Herman J. Bierens

Econometric Model Specification

Consistent Model Specification Tests and Semi-Nonparametric Modeling and Inference

World Scientific

Econometric Model Specification

Consistent Model Specification Tests and Semi-Nonparametric Modeling and Inference

Herman J. Bierens
Pennsylvania State University, USA

World Scientific

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI • TOKYO

Published by

World Scientific Publishing Co. Pte. Ltd.
5 Toh Tuck Link, Singapore 596224
USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601
UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

Library of Congress Cataloging-in-Publication Data
Names: Bierens, Herman J., 1943–   , author.
Title: Econometric model specification : consistent model specification tests
and semi-nonparametric modeling and inference / Herman J Bierens
(Pennsylvania State University, USA).
Description: New Jersey : World Scientific, 2016.
Identifiers: LCCN 2016023598 | ISBN 9789814740500 (hc : alk. paper)
Subjects: LCSH: Econometrics. | Econometric models.
Classification: LCC HB139 .B526 2016 | DDC 330.01/5195--dc23
LC record available at https://lccn.loc.gov/2016023598

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

Desk Editors: Anthony Alexander/Alisha Nguyen

Typeset by Stallion Press
Email: enquiries@stallionpress.com

Printed in Singapore

PREFACE

In empirical economic and econometric research, parametric econometric models are used to analyze the interaction of economic variables of interest for the main purpose of validating the corresponding economic theory, making predictions, and/or conducting policy evaluations. The selection of the variables in the model is based on data availability and economic theory, but economic theory is usually not specific enough to determine the functional form of the model involved. Therefore, in practice parametric econometric models are usually specified such that the parameters can be estimated without too much effort.

According to Karl Popper (*The Logic of Scientific Discovery*), scientific theories should be falsifiable, and if such a theory is falsified it should be replaced with a (hopefully) better falsifiable theory. Nowadays economic theories are formulated in the form of models, so that in order for economics to be a science, economic models should be falsifiable, and since falsification of economic models can only be done via their translation to econometric models, the latter models should be falsifiable as well. So the following two questions arise: (1) How can we falsify parametric econometric models? (2) What can be done if the model is falsified? These questions are just the themes of this book.

The first part of this book presents and reviews my work (some jointly with coauthors) on testing the validity of parametric conditional expectation models (also known as regression models), and conditional distribution models. These tests have the following two desirable properties, among others. First, they are consistent, i.e., they have asymptotic power one against the general alternative that the null hypothesis of model correctness is false. Thus, rejection means that the model is falsified. Second, these tests have nontrivial power against root n local alternatives, where n is the sample size. The five reprinted articles involved are accompanied by addendums in order to bring these tests up-to-date and make them fully operational. In particular, it is shown how to derive upper bounds of the critical values, and how to derive the actual critical values via bootstrap methods. Moreover, these addendums provide detailed proofs of all the results, which make this book uniquely self-contained, and a useful resource for graduate students and researchers interested in model specification issues.

The second part of the book is devoted to semi-nonparametric (SNP) models, i.e., models that are partially parametrized and the unspecified part is modeled via infinite series expansions. If a parametric econometric model is falsified, a possible way to repair it is to incorporate unobserved heterogeneity variables, where the joint distribution of the latter is modeled semi-nonparametrically, or to replace certain parametric functional parts in the model by SNP functional forms. The usefulness of the SNP approach is illustrated in this book by applications to interval-censored duration models, competing risks models and auction models. In the last article in this series, published in 2014, low-level conditions are established for the consistency and asymptotic normality of sieve estimators, using a different approach than in the literature on sieve estimation. The latter paper comes with an extensive appendix with all the proofs not already contained in the main paper, together with a short addendum. These papers will likely be useful to PhD students and researchers interested in SNP models.

Some papers in this book refer to unpublished papers and additional material on web sites: http://econ.la.psu.edu/~hbierens/ or http://grizzly.la.psu.edu/~hbierens/. However, both web sites have been decommissioned. These papers and material are now available from my new web site: http://www.personal.psu.edu/hxb11/.

Acknowledgments

I like to thank my coauthors Werner Ploberger, Jose Carvalho, Hosin Song and Li Wang for their permission to include our joint papers in this book, and Ms. Alisha Nguyen, the economics editor of World Scientific Publishers, for her guidance in preparing this book.

ABOUT THE AUTHOR

Herman J. Bierens (1943) has been a Professor of Economics at the Pennsylvania State University, USA, from 1997 to 2012, and is now Professor Emeritus of Economics. Previously he was a parttime Professor of Econometrics at Tilburg University, the Netherlands (1994–2004), Robert H and Nancy Dedman Trustee Professor of Economics at Southern Methodist University, Dallas (1991–1996), and Professor of Econometrics at the Free University in Amsterdam (1986–1991). He studied mathematics and econometrics at the University of Amsterdam (MS in econometrics, 1972; PhD in economics, 1980). He has been Associate Editor of Econometrica, the Journal of Econometrics and Econometric Reviews, and is Fellow of the Econometric Society (2005). He has written three monographs and more than 40 articles in refereed journals. His research has focused on model specification issues in econometrics, in particular consistent model specification testing of cross-section and time series models, nonparametric estimation, unit root, cointegration and co-trending testing, econometric analysis of dynamic stochastic general equilibrium models, and more recently semi-nonparametric modeling and inference.

CONTENTS

Chapter 1

INTRODUCTION

This book reviews and extends my publications on consistent model specification testing, in Part 1, and semi-nonparametric modeling and inference in Part 2.

In 1982 and 1984, I published two papers in the *Journal of Econometrics*, Bierens (1982, 1984), on the problem how to test the correctness of a linear or nonlinear parametric regression model as a conditional expectation model such that the test has asymptotic power one against all deviations from the null hypothesis, for cross-section and time series regression models, respectively. These papers are the first to show that it is possible to test the validity of regression models consistently. However, at that time I was unable to derive the limiting null distribution of these tests. Therefore, I proposed to use upper bounds of the critical valued on the basis of Chebyshev's inequality for first moments.

These papers were followed up by two papers in *Econometrica*, Bierens (1990) and Bierens and Ploberger (1997). In Bierens (1990), I derived the limiting null distribution of a version of the 1982 test. In the 1997 *Econometrica* paper with Werner Ploberger, we characterized the null distribution further as a weighted sum of independent χ_1^2 distributed random variables, on the basis of which we could derive sharper upper bounds of the critical values. In the latter paper, we also showed that the test is admissible, i.e., there does not exist a uniformly more powerful test, and that the test has nontrivial power against root n local alternatives. More recently, in Bierens and Wang (2012), we extended the previous consistent model specification tests to a consistent test for the validity of parametric conditional distribution models.

In this book, these five papers are accompanied by substantial addendums, in order to bring the papers involved up-to-date, and/or provide additional proofs and results. In particular, I show how to approximate the critical values of the tests in Bierens (1982, 1984) and Bierens and Ploberger (1997) via a bootstrap method.

The second part of this book reviews my recent work on semi-nonparametric modeling and inference. Semi-nonparametric models are models that are only partially parametrized, and the nonparametric part take the form of one or more unknown functions which are approximated via an infinite series expansion. The first three papers in this series, Bierens (2008), Bierens and Carvalho (2007) and Bierens and Song (2012), are applications of semi-nonparametric methods to mixed proportional hazard models, competing risks models and first-price auction models, respectively. In the fourth paper, Bierens (2014b), I derive low-level conditions for consistency and asymptotic normality of sieve estimators of semi-nonparametric maximum likelihood models. Two of these papers, Bierens and Carvalho (2007) and Bierens (2014b), are accompanied by separate appendixes, and the latter paper also by a brief addendum.

In this chapter, I will review each of the papers in this book, and summarize their results.

1. True Models

The notion of a "true model" seems a contradiction, as models are by nature incomplete descriptions of the real world. As I have stated in Bierens (2007), "Most economic theories are partial theories in the sense that only a few related economic phenomena are studied. The analysis of this "partial" theory is justified, explicitly or implicitly, by the ceteris paribus assumption (other things being equal or constant). See Bierens and Swanson (2000) and the references therein. However, when simple economic models of this type are estimated using data which are themselves generated from a much more complex real economy, it is not surprising that they often fit poorly.

Thus, these models do not represent data generating processes, and are not designed to do. The purpose of these models is to gain insight in particular related economic phenomena rather than to describe an actual economy, and to conduct numerical experiments."

However, this argument applies to economic models rather than econometric models. In econometrics, we do not use the *ceteris paribus* assumption, but instead we integrate the "rest of the world" out. In other words, in econometrics we *condition* on the explanatory variables of interest, which under standard conditions yields *tautological* true models. For example, let Y be a dependent variable satisfying $E[Y^2] < \infty$, and let X be a vector of explanatory variables. Then we can decompose Y as

$$Y = g(X) + U, \quad \text{where } g(X) = E[Y|X] \text{ a.s. and } \quad U = Y - E[Y|X]. \tag{1.1}$$

In (1.1) and in the sequel "a.s." stands for "almost surely", which means that the equation involved holds with probability 1.

As is well-known, the function $g(.)$ exists and is Borel measurable [see, for example, Bierens (2004, Theorem 3.10, p. 77)], and by construction, $E[U|X] = 0$ a.s., which is the standard textbook assumption for regression errors. Moreover, $E[U^2] \leq E[Y^2] < \infty$ and $E[U^2|X] = E[Y^2|X] - g(X)^2 = \sigma^2(X)$ a.s., say, which in general is not constant. Thus, (1.1) is a tautological heteroscedastic regression model. Moreover, with $\{(Y_j, X_j)\}_{j=1}^n$ an i.i.d. random sample from (Y, X), the functions $g(.)$ and $\sigma^2(.)$ apply to each pair (Y_j, X_j), i.e.,

$$\begin{aligned} Y_j = g(X_j) + U_j, \quad \text{where } E[U_j|X_j] = 0 \text{ a.s.} \quad \text{and} \\ E[U_j^2|X_j] = \sigma^2(X_j) \text{ a.s.} \end{aligned} \tag{1.2}$$

Thus, under standard conditions true regression models for i.i.d. data exist, and that in general heteroscedasticity is the rule rather than the exception.

In the case of data heterogeneity, the regression function $g(.)$ and conditional variance function $\sigma^2(.)$ may depend on the observation index j, so that (1.2) reads: $Y_j = g_j(X_j) + U_j$, where $E[U_j|X_j] = 0$ a.s. and $E[U_j^2|X_j] = \sigma_j^2(X_j)$ a.s. But now we have only one observation available for g_j, so there is no way we can recover this function from the data. Therefore, the i.i.d. sample assumption is crucial for the existence of a true regression model.

In the time series case, we also have a true regression model, provided that the time series involved are strictly and covariance stationary. In particular, let Y_t be a univariate strictly and covariance stationary time series, and let $X_t \in \mathbb{R}^k$ be a strictly stationary vector time series process of exogenous variables. The case that the model contains contemporaneous exogenous variables as well can easily be accommodated by defining $X_t = X^*_{t+1}$, where the latter is the actual vector of exogenous variables.

Denote $Z_t = (Y_t, X'_t)'$ and let $\mathcal{F}^t_{-\infty}$ be the σ-algebra generated by the sequence $\{Z_{t-j}\}_{j=0}^{\infty}$. Then we can decompose Y_t as $Y_t = G_{t-1} + U_t$, where $G_{t-1} = E\left[Y_t|\mathcal{F}^{t-1}_{-\infty}\right]$ and, by construction, U_t is a martingale difference process with respect to the filtration $\mathcal{F}^{t-1}_{-\infty}$, i.e., $E[|U_t|] < \infty$, U_t is measurable $\mathcal{F}^t_{-\infty}$ and

$$E\left[U_t|\mathcal{F}^{t-1}_{-\infty}\right] = 0 \text{ a.s.} \tag{1.3}$$

In essence, G_{t-1} is a function of all lagged Z_t's, and depends on t only through these lagged Z_t's. More precisely, let $\mathcal{F}^{t-1}_{t-m}$ be the σ-algebra generated by $Z_{t-1}, Z_{t-2}, \ldots, Z_{t-m}$. Then for $m \in \mathbb{N}$, similar to $g(X)$ in (1.1),

$$E\left[Y_t|\mathcal{F}^{t-1}_{t-m}\right] = g_m(Z_{t-1}, Z_{t-2}, \ldots, Z_{t-m}) \text{ a.s.},$$

where $g_m(.)$ is a Borel measurable real function on $\mathbb{R}^{(k+1)\times m}$, which by stationarity does not depend on t, and $G_{t-1} = \lim_{m\to\infty} g_m(Z_{t-1}, Z_{t-2}, \ldots, Z_{t-m})$ a.s. The latter follows from a well-known martingale convergence result. See for example Bierens (2004, Theorem 3.12, p. 80).

Rather than focusing on conditional expectation models, we may also be interested in conditional distribution models, i.e., the conditional distribution function of $Y \in \mathbb{R}^m$ given $X \in \mathbb{R}^k$: $\Pr[Y \leq y|X] = p(y|X)$, say, where $y \in \mathbb{R}^m$. Again, in the i.i.d. sample case the function $p(y|x)$ involved is the same for each observation (Y_j, X_j), i.e., $\Pr[Y_j \leq y|X_j] = p(y|X_j)$, $y \in \mathbb{R}^m$. Also, this is a tautological true model.

2. Validity of Parametric Regression Models

Given a pair $(Y, X) \in \mathbb{R} \times \mathbb{R}^k$ with $E[Y^2] < \infty$, a parametric specification of a regression model for Y takes the general form

$$Y = f(X, \theta_0) + U,$$

where θ_0 is a parameter vector contained in an *a priori* chosen parameter space Θ, $f(x, \theta)$ is an *a priori* chosen real (and usually continuous) function on $\mathbb{R}^k \times \Theta$, and U is the error term. As to the latter, the standard textbook regression *assumption* is that $E[U|X] = 0$ a.s., which is equivalent to the *assumption* that there exists a $\theta_0 \in \Theta$ such that

$$E[Y|X] = f(X, \theta_0) \text{ a.s.} \tag{1.4}$$

If so, this parameter vector θ_0 is determined by

$$\theta_0 = \arg\min_{\theta \in \Theta} E[(Y - f(X, \theta))^2] = \arg\min_{\theta \in \Theta} E[(E[Y|X] - f(X, \theta))^2], \tag{1.5}$$

provided that the expectations involved are finite for at least one $\theta \in \Theta$.

If condition (1.4) does not hold then $f(X, \theta_0)$ with θ_0 defined by (1.5) is merely the "best" approximation of the conditional expectation $E[Y|X]$, best in the sense of the approximation with the smallest mean square error, given the parametric specification $f(x, \theta)$.

In econometrics textbooks, the functional form of parametric econometric models is usually taken as given, with vague reference to economic theory. However, economic theory may tell us which variables to include in a model, and the possible directions of the interactions of explanatory or predetermined variables with the dependent variable(s), but economic theory is usually not specific enough to guide us how to specify the functional form of parametric econometric models. Therefore, in practice, the choice of the functional form of a model is often based on convenience, in the sense that it is chosen such that the model is easy to estimate. The linear regression model is the prime example of such a specification.

In the linear regression case with constant term, $f(x, \theta)$ takes the form

$$f(x, \theta) = (1, x')\theta, \quad \text{with } x \in \mathbb{R}^k, \quad \theta \in \Theta = \mathbb{R}^{k+1}. \tag{1.6}$$

Given that

$$E[Y^2] < \infty, \quad E[X'X] < \infty, \; A = E\left[\begin{pmatrix} 1 & X' \\ X & XX' \end{pmatrix}\right] \text{ is nonsingular,} \tag{1.7}$$

the solution of (1.5) is $\theta_0 = A^{-1}c$, where $c = (E[Y], E[X'Y])'$. Note that $(1, X')\theta_0$ is known as the linear projection of Y on $(1, X')'$, with residual

$$U = Y - (1, X')\theta_0, \tag{1.8}$$

which by the first-order conditions satisfies $E[U] = 0, \; E[X.U] = 0$. Thus, under the conditions (1.7) the linear regression model is also tautological, but is sub-optimal in that the variance of the linear projection residual U is larger than the variance of $Y - E[Y|X]$ if $\Pr[E[Y|X] = (1, X')\theta_0] < 1$.

Nevertheless, given that

Assumption 2.1. *$\{(Y_j, X_j)\}_{j=1}^n$ is a random sample from the distribution of $(Y, X) \in \mathbb{R} \times \mathbb{R}^k$, where Y and X satisfy the conditions in* (1.7), *the OLS estimator $\widehat{\theta}_n$ of θ_0 is strongly consistent and asymptotically normal, i.e.,*

$$\widehat{\theta}_n \xrightarrow{a.s.} \theta_0,$$

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} N_{k+1}(0, A^{-1}BA^{-1}), \; \text{with } B = Var\left(\begin{pmatrix} U \\ U.X \end{pmatrix}\right),$$

provided that the variance matrix B is finite. The latter requires that in addition to (1.7),

Assumption 2.2. *$E[U^2||X||^2] < \infty$, where U is the linear projection residual* (1.8).

2.1. *Consistent model specification tests*

Throughout this section, I will summarize the results in Bierens (1982) and the follow-up papers Bierens (1990) and Bierens and Ploberger (1997) for the linear regression case (1.6) under Assumptions 2.1 and 2.2.

2.1.1. *Chapter 2: The initial integrated conditional moment test*

In the Fall of 1981, as a postdoc in the department of economics of the University of Minnesota, Minneapolis, I started puzzling on the problem how to test, for regression errors U, the null hypothesis

$$H_0 : \Pr(E[U|X] = 0) = 1,$$

consistently against the general alternative hypothesis that H_0 is false:

$$H_1 : \Pr(E[U|X] = 0) < 1.$$

In the early 1980s, it was a common belief among econometricians that "A test of functional form that is consistent against all deviation from the null hypotheses spreads its power so thinly over a continuum of alternatives that there is no power left," as Christopher Sims told me when I showed him my hand-written draft of Bierens (1982). However, contrary to this common belief there exists a sharp distinction between H_0 and H_1, in terms of the Fourier transform of $E[U|X]$ and this distinction can be used to devise a consistent test.

The Fourier transform of $E[U|X]$ takes the form

$$\varphi(\tau) = E\left[E[U|X]\exp(\mathbf{i}.\tau'X)\right] = E\left[U\exp(\mathbf{i}.\tau'X)\right], \quad \tau \in \mathbb{R}^k, \tag{1.9}$$

where $\mathbf{i}$ is the complex number $\sqrt{-1}$. Obviously, under H_0, $\varphi(\tau) = 0$ for all $\tau \in \mathbb{R}^k$, whereas by the uniqueness of Fourier transforms, $\varphi(\tau) \neq 0$ for some $\tau \in \mathbb{R}^k$ under H_1. Moreover, if X is bounded then in the latter case $\varphi(\tau) \neq 0$ for some τ arbitrarily close to the origin of $\mathbb{R}^k$. If X is not bounded we may without loss of generality replace X in (1.9) by $\Phi(X)$, where $\Phi : \mathbb{R}^k \to \mathbb{R}^k$ is a bounded one-to-one mapping with Borel measurable inverse, because then X and $\Phi(X)$ generate the same σ-algebra, which implies that $E[U|X] = E[U|\Phi(X)]$ a.s. Then the Fourier transform of $E[U|X]$ becomes

$$\overline{\varphi}(\tau) = E\left[U\exp(\mathbf{i}.\tau'\Phi(X))\right], \quad \tau \in \mathbb{R}^k,$$

so that under H_1, $\overline{\varphi}(\tau) \neq 0$ for some τ arbitrarily close to the origin of $\mathbb{R}^k$, which by the continuity of $\overline{\varphi}(\tau)$ implies that any open neighborhood of the origin of $\mathbb{R}^k$ contains an open set on which $\overline{\varphi}(\tau) \neq 0$.

Consequently, for a compact subset Υ of $\mathbb{R}^k$ containing the origin of $\mathbb{R}^k$ in its interior we have

$$\int_\Upsilon |\overline{\varphi}(\tau)|^2 \mathrm{d}\tau \begin{cases} = 0 & \text{under } H_0, \\ > 0 & \text{under } H_1. \end{cases}$$

This suggests to test H_0 against H_1 using the Integrated Conditional Moment (ICM) test statistic

$$\widehat{T}_n = \int_\Upsilon \left| \widehat{W}_n(\tau) \right|^2 \mathrm{d}\tau, \text{ where} \tag{1.10}$$

$$\widehat{W}_n(\tau) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \widehat{U}_j \exp(\mathbf{i}.\tau'\Phi(X_j)), \tag{1.11}$$

and the $\widehat{U}_j$'s are the least squares residuals.

In the linear regression case, $\widehat{U}_j = U_j - (1, X_j')(\widehat{\theta}_n - \theta_0)$, and by Assumptions 2.1–2.2,

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) = A^{-1} \frac{1}{\sqrt{n}} \sum_{j=1}^n U_j \begin{pmatrix} 1 \\ X_j \end{pmatrix} + o_p(1),$$

hence,

$$\begin{aligned} \widehat{W}_n(\tau) &= \frac{1}{\sqrt{n}} \sum_{j=1}^n U_j \exp\left(\mathbf{i}.\tau'\Phi(X_j)\right) \\ &\quad - \frac{1}{n} \sum_{j=1}^n (1, X_j') \exp\left(\mathbf{i}.\tau'\Phi(X_j)\right) \sqrt{n}(\widehat{\theta}_n - \theta_0) \\ &= \frac{1}{\sqrt{n}} \sum_{j=1}^n U_j \exp\left(\mathbf{i}.\tau'\Phi(X_j)\right) \\ &\quad - E\left[(1, X') \exp\left(\mathbf{i}.\tau'\Phi(X)\right)\right] \sqrt{n}(\widehat{\theta}_n - \theta_0) \\ &\quad + o_p(1) \\ &= \frac{1}{\sqrt{n}} \sum_{j=1}^n U_j.\rho(\tau, X_j) + o_p(1), \end{aligned} \tag{1.12}$$