi}_n(X_j,t) - (\mathcal{F}_g f)(X_j,t)|^2 \varpi(t)dt \Big\} \tag{2.5}$$

and $\mathcal{B}_n = \big\{\phi(\cdot) = \sum_{l=1}^{k_n} \beta_l q_l(\cdot)\big\}$ 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 $\mathbf{F}_n(t) = \big((\mathcal{F}_g q_{\underline{k_n}})(X_1,t),\dots,(\mathcal{F}_g q_{\underline{k_n}})(X_n,t)\big)'$, the minimum norm estimator of f_B given in (2.5) coincides with $\widehat{f}_{Bn}(\cdot) = q_{\underline{k_n}}(\cdot)'\widehat{\beta}_n$ where

$$\widehat{\beta}_n = \Big(\int \mathbf{F}_n(-t)'\mathbf{F}_n(t)\varpi(t)dt \Big)^{-} \int \mathbf{F}_n(-t)'\mathbf{\Phi}_n(t)\varpi(t)dt$$

and $\mathbf{\Phi}_n(t) = \big(\widehat{\varphi}_n(X_1,t),\dots,\widehat{\varphi}_n(X_n,t)\big)'$.[7] *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.*

[6] The 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 $(\mathcal{F}^{-1}\widetilde{\phi})(z) \equiv (2\pi)^{-d}\int \exp(-itz)\widetilde{\phi}(t)dt$. We also make use of $\overline{(\mathcal{F}_g\phi)(t)} = (\mathcal{F}_g\phi)(-t)$ where $\overline{\phi}$ denotes the complex conjugate of a function ϕ.

[7] The integral transform $\mathcal{F}_g$ of a vector of functions is always understood element-wise, i.e., $(\mathcal{F}_g q_{\underline{k_n}})(X_j,t) = \big((\mathcal{F}_g q_1)(X_j,t),\dots,(\mathcal{F}_g q_{k_n})(X_j,t)\big)'$.

Example 2 (Testing degeneracy under the random coefficients specification). *In the case of an additively separable structure H_{add} (see equation (2.1)), the null hypothesis H_{deg} implies the equality of conditional characteristic functions, i.e.,*

$$E[\exp(itY)|X] = \int \exp\big(itg_1(X, b_{-2})\big) f_{B_{-2}}(b_{-2}) db_{-2} \exp\big(itg_2(X, b_2)\big), \tag{2.6}$$

for each $t \in \mathbb{R}$. Therefore, equation (2.3) holds with

$$\varepsilon(X, t) = E[\exp(itY)|X] - \int \exp\big(itg_1(X, b_{-2})\big) f_{B_{-2}}(b_{-2}) db_{-2} \exp\big(itg_2(X, b_2)\big).$$

Given a partially linear structure $H_{\text{part-lin}}$ (see equation (2.2)), the null hypothesis H_{deg} implies the equality of conditional characteristic functions, i.e., equation (2.3) holds with

$$\varepsilon(X, t) = E[\exp(itY)|X] - \int \exp(itX_1' b_{-2}) f_{B_{-2}}(b_{-2}) db_{-2} \exp\big(itg_2(X, b_2)\big),$$

where the distribution of the random coefficients is identified. Our test, based on the function ε, has power against any failure of hypothesis H_{deg} if the distribution of the random coefficients under the maintained hypothesis H_{add} is identified, i.e., if g_1 and g_2 are linear in X_1 and X_2, respectively, or element-wise transformations of each component of these vectors (see Gautier and Hoderlein [2015]).

To illustrate that our test of degeneracy has power in the random coefficient QUAIDS model $Y = \widetilde{B}_0 + \widetilde{B}_1 X + \widetilde{B}_2 X^2$, note that under the null the conditional first and second moment regressions implied by equation (2.3) already yield that $E[\widetilde{B}_2] = b_2$ and $E[\widetilde{B}_2^2] = b_2^2$, respectively. From this observation we are already in the position to conclude that $\widetilde{B}_2$ is degenerate with $\widetilde{B}_2 = b_2$.

We estimate the function ε by

$$\widehat{\varepsilon}_n(X_j, t) = \widehat{\varphi}_n(X_j, t) - (\mathcal{F}_{g_1} \widehat{f}_{B_{-2},n})(X_j, t) \exp\big(itg_2(X_j, \widehat{b}_{2n})\big),$$

where the estimators $\widehat{f}_{B_{-2},n}$ and $\widehat{b}_{2n}$ are a sieve minimum distance estimators of the p.d.f. $f_{B_{-2}}$ and the parameter b_2, respectively, given by

$$(\widehat{f}_{B_{-2}n}, \widehat{b}_{2n}) \in \underset{(f,b)\in\mathcal{B}_{-2,n}\times\mathcal{B}_2}{\arg\min} \Big\{ \sum_{j=1}^{n} \int \big|\widehat{\varphi}_n(X_j, t) - (\mathcal{F}_{g_1} f)(X_j, t) \exp\big(itg_2(X_j, b)\big)\big|^2 \varpi(t) dt \Big\} \tag{2.7}$$

and $\mathcal{B}_{-2,n} = \big\{\phi(\cdot) = \sum_{l=1}^{k_n} \beta_l q_l(\cdot)\big\}$ is a linear sieve space of dimension $k_n < \infty$ with basis functions $\{q_l\}_{l\geq 1}$ of B_{-2} and $\mathcal{B}_2$ is a compact parameter space. See also Ai and Chen [2003] for sieve minimum distance estimation for finite dimensional parameters and nonparametric *functions. As in the previous example, k_n and m_n increase with sample size n, but we require that m_n increases faster than k_n.*

Example 3 (Testing degeneracy under additive separability alone)**.** *We also present an alternative test of degeneracy under H_{add} (see equation (2.1)) when g_1 depends on covariates X_1 but not on a subvector X_2 of the covariates $X = (X_1', X_2')'$. In this case, we rely on additive separability alone and base our test on*

$$E[\exp(itY)|X] = E\Big[\exp\big(it(Y - g_2(X, b_2))\big)\Big|X_1\Big]\exp\big(itg_2(X, b_2)\big). \tag{2.8}$$

Of course, such a test is only reasonable if the sigma algebra generated by X is not contained in the one generated by X_1. This rules out, for instance, testing degeneracy in the random coefficient QUAIDS model where X is scalar and g_2 is a quadratic function of X.

This test would not require any structure on the first term (despite not depending on X_2), i.e., in equation (2.1) we do neither have to know g_1, nor would have to assume that B_{-2} is finite. In contrast to the setting in Example 2, however, we require b_2 to be point identified, which in the absence of any structure on g_1 may be difficult to establish. There are examples where this structure could be useful. Consider for instance a model which has a complex nonlinear function in X_1, but is linear in X_2, i.e., $Y = g_1(X_1, B_{-2}) + X_2'B_2$, with an unknown function g_1. Suppose a researcher wants to test the null that the random coefficients B_2 has a degenerate distribution. In this case, b_2 is identified by a partially linear mean regression model, since $E[Y|X] = \mu(X_1) + X_2'b_2$, where $\mu(X_1) = E[g_1(X_1, B_{-2})|X_1]$. Evidently, this test requires less structure on the way X_1 enters, but in return suffers from lower power, e.g., if X_1 indeed enters through a random coefficients specification.

Let $\widehat{b}_{2n}$ denote a consistent estimator of the point identified parameter b_2. For instance, under the partially linear structure $H_{part\text{-}lin}$ (see equation (2.2)), we have the moment restriction $E[Y|X] = b_0 + X_1'b_1 + g_2(X, b_2)$ and thus, $\widehat{b}_{2n}$ would coincide with the nonlinear least squares estimator of b_2. We denote $p_{\underline{k_n}}(\cdot) = (p_1(\cdot), \dots, p_{k_n}(\cdot))'$ and $\mathbf{X}_{1n} \equiv \big(p_{\underline{k_n}}(X_{11}), \dots, p_{\underline{k_n}}(X_{1n})\big)'$ which is a $n \times k_n$ matrix. Consequently, we estimate the function ε by

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

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

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. While it might be possible to treat both types of hypotheses under an abstract general testing framework, because of transparency of exposition (at least for applied researchers), we decided to treat both cases separately.

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 $((Y_1, X_1), \ldots, (Y_n, X_n))$ 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 $\sup_{x\in\mathcal{X}} \|p_{\underline{m_n}}(x)\|^2 \leqslant C m_n$ 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, wavelet bases, and Hermite functions (which are orthonormalized Hermite polynomials).[8] 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

[8]When p_l are Hermite functions, it holds due to Crámer's inequality that $\sup_{x\in\mathcal{X}} \|p_{\underline{m_n}}(x)\|^2 \leqslant 1.086\pi^{-1/4} m_n$.

hand, the sequence $(\lambda_n)_{n\geq 1}$ is bounded away from zero if $\{p_l\}_{l\geq}$ 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 1. *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*

$$\int (a_n' p_{\underline{m_n}}(x))^2 \, \mathbb{1}\{f(x) < \lambda_n\}\nu(dx) \leq c \int (a_n' p_{\underline{m_n}}(x))^2 \nu(dx) \tag{2.9}$$

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

Condition (2.9) 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 inequality (2.9), 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 $\int E\big[\delta(V,t)\big|X\big]\varpi(t)dt = 0$ holds.

Definition 1. *Denote by $P_n = E[p_{\underline{m_n}}(X)p_{\underline{m_n}}(X)']$, and define*

$$\mu_{m_n} \equiv \int E\big[|\delta(V,t)|^2 p_{\underline{m_n}}(X)' P_n^{-1} p_{\underline{m_n}}(X)\big]\varpi(t)dt \quad \textit{and}$$
$$\varsigma_{m_n} \equiv \Big(\int\int \Big\|P_n^{-1/2} E\big[\delta(V,s)\overline{\delta(V,t)}p_{\underline{m_n}}(X)\, p_{\underline{m_n}}(X)'\big] P_n^{-1/2}\Big\|_F^2 \varpi(s)\varpi(t)dsdt\Big)^{1/2}.$$

Here, we use the notation $\overline{\phi}$ for the complex conjugate of a function ϕ, and $\|\cdot\|_F$ to denote the Frobenius norm. Alternatively, we could normalize our test statistic using residuals $\exp(itY) - E[\exp(itY)|X]$ rather than $\delta(V,t)$. While this alternative procedure leads to accurate normalization of our test statistic under the null hypothesis, it is not necessarily accurate under alternative models.

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

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 $\varsigma_{m_n} \geq C\sqrt{m_n}$, see Lemma 5.1.

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 $\|\phi\|_\varpi = \big(\int E|\phi(X,t)|^2\varpi(dt)\big)^{1/2}$ and the linear sieve space $\Phi_n \equiv \big\{\phi : \phi(\cdot) = \sum_{l=1}^{m_n}\beta_l p_l(\cdot)\big\}$. Moreover, $\|\cdot\|$ and $\|\cdot\|_\infty$, respectively, denote the Euclidean norm and the supremum norm. Let us introduce $A_n = \int E[(\mathcal{F}_g q_{\underline{k_n}})(X,-t)(\mathcal{F}_g q_{\underline{k_n}})(X,t)']\varpi(t)dt$ and its empirical analog $\widehat{A}_n = n^{-1}\int \mathbf{F}_n(-t)'\mathbf{F}_n(t)\varpi(t)dt$ (see also Example 1). In the following, we introduce a strictly positive, nonincreasing sequence $(\tau_n)_{n\geq 1}$ such that $\tau_n\|A_n^-\|^2 = O(1)$.

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 $n\|\Pi_{m_n}\varphi - \varphi\|_\varpi^2 = o(\sqrt{\tau_n m_n})$ and $\|\Pi_{m_n}\varphi - \varphi\|_\infty = O(1)$. (iii) It holds $k_n \log n = o(\tau_n\sqrt{m_n})$. (iv) It holds $P\big(rank(A_n) = rank(\widehat{A}_n)\big) = 1 + o(1)$. (v) There exists a constant $C > 0$ such that $\sum_{l\geq 1}\big(\int_{\mathbb{R}^{d_b}}\phi(b)q_l(b)db\big)^2 \leq C\int_{\mathbb{R}^{d_b}}\phi^2(b)db$ 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 $\mathcal{I}_g \equiv \big\{f : f \text{ is a p.d.f. with } \varphi = \mathcal{F}_g f\big\}$. This condition ensures that the bias for estimating any f_B in the identified set $\mathcal{I}_g$ is asymptotically negligible. In the linear case, Hermite functions are eigenfunctions of the Fourier transform $\mathcal{F}$ and hence, Assumption 4 (i) is equivalent to imposing a sufficiently small approximation error $\Pi_{k_n} f_B - f_B$. In the following, we present primitive conditions when Assumption 4 (i) holds also for any nonlinear function g and, in particular, is satisfied for, e.g., quadratic functions. We observe that

$$\|\mathcal{F}_g(\Pi_{k_n} f_B - f_B)\|_\varpi \leq \|\mathcal{F}_g\|_\varpi \int_{\mathbb{R}^d}\big|\Pi_{k_n} f_B(b) - f_B(b)\big|db,$$

where we introduced the operator norm given by

$$\begin{aligned}\|\mathcal{F}_g\|_\varpi^2 &\equiv \sup_{\phi\in L^1(\mathbb{R}^{d_b}),\,\int|\phi(b)|db=1} \int E\Big|\int \exp(itg(X,b))\phi(b)db\Big|^2\varpi(t)dt\\ &\leq \sup_{\phi\in L^1(\mathbb{R}^{d_b}),\,\int|\phi(b)|db=1} \int \Big(\int |\phi(b)|db\Big)^2\varpi(t)dt\\ &= \int \varpi(t)dt,\end{aligned}$$

using that $|\exp(itg(X,b))| \leq 1$. The sieve approximation error imposed in Assumption 4 (i) is thus less restrictive than assuming

$$\sqrt{n}\int_{\mathbb{R}^{d_b}} \big|\mathit{\Pi}_{k_n} f_B(b) - f_B(b)\big|db = o(\sqrt{m_n}),$$

for any $f_B \in \mathcal{I}_g$. For instance, if f_B and its approximation $\mathit{\Pi}_{k_n} f_B$ belong to a compact subset of $\mathbb{R}^{d_b}$ and $\|\mathit{\Pi}_{k_n} f_B - f_B\|_\infty = O(k_n^{-s/d_b})$, which is satisfied for B-splines or trigonometric basis functions, we obtain the rate restriction $nk_n^{-2s/d_b} = o(\sqrt{m_n})$, which imposes a lower bound on the dimension parameters k_n and m_n. If in addition $\tau_n^{-1} = O(1)$, Assumptions 4 (i) and (iii) are satisfied if $m_n \sim n^\zeta$ and $k_n \sim n^\kappa$ where $d_b(1-\zeta/2)/(2s) < \kappa < \zeta/2$.[9] We thus require $\zeta > 2d_b/(2s+d_b)$, so s has to increase with dimension d_b, which reflects a curse of dimensionality. In this case, Assumption 4 (ii), which determines the sieve approximation error for the function φ, automatically holds if $\|\mathit{\Pi}_{m_n}\varphi - \varphi\|_\varpi = O(m_n^{-s/d_x})$ and we may choose κ to balance variance and bias, i.e., $\kappa = d_x/(2s+d_x)$.[10] For further discussion and examples of sieve bases, we refer to Chen [2007].

Assumption 4 (iii) has the interpretation of an overidentification restriction imposed on the finite dimensional approximations and requires that there are more moment restriction (captured by m_n) than unknown parameters (captured by the dimension of the sieve space $\mathcal{B}_n$ given by k_n).

Assumption 4 (iv) ensures that the sequence of generalized inverse matrices is bounded and imposes a rank condition. This condition is sufficient and necessary for convergence in probability of generalized inverses of random matrices with fixed dimension, for further discussions and sufficient conditions see Andrews [1987] for the comparable case of generalized Wald tests. Note that Assumption 4 (iv) is more involved than the corresponding assumption in Andrews (1987) due to increasing dimensions of A_n. In 4 (iii) we also restrict the dimension of A_n determined by k_n relative to the size of $\|A_n^-\|$.

[9] We 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 $n \geq 1$.

[10] This 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).

Assumption 4 (v) is satisfied if $\{q_l\}_{l\geq 1}$ forms a Riesz basis in $L^2(\mathbb{R}^{d_b}) \equiv \big\{\phi : \int_{\mathbb{R}^{d_b}} \phi^2(s)ds < \infty\big\}$. The following result establishes asymptotic normality of our standardized test statistic.

Theorem 2.1. *Let Assumptions 1–4 hold with $\delta(V,t) = \exp(itY) - (\mathcal{F}_g f_B)(X,t)$. Then, under H_{mod} we obtain*

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

Remark 2.1 (Estimation of Critical Values)**.** *The asymptotic result of the previous theorem depends on unknown population quantities. As we see in the following, the critical values can be easily estimated. We define $\delta_n(V,t) = \exp(itY) - (\mathcal{F}_g \widehat{f}_{Bn})(X,t)$, and*

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

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

$$\widehat{\mu}_{m_n} = \int tr\Big(\big(\mathbf{X}_n'\mathbf{X}_n\big)^{-1/2}\mathbf{X}_n'\, diag(\sigma_n(t,t))\,\mathbf{X}_n\big(\mathbf{X}_n'\mathbf{X}_n\big)^{-1/2}\Big)\varpi(t)dt$$

and

$$\widehat{\varsigma}_{m_n} = \Big(\int\int \Big\|\big(\mathbf{X}_n'\mathbf{X}_n\big)^{-1/2}\mathbf{X}_n'\, diag(\sigma_n(s,t))\,\mathbf{X}_n\big(\mathbf{X}_n'\mathbf{X}_n\big)^{-1/2}\Big\|_F^2 \varpi(s)\varpi(t)dsdt\Big)^{1/2}.$$

Proposition 2. *Under the conditions of Theorem 2.1, we obtain*

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

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 2.1, Proposition 2, and Lemma 5.1.

Corollary 2.1. *Under the conditions of Theorem 2.1, we obtain*

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

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 4 (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 $\widetilde{q}_l(\cdot) \equiv (-i)^{-l} q_l(\cdot)$ and $\mathbf{X}_n(t) = \big(\widetilde{q}_{\underline{k_n}}(tX_1)', \dots, \widetilde{q}_{\underline{k_n}}(tX_n)'\big)'$. Thus, the estimator of f_B given in (2.5) *simplifies to $\widehat{f}_{Bn}(\cdot) = q_{\underline{k_n}}(\cdot)'\widehat{\beta}_n$ where*

$$\widehat{\beta}_n = \operatorname*{argmin}_{\beta \in \mathbb{R}^{k_n}} \sum_{j=1}^n \int \big|\widehat{\varphi}_n(X_j, t) - \widetilde{q}_{\underline{k_n}}(tX_j)'\beta\big|^2 \varpi(t) dt. \tag{2.10}$$

An explicit solution of (2.10) *is given by*

$$\widehat{\beta}_n = \Big(\int \mathbf{X}_n(-t)'\mathbf{X}_n(t)\varpi(t)dt \Big)^{-} \int \mathbf{X}_n(-t)'\mathbf{\Phi}_n(t)\varpi(t)dt$$

where $\mathbf{\Phi}_n(t) = \big(\widehat{\varphi}_n(X_1, t), \dots, \widehat{\varphi}_n(X_n, t)\big)'$. 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 under the random coefficient specification for the model

For testing degeneracy, Theorem 2.1 is not directly applicable as the required sieve approximation error in Assumption 4 (i) is here not satisfied in general. In contrast, we will impose an approximation condition on the function $\widetilde{g}(x, t, b) \equiv \exp(itg_2(x, b))$ where b belongs to the parameter space $\mathcal{B}_2$.

Let us introduce a $(k_n \cdot l_n)$–dimensional vector valued function χ_n given by $\chi_n(x, t) = (\mathcal{F}_{g_1} q_{\underline{k_n}})(x, t) \otimes \widetilde{p}_{\underline{l_n}}(x, t)$, where $\otimes$ denotes the Kronecker product and $\widetilde{p}_{\underline{l_n}}$ is a l_n–dimensional vector of complex valued basis functions used to approximate $\widetilde{g}(\cdot, \cdot, b)$. For instance, if $g_2(x, b) = \phi(x)\psi(b)$ then approximation conditions can be easily verified due to $\widetilde{g}(x, t, b) = \sum_{l \geq 0} \widetilde{p}_l(x, t)\psi(b)^l$ where $\widetilde{p}_l(x, t) = \big(it\phi(x)\big)^l / l!$. Let us introduce $\mathcal{A}_n = \int E[\chi_n(X, -t)\chi_n(X, t)']\varpi(t)dt$ and its empirical analog $\widehat{\mathcal{A}_n} = n^{-1} \int \sum_{j=1}^n \chi_n(X_j, -t)\chi_n(X_j, t)'\varpi(t)dt$. Recall that $\mathcal{B}_{-2,n} = \big\{\phi(b) = \sum_{l=1}^{k_n} \beta_l q_l(b)$ for $b \in \mathbb{R}^{d_{b_2}}\big\}$ where d_{b_2} denotes the dimension of b_2 and let $\mathcal{G}_{2,n} = \big\{\phi(x, t) = \sum_{l=1}^{l_n} \beta_l \widetilde{p}_l(x, t)\big\}$. In the following, we introduce a strictly positive, nonincreasing sequence

$(\tau_n)_{n\geq 1}$ such that $\tau_n\|\mathcal{A}_n^-\|^2 = O(1)$.

Assumption 5. *(i) The hypothesis H_{add} holds. (ii) The set of parameters b_2 satisfying* (2.6) *belongs to a compact parameter space $\mathcal{B}_2 \subset \mathbb{R}^{d_{b_2}}$. (iii) For any $b \in \mathcal{B}_2$ there exists $\Pi_{l_n}\widetilde{g}(\cdot,\cdot,b) \in \mathcal{G}_{2,n}$ satisfying $n\|\Pi_{l_n}\widetilde{g}(\cdot,\cdot,b) - \widetilde{g}(\cdot,\cdot,b)\|_\varpi^2 = o(\sqrt{m_n})$. (iv) For any p.d.f. $f_{B_{-2}}$ satisfying* (2.6) *there exists $\Pi_{k_n}f_{B_{-2}} \in \mathcal{B}_{-2,n}$ such that $n\|\mathcal{F}_{g_1}(\Pi_{k_n}f_{B_{-2}} - f_{B_{-2}})\|_\varpi^2 = o(\sqrt{m_n})$. (v) It holds $k_n l_n \log n = o(\tau_n\sqrt{m_n})$. (vi) It holds $P\big(rank(\mathcal{A}_n) = rank(\widehat{\mathcal{A}_n})\big) = 1 + o(1)$. (vii) There exists a constant $C > 0$ such that $\sum_{l,l'\geq 1}\langle \mathcal{F}_{g_1}q_l \cdot \widetilde{p}_l, \phi\rangle^2 \leq C\|\phi\|_\varpi^2$ for all functions ϕ with $\|\phi\|_\varpi < \infty$.*

Assumption 5 (i) states the maintained hypothesis of an additive structure of g given in equation (2.1). Assumption 5 (iii) states an asymptotic condition of the sieve approximation error for $\widetilde{g}(\cdot,\cdot,b)$ for any b in the parameter space $\mathcal{B}_2$. By doing so, we impose regularity conditions on the integral transform $\mathcal{F}_{g_2}$ of the Dirac measure at b but not on the Dirac measure itself. For instance, if again $g_2(x,b) = \phi(x)\psi(b)$ and $\widetilde{p}_l(x,t) = \big(it\phi(x)\big)^l/l!$ for $l \geq 1$ then $\|\Pi_{l_n}\widetilde{g}(\cdot,\cdot,b) - \widetilde{g}(\cdot,\cdot,b)\|_\varpi \leq C/(l_n+1)!$ for some constant $C > 0$, provided that $E[\phi^{l_n}(X)]\,\psi^{l_n}(b)\int t^{l_n}\varpi(t)dt$ is bounded. Assumption 5 (iv) requires an appropriate sieve approximation error only for any nondegenerate p.d.f. $f_{B_{-2}}$ satisfying (2.6). This assumption is a modification of Assumption 4 (i), which does not hold under $H_{\deg}$ as degenerate distributions cannot be accurately approximated by basis functions. Assumption 5 (v) restricts the magnitude of k_n also relative to the dimension parameter l_n, which is not too restrictive as the dimension k_n is used to approximate a lower dimensional p.d.f. than in Theorem 2.1. Assumption 5 (vi) and (vii), respectively, are closely related to Assumption 4 (iv) and (v).

Theorem 2.2. *Let Assumptions 1–3, 4 (ii), and 5 be satisfied with $\delta(V,t) = \exp(itY) - (\mathcal{F}_{g_1}f_{B_{-2}})(X,t)\,\widetilde{g}(X,t,b_2)$. Then, under H_{deg} we obtain*

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

The critical values can be estimated as in Remark 2.1 but where now $\delta_n(V,t) = \exp(itY) - (\mathcal{F}_{g_1}\widehat{f}_{B_{-2n}})(X,t)\,\widetilde{g}(X,t,\widehat{b}_{2n})$. The following result shows that, by doing so, the asymptotic distribution of our standardized test statistic remains unchanged. This corollary follows directly from Theorem 2.2 and the proof of Proposition 2; hence we omit its proof.

Corollary 2.2. *Under the conditions of Theorem 2.2 it holds*

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

Remark 2.2 (Comparison to Andrews [2001])**.** *It is instructive to compare our setup and* *results to Andrews [2001], who considers the random coefficient model:*

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

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

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. Such an equivalent characterization 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].*

2.3.4 Testing degeneracy under additive separability alone

We now establish the asymptotic distribution of our test of degeneracy based on separability but not full knowledge of g_1 (see Example 3). We introduce the function $h(\cdot, t) = E[\exp(it(Y - g_2(X, b_2))|X_1 = \cdot]$ and a linear sieve space $\mathcal{H}_n \equiv \{\phi : \phi(x_1) = \sum_{l=1}^{k_n} \beta_l p_l(x_1) \text{ for } x_1 \in \mathbb{R}^{d_{x_1}}\}$ where d_{x_1} denotes the dimension of X_1. The series least squares estimator of h is denoted by $\widehat{h}_n(\cdot) = p_{\underline{k_n}}(\cdot)'\big(\mathbf{X}_{1n}'\mathbf{X}_{1n}\big)^{-1}\mathbf{X}_{1n}'\mathbf{U}_n$ where $\mathbf{U}_n = \big(\exp(it(Y_1 - g_2(X_1, \widehat{b}_{2n}))), \dots, \exp(it(Y_n - g_2(X_n, \widehat{b}_{2n})))\big)'$ and $\widehat{b}_{2n}$ denotes an estimator of b_2. Recall the notation $\widetilde{g}(x, t, b) \equiv \exp(itg_2(x, b))$ for $b \in \mathcal{B}_2$. Below we denote the vector of partial derivatives of $\widetilde{g}$ with respect to b by $\widetilde{g}_b$.

Assumption 6. *(i) The hypothesis* H_{add} *holds, where* g_1 *need not to be known except that it does not depend on* X_2. *(ii) There exists* $\Pi_{k_n} h \in \mathcal{H}_n$ *such that* $n\|\Pi_{k_n} h - h\|_{\varpi}^2 = o\big(\sqrt{m_n}\big)$. *(iii) The parameter* b_2 *is point identified and belongs to the interior of a compact parameter space* $\mathcal{B}_2 \subset \mathbb{R}^{d_{b_2}}$. *(iv) There exists an estimator* $\widehat{b}_{2n}$ *such that* $\sqrt{n}(\widehat{b}_{2n} - b_2) = O_p(1)$ *(v) The function* $\widetilde{g}$ *is partially differentiable with respect to* b *and* $\int E \sup_{b \in \mathcal{B}_2} \|\widetilde{g}_b(X, t, b)\|^2 \varpi(t) dt < \infty$. *(vi) It holds* $k_n = o(\sqrt{m_n})$.

Assumption 6 (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 $\widehat{h}_n$. To see this, let $\|\Pi_{k_n} h - h\|_{\varpi} = O(k_n^{-s/d_{x_1}})$ for some constant $s > 0$. Assumptions 6 (ii) and (vi) 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 $2d_{x_1} \geq d_x$. Under a partially linear structure $H_{\text{part-lin}}$, Assumptions 6 (iv) is automatically satisfied if $\widehat{b}_{2n}$

coincides with the nonlinear least squares estimator. If g_2 is linear, Assumption 6 (*iv*) holds true if $E\|X\|^2 < \infty$ and $\int t^2\, \varpi(t)dt < \infty$.

Theorem 2.3. *Let Assumptions 1–3, 4 (ii), and 6 hold, with* $\delta(V,t) = \exp(itY) - h(X_1,t)\widetilde{g}(X,t,b_2)$. *Then, under* H_{deg} *we obtain*

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

The critical values can be estimated as in Remark 2.1 but where now $\delta_n(V,t) = \exp(itY) - \widehat{h}_n(X_1,t)\exp(itg_2(X,\widehat{b}_{2n}))$. The following result shows that, by doing so, the asymptotic distribution of our standardized test statistic remains unchanged. This corollary follows directly from Theorem 2.3 and the proof of Proposition 2; hence we omit its proof.

Corollary 2.3. *Under the conditions of Theorem 2.3 it holds*

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

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 for all random parameters B. In case of testing functional form restrictions, this is equivalent to a failure of the null hypothesis H_{mod}, i.e., $P\big(Y \neq g(X,B) \quad \text{for all distributions of random parameters } B\big) > 0$. 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_1 and g_2 as discussed in Example 2). The next proposition shows that our test of functional form restrictions has the ability to reject a failure of the null hypothesis H_{mod} with probability one as the sample size grows to infinity.

Proposition 3. *Suppose that* H_{mod} *is false and let Assumptions 1–4 be satisfied. Consider a sequence* $(\gamma_n)_{n\geq 1}$ *satisfying* $\gamma_n = o(n\varsigma_{m_n}^{-1})$. *Then, we have*

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

Recall that under Assumption 3 we have $\varsigma_{m_n} \geq C\sqrt{m_n}$ (see Lemma 5.1). Hence, under this assumption, the rate requirement $\gamma_n = o(n\varsigma_{m_n}^{-1})$ implies $\gamma_n = o(n/\sqrt{m_n})$ which implies $\gamma_n^{-1} = o(1)$.

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 $\mathcal{I}_g = \big\{f : f \text{ is a p.d.f. with } \varphi = \mathcal{F}_g f\big\}$. We assume that $\mathcal{I}_g$ is not empty and denote by f_B^* the p.d.f. in $\mathcal{I}_g$ with minimal norm. We consider the following sequence of local alternatives

$$\varphi_n = \mathcal{F}_g\big(f_B^* + \Delta\sqrt{\varsigma_{m_n}/n}\big), \tag{2.11}$$

for some function $\Delta \in L^1(\mathbb{R}^{d_b}) \cap L^2(\mathbb{R}^{d_b})$. Here, we assume that Δ is such that $f_B^* + \Delta\sqrt{\varsigma_{m_n}/n}$ does not belong to the identified set $\mathcal{I}_g$ and need not to be a density. We also note that the p.d.f. f_B^* coincides with the minimal norm solution of $\|\varphi_n - \mathcal{F}_g f\|_\varpi$ as n tends to infinity. The next result establishes asymptotic normality under (2.11) of the standardized test statistic S_n for testing functional form restrictions.

Proposition 4. *Let the assumptions of Theorem 2.1 be satisfied. Then, under* (2.11) *we obtain*

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

As we see from Proposition 4, our test can detect linear alternatives at the rate $\sqrt{\varsigma_{m_n}/n}$. Results for testing degeneracy follow similarly. In the following, we thus study deviations from the hypothesis of degeneracy only under the maintained hypothesis $H_{\text{lin}} : Y = B_0 + B_1'X_1 + B_2'X_2$. Under the maintained hypothesis of linearity, any deviation between the 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

$$f_{Bn} = f_{B_{deg}} + \Delta\sqrt{\varsigma_{m_n}/n}, \tag{2.12}$$

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 (2.12) 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 (2.12) for the standardized test statistic S_n for testing degeneracy. This corollary follows by similar arguments used to establish Proposition 4 and hence we omit the proof.

Corollary 2.4. *Let the assumptions of Theorem 2.3 be satisfied. Then, under* (2.12) *we obtain*

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

3 Monte Carlo Experiments

In this section, we study the finite-sample performance of our test by presenting the results of a Monte Carlo simulation study. 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.

3.1 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

$$Y = \eta B_0 + X B_1, \tag{3.1}$$

the quadratic model

$$Y = c_1(\eta B_0 + X B_1 + X^2 B_2), \tag{3.2}$$

or the nonlinear model

$$Y = c_2(\eta B_0 + X B_1 + \sqrt{|X|}\, B_2), \tag{3.3}$$

where the constant η is either 0.7 or 1. Here, the normalization constants c_1 and c_2 ensure that the dependent variables in models (3.1)–(3.3) have the same variance.[11] Note that the random coefficient density f_B is neither point identified in model (3.2) nor in model (3.3). 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, testing linearity in the

[11] This normalization ensures that large empirical rejection probabilities are not only driven by a large variance of the alternative models (see, for instance, Blundell and Horowitz [2007]).

heterogeneous QUAIDS model (3.2), where the first two conditional moments yield $E[B_2] = 0$ and $Var(B_2) = 0$. Consequently, $P\big(\int |\varepsilon(X,t)|^2 \varpi(t)dt \neq 0\big) > 0$ if and only if $P(B_2 \neq 0) > 0$. In the finite sample experiment, we also observe that our testing procedure is able to detect such deviations.

rows	*Null Model* H_{mod}	*Alt. Model* *True DGP*	η	k_n	*Empirical Rejection probabilities using* $m_n = 8$	$m_n = 12$	$m_n = 16$
1	(3.1)		0.7	5	0.041	0.006	0.004
2	(3.2)			7	0.120	0.063	0.020
3	(3.1)	(3.2)		5	0.958	0.727	0.561
4	(3.1)	(3.3)			0.645	0.233	0.129
5	(3.2)	(3.1)		7	0.935	0.780	0.558
6	(3.2)	(3.3)			0.990	0.903	0.734
7	(3.1)		1	5	0.093	0.014	0.002
8	(3.2)			7	0.290	0.120	0.051
9	(3.1)	(3.2)		5	0.876	0.513	0.327
10	(3.1)	(3.3)			0.550	0.146	0.053
11	(3.2)	(3.1)		7	0.994	0.952	0.837
12	(3.2)	(3.3)			0.996	0.966	0.866
13	(3.1)		0.7	6	0.019	0.003	0.001
14	(3.2)			9	0.140	0.049	0.023
15	(3.1)	(3.2)		6	0.887	0.539	0.313
16	(3.1)	(3.3)			0.524	0.161	0.064
17	(3.2)	(3.1)		9	0.938	0.778	0.581
18	(3.2)	(3.3)			0.986	0.893	0.756
19	(3.1)		1	6	0.042	0.004	0.003
20	(3.2)			9	0.292	0.103	0.042
21	(3.1)	(3.2)		6	0.847	0.465	0.261
22	(3.1)	(3.3)			0.364	0.085	0.037
23	(3.2)	(3.1)		9	0.991	0.952	0.833
24	(3.2)	(3.3)			0.994	0.957	0.859

Table 1: *Rows 1,2,7,8, 13, 14, 19, 20 depict the empirical rejection probabilities if H_{mod} holds true, the rows 3–6, 9–12, 15–18, 21–24 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. Column 4 depicts the values of the varying dimension parameters k_n. Columns 5–7 depict the empirical rejection probabilities for the nominal level* 0.05.

The test is implemented using Hermite functions, and uses the standardization described in Remark 2.1. When (3.1) is the true model, we estimate the random coefficient density as described in Example 4, where we make use of the fact that the Hermite functions are eigenfunctions of the Fourier transform. If (3.2) is the true model, the integral transform $\mathcal{F}_g$ is computed using numerical integration. In both cases, the weighting function ϖ is given by the standard normal p.d.f., following Su and White [2007] and Chen and Hong [2010], or following Chen et al. [2013] by the uniform p.d.f. with support $[-2, 2]$. We also tried different weighting functions and found, similarly to Chen and Hong [2010], that the results of our finite sample analysis are not sensitive when these functions have support on the whole real line. For finite support weight functions the results are equally sensitive and thus we report the empirical rejection probabilities of our tests using the uniform weights only in the Supplementary Material.

Our test statistic is implemented using a varying number of Hermite functions to analyze its sensitivity to that dimension parameter choice. If (3.1) is the correct model, we use either $k_n = 5\,(= 3 + 2)$ or $k_n = 6\,(= 3 + 3)$ Hermite functions to estimate the density of the bivariate random coefficients (B_0, B_1). If (3.2) is the correct model, we have an additional dimension which accounts for the nonlinear part. Here, the choice of Hermite basis functions is either $k_n = 7\,(= 3 + 2 \cdot 2)$ or $k_n = 9\,(= 3 + 2 \cdot 3)$. In both cases we vary the dimension parameter m_n between 8, 12, and 16.

The empirical rejection probabilities of our tests are shown in Table 1 at the nominal level 0.05. 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, under the correct functional form restrictions, is accurate for $m_n = 8$ if the linear model is the correct model (see rows 1, 7, 13, and 19) while for the correct quadratic model we require a large value of m_n to obtain accurate finite sample coverage (see rows 2, 8, 14, and 20). This is not surprising but in line with our theory, where we require m_n to be larger than k_n and the quadratic model requires a larger choice of k_n.

From Table 1 we see that the empirical rejections probabilities become larger as the parameter η increases. 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 (3.2) more often than the nonlinear model (3.3). From rows 5, 11 and 6, 12 we see that our test rejects the nonlinear model (3.3) slightly more often than the linear model (3.1).

Note that m_n could be any integer larger than *const.* $\times\, k_n^2$ that is smaller than $n^{1/2}$ (up to logs). The range of admissible dimension parameters for this minimization-maximization routine reflects the dimension restrictions imposed in Theorem 2.1 and the consistency results thereafter, i.e., $m_n^2 \log n = o(n)$ and $k_n = o(\sqrt{m_n})$.[12] From Table 1 we see that the condition $k_n = o(\sqrt{m_n})$ might be too restrictive in finite samples when Hermite functions are used.[13] We thus modify the range of possible dimension parameters to ensure accurate finite sample coverage. 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 $\min_{1 \leq k_n < 2n^{1/4}} \max_{k_n < m_n < \sqrt{n}} \{s(k_n, m_n)\}$. The intuition behind this criterion is that we choose k_n to have a good model fit and to choose m_n such that the finite sample power of the test statistic is maximized. For instance, as we see from Table 1, if $\eta = 1$ and (3.1) is the correct model, the principle yields $k_n = 6$ and $m_n = 8$ which implies an empirical rejection probability of 0.042 (see row 19). The minimum-maximum principle also ensures that k_n is always smaller than m_n and thus precludes inaccurate finite sample coverage in the quadratic model due to too small m_n as we see in rows 2, 8, 14, and 20. For instance, if $\eta = 0.7$ and (3.2) is the true model the principle yields $k_n = 9$ and $m_n = 12$ leading to the empirical rejection probability of 0.049 (see row 14). Yet for larger values of η, i.e. if $\eta = 1$ and (3.2) is the true model, the principle yields again $k_n = 9$ and $m_n = 12$ leading to the empirical rejection probability of 0.103. Thus, the testing procedure works generally well but leads in some cases to overrejection (see row 20).

When we consider different data generating processes, such as a cubic polynomial with random coefficients, we find that our test of linearity leads to empirical rejection probabilities which are close to one for all nominal levels considered. Hence, these results are not reported here. Regarding consistency of the test statistic, we conduct experiments with increasing sample sizes. We find a slight tendency of our test statistic to under-reject for small η, see in Table 1 in rows 1, 2, 13, and 14. However, this under-rejection diminishes as we increase the sample size to $n = 1000$. Not surprisingly, when $n = 1000$ also the empirical rejection probabilities in in alternative models increase.

Recommendation on choice of tuning parameters. In the following, based on the theoretical results and the Monte Carlo investigation we provide a recommendation on the choice of weighting function and dimension parameters to implement the test in practice.

- Concerning the weighting function ϖ, choosing a standard normal p.d.f. performs well in many different settings, and should probably be considered as a benchmark. However, the results in the simulation section suggest that the choice of weighting is immaterial, as the results do not appear to be sensitive.

- In contrast, the test appears to be significantly more sensitive to the choice of dimension

[12]For simplicity, we assume here the the minimal eigenvalues of the associated matrices are uniformly bounded away from zero.

[13]This rate requirement is not too restrictive for B-spline basis functions as we see below.

parameters k_n and m_n. In particular, the test appears more sensitive to the choice of k_n than to the choice of m_n. We recommend to choose the dimension parameters k_n and m_n according to the minimum-maximum principle as proposed above, i.e., choose m_n to maximize the finite sample power of the test, and k_n to minimize the specification error.

3.2 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

$$A = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} \quad \text{and} \quad A_\rho = \begin{pmatrix} 2 & \rho \\ \rho & 2 \end{pmatrix},$$

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

$$Y = 0.25 \cdot B_2^\kappa + B_1^\kappa X_1 + X_2,$$

if the null hypothesis H_{deg} holds, where the constant κ is either 1 or 2 in the experiments below. For the alternative, we generate the dependent variable Y using

$$Y = 0.25 \cdot B_2^\kappa + B_1^\kappa X_1 + \eta B_2^\kappa X_2,$$

for some constants $\eta > 0$, and κ which vary in the simulations below.

The test is implemented as described in Example 2 with B–splines, and uses the standardization described in Remark 2.1 with $\delta_n(V, t) = \exp(itY) - \widehat{h}_n(X_1, t) \exp(itg_2(X, \widehat{b}_{2n}))$. This means that we use the more general test that allows for a nearly arbitrary specification in the remaining model $Y - g_2(X_2, b_2)$. We focus in the simulation on this specification, because it has arguably less power than the more specific one that imposes in addition the linear random coefficients structure. However, as will be evident from the results below, this test already has very good power properties, implying that separating the term involving the fixed coefficient turns out to already be a powerful device in testing. To estimate the restricted conditional characteristic function, we use B–splines of order 2 with one or two knots (hence, $k_n = 4$ or $k_n = 5$), and for the unrestricted one a tensor-product of these B–spline basis functions (hence, $m_n = 16$, $m_n = 20$, or $m_n = 25$). We do not consider larger values for k_n, because we want to ensure the requirement $k_n^2 \leq m_n$, see also the minimum maximum principle below.

The empirical rejection probabilities for testing degeneracy are shown in Table 2 at the nominal level 0.05. Again we normalize the models to ensure that the null and alternative have

rows	κ	ρ	Alt. Model η	*Empirical Rejection probabilities using* $k_n = 4$, $m_n = 16$	$k_n = 4$, $m_n = 20$	$k_n = 5$, $m_n = 20$	$k_n = 5$, $m_n = 25$	*Andrews' Test*
1	1	1		0.053	0.016	0.017	0.003	0.060
2			0.3	0.184	0.135	0.009	0.009	0.974
3			0.5	0.489	0.419	0.107	0.076	0.999
4			0.7	0.783	0.709	0.400	0.344	1.000
5		1.5		0.110	0.019	0.018	0.003	0.070
6			0.3	0.352	0.269	0.032	0.024	0.976
7			0.5	0.760	0.691	0.291	0.229	1.000
8			0.7	0.953	0.946	0.692	0.660	1.000
9	2	1		0.102	0.050	0.058	0.032	0.220
10			0.3	0.461	0.389	0.344	0.335	0.822
11			0.5	0.852	0.810	0.786	0.756	0.983
12			0.7	0.984	0.969	0.970	0.945	0.996
13		1.5		0.178	0.083	0.103	0.046	0.190
14			0.3	0.601	0.507	0.512	0.448	0.801
15			0.5	0.925	0.890	0.894	0.865	0.984
16			0.7	0.995	0.991	0.991	0.978	1.000

Table 2: *The first row depicts the empirical rejection probabilities under degeneracy of the coefficient of X_2, the rows 2–4, 6–8, 10–12, and 14–16 show the finite sample power of our tests against various alternatives. Column 1 depicts the value of κ in the correct and alternative models. Column 2 specifies the covariance of B_1 and B_2 for the alternative models. Column 3 depicts the value of η in the correct model and is empty if the null model is correct. Columns 4–7 depict the empirical rejection probabilities for the nominal level* 0.05. *Column 8 depicts the empirical rejection probabilities using the quasi-likelihood ratio test proposed by Andrews [2001].*

the same variance. The differences between the nominal and empirical rejection probabilities are small under a fixed coefficient for X_2, as is obvious from the first row. In Table 2, we also see from rows 2–4, 6–8, 10–12, and 14–16 that our test rejects the alternative model more 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 and 10–12 with 14–16.

In case of B-spline basis functions we need to be less restrictive regarding the size of the dimension parameters and recommend the following criterion. Again if $s(k_n, m_n)$ denotes the value of the test statistic, we consider the modified minimum-maximum principle $\min_{1 \leq k_n < 2n^{1/4}} \max_{k_n^2 \leq m_n < 2\sqrt{n}} \{s(k_n, m_n)\}$. For instance, as we see from Table 2, if $\kappa = 2$, $\rho = 1$ and B_2 is deterministic, this principle yields $k_n = 5$ and $m_n = 20$ which implies an empirical rejection probability of 0.058 (see row 9). Yet when ρ is larger, i.e., $\rho = 1.5$ and again $\kappa = 2$, the minimum-maximum principle leads an empirical rejection probability of 0.103 and thus, the testing procedure leads occasionally to overrejection.

In Table 2, we compare our testing procedure to the quasi-likelihood ratio test proposed by Andrews [2001]. In both settings, conditional mean independence of random slope and intercept parameters is violated. We see that this violation of Andrews [2001] framework leads inaccurate empirical rejection probabilities, in particular, in the second case. We see that for $\kappa = 1$, the quasi-likelihood ratio test of Andrews [2001] is more powerful than the normalized statistic S_n. When $\kappa = 2$, however, the statistic of Andrews has inaccurate finite sample coverage, see rows 9 and 13, due to misspecification.

Recommendation on choice of tuning parameters. In the following, based on the theoretical results and the Monte Carlo investigation we provide a recommendation on the choice of weighting function and dimension parameters to implement the test in practice.

- As above we recommend choosing the weighting function ϖ to be the standard normal p.d.f.
- In contrast, the test appears to be significantly more sensitive to the choice of dimension parameters k_n and m_n. In particular, the test appears more sensitive to the choice of k_n than to the choice of m_n. We recommend to choose the dimension parameters k_n and m_n according to the modified minimum-maximum principle as proposed above, i.e., choose m_n to maximize the finite sample power of the test, and k_n to minimize the specification error.

4 Application

4.1 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 expenditure[14]. The model is linear in log prices and a term that involves log total expenditure over a nonlinear price index that depends on parameters of the utility function. In applications, one frequent shortcut is to replace this utility dependent price index by a conventional price index (e.g., Laspeyres), another is that

[14] The 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].

homogeneity of degree zero is imposed, which means that all prices and total expenditure are relative to a price index, resulting in an entirely linear model.

A popular extension of 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 fixed. 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. However, we want to allow for preference heterogeneity, and hence consider the following model:

$$BSF_i = B_{0i} + B_{1i}\log(TotExp_i) + B_{2i}\big(\log(TotExp_i)\big)^2 + b_4 W_{1i} + b_5 W_{2i}. \tag{4.1}$$

Unobserved heterogeneity is reflected in the three random coefficients B_{0i}, B_{1i} and B_{2i}. 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}. These principal components are obtained through two different linear combinations of the original covariates S_i, i.e., $W_{1i} = \lambda_1' S_i$ and $W_{2i} = \lambda_2' S_i$, where λ_1, λ_2 are the first two loadings, and are computed using the R command `princomp`. While including 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. Moreover, the additive specification can be justified as letting the mean of the random intercept B_{0i} depend on covariates.

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 3. For testing functional form restrictions, our test is implemented as described in Example 1, where in the linear case we employ the estimation procedure in Example 4. In both cases, we choose the dimension parameters k_n and m_n by the minimum-maximum principle explained in the Monte Carlo section.

4.2 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. Since the data are similar to the data used in Hoderlein (2011), for brevity of exposition we refer to this paper for additional descriptive statistics, especially regarding household covariates.

	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

4.3 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, which means that mean effects are rather insignificant. A potential role of the nonlinear term more generally is, however, picked up by our procedure. Table 3 shows the different values of the test statistics and p-values. We evaluate results using a nominal level of 0.05. As we see from Table 3, our test fails to reject the model (4.1) with degenerate B_{2i} but rejects the linear random coefficient model where $B_{2i} = 0$. Not surprisingly, 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.[15] It is interesting to note that the procedure selects higher order basis functions to account for the random coefficient of the quadratic term. Since higher order basis functions are required to estimate sharp peaks, this also supports the hypothesis that the marginal p.d.f. B_2 is akin to a Dirac measure (i.e., the distribution is

[15]Consequently, we choose k_n to minimize the value of our statistic and m_n to maximize the value the test statistic over the range $1 \leq k_n \leq 2n^{1/4}$ and $k_n < m_n < \sqrt{n}$.

degenerate), or very close it. We also performed a quasi-likelihood ratio test of degeneracy as proposed by Andrews [2001] and obtained the value of the test statistic of 0.243 with critical value 2.706.[16] Consequently, again we fail to reject the hypothesis of a quadratic model with fixed coefficient on the quadratic term.

Null Hypothesis	linear RC	quadratic RC	RC with fixed coeff. on quadratic term in TotExp
value of test	2.1289	1.4200	1.4029
p-values	0.0166	0.0778	0.0803

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 (4.1)), a quadratic random coefficient model (i.e., random B_{2i} in (4.1)), or a random coefficient model with degenerate coefficient on the quadratic term (i.e., $B_{2i} = b_2$ in (4.1) 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 of the functional form assumptions, but because of a violation of exogeneity of total expenditure. Therefore, we follow Imbens and Newey [2009], and model 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], who propose a life-cycle structural model that yields this specification. Generally, the control function U would have to be estimated in a first stage. Since the

[16] Note that all the hypotheses under consideration would have been rejected under a nominal level of 0.1. However, given the recent discussion of lowering the significance levels, see Benjamin et al. [2017], we feel that a level of 0.1 is not meaningful here.

theory involving pre-estimation is beyond the scope of this paper, we do not adjust for estimation error in this variable, which may lead to a higher variance (depending on the smoothness assumptions).

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	1.3747
p–values	0.0194	0.0810	0.0846

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}.

5 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 $\langle\phi, \psi\rangle_X \equiv E[\phi(X)\psi(X)]$. For any integer $m \geq 1$, $\mathcal{I}_n$ denotes the $m_n \times m_n$ dimensional identity matrix. Recall the notation $P_n = E[p_{\underline{m_n}}(X)p_{\underline{m_n}}(X)']$.

Proofs of Section 2.

Proof of Proposition 1. 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

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

Consequently, we obtain $\lambda_n \mathcal{I}_n \lesssim P_n$. $\square$

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

Lemma 5.1. *Under Assumption 2 (ii) it holds* $\sqrt{m_n} \lesssim \varsigma_{m_n}$.

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

$$\begin{aligned}\varsigma_{m_n}^2 &= \int\int \Big\| P_n^{-1/2} E\big[\delta(V,s)\overline{\delta(V,t)} p_{\underline{m_n}}(X)\, p_{\underline{m_n}}(X)'\big] P_n^{-1/2} \Big\|_F^2 \varpi(s)\varpi(t)dsdt\\ &\geq \lambda_n^{-2} \sum_{l=1}^{m_n} \int\int \Big| E\big[\delta(V,s)\overline{\delta(V,t)} p_l^2(X)\big]\Big|^2 \varpi(s)\varpi(t)dsdt\\ &\geq \lambda_n^{-2} \sum_{l=1}^{m_n} \Big(E\big[|\int \delta(V,t)\varpi(t)dt|^2 p_l^2(X)\big]\Big)^2 && \text{(by Jensen's inequality)}\\ &\geq C\lambda_n^{-2} \sum_{l=1}^{m_n} \big(E[p_l^2(X)]\big)^2 && \text{(by Assumption 3)}\\ &= Cm_n.\end{aligned}$$

$\square$

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 $\widehat{\gamma}_n(t) \equiv (n\widehat{P}_n)^{-1}\sum_j \exp(itY_j)p_{\underline{m_n}}(X_j)$. 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 its population counterpart $A_n = E\big[\int (\mathcal{F}_g q_{\underline{k_n}})(X,-t)(\mathcal{F}_g q_{\underline{k_n}}')(X,t)\varpi(t)dt\big]$. Recall the definition $\widehat{\beta}_n = (n\widehat{A}_n)^{-}\sum_j \int (\mathcal{F}_g q_{\underline{k_n}})(X_j,-t)\widehat{\varphi}_n(X_j,t)\varpi(t)dt$ and further, we introduce $\beta_n = A_n^{-}\int E[(\mathcal{F}_g q_{\underline{k_n}})(X,-t)\varphi(X,t)]\varpi(t)dt$.

Proof of Theorem 2.1. We make use of the decomposition

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

Consider I_n. We conclude

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

Since $(\varPi_{m_n}\varphi(X,t) - \varphi(X,t))p_{\underline{m_n}}(X)$ is a centered random variable for all t it is easily seen that $B_{1n} = \lambda_n^{-1}\int \big\|n^{-1/2}\sum_j \big(\exp(itY_j) - \varphi(X_j,t)\big)p_{\underline{m_n}}(X_j)\big\|^2\varpi(t)dt + o_p(1)$. Thus, Lemma 5.2 yields $(\sqrt{2}\varsigma_{m_n})^{-1}\big(B_{1n} - \mu_{m_n}\big) \xrightarrow{d} \mathcal{N}(0,1)$. To show that $B_{2n} = o_p(\sqrt{m_n})$ note that

$$\|\widehat{P}_n^{-1} - \lambda_n^{-1}\mathcal{I}_n\| \le \lambda_n^{-1}\|(\widehat{P}_n/\lambda_n)^{-1}\|\|\mathcal{I}_n - \widehat{P}_n/\lambda_n\| = O_p\Big(\sqrt{(m_n\log n)/(n\lambda_n^2)}\Big)$$

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

$1 \le l \le m_n$, we deduce

$$n^{-1}\int E\Big\|\sum_j \big(\exp(itY_j) - \Pi_{m_n}\varphi(X_j,t)\big)p_{\underline{m_n}}(X_j)\Big\|^2\varpi(t)dt$$
$$\lesssim \int \varpi(t)dt\, E\|p_{\underline{m_n}}(X)\|^2 + \sup_{x\in\mathcal{X}}\|p_{\underline{m_n}}(x)\|^2\sum_{l=1}^{m_n}\int \big\langle\varphi(\cdot,t),p_l\big\rangle_X^2\varpi(t)dt\, E[p_l^2(X)]$$
$$\lesssim m_n\lambda_n. \tag{5.1}$$

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

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

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

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

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

$$n(\overline{\widehat{\beta}_n - \beta_n})'\widehat{A}_n(\widehat{\beta}_n - \beta_n) \lesssim n(\overline{\widehat{\beta}_n - \widetilde{\beta}_n})'\widehat{A}_n(\widehat{\beta}_n - \widetilde{\beta}_n) + n(\overline{\widetilde{\beta}_n - \beta_n})'\widehat{A}_n(\widetilde{\beta}_n - \beta_n)$$
$$\lesssim \Big\|n^{-1/2}\sum_j\int(\mathcal{F}_g q_{\underline{k_n}})(X_j,-t)\big(\widehat{\varphi}_n(X_j,t) - \varphi(X_j,t)\big)\varpi(t)dt\Big\|^2\|\widehat{A}_n^-\|$$
$$+ n\Big\|\int E\big[(\mathcal{F}_g q_{\underline{k_n}})(X,-t)\varphi(X,t)\big]\varpi(t)dt\Big\|^2\|\widehat{A}_n^- - A_n^-\|^2\|\widehat{A}_n\|$$
$$+ \Big\|n^{-1/2}\sum_j\int\Big((\mathcal{F}_g q_{\underline{k_n}})(X_j,-t)\varphi(X_j,t) - E\big[(\mathcal{F}_g q_{\underline{k_n}})(X,-t)\varphi(X,t)\big]\Big)\varpi(t)dt\Big\|^2\|A_n^-\|^2\|\widehat{A}_n\|.$$

From Lemma 5.3 we have $\|\widehat{A}_n^- - A_n^-\| = O_p\big(\sqrt{(\log n)k_n/(n\tau_n)}\big)$. By Assumption 4 (v) it holds $\sqrt{\tau_n}\|A_n^-\| = O(1)$ and thus, $\|\widehat{A}_n^-\| \le \|\widehat{A}_n^- - A_n^-\| + \|A_n^-\| = O_p(\tau_n^{-1/2})$. Thereby, it is sufficient to consider

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

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

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

where $\mathcal{F}_g^*$ is the adjoint operator of $\overline{\mathcal{F}_g}$ given by $(\mathcal{F}_g^*\phi)(b)=\int E[\exp(-itg(X,b))\phi(X,t)]\varpi(t)dt$. Consequently, we have $n\overline{(\widehat{\beta}_n-\beta_n)}'\widehat{A}_n(\widehat{\beta}_n-\beta_n)=O_p\big((\log n)k_n/\tau_n+n\|\mathit{\Pi}_{m_n}\varphi-\varphi\|_\varpi^2/\sqrt{\tau_n}\big)=o_p(\sqrt{m_n})$ and, in particular, $III_n=o_p(\sqrt{m_n})$. Consider II_n. From above we infer $n\|\widehat{\beta}_n-\beta_n\|^2=$

$O_p\big((\log n)k_n/\tau_n + n\|\Pi_{m_n}\varphi - \varphi\|_\varpi^2/\sqrt{\tau_n}\big)$. Thereby, we obtain

$$
\begin{aligned}
|II_n|^2 \lesssim \Big| \int \sum_j \big(\exp(itY_j) - \Pi_{m_n}\varphi(X_j,t)\big) p_{\underline{m_n}}(X_j)' \big\langle \varphi(\cdot,t) - (\mathcal{F}_g \Pi_{k_n} f_B)(\cdot,t), p_{\underline{m_n}} \big\rangle_X \varpi(t)dt \Big|^2 \\
+ \Big| \int \sum_j \big(\exp(itY_j) - \Pi_{m_n}\varphi(X_j,t)\big) p_{\underline{m_n}}(X_j)' \big\langle (\mathcal{F}_g(\Pi_{k_n} f_B - \widehat{f}_{Bn})(\cdot,t), p_{\underline{m_n}} \big\rangle_X \varpi(t)dt \Big|^2 + o_p(m_n) \\
\lesssim n \int E\big|\Pi_{m_n}\varphi(X,t) - (\Pi_{m_n}\mathcal{F}_g \Pi_{k_n} f_B)(X,t)\big|^2 \varpi(t)dt \\
+ n \int E\big\|\big(\exp(itY) - \Pi_{m_n}\varphi(X,t)\big)(\Pi_{m_n}\mathcal{F}_g q_{\underline{k_n}})(X,t)\big\|^2 \varpi(t)dt\, \|\widehat{\beta}_n - \beta_n\|^2 + o_p(m_n) \\
= O_p\Big(n\|\mathcal{F}_g(\Pi_{k_n} f_B - f_B)\|_\varpi^2 + k_n((\log n)k_n/\tau_n + n\|\Pi_{m_n}\varphi - \varphi\|_\varpi^2/\sqrt{\tau_n})\Big) + o_p(m_n) \\
= o_p(m_n)
\end{aligned}
$$

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

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

Proof of Proposition 2. To keep the presentation of the proof simple, we do not consider estimation of P_n in $\widehat{\varsigma}_{m_n}$ and $\widehat{\mu}_{m_n}$. We make use of the relationship

$$
\begin{aligned}
\delta_n(\cdot,s)\delta_n(\cdot,t) - \widetilde{\delta}_n(\cdot,s)\widetilde{\delta}_n(\cdot,t) =& \widetilde{\delta}_n(\cdot,s)\big((\mathcal{F}_g\widehat{f}_{Bn})(\cdot,t) - (\mathcal{F}_g\Pi_{k_n} f_B)(\cdot,t)\big) \\
&+ \delta_n(\cdot,t)\big((\mathcal{F}_g\widehat{f}_{Bn})(\cdot,s) - (\mathcal{F}_g\Pi_{k_n} f_B)(\cdot,s)\big).
\end{aligned}
$$

Observe

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

We conclude

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

Here, we used $\|\widehat{\beta}_n-\beta_n\|^2=O_p((\log n)k_n/\tau_n+n\|\varPi_{m_n}\varphi-\varphi\|_\varpi^2/\sqrt{\tau_n})$ which can be seen as in the proof of Theorem 2.1. Since $I_n=o_p(1)$ we conclude

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

by using $\log nk_n=o(\tau_n\sqrt{m_n})$. Finally, it is easily seen that $\varsigma_{m_n}^2-\int\int\|\widetilde{\Sigma}_{m_n}(s,t)\|^2\varpi(s)ds\varpi(t)dt=o_p(1)$, which proves $\varsigma_{m_n}\widehat{\varsigma}_{m_n}^{-1}=1+o_p(1)$. In particular, convergence of the trace of $\widehat{\varSigma}_{m_n}(t,t)$ to the trace of $\varSigma_{m_n}(t,t)$ follows by using $|\widehat{\mu}_{m_n}-\mu_{m_n}|^2\leq m_n\int\|\widehat{\varSigma}_{m_n}(t,t)-\varSigma_{m_n}(t,t)\|_F^2\varpi(t)dt=o_p(m_n)$. □

Proof of Theorem 2.2. Let us introduce $\alpha_n=(n\mathcal{A}_n)^-\int E[\chi_n(X,t)\varphi(X,-t)]\varpi(t)dt$ and the estimator

$$
\widehat{\alpha}_n=(n\widehat{\mathcal{A}_n})^-\int\sum_j\chi_n(X_j,-t)\widehat{\varphi}_n(X_j,t)\varpi(t)dt.
$$

We prove in the following that

$$\sum_j \int \left|\widehat{\varphi}_n(X_j,t) - (\mathcal{F}_{g_1}\widehat{f}_{B_{-2},n})(X_j,t)\,\widetilde{g}(X_j,t,\widehat{b}_{2n})\right|^2 \varpi(t)dt$$
$$= \sum_j \int \left|\widehat{\varphi}_n(X_j,t) - \chi_n(X_j,t)\widehat{\alpha}_n\right|^2 \varpi(t)dt + o_p(\sqrt{m_n}).$$

By the definition of the estimator $\widehat{b}_{2n}$ in (2.7) we obtain

$$\sum_j \int \left|\widehat{\varphi}_n(X_j,t) - (\mathcal{F}_{g_1}\widehat{f}_{B_{-2},n})(X_j,t)\,\widetilde{g}(X_j,t,\widehat{b}_{2n})\right|^2 \varpi(t)dt$$
$$\leq \sum_j \int \left|\widehat{\varphi}_n(X_j,t) - (\mathcal{F}_{g_1}\widehat{f}_{B_{-2},n})(X_j,t)\,\widetilde{g}(X_j,t,b_2)\right|^2 \varpi(t)dt \tag{5.2}$$

for any $b_2 \in \mathcal{B}_2$ satisfying (2.6). By the definition of the least squares estimator $\widehat{\alpha}_n$ and the triangular inequality we obtain

$$\begin{aligned}
&\sqrt{\sum_j \int \left|\widehat{\varphi}_n(X_j,t) - (\mathcal{F}_{g_1}\widehat{f}_{B_{-2},n})(X_j,t)\widetilde{g}(X_j,t,\widehat{b}_{2n})\right|^2 \varpi(t)dt} \\
&\geq \sqrt{\sum_j \int \left|\widehat{\varphi}_n(X_j,t) - (\mathcal{F}_{g_1}\widehat{f}_{B_{-2},n})(X_j,t)\mathit{\Pi}_{l_n}\widetilde{g}(X_j,t,\widehat{b}_{2n})\right|^2 \varpi(t)dt} \\
&\quad - \sqrt{\sum_j \int \left|(\mathcal{F}_{g_1}\widehat{f}_{B_{-2},n})(X_j,t)\big(\mathit{\Pi}_{l_n}\widetilde{g}(X_j,t,\widehat{b}_{2n}) - \widetilde{g}(X_j,t,\widehat{b}_{2n})\big)\right|^2 \varpi(t)dt} \\
&\geq \sum_j \int \left|\widehat{\varphi}_n(X_j,t) - \chi_n(X_j,t)'\widehat{\alpha}_n\right|^2 \varpi(t)dt - O_p\Big(\sqrt{n}\max_{b\in\mathcal{B}_2} \|\mathit{\Pi}_{l_n}\widetilde{g}(\cdot,\cdot,b) - \widetilde{g}(\cdot,\cdot,b)\|_\varpi\Big) \\
&= \sum_j \int \left|\widehat{\varphi}_n(X_j,t) - \chi_n(X_j,t)'\widehat{\alpha}_n\right|^2 \varpi(t)dt - o_p(m_n^{1/4}).
\end{aligned}$$

Consequently, applying again the triangular inequality together with inequality (5.2) yields

$$
\begin{aligned}
&\Big|\sqrt{\sum_j \int \big|\widehat{\varphi}_n(X_j,t) - (\mathcal{F}_{g_1}\widehat{f}_{B_{-2},n})(X_j,t)\widetilde{g}(X_j,t,\widehat{b}_{2n})\big|^2 \varpi(t)dt} \\
&\quad - \sqrt{\sum_j \int \big|\widehat{\varphi}_n(X_j,t) - \chi_n(X_j,t)'\widehat{\alpha}_n\big|^2 \varpi(t)dt}\Big| \\
&\quad \le \sqrt{\sum_j \int \big|\widehat{\varphi}_n(X_j,t) - (\mathcal{F}_{g_1}\widehat{f}_{B_{-2},n})(X_j,t)\widetilde{g}(X_j,t,b_2)\big|^2 \varpi(t)dt} \\
&\quad - \sqrt{\sum_j \int \big|\widehat{\varphi}_n(X_j,t) - \chi_n(X_j,t)'\widehat{\alpha}_n\big|^2 \varpi(t)dt} + o_p(m_n^{1/4}) \\
&\quad \le \sqrt{\sum_j \int \big|\mathit{\Pi}_{l_n}\widetilde{g}(X_j,t,b_2) - \widetilde{g}(X_j,t,b_2)\big|^2 \varpi(t)dt} + \sqrt{n}\,\|\widehat{\alpha}_n - \alpha_n\| + o_p(m_n^{1/4}) \\
&\quad = n\|\mathit{\Pi}_{l_n}\widetilde{g}(\cdot,\cdot,b_2) - \widetilde{g}(\cdot,\cdot,b_2)\|_\varpi^2 + O_p\Big(\sqrt{k_n l_n(\log n)/\tau_n} + \sqrt{n}\tau_n^{-1/4}\,\|\mathit{\Pi}_{m_n}\varphi - \varphi\|_\varpi\Big) + o_p(m_n^{1/4}) \\
&\quad = o_p(m_n^{1/4}),
\end{aligned}
$$

as in the proof of Theorem 2.1. Now following line by line the proof of Theorem 2.1 and using

$$
\begin{aligned}
&\sum_j \int \big|\mathit{\Pi}_{m_n}\varphi(X_j,t) - (\mathcal{F}_{g_1}\mathit{\Pi}_{k_n}f_{B_{-2}})(X_j,t)\mathit{\Pi}_{l_n}\widetilde{g}(X_j,t,b_2)\big|^2 \varpi(t)dt \\
&\quad \lesssim n\|\mathit{\Pi}_{m_n}\varphi - \varphi\|_\varpi^2 + n\|\mathcal{F}_{g_1}\mathit{\Pi}_{k_n}f_{B_{-2}} - \mathcal{F}_{g_1}f_{B_{-2}}\|_\varpi^2 + n\|\mathit{\Pi}_{l_n}\widetilde{g}(\cdot,\cdot,b_2) - \widetilde{g}(\cdot,\cdot,b_2)\|_\varpi^2 + o_p(\sqrt{m_n}) \\
&\quad = o_p(\sqrt{m_n}),
\end{aligned}
$$

the result follows. □

Proof of Theorem 2.3. We make use of the decomposition

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

where we used $\big\langle h(\cdot,t)\widetilde{g}(\cdot,t,b_2), p_{\underline{m_n}}\big\rangle_X = \big\langle \varphi(\cdot,t), p_{\underline{m_n}}\big\rangle_X$. Consider I_n. As in the proof of

Theorem 2.1 we have

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

Thus, Lemma 5.2 yields $(\sqrt{2}\varsigma_{m_n})^{-1}\big(I_n - \mu_{m_n}\big) \xrightarrow{d} \mathcal{N}(0,1)$. Consider III_n. Since $|\widetilde{g}(X_j,t,b)| \leq 1$ for all b we evaluate

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

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

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

by a Taylor series expansion, where $\widetilde{b}_{2n}$ is between $\widehat{b}_{2n}$ and b_2. As in relation (5.1), from $E[\big(\Pi_{k_n}h(X,t) - \exp(it(Y - g_2(X,b_2)))\big)p_{\underline{k_n}}(X)] = 0$ we deduce

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

Further, since $\int E\sup_{b\in\mathcal{B}_2}\big\|\widetilde{g}_b(X,t,b)\big\|^2\varpi(t)dt\leq C$ we have

$$
\begin{aligned}
E\Big(\sum_{l=1}^{k_n}\int\big\|n^{-1}\sum_j \exp(itY_j)\widetilde{g}_b(X_j,t,\widetilde{b}_{2n})p_l(X_j)\big\|^2\varpi(t)dt\Big)^{1/2}\\
&\leq E\Big[\|p_{\underline{k_n}}(X)\|\Big(\int\big\|\widetilde{g}_b(X,t,\widetilde{b}_{2n})\big\|^2\varpi(t)dt\Big)^{1/2}\Big]\\
&\leq \big(E\|p_{\underline{k_n}}(X)\|^2\big)^{1/2}\Big(\int E\sup_{b\in\mathcal{B}_2}\big\|\widetilde{g}_b(X,t,b)\big\|^2\varpi(t)dt\Big)^{1/2}\\
&= O\big(\sqrt{\lambda_n k_n}\big).
\end{aligned}
$$

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

$$
\begin{aligned}
|II_n|\lesssim{}&\Big|\int\sum_j\big(\exp(itY_j)-\mathit{\Pi}_{m_n}\varphi(X_j,t)\big)p_{\underline{m_n}}(X_j)'\big\langle\mathit{\Pi}_{m_n}\varphi(\cdot,t)-\mathit{\Pi}_{k_n}h(\cdot,t)\widetilde{g}(\cdot,t,b_2),p_{\underline{m_n}}\big\rangle_X\varpi(t)dt\Big|\\
&+\Big|\int\sum_j\big(\exp(itY_j)-\mathit{\Pi}_{m_n}\varphi(X_j,t)\big)p_{\underline{m_n}}(X_j)'\big\langle\mathit{\Pi}_{k_n}h(\cdot,t)\widetilde{g}(\cdot,t,b_2)-\widehat{h}_n(\cdot,t)\widetilde{g}(\cdot,t,\widehat{b}_{2n}),p_{\underline{m_n}}\big\rangle_X\varpi(t)dt\Big|\\
&+o_p(\sqrt{m_n})\\
={}&O_p(\sqrt{n}\|\mathit{\Pi}_{k_n}h-h\|_\varpi)+o_p(\sqrt{m_n})\\
&+O_p\Big(\Big(k_n\int E\sup_{b\in\mathcal{B}_2}\big\|\sum_{l=1}^{m_n}p_l(X)\big\langle\widetilde{g}_b(\cdot,t,b)p_{\underline{k_n}}',p_l\big\rangle_X\big\|^2\varpi(t)dt\Big)^{1/2}\Big)\\
={}&o_p(\sqrt{m_n}),
\end{aligned}
$$

using that $\int E\sup_{b\in\mathcal{B}_2}\big\|\sum_{l=1}^{m_n}p_l(X)\big\langle\widetilde{g}_b(\cdot,t,b)p_{\underline{k_n}}',p_l\big\rangle_X\big\|^2\varpi(t)dt\leq\sum_{l=1}^{k_n}E[p_l^2(X)]=O(k_n)$, which proves the result. □

In the following, recall the definition of f_B^* satisfying $\|\mathcal{F}_gf_B^*-\varphi\|_\varpi\leq\|\mathcal{F}_gf-\varphi\|_\varpi$ for all p.d.f. f.

Proof of Proposition 3. For the proof it is sufficient to show $S_n\geq C\|\mathcal{F}_gf_B^*-\varphi\|_\varpi^2+o_p(1)$. The proof of Theorem 2.1 together with the basic inequality $(a-b)^2\geq a^2-b^2$ implies that

$$
\begin{aligned}
S_n&=\lambda_n^{-1}\sum_{l=1}^{m_n}\int\Big|n^{-1}\sum_j(\exp(itY_j)-(\mathcal{F}_gf_B^*)(X_j,t))p_l(X_j)\Big|^2\varpi(t)dt+o_p(1)\\
&\gtrsim\sum_{l=1}^{m_n}\int\big|E[(\exp(itY)-(\mathcal{F}_gf_B^*)(X,t))p_l(X)]\big|^2\varpi(t)dt+o_p(1)\\
&\gtrsim\|\mathcal{F}_gf_B^*-\varphi\|_\varpi^2+o_p(1),
\end{aligned}
$$

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

Proof of Proposition 4. Following the proof Theorem 2.1, it is easily seen that

$$\begin{aligned} nS_n =&\lambda_n^{-1}\sum_{l=1}^{m_n}\int\Big|n^{-1/2}\sum_j (Y_i-\varphi(X_j,t))p_l(X_j)\Big|^2\varpi(t)dt\\ &+\sum_j\int\big|(\mathcal{F}_g\mathit{\Pi}_{k_n}f_B^*)(X_j,t)-\mathit{\Pi}_{m_n}\varphi(X_j,t)\big|^2\varpi(t)dt+o_p(\sqrt{m_n}). \end{aligned}$$

Further, under the sequence of local alternatives (2.11), we calculate

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

which proves the result. □

Technical Appendix.

Lemma 5.2. *Let Assumptions 1–3 hold true. Then*

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

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

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

We observe

$$\begin{aligned}
E|I_n-\mu_{m_n}|^2 &= Var\Big((\lambda_n n)^{-1}\sum_{l=1}^{m_n}\sum_j\int\Big|\delta(V_j,t)p_l(X_j)\Big|^2\varpi(t)dt\Big)\\
&\le \lambda_n^{-2}n^{-1}E\Big[\int\big|\delta(V,t)\big|^4\varpi(t)dt\Big(\sum_{l=1}^{m_n}p_l^2(X)\Big)^2\Big]\\
&\lesssim \sup_{x\in\mathcal{X}}\|p_{\underline{m_n}}(x)\|^2\lambda_n^{-2}n^{-1}\sum_{l=1}^{m_n}E[p_l^2(X)]\lesssim m_n^2n^{-1}\lambda_n^{-1}=o(1)
\end{aligned}$$

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

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

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 $\sum_{j=1}^\infty E|Q_{nj}|^2\le 1$ observe that

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

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

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

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

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

Lemma 5.3. *Under the conditions of Theorem 2.1 it holds*

$$\|\widehat{A}_n^- - A_n^-\|=O_p\Big(\sqrt{(\log n)k_n/(n\tau_n)}\Big).$$

Proof. On the set $\Omega \equiv \Big\{\|A_n^-\|\|\widehat{A}_n - A_n\| < 1/4, \quad \operatorname{rank}(A_n) = \operatorname{rank}(\widehat{A}_n)\Big\}$, 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:

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

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 $I_{k_n} - A_nA_n^-$ and $I_{k_n} - \widehat{A}_n\widehat{A}_n^-$ as projections have operator norm bounded by one, we obtain

$$\begin{aligned}\|\widehat{A}_n^- - A_n^-\|\, 1_\Omega &= \big(\|\widehat{A}_n^-\|\|\widehat{A}_n - A_n\|\|A_n^-\| + \|\widehat{A}_n^-\|^2\|\widehat{A}_n - A_n\| + \|A_n^-\|^2\|\widehat{A}_n - A_n\|\big)\, 1_\Omega\\ &\le 3\,\|\widehat{A}_n - A_n\| \max\Big\{\|A_n^-\|^2, \|\widehat{A}_n^-\|^2\, 1_\Omega\Big\}.\end{aligned}$$

By Theorem 3.2 of Chen et al. [1996] it holds $\|\widehat{A}_n^-\|\, 1_\Omega \le 3\|A_n^-\| = O(\tau_n^{-1/2})$. Consequently, Lemma 6.2 of Belloni et al. [2015] yields $\|\widehat{A}_n^- - A_n^-\|\, 1_\Omega = O_p\big(\sqrt{k_n(\log n)/(n\tau_n)}\big)$. The assertion follows by $1_\Omega = 1$ with probability approaching one. □

References

C. Ai and X. Chen. Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica*, 71:1795–1843, 2003.

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.

D. J. Benjamin, J. O. Berger, M. Johannesson, B. A. Nosek, E.-J. Wagenmakers, R. Berk, K. A. Bollen, B. Brembs, L. Brown, C. Camerer, et al. Redefine statistical significance. *Nature Human Behaviour*, page 1, 2017.

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.

C. Breunig and S. Hoderlein. Supplement to "specification testing in random coefficient models". Technical report, 2017.

B. Chen and Y. Hong. Characteristic function–based testing for multifactor continuous-time markov models via nonparametric regression. *Econometric Theory*, 26(04):1115–1179, 2010.

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.

S. X. Chen, L. Peng, L. Y. Cindy, et al. Parameter estimation and model testing for markov processes via conditional characteristic functions. *Bernoulli*, 19(1):228–251, 2013.

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. Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals. *Econometrica*, 80(1):277–321, 01 2012.

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. Technical report, 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*, 188(1):135–149, 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. *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. *Pesaran, H. and M. Wickens (Eds.), 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.

L. Su and H. White. A consistent characteristic function-based test for conditional independence. *Journal of Econometrics*, 141(2):807–834, 2007.

P. A. Swamy. Efficient inference in a random coefficient regression model. *Econometrica*, pages 311–323, 1970.