Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Specification testing for transformation models with an application to generalized accelerated failure-time models[☆]

Arthur Lewbel[a,*], Xun Lu[b], Liangjun Su[c]

[a] Department of Economics, Boston College, United States

[b] Department of Economics, Hong Kong University of Science and Technology, Hong Kong

[c] School of Economics, Singapore Management University, Singapore

ARTICLE INFO

Article history:
Received 2 May 2013
Received in revised form
22 July 2014
Accepted 13 September 2014
Available online 17 October 2014

JEL classification:
C12
C14

Keywords:
Additivity
Control variable
Endogenous variable
Monotonicity
Nonparametric nonseparable model
Hazard model
Specification test
Transformation model
Unobserved heterogeneity

ABSTRACT

This paper provides a nonparametric test of the specification of a transformation model. Specifically, we test whether an observable outcome Y is monotonic in the sum of a function of observable covariates X plus an unobservable error U. Transformation models of this form are commonly assumed in economics, including, e.g., standard specifications of duration models and hedonic pricing models. Our test statistic is asymptotically normal under local alternatives and consistent against nonparametric alternatives violating the implied restriction. Monte Carlo experiments show that our test performs well in finite samples. We apply our results to test for specifications of generalized accelerated failure-time (GAFT) models of the duration of strikes.

1. Introduction

Consider a scalar observable outcome Y, a $d_x \times 1$ vector of observable covariates of interest X, and a scalar unobservable cause or error U. Our goal of this paper is to test the following hypotheses[1]:

$\mathbb{H}_{10}$: There exist two measurable functions $G : \mathbb{R} \to \mathbb{R}$ and $H_1 : \mathbb{R}^{d_x} \to \mathbb{R}$ such that
$Y = G[H_1(X) + U]$ a.s., and G is strictly monotonic;

$\mathbb{H}_{1A}$: $\mathbb{H}_{10}$ is false.

Specifications that are monotonic functions of additive models have been called "transformation models" (e.g. Chiappori et al., 2013), or "transformed additively separable models" (e.g. Jacho-Chávez et al., 2010), or "generalized additive models with unknown link function" (e.g. Horowitz, 2001; Horowitz and Mammen, 2004).

Broadly speaking, there are two kinds of transformation models that are common in the economics literature. The first type assumes that Y and X are observable, U is unobservable, and the link function $G(\cdot)$ may be known or unknown. Our paper belongs to this category. Ridder (1990), Horowitz (1996), Ekeland et al. (2004), Ichimura and Lee (2011), and Chiappori et al. (2013) discuss identification and estimation for transformation models of this

☆ Halbert White inspired this project, brought us together to work on it, and provided substantial advice, discussion, and enthusiasm. We deeply mourn his passing. We gratefully thank the co-editor Jianqinq Fan, the associate editor, and three anonymous referees for their many constructive comments. We are also indebted to Songnian Chen, Yanqin Fan, and Shakeeb Khan for helpful comments and suggestions. Lu acknowledges support from Hong Kong University of Science and Technology under grant number FSECS13BM02. Su acknowledges support from the Singapore Ministry of Education for Academic Research Fund under grant number MOE2012-T2-2-021.

* Corresponding author. Tel.: +1 617 552 3678; fax: +1 617 552 2308.
E-mail address: lewbel@bc.edu (A. Lewbel).

1 The error term U is a scalar under the null $\mathbb{H}_{10}$, however, under the alternative $\mathbb{H}_{1A}$, Y could be a function of X and a vector of unobservable errors $\mathbb{U}$. For example, the alternative might include models with random coefficients. More generally, we could write $\mathbb{H}_{1A}$ as $Y = R(X, \mathbb{U})$, and then under the null there would exist a scalar valued function H_2 such that $U = H_2(\mathbb{U})$. We later define a more specific set of alternatives that our test has power against.

http://dx.doi.org/10.1016/j.jeconom.2014.09.008
0304-4076/

category. In this class of models, the functions G and H_1 and the distribution of U are identified and estimated. In the second type of transformation model, both X and U are observable, and Y is an object that can be estimated such as a conditional mean or quantile function. Horowitz (2001), Horowitz and Mammen (2004, 2007, 2011), Horowitz and Lee (2005), and Jacho-Chávez et al. (2010) provide identification and estimation results for this second kind of transformation model, while Gozalo and Linton (2001) consider specification tests for such models. See also Horowitz (2014) for a recent survey on the latter class of models.

The transformation models under our null are commonly used (and hence assumed to hold) in a wide range of economic applications. For example, they are often used to study duration data (see, e.g. Heckman and Singer, 1984; Keifer, 1988; Mata and Portugal, 1994; Engle, 2000; Abbring et al., 2008). In particular generalized accelerated failure-time (GAFT) models, which includes accelerated failure-time (AFT) models, proportional hazard (PH) models, and mixed proportional hazard (MPH) models as special cases, are all examples of models that satisfy our null hypothesis. The MPH specification in particular is a widely used class of duration data specifications (for a review, see Van den Berg, 2001).

Despite its popularity, economic theory rarely justifies the MPH or other GAFT specifications. For example, Van den Berg (2001, p. 3400) points out that "the MPH model specification is not derived from economic theory and it remains to be seen whether the MPH specification is actually able to capture important theoretical relations." He also provides some specific economic examples where the MPH specification is violated. In their microeconometrics textbook, Cameron and Trivedi (2005, p. 613) say that "the multiplicative heterogeneity assumption [in MPH models] is also rather special, but it is mathematically convenient…"Given the popularity (and the limitations) of GAFT models, especially MPH models, it is obvious that a formal specification test of these models would be useful for empirical research. While some specification tests for certain parametric forms of duration models exist (see, e.g. Fernandes and Grammig, 2005), to the best of our knowledge, ours is the first that specifically tests for some testable implications of the general specification of GAFT models.[2]

Another major set of applications of transformation model specifications where U is unobservable are hedonic models (see, e.g. Ekeland et al., 2004; Heckman et al., 2005). Here again, we believe that our paper is the first to provide a general specification test for this class of transformation models.

A conditional exogeneity assumption is imposed to test $\mathbb{H}_{10}$, i.e., we assume that U and X are conditionally independent, conditioning on an observable covariate vector Z. This is analogous to the conditional unconfoundedness assumption in the treatment effect literature, and to the assumptions required for use of control function type methods of dealing with endogeneity (See, e.g. Heckman and Robb, 1986 and Blundell and Powell, 2003. In a control function setting Z would be the errors obtained after regressing endogenous elements of X on a vector of instruments.) Chiappori et al. (2013) provide a nonparametric estimator for the transformation model under similar assumptions. Our test allows for a covariate vector Z, but unlike some other estimators and tests (see below), we do not require a vector Z, i.e., our test can also be applied when U and X are unconditionally independent and no other relevant covariates are observed.

We first show that if the data are generated by a transformation model, i.e., if $\mathbb{H}_{10}$ holds, then the ratio of the derivatives with respect to Y and to X of the conditional CDF of Y given (X, Z) can be written as a product of functions of X and Y.[3] We then use local polynomial methods to estimate these derivatives, and construct test statistics based on the L_2 distance between restricted and unrestricted estimators of this ratio of derivatives. We show that our test statistic is asymptotically normal under the null and under a sequence of Pitman local alternatives and is consistent against the alternatives violating the implied restriction. To facilitate the application of our test, we use subsampling to obtain p-values or critical values. We also evaluate our test both in a Monte Carlo setting, and in an empirical application concerning duration of strikes by manufacturing workers.

Our null $\mathbb{H}_{10}$ is weaker than additive separability but stronger than monotonicity. Lu and White (2014) and Su et al. (forthcoming) propose tests for additive separability under the same conditional exogeneity assumption that U is independent of X given Z. Specifically, they test whether there exists an unknown measurable function G_1 such that

$$Y = G_1(X) + U \quad \text{a.s.}$$

Testing $\mathbb{H}_{10}$ is more general than testing for separability, since our null is equivalent to additive separability in the special case where G is known to be the identity function. Hence if we reject our $\mathbb{H}_{10}$, then we also reject their additive separability.

Hoderlein et al. (2014, HSWY hereafter) test for monotonicity under a conditional exogeneity assumption. HSWY test whether there exists a function R such that

$$Y = R(X, U)$$

where R is strictly monotonic in its second argument. Our null is stronger than monotonicity, so if the HSWY test rejects monotonicity, then our null $\mathbb{H}_{10}$ is also rejected. Our null $\mathbb{H}_{10}$ combines monotonicity with the additional restriction that the observable X and unobservable U are additively separable under a transformation function G. Our test exploits this additivity restriction, and so should be generally stronger than HSWY for testing our null $\mathbb{H}_{10}$. Also, the HSWY test requires that Z not be empty, while our test of $\mathbb{H}_{10}$ can be applied even if we have no conditioning covariates Z.

The rest of the paper is organized as follows. In Section 2, we propose and motivate our test. In Section 3, we show that our test statistics are asymptotically normal under the null, and we analyze their global and local power. In Section 4, we conduct some Monte Carlo simulations to evaluate the finite sample performance of our test statistics. In Section 5, we provide an empirical application to test the specification of GAFT models in data on the durations of strikes. In Section 6, we discuss extensions to other closely related hypotheses. Section 7 concludes. The proofs of the main results in the paper are relegated to Appendices A–C and those for the technical lemmas are given in the supplementary material (see Appendix D).

2. A specification test for transformation models

In this section, we describe implications of $\mathbb{H}_{10}$ that are used to motivate our test construction, and then describe our proposed test statistic.

[2] Recently, Chiappori et al. (2013) provide a nonparametric test, not for the transformation model specification itself, but for a conditional exogeneity assumption within the context of a transformation model. Still, their test might be interpreted as a model specification test. See Remark 2.6 in Section 2.1 for details.

[3] Horowitz (1996) considers the estimation of the semiparametric model under our null, where the function H_1 takes a parametric form (unlike our nonparametric case) and without covariates Z. His estimator also relies on the implication that the ratio of the derivatives is a multiplicative function of X and Y.