Specification Test for Missing Functional Data*

Federico A. Bugni
Department of Economics
Duke University
Email: federico.bugni@duke.edu

September 12, 2011

Abstract

Economic data are frequently generated by stochastic processes that can be modeled as realizations of random functions (functional data). This paper adapts the specification test for functional data developed by Bugni, Hall, Horowitz, and Neumann (2009) to the presence of missing observations. By using a worst case scenario approach, our method is able to extract the information available in the observed portion of the data while being agnostic about the nature of the missing observations. The presence of missing data implies that our test will not only result in the rejection or lack of rejection of the null hypothesis, but it may also be inconclusive.

Under the null hypothesis, our specification test will reject the null hypothesis with a probability that, in the limit, does not exceed the significance level of the test. Moreover, the power of the test converges to one whenever the distribution of the observations conveys that the null hypothesis is false.

Monte Carlo evidence shows that the test may produce informative results (either rejection or lack of rejection of the null hypothesis) even under the presence of significant amounts of missing data. The procedure is illustrated by testing whether the Burdett-Mortensen labor market model is the correct framework for wage paths constructed from the NLSY79 survey.

1 Introduction

Economic data are frequently generated by stochastic processes that can be modelled as occurring in continuous time. The data may then be treated as realizations of random functions (functional data). Examples include wage paths, asset prices, or returns. In this case, economic theory may

*I am indebted to my advisors, Joel Horowitz, Rosa Matzkin, and Elie Tamer for their guidance and support. I thank the co-editor and two anonymous referees for comments and suggestions that have significantly helped to improve this paper. I also thank comments and suggestions from Audra Bowlus, Ivan Canay, Jon Gemus, Silvia Glaubach, Shakeeb Khan, Ivana Komunjer, George Neumann, and participants of the 2009 LACEA/LAMES conference in Buenos Aires, the 2009 Triangle Econometrics Conference, and the 2010 Canadian Econometrics Group Conference. Financial support from the Robert Eisner Memorial Fellowship and the Dissertation Year Fellowship is gratefully acknowledged. Erik Vogt provided excellent research assistance. All errors are my own.

provide a parametric model for the data, i.e., a stochastic process which is known up to a finite dimensional parameter that constitutes a candidate for the true process that generated the data. In such cases, a natural research question is whether the parametric model is the right model for the data, that is, whether there is a parameter value for which the model is the data generating process. This type of hypothesis test is referred to as a *specification test*.

In a recent paper, Bugni, Hall, Horowitz, and Neumann (2009) (hereafter, referred to as BHHN) developed the first method for carrying out a specification test for functional data [1]. Their contribution constitutes the generalization of the Cramér-von Mises[2] specification test to the distribution of random functions that depend on an unknown finite-dimensional parameter vector. Their procedure contributes to the literature by introducing functional data approaches to specification testing in econometrics and by developing parametric bootstrap methods that enable its implementation.

The specification test in BHHN does not allow for the existence of missing observations. Both the theoretical results and the empirical implementation of the test require the researcher to observe a sample of independent and identically distributed (i.i.d.) functions. This does not only forbid functions to be missing, but it also forbids functions from being unobserved in certain sections of its domain. Unfortunately, this is a strong restriction: missing data is a pervasive problem in most data samples and functional data samples are no exception. The particular feature of functional data is that observations can present missing sections, rather than being completely unobserved.

One might wonder if the specification test developed by BHHN can still be applied to a functional data sample with missing observations by eliminating observations that present missing sections. There are two reasons why this procedure should be avoided. First, the results of this test cannot be extrapolated to the distribution of the data unless we assume that the observed data is a representative sample of the general data, that is, unless missing data are missing *at random*[3]. If this assumption fails, our test results will be contaminated by sample selection bias, which invalidates our test results. Second, in the specific case of functional data, eliminating observations that have some missing sections will eliminate valuable information contained in their non-missing sections.

The objective of this paper is to provide a specification test that can be applied to functional data that has missing observations. In order to deal with the missing data problem, we adopt a worst case scenario approach in the spirit of Manski (1989), which is able to extract information about the observed data and still be agnostic about the nature of the unobserved data.

[1] As in BHHN, we focus exclusively on specification tests in functional data settings, i.e., infinite dimensional data setting. In certain restricted infinite dimensional data settings, Cuesta-Albertos, Fraiman, and Ransford (2006) and Cuesta-Albertos, del Barrio, Fraiman, and Matrán (2007) develop alternative specification tests using different methodologies (projections). On the other hand, in finite dimensional data settings, Andrews (1997), Zheng (2000), and Bierens and Wang (2010) propose consistent specification tests for the conditional CDF. For an excellent survey of this literature, see Bierens (2009).

[2] BHHN also show how to implement the functional data analogue of the Kolmogorov-Smirnov specification test for functional data. Nevertheless, the Cramér-von Mises test is preferred to the Kolmogorov-Smirnov test since it tends to be more powerful in finite-dimensional settings and it is easier to compute in the infinite-dimensional setting.

[3] See, e.g., Heckman (1979), Manski (1989), and Manski (2003).

Instead of developing a new hypothesis testing procedure for missing functional data, one could consider applying the hypothesis testing procedure of BHHN to the subset of the data that has no missing sections. When compared to this procedure, the hypothesis test developed in this paper presents two clear advantages. First, by avoiding untestable assumptions about the nature of the missing data, our hypothesis test produces conclusions that are valid regardless of the (unobserved) features of the distribution of the missing data. Second, in the presence of functional data where functions in the data have missing sections and non-missing sections (that is, the functions are partly unobserved), our hypothesis testing procedure is able to extract all of the information contained in the non-missing sections while still being agnostic about the nature of the missing sections. While the first advantage is common to every inferential procedure that adopts the worst case scenario approach to missing data, the second advantage is relatively exclusive to functional data analysis.

Unfortunately, our hypothesis testing procedure has an unavoidable cost. Without assumptions about the nature of the missing data, the test statistic is *partially or set identified*, that is, it can only be restricted to an interval, even asymptotically. This implies that the resulting hypothesis test may be inconclusive, i.e., as a conclusion of the hypothesis test one cannot decide whether to reject or not to reject the null hypothesis.

Even though the present paper focuses on adapting a relatively specific hypothesis testing problem to the presence of missing data, one could envision that the ideas and techniques developed in this paper could be applied to other settings that are also partially identified[4]. In this sense, we believe that this paper constitutes our first contribution in a research agenda regarding specification testing in partially identified models.

The remaining of the paper is organized as follows. Section 2 describes the hypothesis test when there are no missing observations. Section 3 studies the identification problem posed by missing data, which is the basis of our hypothesis test. In Section 4, we introduce our hypothesis test and analyze its theoretical properties. Section 5 presents Monte Carlo evidence about the performance of our test. In Section 6, we use our methodology to test whether a random sample of wage paths constructed from the NLSY79 data is distributed according to the Burdett-Mortensen labor market model. Finally, Section 7 concludes the paper. All of the proofs of the paper are collected in the appendix.

2 Specification test without missing functional data

We begin by describing the nature of the specification test when there are no missing observations. Even though this will not be the setup we are ultimately interested in, it constitutes an adequate

[4] For example, it would be interesting to develop specification tests in the context of partially identified models defined by moment inequalities and equalities. While there is an extensive literature on inference for the parameters in these kinds of models (see, e.g., Beresteanu and Molinari (2008), Rosen (2006), Chernozhukov, Hong, and Tamer (2007), Romano and Shaikh (2008), Stoye (2009), Andrews and Soares (2010), Canay (2010), and Bugni (2010), among many others), there are no hypothesis testing procedures that are specially designed to test the specification of the underlying economic model.

starting point to introduce the elements of the specification test for our functional data.

The following notation will be used throughout the paper. Let $\mathcal{I}$ denote an arbitrary bounded subset in the real line, which will constitute the (common) domain of the random functions in our specific application[5]. Let $\mathbb{R}^{\mathcal{I}}$ denote the space of all functions that map $\mathcal{I}$ on $\mathbb{R}$, let $L_2(\mathcal{I})$ denote the space of square integrable functions that map $\mathcal{I}$ on $\mathbb{R}$, let $C(\mathcal{I})$ denote the space of continuous functions that map $\mathcal{I}$ on $\mathbb{R}$, and let $C'(\mathcal{I})$ denote the space of functions that map $\mathcal{I}$ on $\mathbb{R}$ that are continuous except for a countable subset of $\mathcal{I}$. Finally, the probability space is denoted by $(\Omega, \mathcal{B}, P)$, and it is assumed to be complete.

The formal structure of the specification test is the following. Our data are a random sample of size n of functions, denoted by $\mathcal{X}_n = \{X_1, \ldots, X_n\}$, which are distributed according to a separable[6] stochastic process in $L_2(\mathcal{I})$, almost surely, denoted by $X : \Omega \to L_2(\mathcal{I})$. An econometric model conjectures that the data generating process behaves according to a certain separable stochastic process in $L_2(\mathcal{I})$, almost surely, denoted by $Y_\theta : \Omega \to L_2(\mathcal{I})$, which is known up to a finite-dimensional parameter θ that belongs to a parameter space $\Theta \subseteq \mathbb{R}^p$ ($p < \infty$). Our objective is to conduct a hypothesis test to decide whether the model $\{Y_\theta : \theta \in \Theta\}$ is a correct specification for X, or not, i.e.,

$$\begin{aligned} &H_0 : \exists \theta \in \Theta, \text{ such that } X \text{ and } Y_\theta \text{ are equally distributed}, \\ &H_1 : \nexists \theta \in \Theta, \text{ such that } X \text{ and } Y_\theta \text{ are equally distributed}. \end{aligned} \tag{2.1}$$

We now describe the BHHN specification test to implement the hypothesis test in Eq. (2.1), as it constitutes the basis of the specification test developed in this paper. In order to develop a hypothesis testing procedure, BHHN compare the cumulative distribution function (CDF) of the data and the model. Formally, if $Z : \Omega \to L_2(\mathcal{I})$ is a stochastic process in $(\Omega, \mathcal{B}, P)$, the CDF of Z evaluated at any (non-stochastic) $x \in \mathbb{R}^{\mathcal{I}}$ is defined as follows:

$$F_Z(x) = P^*\left(\omega \in \Omega : \{Z(\omega, t) \leq x(t), \ \forall t \in \mathcal{I}\}\right), \tag{2.2}$$

where P^* is the outer probability[7] associated to the measure P. Notice that the only difference between Eq. (2.2) and the traditional definition of CDF for a random vector is that we need to take into account the measurability issues due to the fact that $\mathcal{I}$ is possibly an uncountable set. For notational convenience, every $x \in \mathbb{R}^{\mathcal{I}}$ and every parameter value $\theta \in \Theta$, we let $F_Y(x|\theta)$ denote the CDF of Y_θ (rather than $F_{Y_\theta}(x)$).

In order to conduct the hypothesis test in Eq. (2.1), BHHN utilize the following idea. Under the null hypothesis, there exists a parameter value $\theta \in \Theta$ such that $F_X(x) = F_Y(x|\theta)$ for all $x \in L_2(\mathcal{I})$

[5]In most of our applications, the random functions are functions of time and so $\mathcal{I}$ constitutes a time interval, which we can normalize to be the unit interval.

[6]See Definition A.1 in the appendix. By Theorem 2.8.1 in Ito (2006), any stochastic process has a separable modification, so restricting the stochastic processes to separable ones does not restrict the applicability of our analysis.

[7]In a probability space $(\Omega, \mathcal{B}, P)$, the outer probability measure P^* associated to the probability measure P is the function $P^* : \Omega \to \mathbb{R}$ such that $P^*(S) = \inf_{B \in \mathcal{B}, S \subseteq B} P(B)$ for all $S \in \Omega$.

and, under the alternative hypothesis, no such parameter value exists [8]. Let μ be a non-degenerate probability measure on $L_2(\mathcal{I})$. [9] BHHN measure distance between the distributions of X and Y_θ with functional-data analogue of the Cramér-von Mises two-sample (population) statistic, given by:

$$T(X, Y_\theta) = \int (F_X(x) - F_Y(x|\theta))^2 \, d\mu(x). \tag{2.3}$$

We assume that there is a unique parameter value that minimizes $T(X, Y_\theta)$, denoted by θ_0. The key insight of the BHHN test is that the minimized value of the test statistic, $T(X, Y_{\theta_0})$, can then be used to distinguish between the null and alternative hypothesis, namely:

$$\begin{aligned} H_0 &: T(X, Y_{\theta_0}) = 0, \\ H_1 &: T(X, Y_{\theta_0}) > 0. \end{aligned} \tag{2.4}$$

The hypothesis test developed by BHHN is implemented by estimating the parameter θ_0, replacing cumulative distribution functions by their sample analogues and computing integrals by using Monte Carlo integration methods. The asymptotic distribution of the test statistic is approximated using the bootstrap. Formally, the test involves the following sequence of steps.

1. Use the data sample, $\mathcal{X}_n$, to:

   a. Estimate the parameter θ_0 root-n-consistently and denote it by $\hat{\theta}_0$.

   b. Estimate F_X by sample analogue estimation, i.e., for every $x \in L_2(\mathcal{I})$:

   $$\hat{F}_X(x) = \tfrac{1}{n} \textstyle\sum_{i=1}^{n} 1(X_i(t) \le x(t), \forall t \in \mathcal{I}).$$

2. Use $\hat{\theta}_0$ to construct an i.i.d. sample of size m of $Y_{\hat{\theta}_0}$, denoted by $\mathcal{Y}_{\hat{\theta}_0,m} \equiv \{Y_{\hat{\theta}_0,i}\}_{i=1}^{m}$. Use $\mathcal{Y}_{\hat{\theta}_0,m}$ to estimate $F_Y(\cdot|\hat{\theta}_0)$ by sample analogue estimation, i.e., for every $x \in L_2(\mathcal{I})$:

   $$\hat{F}_Y(x|\hat{\theta}_0) = \tfrac{1}{m} \textstyle\sum_{i=1}^{m} 1(Y_{\hat{\theta}_0,i}(t) \le x(t), \forall t \in \mathcal{I}).$$

3. Compute the test statistic:

   $$\hat{T}(X, Y_{\hat{\theta}_0}) = \tfrac{1}{V} \textstyle\sum_{j=1}^{V} (\hat{F}_X(Z_j) - \hat{F}_Y(Z_j|\hat{\theta}_0))^2,$$

   where $\mathcal{Z}_V \equiv \{Z_j\}_{j=1}^{V}$ is an i.i.d. sample of μ, and $\hat{F}_X$ and $\hat{F}_Y(\cdot|\hat{\theta}_0)$ are as in steps 1.b and 2, respectively.

4. Repeat the following many times:

[8] Since X and Y_θ are assumed to be separable stochastic processes, equality of their CDFs for all $x \in L_2(\mathcal{I})$ is sufficient to establish that X and Y_θ have the same finite dimensional distributions, which is sufficient to establish that these two stochastic processes are weakly equivalent.

[9] For example, μ can be the Gaussian process described in BHHN.

a. Construct a bootstrap sample of size n of $Y_{\hat{\theta}_0}$ and denote it by $\mathcal{X}_n^* \equiv \{X_i^*\}_{i=1}^n$.

b. Use $\mathcal{X}_n^*$ as in step 1, i.e., estimate the parameter θ_0 root-n-consistently (denoted by $\hat{\theta}_0^*$) and estimate F_{X^*} by sample analogue estimation.

c. Use $\hat{\theta}_0^*$ as in step 2, i.e., use it to construct $\mathcal{Y}_{\hat{\theta}_0^*,m}$, and use it to estimate $F_Y(\cdot|\hat{\theta}_0^*)$ by sample analogue estimation.

d. Compute the test statistic from the bootstrapped data as in step 3 and denote it by $\hat{T}(X^*, Y_{\hat{\theta}_0^*})$.

5. Denote by $t^*_{\hat{\theta}_0}(1-\alpha)$ the $(1-\alpha)$ quantile of the simulated statistics $n\hat{T}(X^*, Y_{\hat{\theta}_0^*})$ in step 4.

6. Decide the outcome of the test in the following way:

Outcome	Decision
$t^*_{\hat{\theta}_0}(1-\alpha) < n\hat{T}(X, Y_{\hat{\theta}_0})$	Reject H_0
$n\hat{T}(X, Y_{\hat{\theta}_0}) \leq t^*_{\hat{\theta}_0}(1-\alpha)$	Do not reject H_0

3 Identification analysis with missing functional data

We now consider how missing data in the functional data sample affects the BHHN specification test. It does so in two ways. First, missing data may affect our ability to consistently estimate the parameter θ_0 (step 1.a). This will certainly be the case if our estimator is obtained by maximum likelihood method based on the value of all the observations in the interval $\mathcal{I}$. If we cannot compute the estimate, we cannot estimate the CDF according to the model (step 2), we cannot compute the test statistic (step 3), and we will also be unable to simulate the critical value (step 4). Second, missing data will prevent us from identifying and, hence, estimating the distribution of the data, denoted by F_X (step 1.b).

The first problem may be avoided if we manage to estimate the parameter root- n-consistently in spite of the missing data problem. For example, suppose that our sample consists of observations of sample paths of an economic phenomenon over two years and there are missing data only during the second year (sample attrition). Then, it may be possible to estimate the parameter root-n-consistently using exclusively the information of the first year, where the sample is completely observed. In comparison, the second problem is unavoidable. If we are unwilling to make assumptions about the nature of the missing data, any period of unobserved data for functions in our sample implies that the distribution of the data is not identified. For the remainder of the paper, we will assume that the first problem can be avoided, i.e., we will assume there is a root-n-consistent estimator of θ_0 (despite the missing data), and we will focus the analysis on dealing with the second problem[10].

[10] It is conceptually possible to extend our analysis to consider the case when θ_0 is partially identified and, thus, to a case where no root-n -consistent estimator of θ_0 is available. However, this extension is likely to be computationally

In order to conduct the identification analysis, we now provide structure to the missing data problem. We interpret each of our observations as being the realization of a function over an interval of time, which we have denoted by $\mathcal{I}$. Before the introduction of missing data, we assume that there is an i.i.d. random sample of functions, referred to as the *underlying* sample and denoted by $\mathcal{X}_n = \{X_i\}_{i=1}^n$, each of them distributed according to F_X. This underlying sample is not directly observed. Instead, we observe a sample of data that is the result of an unspecified missing data process affecting the underlying sample. Specifically, for each observation $i = 1, \ldots, n$, instead of observing the underlying function $X_i : \Omega \to L_2(\mathcal{I})$, the researcher observes the pair $(O_i, X_i') : \Omega \to \mathcal{I} \times L_2(O_i)$, where the first component, $O_i : \Omega \to \mathcal{I}$, is the subset of the interval $\mathcal{I}$ where the underlying function is observed and the second component, $X_i' : \Omega \to L_2(O_i)$, is the value of the underlying stochastic process where it is observed (i.e. $X_i'(t) = X_i(t)$ for all $t \in O_i$). The *observed* sample is denoted by $\mathcal{X}_n' = \{(O_i, X_i')\}_{i=1}^n$. We further assume that the mapping between the underlying sample, $\mathcal{X}_n = \{X_i\}_{i=1}^n$, and the observed sample, $\mathcal{X}_n' = \{(O_i, X_i')\}_{i=1}^n$, is i.i.d. distributed across individuals and, thus, the resulting observed sample $\mathcal{X}_n' = \{(O_i, X_i')\}_{i=1}^n$ is an i.i.d. sample[11]. It is important to point out that we are being completely agnostic about how missing data affects each of the observations of the underlying random sample other than the fact that the resulting sample is i.i.d. In particular, the distribution of those functions that are fully observed is allowed to be different from the distribution of the i.i.d. sample of the underlying data, i.e., the data are not assumed to be missing at random.

We now impose one further assumption on the structure of missing data patterns. Recall that we denoted by $\mathcal{I}$ the interval of time where the functions are defined. We assume that there is a partition[12] of $\mathcal{I}$ into J sub-intervals ($J < \infty$), denoted by $\{\mathcal{I}_j\}_{j=1}^J$, that is fine enough so that for each function in the sample, a function can only be completely observed or completely unobserved within each of these J sub-itervals. In other words, none of the functions of the data should be partly observed and partly unobserved in any of the sub-intervals. In our notation, for every $i = 1, \ldots, n$ and $j = 1, \ldots, J$, only one of the following occur: $\mathcal{I}_j \subseteq O_i$ or $\mathcal{I}_j \subseteq \mathcal{I}/O_i$. This assumption limits the richness of the missing data process: it is restrictive in the sense that the sub-intervals are not allowed to be random and their cardinality is finite and not allowed to be a function of the sample size. For illustration purposes, we note that the restriction is satisfied in our empirical application with NLSY79 data. In this application, $\mathcal{I}$ is an interval of time spanning several years and the sample of wage functions can be missing or not at a weekly frequency, i.e., the set $\{\mathcal{I}_j\}_{j=1}^J$ is composed of the subset of intervals of weekly length in $\mathcal{I}$. As a consequence of this restriction on the missing data process, there are (only) 2^J possible missing data patterns. We can characterize each missing data pattern by the subset of $\mathcal{I}$ where the functions are observed, i.e., for $j = 1, \ldots, 2^J$, we denote by S_j the subset of $\mathcal{I}$ where functions are observed in the j^{th} missing data

challenging and, more importantly, to result in an uninformative hypothesis test. For this reasons, this extension is omitted.

[11] The i.i.d. assumption of the observed data is necessary for inferential purposes, see, e.g., Manski (2003).

[12] A collection of sub-intervals $\{\mathcal{I}_j\}_{j=1}^J$ is a partition of $\mathcal{I}$ if and only if $\mathcal{I}_j$ is non-empty for all $j = 1, \ldots, K$, $\mathcal{I} = \cup_{j=1}^J \mathcal{I}_j$, and $\mathcal{I}_{j_1} \cap \mathcal{I}_{j_2} = \varnothing$ for $j_1 \neq j_2$, $j_1, j_2 = 1, \ldots, J$.

pattern. Without loss of generality, we define the first missing data pattern to be the one without missing data, i.e., $S_1 = \mathcal{I}$.

The structure imposed so far is summarized in the following assumption.

Assumption 1. *The data generating process is defined on a complete probability space* $(\Omega, \mathcal{B}, P)$ *and it satisfies the following conditions:*

a. *The underlying data,* $\mathcal{X}_n = \{X_i\}_{i=1}^n$*, is an i.i.d. sample distributed according to a separable stochastic process in* $L_2(\mathcal{I})$*, almost surely. A generic observation of* $\mathcal{X}_n$ *is denoted by* X *and its CDF is denoted by* F_X*.*

b. *The observed data,* $\mathcal{X}'_n = \{(O_i, X'_i)\}_{i=1}^n$*, is an i.i.d. sample and is the result of missingness affecting* $\mathcal{X}_n$*. For every* $i = 1, \ldots, n$*, we observe the pair* (O_i, X'_i)*, where the first component,* O_i*, is the subset of* $\mathcal{I}$ *where the underlying stochastic process* X_i *is observed and the second component,* X'_i*, is the realization of the underlying stochastic process* X_i *on* O_i*. A generic observation of* $\mathcal{X}_n$ *is denoted by* (O, X')*.*

c. *There exists a partition of* $\mathcal{I}$ *into* J *sub-intervals* ($J < \infty$)*,* $\{\mathcal{I}_j\}_{j=1}^J$*, so that for every* $i = 1, \ldots, n$ *and* $j = 1, \ldots, J$*, either* $\mathcal{I}_j \subseteq O_i$ *or* $\mathcal{I}_j \subseteq \mathcal{I}/O_i$*.*

The derivation of the test proceeds as follows. The first step is to derive the identified set of the CDF of X, F_X, in the presence of missing functional data. The second step is to use this set to derive worst case scenario bounds for the population test statistic, $T(X, Y_{\theta_0})$. In these two steps, we assume that we know the population from where the observed data is sampled (of course, missing data is still unobserved) and, as a consequence, we compute the population version of these worst case scenario bounds. In the third and final step, we use sample analogue estimation and Monte Carlo integration to estimate the worst case scenario bounds, which allows us to implement our specification test.

3.1 Identified set for the CDF

The objective of this section is to characterize the identified set of F_X under the presence of missing data, which we denote by $\mathcal{H}(F_X)$.

Our first result illustrates the nature of the identification problem generated by the missing data. Specifically, the result provides an expression for $F_X(x)$ for any $x \in C'(\mathcal{I})$.[13]

Lemma 3.1. *Assume Assumption 1. For any* $x \in C'(\mathcal{I})$*:*

$$F_X(x) = \left\{ \begin{array}{l} P(\{X'(t) \leq x(t), \forall t \in \mathcal{I}\} \cap \{O = S_1\}) + \\ \sum_{j=2}^{2^J} P(\{X'(t) \leq x(t), \forall t \in S_j\} \cap \{X(t) \leq x(t), \forall t \in \mathcal{I}/S_j\} \cap \{O = S_j\}) \end{array} \right\}. \quad (3.1)$$

[13] The reason for restricting to functions in $C'(\mathcal{I})$ is that, in this case, the event $\{X(\omega, t) \leq x(t), \forall t \in \mathcal{I}\}$ is measurable (see Lemma A.3 in the appendix). The measurability is useful because it allows us to provide precise expressions in terms of probability, instead of bounds in terms of outer probability. As we show Lemma 3.2, the measurability is not necessary to establishing worst case scenario bounds. It is, however, necessary to establish that these bounds are sharp.

Moreover, for every $j = 2, \ldots, 2^J$, one of the following occurs:

- *If $P(\{X'(t) \le x(t), \forall t \in S_j\} \cap \{O = S_j\}) = 0$, then:*

$$P\left(\left\{X'(t) \le x(t), \forall t \in S_j\right\} \cap \{X(t) \le x(t), \forall t \in \mathcal{I}/S_j\} \cap \{O = S_j\}\right) = 0.$$

- *If $P(\{X'(t) \le x(t), \forall t \in S_j\} \cap \{O = S_j\}) > 0$, then:*

$$P\left(\left\{X'(t) \le x(t), \forall t \in S_j\right\} \cap \{X(t) \le x(t), \forall t \in \mathcal{I}/S_j\} \cap \{O = S_j\}\right) = \left\{ \begin{array}{c} P(\{X(t) \le x(t), \forall t \in \mathcal{I}/S_j\} \mid \{X'(t) \le x(t), \forall t \in S_j\} \cap \{O = S_j\}) \\ \times P(\{X'(t) \le x(t), \forall t \in S_j\} \mid \{O = S_j\}) P(O = S_j) \end{array} \right\}.$$

Lemma 3.1 expresses $F_X(x)$ as the sum of finitely many terms which, in general, are not point identified. In particular, the distribution of observables allows us to identify the frequency of each missing data pattern (i.e. $P(O = S_j)$ for $j = 1, \ldots, 2^J$) and the distribution of the random functions where these are observed (i.e. $P(\{X'(t) \le x(t), \forall t \in S_j\} | O = S_j)$ for $j = 1, \ldots, 2^J$), but is completely silent about the (conditional) distribution of the random functions where these functions are unobserved (i.e. $P(\{X(t) \le x(t), \forall t \in \mathcal{I}/S_j\} | \{O = S_j\} \cap \{X'(t) \le x(t), \forall t \in S_j\})$ for $j = 2, \ldots, 2^J$). We obtain bounds for the distribution of the data by imposing logical bounds to all the expressions that are not identified.

Our next result establishes worst case scenario bounds (WCSBs) for $F_X(x)$ for any $x \in \mathbb{R}^{\mathcal{I}}$, which constitute the basic building block to characterize $\mathcal{H}(F_X)$.

Lemma 3.2. *Assume Assumption 1. Define the functions $F_X^L : \mathbb{R}^{\mathcal{I}} \to \mathbb{R}$ and $F_X^H : \mathbb{R}^{\mathcal{I}} \to \mathbb{R}$ as follows. For every $x \in \mathbb{R}^{\mathcal{I}}$:*

$$\begin{aligned} F_X^L(x) &\equiv P^*(\{X'(t) \le x(t), \forall t \in \mathcal{I}\} \cap \{O = \mathcal{I}\}), \\ F_X^H(x) &\equiv \textstyle\sum_{j=1}^{2^J} P^*(\{X'(t) \le x(t), \forall t \in S_j\} \cap \{O = S_j\}). \end{aligned} \tag{3.2}$$

For any $x \in \mathbb{R}^{\mathcal{I}}$, $F_X(x)$ satisfies the following WCSBs:

$$F_X(x) \in \left[F_X^L(x), F_X^H(x)\right]. \tag{3.3}$$

Moreover, for $x \in C'(\mathcal{I})$, these bounds are sharp, i.e., $F_X(x)$ cannot be restricted any further.

The bounds described in Lemma 3.2 are "pointwise" bounds in the sense that they are established by conducting a worst case scenario analysis for F_X evaluated at $x \in \mathbb{R}^{\mathcal{I}}$. Notice that when the data are completely observed (i.e. $P(\{O = \mathcal{I}\}) = 1$), then $F_X^L = F_X^H$ and, as expected, F_X is point identified. Finally, notice that measurability is not necessary to establish (informative) WCSBs, i.e., we can provide (informative) bounds on $F_X(x)$ in terms of outer probability measure P^*. Nevertheless, restricting attention to $x \in C'(\mathcal{I})$ implies that the events under consideration are measurable, and this, in turn, is instrumental in establishing that the WCSBs are sharp.

The following step is to use the previous findings to characterize $\mathcal{H}(F_X)$. By definition, this set is composed of all the functions that are logically possible candidates for the CDF of X according to the observed data. As we now argue, providing an exact characterization of this set is a computationally complicated problem. We now explain why.

Notice that F_X needs to satisfy the restrictions imposed by the fact that it is the CDF of a separable stochastic process in $L_2(\mathcal{I})$. Denote by Γ the space of CDFs for separable stochastic processes in $L_2(\mathcal{I})$, i.e., for any $G \in \Gamma$, there exists a separable stochastic process $Y : \Omega \to L_2(\mathcal{I})$ such that:

$$G(x) = P^*(\omega \in \Omega : \{Y(\omega, t) \leq x(t), \ \forall t \in \mathcal{I}\}), \text{ for all } x \in \mathbb{R}^{\mathcal{I}}.$$

It is well known that the space of CDFs of random variables in finite dimensional Euclidean spaces can be characterized by a set of defining properties[14]. It is possible to extend these results for separable stochastic processes in $L_2(\mathcal{I})$.[15] Our analysis reveals that these restrictions are computationally hard to impose, which implies that Γ is a relatively hard space to work with. To complicate things even further, F_X also needs to satisfy "pointwise" bounds derived in Lemma 3.2. From our arguments, it follows that:

$$\mathcal{H}(F_X) \subseteq \left\{\Gamma \cap \left\{G : F_X^L(x) \leq G(x) \leq F_X^H(x) \ \forall x \in \mathbb{R}^{\mathcal{I}}\right\}\right\}.$$

We show in the appendix that the reverse inclusion does not hold in general (See Lemma A.4). Intuitively speaking, the restrictions imposed by Γ and the restrictions imposed by "pointwise" bounds derived in Lemma 3.2 interact with each other to restrict the space of functions by more than their intersection. Since the restrictions imposed by Γ are already relatively hard to work with, this makes $\mathcal{H}(F_X)$ is a computationally hard set to characterize. This is unfortunate because, as we reveal next, $\mathcal{H}(F_X)$ is instrumental in characterizing the WCSBs of the object we are ultimately interested in, namely, the population version of the test statistic. The next subsection develops these WCSBs and provides a feasible way to deal with these complications.

3.2 Bounds for the population test statistic

The objective of this section is to develop WCSBs for the population version of the test statistic. The following result characterizes the sharp version of the WCSBs.

[14] For one-dimensional random variables see, e.g., Billingsley (1995), Theorem 14.1. For multi-dimensional random variables see, e.g., Rao (1995), page 4, or Gihman and Skorohod (1974), page 13.

[15] This extension is based on the Daniell-Kolmogorov extension theorem (see, e.g., Rao (1995), pages 4-15). For reasons of brevity, this analysis is not included in the paper. The formal arguments are available from the author, upon request.

Lemma 3.3. *Assume Assumption 1. Let $T_L(X, Y_{\theta_0})$ and $T_H(X, Y_{\theta_0})$ be defined as follows:*

$$\begin{aligned} T_L(X, Y_{\theta_0}) &= \inf_{G \in \mathcal{H}(F_X)} \int (G(x) - F_Y(x|\theta_0))^2 \, d\mu(x), \\ T_H(X, Y_{\theta_0}) &= \sup_{G \in \mathcal{H}(F_X)} \int (G(x) - F_Y(x|\theta_0))^2 \, d\mu(x). \end{aligned} \tag{3.4}$$

Then, the population test statistic satisfies the following WCSBs:

$$T_L(X, Y_{\theta_0}) \leq T(X, Y_{\theta_0}) \leq T_H(X, Y_{\theta_0}). \tag{3.5}$$

Furthermore, without additional information, these WCSBs are sharp, i.e., the population test statistic cannot be restricted any further.

The presence of missing data opens a gap between the worst case scenario lower and upper bounds for the population test statistic. Lemma 3.3 provides an implicit characterization a sharp version of these bounds and, thus, represents the best we can offer with the available information. The computation of these bounds requires solving an infinite dimensional constrained optimization problem where the constraint set is $\mathcal{H}(F_X)$. As we have already explained, $\mathcal{H}(F_X)$ is a complex space to work on and, as a consequence, this makes the computation of the sharp WCSBs intractable to compute or estimate. In order to circumvent this problem, we consider alternative ways of imposing bounds that are informative about the population test statistic but are amenable to compute and estimate. These bounds will be referred to as *alternative WCSBs*.

Our strategy to obtain alternative WCSB is to replace the constraint set $\mathcal{H}(F_X)$ in the optimization problems of Eq. (3.4) by a superset of $\mathcal{H}(F_X)$, generically denoted by $\mathcal{H}'(F_X)$, which produces a tractable optimization problem. This replacement is expected to result in loss of information, i.e., the alternative WCSBs might not be as restrictive as the sharp WCSBs. Nevertheless, given the computational complexity of the sharp WCSBs, we consider alternative WCSBs to be our only feasible option. The following result formalizes the concept of alternative WCSBs and describes their basic property.

Lemma 3.4. *Assume Assumption 1. Let $T'_L(X, Y_{\theta_0})$ and $T'_H(X, Y_{\theta_0})$ be defined as follows:*

$$\begin{aligned} T'_L(X, Y_{\theta_0}) &= \inf_{G \in \mathcal{H}'(F_X)} \int (G(x) - F_Y(x|\theta_0))^2 \, d\mu(x), \\ T'_H(X, Y_{\theta_0}) &= \sup_{G \in \mathcal{H}'(F_X)} \int (G(x) - F_Y(x|\theta_0))^2 \, d\mu(x), \end{aligned} \tag{3.6}$$

where $\mathcal{H}'(F_X)$ is such that $\mathcal{H}(F_X) \subseteq \mathcal{H}'(F_X)$. Then, the population test statistic satisfies the following WCSBs:

$$T'_L(X, Y_{\theta_0}) \leq T(X, Y_{\theta_0}) \leq T'_H(X, Y_{\theta_0}). \tag{3.7}$$

These bounds are referred to as alternative WCSBs.

The choice of the constraint set $\mathcal{H}'(F_X)$ in Eq. (3.6) generates a trade-off between informational content of the bounds and their tractability. Choosing a smaller constraint set produces more informative (i.e. tighter) bounds but also typically makes the computation of these bounds harder. As we have already explained, choosing $\mathcal{H}'(F_X) = \mathcal{H}(F_X)$ implies that the associated bounds are not computationally feasible. In practice, we found that an adequate solution to this trade-off is achieved by adopting the following constraint set:

$$\mathcal{H}'(F_X) = \left\{G : F_X^L(x) \leq G(x) \leq F_X^H(x)\ \forall x \in \mathbb{R}^{\mathcal{I}}\right\}. \tag{3.8}$$

Henceforth, for the remainder of this paper, we exclusively refer to alternative WCSBs as the ones that correspond to the constraint set in Eq. (3.8). In this case, these alternative WCSBs have the following closed form solution:

$$\begin{aligned} T_L'(X, Y_{\theta_0}) &= \int \left\{ \begin{array}{c} 1\left[F_Y(x|\theta_0) < F_X^L(x)\right]\left(F_X^L(x) - F_Y(x|\theta_0)\right)^2 + \\ 1\left[F_Y(x|\theta_0) > F_X^H(x)\right]\left(F_X^H(x) - F_Y(x|\theta_0)\right)^2 \end{array} \right\} d\mu(x), \\ T_H'(X, Y_{\theta_0}) &= \int \max\left\{\left(F_X^L(x) - F_Y(x|\theta_0)\right)^2, \left(F_X^H(x) - F_Y(x|\theta_0)\right)^2\right\} d\mu(x) \end{aligned} \tag{3.9}$$

and are therefore straightforward to compute. Furthermore, our formal results in later sections reveal that these WCSBs are straightforward to estimate consistently (See Theorem A.6 in Section 4.2). The computational simplicity of the proposed alternative WCSBs comes at a cost of a potential lack of sharpness. In fact, we verify in the appendix that it is possible that neither the lower alternative WCSB nor the upper alternative WCSB are non-sharp (See Lemma A.5). In light of these findings, the reader might wonder whether the proposed bounds are too wide to be of practical relevance. We verify through Monte Carlo simulations and with our empirical application that the alternative WCSB can still be quite informative about the specification of the data generating process.

4 Specification test with missing functional data

This section utilizes the identification analysis of Section 3 to develop our specification test described in Eq. (2.1) under the presence of missing functional data. We begin by introducing several assumptions that are also assumed in BHHN.

Assumption 2. *The parameter space is denoted by $\Theta \subseteq \mathbb{R}^p$ for some $p < \infty$. Moreover:*

a. θ_0 is uniquely defined as follows: $\theta_0 = \arg\min_{\theta\in\Theta}\{T(X, Y_\theta)\}$.

b. $\hat{\theta}_0$ is an estimator for θ_0, that is exclusively a function of $\mathcal{X}_n$[16] and has the following asymp-

[16] The estimator is not allowed to depend on the random variables used to estimate the CDF of the model or to approximate the integrals (see Assumption 4). This restriction could be easily dispensed at the expense of stronger assumptions.

totic representation:

$$n^{1/2}(\hat{\theta}_0 - \theta_0) = n^{-1/2} \sum_{i=1}^{n} \Lambda\left(X_i\right) + o_p\left(1\right),$$

where Λ *is a* p*-dimensional function such that* $\mathbb{E}\left(\Lambda\left(X\right)\right) = 0$ *and* $cov\left(\Lambda\left(X\right)\right)$ *is non-singular and* $\int \Lambda\left(x\right)' \Lambda\left(x\right) d\mu\left(x\right) < \infty$.

Assumption 3. $\partial F_Y\left(\cdot|\theta\right)/\partial\theta$ *exists for all* θ *in an open set* $\mathcal{O}$ *that contains* θ_0*. Moreover:*

$$\sup_{\theta \in \mathcal{O}} \int \frac{\partial F_Y(x|\theta)}{\partial \theta'} \frac{\partial F_Y(x|\theta)}{\partial \theta} d\mu\left(x\right) < \infty,$$

$$\lim_{\varepsilon \to 0} \int \sum_{i,j=1,\ldots,p} \sup_{\|\theta-\theta_0\| \leq \varepsilon} \left| \left[\frac{\partial F_Y(x|\theta)}{\partial \theta_i} - \frac{\partial F_Y(x|\theta_0)}{\partial \theta_i}\right] \left[\frac{\partial F_Y(x|\theta)}{\partial \theta_j} - \frac{\partial F_Y(x|\theta_0)}{\partial \theta_j}\right] \right| d\mu\left(x\right) = 0.$$

Assumption 4. μ *is the measure induced by the following process:*

$$Z\left(t\right) = \sum_{l=1}^{\infty} \rho_l N_l \phi_l\left(t\right), \tag{4.1}$$

where $\{N_l\}_{l=1}^{\infty}$ *is a sequence of i.i.d. standard normal random variables and there are constants* $C > 0$ *and* $d > 1$ *such that* $0 < |\rho_l| \leq Cl^{-d}$ *and* $\phi_l\left(t\right) = \sqrt{2}\sin\left(l\pi t\right)$ *for all* $l \in \mathbb{N}$.

Assumption 5. *The test is implemented using the following approximations:*

a. *For any* $\theta \in \Theta$ *and* $x \in \mathbb{R}^{\mathcal{I}}$, $F_Y\left(x|\theta\right)$ *is approximated with its sample analogue, i.e.,*

$$\hat{F}_Y(x|\theta) = \frac{1}{m} \sum_{i=1}^{m} 1\left(Y_{\theta,i}\left(t\right) \leq x\left(t\right), \forall t \in \mathcal{I}\right),$$

where $\mathcal{Y}_{\theta,m} \equiv \{Y_{\theta,i}\}_{i=1}^{m}$, *is an i.i.d. sample of size* m *distributed according to* Y_θ.

b. *Integrals with respect to measure* μ *are approximated by Monte Carlo integration. In other words, for any function* $\Upsilon : L_2\left(\mathcal{I}\right) \to \mathbb{R}$, $\int \Upsilon\left(x\right) d\mu\left(x\right)$ *is approximated by* $\frac{1}{V} \sum_{j=1}^{V} \Upsilon\left(Z_j\right)$, *where* $\mathcal{Z}_V \equiv \{Z_j\}_{j=1}^{V}$ *is and i.i.d. sample of size* V *distributed according to the measure* μ.

c. *For every* $\theta \in \Theta$, $\mathcal{X}_n$, $\mathcal{Y}_{\theta,m}$, *and* $\mathcal{Z}_V$ *are independent samples.*

We note that these assumptions are slight variations of conditions required by BHHN. Assumption 2.a indicates that θ_0 is the unique minimizer of the population test statistic $T\left(X, Y_\theta\right)$. Since $T\left(X, Y_\theta\right)$ is not point identified due to the missing data problem, it is not clear from Assumption 2.a whether θ_0 is identified or whether it can be root-n-consistently estimated. Nevertheless, Assumption 2.b indicates that missing data does not preclude our ability to estimate the parameter θ_0 root-n-consistently, which implicitly assumes that this parameter is point identified. As we have already explained, this is a reasonable assumption when the sample is fully observed during a certain period, as this information can be sufficient to estimate the parameter θ_0 root-n-consistently.

The NLSY79 survey used in our empirical application is a perfect example of a situation where this assumption is satisfied. In this survey, a sample of n individuals are randomly selected from the population and are asked to describe their wage path during the previous year (among numerous other questions). In order to complete the evolution of the wage path over, say, 10 years, the same n individuals are revisited in the following years. Unfortunately, during these subsequent years, some of these individuals are unavailable due to sample attrition. As a consequence, the n observations are fully observed during the first year but a subset of them is affected by missing sections starting from year 2 onwards. However, in certain models, such as the Burdett-Mortensen labor market model, one year of data for the n observations in the sample is enough to estimate the parameter vector in a root-n-consistent fashion using any standard extremum estimation method, such as maximum likelihood[17]. As a consequence, the presence of missing data does not preclude our ability of estimating the parameter vector root-n-consistently.

Assumption 3 requires the model to be sufficiently smooth with respect to the parameter value. Assumption 4 defines the bounded and non-degenerate measure μ on $L_2(\mathcal{I})$ used to measure the distance between the CDF of the model and the CDF the data. The measure is constructed using a sequence of sine functions as a basis of the functional space, but it is possible to use different measures by changing the class of basis functions[18] . Assumption 5 describes the methodology that will be used to approximate the CDF of the model and the integrals with respect to the measure μ. The methodology is based on sample analogue estimation and is implicitly assumed in BHHN. The value of making this assumption explicit is to reveal that the performance of the hypothesis test depends on the parameters that control the quality of the approximation, m and V, as well as the sample size, n. As we explain in Section 4.2, our analysis is based on asymptotic arguments and it requires that the sample size and quality of the approximation to diverge to infinity. We defer comments regarding the choice of the quality of the approximation to that section.

To conclude the section, we note that simulating from the measure μ is practically unfeasible, as Eq. (4.1) expresses that it is the measure induced by an infinite sum of random variables. In order to circumvent this issue, BHHN propose replacing the measure μ with a finite dimensional approximation, i.e., for some $M \in \mathbb{N}$, replace μ with μ_M, which is the measure induced by $Z(t) = \sum_{l=1}^{M} \rho_l N_l \phi_l(t)$, where $\{N_l\}_{l=1}^{M}$, $\{\rho_l\}_{l=1}^{M}$, and $\{\phi_l\}_{l=1}^{M}$ are as in Assumption 4. Section 3.3 in BHHN shows that this replacement has no effect on the asymptotic behavior of the test provided that $M \to \infty$. Their arguments can be used to establish the same result for the test considered in this paper.

In their simulations, BHHN implement their hypothesis test for relatively low values of M (e.g.

[17] In this case, $\hat{\theta}_0 = \arg\max_{\theta\in\Theta}\{n^{-1}\sum_{i=1}^{n}\ln f(X_i|\theta)\}$, where $f(X|\theta)$ denotes the likelihood of observing the function X during the first year and the parameter is θ. Notice that even though $X = \{X(t),\, t \in \mathcal{I}\}$ is a random function over 10 years (possibly with missing sections), only the first year is used to construct the likelihood function. Under mild regularity assumptions, the maximum likelihood estimator can be shown to satisfy Assumption 2.b with $\Lambda(X) \equiv \partial \ln f(X|\theta_0)/\partial\theta$.

[18] For example, we could form a basis by using a sequence of cosine functions, a sequence of both sine and cosine functions, or a sequence of polynomials.

$M = 4$), and obtain excellent finite sample results in terms of size and power. While they do not formally analyze the asymptotic properties of their test implemented with measure μ_M for a fixed M, it is not hard to conjecture what these properties are. First, their hypothesis test implemented with the measure μ_M (instead of μ) is asymptotically size correct. This is because the critical value of the test is computed with the bootstrap, which provides size correctedness regardless of the measure employed to compute the test statistic. Second, it is not hard to see that the test implemented with measure μ_M will be consistent against fixed alternative hypotheses that belong to a certain set. Furthermore, it is clear that this set of alternative hypotheses is a subset of the corresponding set for the test implemented with measure μ . We now explain why. The BHHN hypothesis test considers the CDF of the data and the CDF of the model to be equal if and only if they coincide for every function in the measure used to implement the test. As the support of the measure μ_M is a subset of the support of the measure μ, the test implemented with measure μ is able to detect deviations from the null hypothesis that go undetected for the test implemented with measure μ_M. However, if a given alternative hypothesis can be detected by the measure μ_M, then the corresponding BHHN test will be consistent [19]. This explains why the Monte Carlo evidence in BHHN reveals that the hypothesis test can have excellent performance for relatively low values of M. The Monte Carlo evidence presented in Section 5 reveals that our hypothesis test is in line with these findings.

4.1 Implementation of the test

Our specification test for missing functional data is given by the following steps:

1. Use the data sample, $\mathcal{X}_n$, to:

   a. Estimate θ_0 using an estimation procedure that is root-n-consistent under the presence of missing data (recall that this is possible according to Assumption 2). Denote this estimate $\hat{\theta}_0$.

   b. Estimate the upper and lower bounds for F_X, denoted by $\hat{F}_X^L$ and $\hat{F}_X^H$, respectively. For every $x \in \mathbb{R}^{\mathcal{I}}$, these are given by:

$$\hat{F}_X^L(x) = \hat{P}^*\left(\{X'(t) \leq x(t), \forall t \in O\} \cap \{O = \mathcal{I}\}\right),$$
$$\hat{F}_X^H(x) = \sum_{j=1}^{2^J} \hat{P}^*\left(\{X'(t) \leq x(t), \forall t \in S_j\} \cap \{O = S_j\}\right).$$

   For every $x \in \mathbb{R}^{\mathcal{I}}$ and $j = 1, \ldots, J$, $\hat{P}^*\left(\{X'(t) \leq x(t), \forall t \in S_j\} \cap \{O = S_j\}\right)$ is the sample analogue estimator, given by:

$$\hat{P}^*\left(\{X'(t) \leq x(t), \forall t \in S_j\} \cap \{O = S_j\}\right) = n^{-1} \sum_{i=1}^{n} 1\left(X_i'(t) \leq x(t), \forall t \in S_j\right) 1\left(O_i = S_j\right).$$

[19] An analogous result could be established for power against sequences of local alternative hypotheses.

2. Use $\hat{\theta}_0$ to construct an i.i.d. sample of size m of $Y_{\hat{\theta}_0}$, denoted by $\mathcal{Y}_{\hat{\theta}_0,m} \equiv \{Y_{\hat{\theta}_0,i}\}_{i=1}^{m}$. Use $\mathcal{Y}_{\hat{\theta}_0,m}$ to estimate $F_Y(\cdot|\hat{\theta}_0)$ by sample analogue estimation, i.e., for every $x \in L_2(\mathcal{I})$:

$$\hat{F}_Y(x|\hat{\theta}_0) = \tfrac{1}{m}\sum_{i=1}^{m} 1(Y_{\hat{\theta}_0,i}(t) \leq x(t), \forall t \in \mathcal{I}).$$

3. Compute the alternative WCSBs for the test statistic:

$$\begin{aligned} \hat{T}'_L(X,Y_{\hat{\theta}_0}) &= \tfrac{1}{V}\sum_{j=1}^{V}\left\{\begin{array}{c} 1[\hat{F}_Y(Z_j|\hat{\theta}_0) < \hat{F}_X^L(Z_j)](\hat{F}_X^L(Z_j) - \hat{F}_Y(Z_j|\hat{\theta}_0))^2 + \\ 1[\hat{F}_Y(Z_j|\hat{\theta}_0) > \hat{F}_X^H(Z_j)](\hat{F}_X^H(Z_j) - \hat{F}_Y(Z_j|\hat{\theta}_0))^2 \end{array}\right\}, \\ \hat{T}'_H(X,Y_{\hat{\theta}_0}) &= \tfrac{1}{V}\sum_{j=1}^{V}\max\{(\hat{F}_X^L(Z_j) - \hat{F}_Y(Z_j|\hat{\theta}_0))^2, (\hat{F}_X^H(Z_j) - \hat{F}_Y(Z_j|\hat{\theta}_0))^2\}, \end{aligned} \quad (4.2)$$

where $\mathcal{Z}_V \equiv \{Z_j\}_{j=1}^{V}$ is an i.i.d. sample of μ, $\hat{F}_X^L$ and $\hat{F}_X^H$ are as in step 1.b, and $\hat{F}_Y(\cdot|\hat{\theta}_0)$ is as in step 2.

4. Repeat the following many times:
   a. Construct a bootstrap sample of size n of $Y_{\hat{\theta}_0}$ and denote it by $\mathcal{X}_n^* \equiv \{X_i^*\}_{i=1}^{n}$.
   b. Use $\mathcal{X}_n^*$ to:
      i. Estimate the parameter θ_0 root-n-consistently as in step 1.a. Denote this estimate by $\hat{\theta}_0^*$.
      ii. Estimate F_{X^*} by sample analogue estimation, i.e., for every $x \in L_2(\mathcal{I})$:

$$\hat{F}_{X^*}(x) = \tfrac{1}{n}\sum_{i=1}^{n} 1(X_i^*(t) \leq x(t), \forall t \in \mathcal{I}).$$

   c. Use $\hat{\theta}_0^*$ as in step 2, i.e., use it to construct $\mathcal{Y}_{\hat{\theta}_0^*,m}$, and use this to estimate $F_Y(\cdot|\hat{\theta}_0^*)$ by sample analogue estimation.
   d. Compute the test statistic from the bootstrapped data as follows:

$$\hat{T}(X^*,Y_{\hat{\theta}_0^*}) = \tfrac{1}{V}\sum_{j=1}^{V}(\hat{F}_{X^*}(Z_j) - \hat{F}_Y(Z_j|\hat{\theta}_0^*))^2,$$

   where $\mathcal{Z}_V \equiv \{Z_j\}_{j=1}^{V}$ is an i.i.d. sample of μ, and $\hat{F}_{X^*}$ and $\hat{F}_Y(\cdot|\hat{\theta}_0^*)$ are as in steps 4.b and 4.c, respectively.

5. Denote by $t^*_{\hat{\theta}_0}(1-\alpha)$ the $(1-\alpha)$ quantile of the simulated statistics $n\hat{T}(X^*,Y_{\hat{\theta}_0^*})$ in step 4.

6. Decide the outcome of the test in the following way:

Outcome	Decision
$t^*_{\hat{\theta}_0}(1-\alpha) < n\hat{T}'_L(X,Y_{\hat{\theta}_0}) \leq n\hat{T}'_H(X,Y_{\hat{\theta}_0})$	Reject H_0
$n\hat{T}'_L(X,Y_{\hat{\theta}_0}) \leq n\hat{T}'_H(X,Y_{\hat{\theta}_0}) \leq t^*_{\hat{\theta}_0}(1-\alpha)$	Do not reject H_0
$n\hat{T}'_L(X,Y_{\hat{\theta}_0}) \leq t^*_{\hat{\theta}_0}(1-\alpha) < n\hat{T}'_H(X,Y_{\hat{\theta}_0})$	Inconclusive

As we have shown in Section 3.2, the presence of missing data opens a gap between the population lower and upper bounds for the population test statistic. The gap at a population level gets mapped into its sample analogue and, thus, the hypothesis test has a region where the test is inconclusive. This is an undesired but unavoidable consequence of having missing data and imposing no assumptions regarding their distribution[20].

The third step of the hypothesis test presents closed form solutions for the alternative WCSBs (Eq. (4.2)). These bounds are the sample analogues of the population version of the alternative WCSBs in Eq. (3.9), i.e., they are the direct result of constraining the CDF of the data to the sample analogue estimator of $\mathcal{H}'(F_X)$. Recall from the discussion in Sections 3.2 and 3.1 that $\mathcal{H}'(F_X)$ is a superset of the sharp identified set for the CDF of the data that was chosen due to computational simplicity. As we can see from our implementation, this choice actually results in closed form solutions for the sample alternative WCSBs. In practice, we could have estimated sharper (i.e. more informative) versions of alternative WCSBs by incurring large computational costs[21]. Nevertheless, our experience from implementing our bounds in Monte Carlo simulations and in the empirical application reveal that the WCSBs proposed in this paper are not only simple to implement but are also extremely informative about the hypothesis of interest.

4.2 Properties of the test

As in any other hypothesis test, the properties of the hypothesis test depend on whether the null hypothesis is true or false, i.e., whether $T(X, Y_{\theta_0}) = 0$ or $T(X, Y_{\theta_0}) > 0$. Moreover, in the presence of missing data, the properties of the hypothesis test depend on the associated identification problem. The following table describes all the possibilities:

	H_0 is true ($T(X, Y_{\theta_0}) = 0$)	H_0 is false ($T(X, Y_{\theta_0}) > 0$)
$T'_L(X, Y_{\theta_0}) = 0$, $T'_H(X, Y_{\theta_0}) = 0$	case 1	impossible
$T'_L(X, Y_{\theta_0}) = 0$, $T'_H(X, Y_{\theta_0}) > 0$	case 2	case 3
$T'_L(X, Y_{\theta_0}) > 0$, $T'_H(X, Y_{\theta_0}) > 0$	impossible	case 4

The columns of the table represent the unknown truth that we are interested in learning and the rows represent the truth that can be identified from the population affected by missing data. The first row (case 1) corresponds to the case when the null hypothesis is true and this can be identified from the population. The last row (case 4) corresponds to the case when the null hypothesis is false and this can be identified from the population. Finally, the middle row (cases 2 and 3) represents

[20] Since lack of rejection of H_0 does not imply that we accept H_0, one could correctly relabel the "Inconclusive" outcome as "Do not reject H_0" outcome. After this relabeling, the resulting hypothesis test can be formally interpreted as a hypothesis test in which the null hypothesis is that the distribution of the model belongs to the identified set of the distribution of the data, and the alternative hypothesis is the negation of the null hypothesis. I thank an anonymous referee for suggesting this interpretation of the test.

[21] In particular, we have considered alternative WCSBs that incorporated the fact that the CDF has to be a weakly increasing function. In practice, this restrictions resulted in WCSBs that were extremely costly to compute and only slightly sharper.

the situation where we cannot decide if the null hypothesis is true or not, even if we knew the data generating process of the observed data.

We now provide results regarding the asymptotic performance of our test. From Section 4.1, it is clear that the performance of the test depends on the sample size, n, and on the quality of the approximation procedure described in Assumption 5, controlled by m and V. Our analysis is asymptotic in the sense that all of these parameters are required to diverge to infinity. Moreover, we consider the situation when these parameters are assumed to diverge to infinity in turns, i.e., we study the performance when $V \to \infty$, $m \to \infty$, and, then, $n \to \infty$. It is not hard to show that the order in which the three sequences are assumed to diverge to infinity is not relevant, as long as they diverge to infinity in turns. Nevertheless, assuming that $V \to \infty$ and $m \to \infty$ occur before $n \to \infty$ occurs is not a problematic assumption as the researcher is always in control of the simulation error (for any sample size) [22]. Finally, given the assumptions in BHHN, we suspect that the sequential application of limits is implicitly behind their results[23].

Theorem 4.1. *Assume Assumptions 1-5. If* $T(X, Y_{\theta_0}) = 0$, *then:*

$$\limsup_{n\to\infty} \limsup_{m\to\infty} \limsup_{V\to\infty} P\left(t^*_{\hat{\theta}_0}(1-\alpha) < n\hat{T}'_L(X, Y_{\hat{\theta}_0}) \leq n\hat{T}'_H(X, Y_{\hat{\theta}_0})\right) \leq \alpha.$$

Theorem 4.1 implies that the hypothesis test is asymptotically correct in level but may result in conservative inference. From the analysis in Section 3, we know that when there are no missing data, the upper and lower bounds collapse and our hypothesis testing procedure coincides with the one in BHHN, which result in non-conservative inference. Hence, our hypothesis test may be conservative, but this is solely due to the presence of missing data.

Theorem 4.2. *Assume Assumptions 1-5. Suppose that for some* $D > 0$, $T'_H(X, Y_{\theta_0}) = D > 0$ *and* $T(X, Y_{\theta_0}) = 0$, *i.e., case 2 occurs. Then:*

$$\lim_{n\to\infty} \lim_{m\to\infty} \lim_{V\to\infty} P\left(n\hat{T}'_L(X, Y_{\hat{\theta}_0}) \leq n\hat{T}'_H(X, Y_{\hat{\theta}_0}) \leq t^*_{\hat{\theta}_0}(1-\alpha)\right) = 0.$$

Theorem 4.2 reveals that when the null hypothesis is true but the WCSBs do not cooperate with this information, then the probability of making the right decision (not rejecting the null hypothesis)

[22] By imposing additional assumptions, it is conceivable to extend the asymptotic results to the case when $(V, m, n) \to (\infty, \infty, \infty)$. However, we avoid doing this to keep our assumptions as close to the conditions in BHHN as possible and to simplify our analysis.

[23] In practice, the researcher needs to make specific choices of m and V for a particular sample size of n. It is not hard to derive from our formal results that the error of the approximation in Assumption 5 is reduced as m and V increase, and, in this sense, the researcher should choose these parameters as high as possible (ideally, they should be set to infinity). As the researcher increases m and V, one expects that our asymptotic results become a better approximation of the finite sample behavior. Unlike with other parameter choices in econometrics (e.g., bandwidth parameter in kernel estimation or number of sieve terms in sieve estimation), the choice of these parameters presents no trade-off for the inferential problem: a higher value represents an improvement in the accuracy of the results. In this sense, in practice, m and V should be set as high as the computationally feasible. Our Monte Carlo evidence suggests that the finite sample results are robust to changes in these parameters, provided that they are set high enough, relative to the sample size.

converges to zero. This is an undesirable but inevitable result of undertaking an agnostic approach in the presence of missing data, as we only learn features of the population through the information conveyed by the WCSBs.

Theorem 4.3. *Assume Assumptions 1-5. Suppose that for some* $D > 0$, $T(X, Y_{\theta_0}) \geq T'_L(X, Y_{\theta_0}) = D > 0$, *i.e., case 4 occurs. Then:*

$$\lim_{n\to\infty}\lim_{m\to\infty}\lim_{V\to\infty} P\left(t^*_{\hat{\theta}_0}(1-\alpha) < n\hat{T}'_L(X, Y_{\hat{\theta}_0}) \leq n\hat{T}'_H(X, Y_{\hat{\theta}_0})\right) = 1.$$

Theorem 4.3 shows that whenever the null hypothesis is false and the worst case scenario bounds contain this information, then, the probability of making the right decision (rejecting) converges to one. In other words, the test is consistent against fixed alternative hypotheses, provided that this information is revealed by the observed population.

In order to provide a full characterization of the properties of the hypothesis test for fixed hypotheses, we should consider case 3, i.e., for a fixed $D > 0$, $T(X, Y_{\theta_0}) = D > 0$ and $T'_L(X, Y_{\theta_0}) = 0$. In this case, the asymptotic behavior of the probability of making the correct decision (i.e. rejecting the null hypothesis) depends on the underlying parameters of the data generating process. In fact, parts 1 and 2 of Lemma A.7 in the appendix reveal that it is possible that this probability converges to zero or to a positive number, depending on the features of the distribution. From this it follows that a general characterization of the properties of this test under case 3 appears to be an extremely complicated problem that is out of the scope of this paper.

We conclude the section by considering the behavior of the test under sequences of local alternative hypotheses, i.e., $T(X, Y_{\theta_0}) = D_n$ with $D_n > 0$ and $D_n \to 0$ as $n \to \infty$. By definition of the alternative WCSBs, $T'_L(X, Y_{\theta_0}) \leq T(X, Y_{\theta_0})$ and, thus, $T'_L(X, Y_{\theta_0}) \to 0$ as $n \to \infty$. The local power properties of the test depend on the underlying parameters of the data generating process that govern the rate at which $T'_L(X, Y_{\theta_0})$ vanishes. On the one hand, consider the extreme case when there are no missing data. In this situation, $T'_L(X, Y_{\theta_0}) = T(X, Y_{\theta_0})$ and, consequently, they both vanish at the same rate. Without missing data, our hypothesis test coincides exactly with the one developed by BHHN and, as a result, the test procedure exhibits non-trivial power against local alternatives with $D_n = Dn^{-1}$ for some $D > 0$.[24] On the other hand, a significant amount of missing data can result in $T'_L(X, Y_{\theta_0})$ vanishing much faster than $T(X, Y_{\theta_0})$, resulting in little or no power against local alternative hypotheses. In fact, Part 3 of Lemma A.7 in the appendix shows that when the missing data problem becomes overwhelming, the probability of rejection of the null hypothesis converges to zero. Between these two polar cases, there are a plethora of possibilities that depend on the underlying features of the data generating process. From this, we conclude that a general characterization of the local power properties of the test appears to be a formidable task, but it is not hard to understand which are the forces at play. The informational content of the

[24]According to Eq. (2.3), $D_n = Dn^{-1}$ for some $D > 0$ is equivalent to $F_X(x) = F_Y(x|\theta) + n^{-1/2}H(x)$ for some function H that satisfies: $D = \int (H(x))^2 \, d\mu(x) > 0$. The result then follows immediately from Theorem 3.1 in BHHN.

local alternative hypothesis works in favor of producing non-trivial rejection rates but the partial identification problem caused by the missing data works in the opposite direction and can even overturn its effect.

5 Monte Carlo simulations

In this section, we describe a Monte Carlo simulation used to study the performance of our specification test. The framework for these simulations is a two-sector version of the Burdett-Mortensen labor market model. We now succinctly describe the setup of the Q-sector version of this model, and readers who are interested in the details can consult Burdett and Mortensen (1998) or Mortensen (2003).

The Burdett-Mortensen labor market model is a general equilibrium model that describes how firms and workers match to produce a single homogeneous good. In the Q-sector version of the model, there are Q types of firms which differ in terms of (exogenous) productivity, i.e., there is a vector of productivities denoted by $\{\pi_i\}_{i=1}^{Q}$ such that $0 < \pi_1 < \ldots < \pi_Q$. Firms are identical except in their productivity level. In the economy there is a unit mass of firms and, for every $i = 1, \ldots, Q$, there is a γ_i fraction of firms that are of productivity π_i. In order to produce the good, firms need to form a match with workers. This matching process is affected by frictions, reflecting the fact that it takes time and effort for unemployed workers and for vacant firms to discover each other and agree to produce.

There is a unit mass of workers. From the point of view of the worker, the dynamics are as follows. At each point in time, workers in this economy can be employed (matched with a firm) or unemployed (unmatched). At a Poisson rate λ_0, unemployed workers receive a job offer with a wage distributed according to an endogenous offer distribution. In equilibrium, unemployed workers will only receive offers that are higher than their reservation wage, denoted r, and will hence be immediately accepted. Employed workers receive two types of shocks. First, at a Poisson rate λ_1, they receive a new job offer, which they will only accept if it represents an improvement to the current wage rate. Second, at a Poisson rate δ, they receive a shock that destroys their current match and leaves them immediately unemployed.

The search in this model is not directed, i.e., workers search for firms without knowing their productivity. Upon meeting with a worker, the firm proposes a wage contract, which the worker can accept or reject. Firms propose wage offers to maximize the profits of production. This model endogenously produces a continuous distribution of equilibrium wages, which has a known closed form[25].

For our simulations, we randomly extract n workers from the population and follow their (equilibrium) wage paths over ten years[26]. We introduce missing data as follows: if an observation is affected by missing data, the wage path will be unobserved during the last five years. In this sense,

[25] For example, the formula for the density can be found in Mortensen (2003), Eq. (3.19).
[26] In other words, $\mathcal{I}$ is a time interval that is 10 years long.

the data are affected by attrition during the last 5 years of the sampling period[27]. In order to study how missing data affects the behavior of the test statistic, we perform simulations with different percentages of missing data and, as a benchmark, we include the case when the test has no missing data. Since the dataset is completely observed during the first 5 years, it is possible to use only these years to estimate the parameter root-n-consistently using maximum likelihood estimation.

We implement simulations with $n = 100$, $T = 530$ (530 weeks, i.e., 10 years), $m = 500$, and $V = 500$.[28] For fixed values of n, we implemented simulations with different values of m and V and obtained similar results, which leads us to believe that the performance is relatively insensitive to these parameters. We compute critical values using the bootstrap procedure with $S = 200$ simulations. Finally, following BHHN, we approximate the measure μ with μ_M for $M = 5$.[29] Simulations with other values of M produced qualitatively similar results.

5.1 Simulations under the null hypothesis

We begin by presenting the result of simulations under the null hypothesis. The parameter values for our simulations under the null are the following: $\lambda_0 = 0.03$, $\lambda_1 = 0.01$, $\delta = 0.0035$, $r = 100$, $Q = 2$ (i.e., two-sector model), $\pi_1 = 500$, $\pi_2 = 1000$, $\gamma_1 = \gamma_2 = 0.5$ (i.e., there is an equal proportion of low and high productive firms). The values for the parameters λ_0, λ_1, and δ are chosen in line with the findings of Bowlus, Kiefer, and Neumann (2001).

Table 1 describes the results of 1000 simulations. For each significance level, we compute the percentage of simulations for which the test results in rejection, lack of rejection, or inconclusive. Recall that with no missing data, our hypothesis test is identical to the BHHN test. Thus, we expect the test to be asymptotically size correct when there are no missing observations. The table reveals that this is indeed the case: the true and the nominal probability of rejection are extremely close. From Theorem 4.1, we expect our hypothesis test to be correct in level (but possibly conservative) even under the presence of missing data. Our simulations confirm the theoretical findings. In particular, as the percentage of missing data increases, both the probability of rejection and the probability of non-rejection decrease steadily, and, consequently, the probability of inconclusive results rises steadily.

5.2 Simulations under the alternative hypothesis

We consider two versions of an alternative hypothesis of the Burdett-Mortensen model. For our first alternative hypothesis, we consider a modification of the model that allows certain job to job

[27] In other words, we can implement the test by partitioning $\mathcal{I}$ into two sub-intervals, $\{\mathcal{I}_1, \mathcal{I}_2\}$, where $\mathcal{I}_1$ represents the first five years and $\mathcal{I}_2$ represents the last five years. According to this, then, $J = 2$.

[28] Notice that m and V are set relatively large with respect to n in the hopes of producing results close to the ones obtained in the asymptotic analysis in Section 4.2.

[29] As we have explained in Section 4, using a relatively small value of M should produce a test that is size correct but may lack power against certain alternatives. Nevertheless, even for these small values of M, our hypothesis test has excellent power against several relevant alternative hypotheses. Furthermore, the results of simulations with $M = 25$ confirm these conjectures.

Percentage of	$\alpha = 10\%$			$\alpha = 5\%$			$\alpha = 1\%$		
missing data	Rej.	No Rej.	Inc.	Rej.	No rej.	Inc.	Rej.	No rej.	Inc.
0%	9.5%	90.5%	0%	4.8%	95.2%	0%	0.7%	99.3%	0%
10%	3.3%	64.5%	33.2%	1.4%	81.3%	17.3%	0%	94.4%	5.6%
25%	0.8%	0%	99.2%	0.2%	2.0%	97.8%	0%	38.1%	61.9%
50%	0.1%	0%	99.9%	0%	0%	100%	0%	0%	100%

Table 1: Results of simulations under the null

transitions to result in wage decreases. To allow this, we alter the Burdett-Mortensen model so that employed agents will accept any new offer, regardless of the wage level. All of the features and parameters of the model remain the same as in the null hypothesis.

Table 2 presents the simulation results in this alternative hypothesis. The hypothesis test is able to strongly reject the null hypothesis when there are no missing observations. As it is expected, the percentage of rejection (inconclusive) results decreases (increases) as the percentage of missing data increases. Nevertheless, the true rejection rates are also significantly high (relative to the nominal level) even when there are significant amounts of missing observations. In this sense, our specification test is still very informative even under the presence of significant amounts of missing data.

Percentage of	$\alpha = 10\%$			$\alpha = 5\%$			$\alpha = 1\%$		
missing data	Rej.	No Rej.	Inc.	Rej.	No rej.	Inc.	Rej.	No rej.	Inc.
0%	92.6%	7.4%	0%	88.3%	11.7%	0%	73.0%	27.0%	0%
10%	84.5%	0.5%	15.0%	77.3%	2.8%	19.9%	55.9%	9.2%	34.9%
25%	68.3%	0%	31.7%	57.4%	0%	42.6%	36.4%	0.3%	63.3%
50%	42.1%	0%	57.9%	30.6%	0%	69.4%	14.8%	0%	85.2%

Table 2: Results of simulations under the first alternative hypothesis

In our second alternative hypothesis, we modify the model to introduce heterogeneity in the workforce. We assume that the sample is equally divided into two types of workers: stable and unstable. The types differ in their transition rates. Specifically, stable workers are characterized by the following parameters: $\lambda_0 = 0.015$, $\lambda_1 = 0.005$, $\delta = 0.00175$, whereas unstable workers are characterized by the following parameters: $\lambda_0 = 0.06$, $\lambda_1 = 0.02$, $\delta = 0.007$.[30] As a consequence, unstable workers will transition more often between jobs and between employment and unemployment than stable workers. We specify half of our workers to be stable and half to be unstable. All of the features and parameters of the model remain the same as in the null hypothesis.

Table 3 presents the simulation results in this alternative hypothesis. Without missing data, the null hypothesis is rejected with significant probability, although the frequency of rejection is not as high as in the first alternative hypothesis. As in the first alternative hypothesis, our specification

[30] These parameter values are chosen so that the average transition rates in the sample are those used for the null hypothesis.

test can still (correctly) reject the null hypothesis under the presence of significant amounts of missing data. Once again, our specification test is still very informative even under the presence of significant amounts of missing data.

Percentage of	$\alpha = 10\%$			$\alpha = 5\%$			$\alpha = 1\%$		
missing data	Rej.	No Rej.	Inc.	Rej.	No rej.	Inc.	Rej.	No rej.	Inc.
0%	68.1%	31.9%	0%	54.6%	45.4%	0%	30.1%	69.9%	0%
10%	57.2%	7.2%	35.6%	44.5%	15.8%	39.7%	22.5%	38.5%	39.0%
25%	46.7%	0%	53.3%	33.3%	0%	66.7%	16.6%	2.5%	80.9%
50%	33.3%	0%	66.7%	22.7%	0%	77.3%	8.4%	0%	91.6%

Table 3: Results of simulations under the second alternative hypothesis

6 Empirical Illustration

In this section, we use the test developed in this paper to test whether the observations of wage processes from the National Longitudinal Survey of Youth, 1979 (NLSY79) are distributed according to the Burdett-Mortensen model described in Section 5.

Our data are composed of young individuals (ages 17 to 22, in our sample), first interviewed in 1979, who are re-interviewed in subsequent years. In every interview year, each individual is asked about their job spells that occurred since the last interview. The first job spell reported in an interview corresponds to the main job spell (called the current/most recent job spell) but the interview process allows up to 5 job spells between interviews. For each job spell, the individual reports the beginning and the end of the job spell at a weekly precision, as well as its wage rate. With this information, we can construct the wage path for each individual from January 1st 1982[31] until December 31st 1991[32]. Given that the time information provided by the NLSY79 is sampled at weekly frequency, then, by construction, each of the wage paths in the data can be either missing or not at the same frequency[33]. We express all wages in terms of weekly remuneration and in terms of 1990 U.S. dollars using the Consumer Price Index[34].

The Burdett-Mortensen model assumes that workers in the economy are ex-ante homogeneous. Even though our sample contains very heterogeneous group of individuals, we restrict attention to a subsample with the same observable characteristics in the hopes of obtaining a homogeneous sample. Following Bowlus, Kiefer, and Neumann (2001), we restrict attention to white males that are High school or GED graduates and who are not in the military sample. This constitutes a

[31] Even though we have data since 1979, we avoid using the first three years of the sample since, during those years, some of the individuals of the sample were less than 20 years of age and their job market opportunities could be different from their older counterparts.

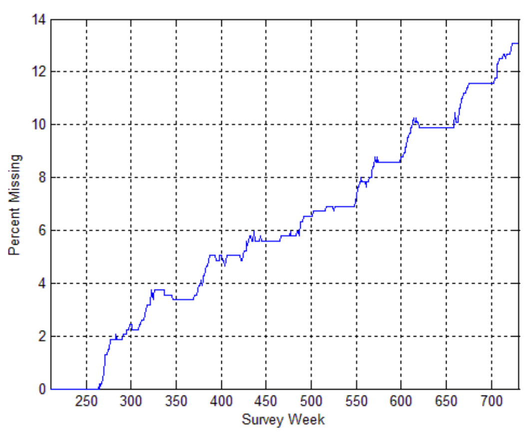
[32] In other words, $\mathcal{I}$ is a time interval starting on January 1st 1982 and ending on December 31st 1991.

[33] In other words, we implement the test by partitioning the interval $\mathcal{I}$ into its $J = 521$ weeks, i.e., we use $\{\mathcal{I}_j\}_{j=1}^{521}$, where, for $j = 1, \ldots, 521$, $\mathcal{I}_j$ is the j^{th} week in $\mathcal{I}$.

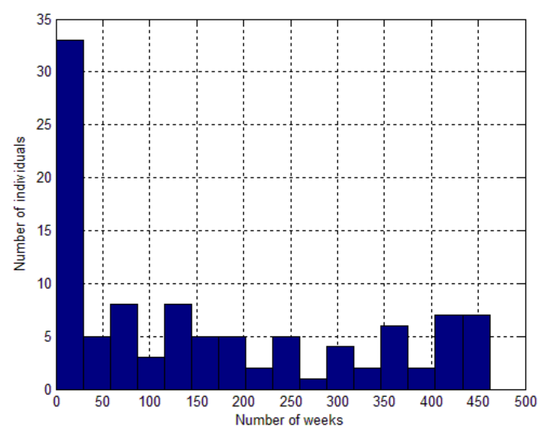
[34] This information is publicly available in the U.S. Bureau of Labor Statistics website.

sample of 816 individuals. We eliminate from the sample individuals who, at any point in the survey, presented problems in their duration data[35] or reported having weekly wages of over a thousand 1990 U.S. dollars[36]. This reduces our representative sample to 589 individuals. Finally, recall that our test requires the parameter of interest to be estimated in a root-n-consistent fashion. As we have explained in Section 3, one way to achieve this goal is to have a period of time where all of the observations of the sample are completely observed. To be able to illustrate our method, we eliminate all individuals that have an episode of missing data during 1982, which represents only 53 individuals or less than 9% of the sample. This produces the sample that will be used in our hypothesis test, composed of $n = 536$ individuals.

Out of our sample of 536 individuals, 433 individuals (80.7%) have no missing wage information and 103 individuals (19.3%) have at least one episode of missing wage information. Moreover, only 6.07% of all the weeks in the sample are missing. In turn, out of the 103 individuals with some missing data, 58 of them (56.3%) suffer from attrition from the sample, i.e., the information of an individual becomes unobserved at a certain point and remains unobserved for the rest of the sample. For the remaining 45 individuals, there are very few episodes that violate attrition. These figures indicate that sample attrition is a common explanation for missing observations in the NLSY79 survey.



(a) Percentage of individuals with missing data.



(b) Number of missing weeks conditional on missing data.

Figure 1: Description of the NLSY79 data

Figure 1a presents the evolution of the percentage of individuals with missing data. The percentage of missing data is (almost) weakly increasing in time. Again, this is indicative that most of the missing data is generated by sample attrition. For those individuals who have missing information, the average number of missing weeks is 166.4, which represents 31.6% of the weeks in

[35]These problems are either negative job spell duration or missing time information.

[36]This type of trimming is also utilized by Bowlus, Kiefer, and Neumann (2001).

our sample. Figure 1b presents the histogram of the number of missing weeks for this subset of individuals. Even though there are individuals with a large number of missing information, most of the individuals in the sample have relatively few missing observations.

We now describe the result of testing whether the Burdett-Mortensen model is the right specification for the wage processes in the NLSY79 survey. With a complete dataset for the first year, we use this year to estimate the parameter root-n-consistently using maximum likelihood estimation. Under the assumption that the missing data are missing at random, BHHN strongly reject the null hypothesis that the four sector Burdett-Mortensen model is the right specification for the data. Applying the techniques developed in this paper, we can conduct the same hypothesis test without requiring this assumption. We implement the specification test for a one, two, three, four, and five-sector Burdett-Mortensen model.

We implement the hypothesis test with $m = 1000$, $V = 1000$, with a bootstrap procedure with $S = 200$ simulations, and, following BHHN, we approximate the measure μ with μ_M for $M = 5$.[37] The results are presented in Table 4, which presents the estimated alternative WCSBs for the test statistic, as well as the 90$^{\text{th}}$, 95$^{\text{th}}$ and 99$^{\text{th}}$ quantiles of the statistic under the null hypothesis. Our specification test strongly rejects each of the specifications of the Burdett-Mortensen model[38]. In other words, the information contained in the sample with missing data is sufficient to reject the model without making ad-hoc assumptions about the nature of the missing observations.

Model	WCSBs		Quantiles under H_0		
	$n\hat{T}'_L(X, Y_{\hat{\theta}_0})$	$n\hat{T}'_H(X, Y_{\hat{\theta}_0})$	$t^*_{\hat{\theta}_0}(90\%)$	$t^*_{\hat{\theta}_0}(95\%)$	$t^*_{\hat{\theta}_0}(99\%)$
One sector	55.92	99.79	0.25	0.34	0.45
Two sectors	47.26	88.02	0.44	0.57	0.77
Three sectors	38.84	77.61	0.51	0.65	1.00
Four sectors	46.59	86.66	0.30	0.35	0.48
Five sectors	32.71	66.75	0.39	0.49	0.95

Table 4: Results of test on NLSY79 data.

7 Conclusion

This paper develops a specification test for functional data that allows for the presence of missing observations. In order to deal with the missing data problem, we adopt a worst case scenario approach which is agnostic about the distribution of the missing data. The specification test adapts the Cramér-von Mises specification test developed in BHHN to the presence of missing data. In order to develop the specification test, we study the identification problem caused by

[37] We repeated the hypothesis test for different values of m, V, and M. Each repetition of the test produces the same qualitative result.

[38] The table presents the outcome of one run of the hypothesis test. Since the outcome of the test is dependent on certain random draws, the test was repeated several times. All of these repetitions resulted in the rejection of the null hypothesis.

missing observations. We show how missing data implies that the distribution of the functional data is partially identified and we derive worst case scenario bounds for the distribution of the Cramér-von Mises statistic proposed by BHHN.

In the presence of missing functional data, one might consider conducting a model specification test by applying the BHHN test with the subset of the sample that presents no missing sections. Relative to this procedure, our hypothesis test presents two important advantages. First, our analysis avoids making untestable assumptions about the nature of the missing data and, consequently, our hypothesis test produces conclusions that are valid regardless of the (unobserved) features of the distribution of the missing data. Second, in the presence of functional data where functions in the data have missing sections and non-missing sections (i.e. functions are partly unobserved), our hypothesis testing procedure is able to extract all of the information contained in the non-missing sections while still being agnostic about the nature of the missing sections.

Our specification test can have three outcomes: rejection of the null hypothesis, lack of rejection of the null hypothesis, or inconclusive. The possibility of an inconclusive result is an undesired but unavoidable consequence of the existence of missing data and our unwillingness to impose assumptions regarding its distribution.

The theoretical properties of our specification test depend not only on whether the null hypothesis is true or false, but also on whether this can be identified from the distribution of observed data. Under the null hypothesis, our specification test will reject the null hypothesis with a probability that, in the limit, does not exceed the significance level of the test. Under the alternative hypothesis, the behavior of the test depends critically on whether this can be learned from the distribution of the observed data. Whenever the distribution of the observed data contains enough information to identify that the null hypothesis is false, our hypothesis test is consistent (i.e. the power of our test converges to one). Monte Carlo evidence reveals that these theoretical results hold in finite samples.

As an empirical illustration, we test whether observations of the wage process in the NLSY79 are distributed according to the Burdett-Mortensen labor market model. In the 1982 - 1991 period, 19.3% of the individuals in the survey are affected by some form of missing data, typically caused by sample attrition. Even under the presence of missing data, our specification test strongly rejects that the Burdett-Mortensen model is the correct framework for the NLSY79 data. This illustration constitutes an ideal application of our specification test, since it delivers informative results even though we adopt a worst case scenario approach about the nature of the missing observations.

References

Andrews, D. (1997): "A Conditional Kolmogorov Test," *Econometrica*, 65(5), 1097–1128.

Andrews, D., and G. Soares (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78(1), 119–157.

Ash, R., and M. Gardner (1975): *Topics in Stochastic Processes*. Academic Press.

Beresteanu, A., and F. Molinari (2008): "Asymptotic Properties for a Class of Partially Identified Models," *Econometrica*, 76(4), 763–814.

Bierens, H. J. (2009): "Consistent Model Specification Tests," Mimeo: Pennsylvania State University.

Bierens, H. J., and L. Wang (2010): "Integrated Conditional Moment Test for Parametric Conditional Distribution," Mimeo: Pennsylvania State University.

Billingsley, P. (1995): *Probability and Measure*. John Wiley and Sons, Inc.

Bowlus, A., N. Kiefer, and G. Neumann (2001): "Equilibrium Search Models and the Transition from School to Work," *International Economic Review*, 42(2), 317–43.

Bugni, F. (2010): "Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set," *Econometrica*, 78(2), 735–753.

Bugni, F., P. Hall, J. Horowitz, and G. Neumann (2009): "Goodness-of-fit test for functional data," *The Econometrics Journal*, 12(s1), S1–S18.

Burdett, K., and D. Mortensen (1998): "Wage Differentials, Employer Size, and Unemployment," *International Economic Review*, 39(2), 257–273.

Canay, I. (2010): "E. L. Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity," *Journal of Econometrics*, 156(2), 408–425.

Chernozhukov, V., H. Hong, and E. Tamer (2007): "Parameter Set Inference in a Class of Econometric Models," *Econometrica*, 75(5), 1243–1284.

Cuesta-Albertos, J., E. del Barrio, R. Fraiman, and C. Matrán (2007): "The random projection method in goodness of fit for functional data," *Computational Statistics and Data Analysis*, 51(10), 4814–4831.

Cuesta-Albertos, J., R. Fraiman, and T. Ransford (2006): "Random projections and goodness-of-fit tests in infinite dimensional spaces," *Bulletin of the Brazilian Mathematical Society*, 37(5), 1–25.

Doob, J. L. (1990): *Stochastic Processes*. Wiley Classics Library.

Gihman, I. I., and A. V. Skorohod (1974): *The Theory of Stochastic Processes I*. Springer-Verlag.

Heckman, J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47(1), 153–61.

Ito, K. (2006): *Essentials of Stochastic Processes*. Translations of Mathematical Monographs, Volume 231.

Manski, C. (1989): "Anatomy of the Selection Problem," *The Journal of Human Resources*, 24(3), 343–360.

——— (2003): *Partial Identification of Probability Distributions*. Springer-Verlag.

Mortensen, D. (2003): *Wage Dispersion: Why are similar workers paid differently?* MIT press.

Rao, M. M. (1995): *Stochastic Processes: General Theory*. Kluwer Academic Publishers.

Romano, J. P., and A. M. Shaikh (2008): "Inference for identifiable Parameters in Partially Identified Econometric Models," *Journal of Statistical Planning and Inference*, 138, 2786–2807.

Rosen, A. (2006): "Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities," Mimeo: University College London.

Royden, H. L. (1988): *Real Analysis*. Prentice-Hall.

Stoye, J. (2009): "More on Confidence Intervals for Partially Identified Parameters," *Econometrica*, 77(4), 299–1315.

Zheng, J. X. (2000): "A Consistent Test of Conditional Parametric Distributions," *Econometric Theory*, 16, 667–691.

A Appendix

A.1 Notation and definitions

- For $x_1, x_2 \in \mathbb{R}^K$, $\|x_1 - x_2\| \equiv \sqrt{(x_1 - x_2)'(x_1 - x_2)}$ and for $G_1, G_2 \in L_2(\mathcal{I})$, $\|G_1 - G_2\|_\mu \equiv \int (G_1(x) - G_2(x))^2 \, d\mu(x)$.
- Let $S \subseteq \Omega$, where $(\Omega, \mathcal{B}, P)$is a complete probability space under consideration. The outer P-measure applied to S is defined as $P^*(S) \equiv \inf_{B \in \mathcal{B}, S \subseteq B} P(B)$, and the inner P-measure applied to S is defined as $P_*(S) \equiv \sup_{C \in \mathcal{B}, C \subseteq S} P(C)$. By arguments in Section 12.6 (Proposition 34) in Royden (1988): $S \in \mathcal{B}$ if and only if $P^*(S) = P_*(S) = P(S)$.
- In order to be able to prove measurability of several events associated to stochastic processes, we assume that the stochastic process of interest is *separable*. This concept was introduced by Doob (1990) (Chapter 2, Section 2) and has been adopted by several references in the literature of stochastic processes, e.g., Ito (2006) (Definition 2.8.1) or Ash and Gardner (1975) (Definition 4.1.2). For we now provide the definition.

Definition A.1 (Separable stochastic process)**.** A stochastic process $\{Y(\omega,t) : t \in \mathcal{I}\} : \Omega \to \mathbb{R}^{\mathcal{I}}$ is *separable* if and only if there exists a countable and dense subset of $\mathcal{I}$, denoted by S, such that:

$$P\left(\forall t \in \mathcal{I} : \left\{ \begin{array}{c} \exists \{s_m\}_{m=1}^{\infty} : s_m \in S,\ s_m \to t,\ \text{such that,} \\ \left\{ \liminf_{m\to\infty} Y(\omega, s_m) \leq Y(\omega,t) \leq \limsup_{m\to\infty} Y(\omega,s_m) \right\} \end{array} \right\}\right) = 1. \tag{A.1}$$

The set S in the definition of a separable stochastic process is referred to as the *separant set.*

A.2 Auxiliary results

Lemma A.1. *Let $Y : \Omega \to \mathbb{R}^{\mathcal{I}}$ be a separable stochastic process in $(\Omega, \mathcal{B}, P)$ with separant set S and let $x \in \mathbb{R}^{\mathcal{I}}$ be a continuous function on a set $M \subseteq \mathcal{I}$. Then:*

$$P\left(\forall t \in M : \left\{ \begin{array}{c} \exists \{s_m\}_{m=1}^{\infty} : s_m \in S,\ s_m \to t,\ \textit{such that,} \\ \left\{ \liminf_{m\to\infty} (Y(\omega, s_m) - x(s_m)) \leq Y(\omega,t) - x(t) \leq \limsup_{m\to\infty} (Y(\omega,s_m) - x(s_m)) \right\} \end{array} \right\}\right) = 1. \tag{A.2}$$

From this, it follows also that:

$$\begin{array}{l} P\left(\forall t \in M : \left\{ \begin{array}{c} \forall \{s_m\}_{m=1}^{\infty} : s_m \in S,\ s_m \to t,\ \textit{such that,} \\ \left\{ \liminf_{m\to\infty} (Y(\omega, s_m) - x(s_m)) > Y(\omega,t) - x(t) \right\} \end{array} \right\}\right) = 0, \\ P\left(\forall t \in M : \left\{ \begin{array}{c} \forall \{s_m\}_{m=1}^{\infty} : s_m \in S,\ s_m \to t,\ \textit{such that,} \\ \left\{ Y(\omega,t) - x(t) > \limsup_{m\to\infty} (Y(\omega, s_m) - x(s_m)) \right\} \end{array} \right\}\right) = 0. \end{array} \tag{A.3}$$

Proof. Notice that both statements in Eq. (A.3) follow immediately from Eq. (A.2). As a consequence, we only show Eq. (A.2). For arbitrary $t \in M$ and $\{s_m\}_{m=1}^{\infty}$, such that $s_m \in S$, $s_m \to t$, it follows that:

$$\begin{aligned} & \left\{ \liminf_{m\to\infty} Y(\omega, s_m) \leq Y(\omega,t) \leq \limsup_{m\to\infty} Y(\omega,s_m) \right\} \\ \subseteq & \left\{ \begin{array}{c} \left\{ \liminf_{m\to\infty} (Y(\omega, s_m) - x(s_m)) = \liminf_{m\to\infty} Y(\omega,s_m) - \limsup_{m\to\infty} x(s_m) \leq Y(\omega,t) - x(t) \right\} \\ \cap \left\{ Y(\omega,t) - x(t) \leq \limsup_{m\to\infty} Y(\omega,s_m) - \liminf_{m\to\infty} x(s_m) = \limsup_{m\to\infty} (Y(\omega,s_m) - x(s_m)) \right\} \end{array} \right\} \\ \subseteq & \left\{ \liminf_{m\to\infty} \{Y(\omega,s_m) - x(s_m)\} \leq Y(\omega,t) - x(t) \leq \limsup_{m\to\infty} \{Y(\omega,s_m) - x(s_m)\} \right\}, \end{aligned}$$

where we have used the continuity of x on M and elementary properties of the lim inf and lim sup operators. From this and the fact that Y is a separable stochastic process in $(\Omega, \mathcal{B}, P)$ with separant set S, Eq. (A.2) follows. □

Lemma A.2. *Let $Y : \Omega \to \mathbb{R}^{\mathcal{I}}$ be a separable stochastic process in $(\Omega, \mathcal{B}, P)$ with separant set S and let $\mathcal{J} \subseteq \mathcal{I}$ be such that $\mathcal{J}/int(\mathcal{J})$ is countable. Then, the stochastic process $Y' : \Omega \to \mathbb{R}^{\mathcal{J}}$ defined by $Y'(\omega,t) = Y(\omega,t)$ for all $t \in \mathcal{J}$ is a separable stochastic process with separant set $S' = \{S \cap int(\mathcal{J})\} \cup \{\mathcal{J}/int(\mathcal{J})\}$.*

Proof. The proof is completed by verifying the requirements in Definition A.1. We divide the proof into steps.

Step 1: Verify that S' satisfies several properties.

- S' is countable. This is because $S' \subseteq S \cup \{\mathcal{J}/int(\mathcal{J})\}$, both of which are countable.
- $S' \subseteq \mathcal{J}$. This follows from $\{S \cap int(\mathcal{J})\} \subseteq int(\mathcal{J}) \subseteq \mathcal{J}$ and $\{\mathcal{J}/int(\mathcal{J})\} \subseteq \mathcal{J}$.
- S' is dense in $\mathcal{J}$. This follows from the fact that S is dense in $\mathcal{I} \supseteq \mathcal{J}$.

Step 2: Show that $\forall t \in \mathcal{J}$:

$$\left\{ \begin{array}{c} \exists \{s_m\}_{m=1}^{\infty} : s_m \in S,\ s_m \to t : \\ \left\{ \liminf_{m\to\infty} Y(\omega, s_m) \leq Y(\omega,t) \leq \limsup_{m\to\infty} Y(\omega, s_m) \right\} \end{array} \right\} \subseteq \left\{ \begin{array}{c} \exists \{s'_m\}_{m=1}^{\infty} : s'_m \in S',\ s'_m \to t : \\ \left\{ \liminf_{m\to\infty} Y'(\omega, s'_m) \leq Y'(\omega,t) \leq \limsup_{m\to\infty} Y'(\omega, s'_m) \right\} \end{array} \right\}.$$

Suppose that the event on the left hand side (LHS) occurs. Then, there are only two cases. In the first case: $t \in \{\mathcal{J}/int(\mathcal{J})\}$. In this case, we set $s'_m = t\ \forall m \in \mathbb{N}$ and then, $s'_m \in S'$ and $Y(\omega, s'_m) = Y(\omega, t)$, $\forall m \in \mathbb{N}$ and so, $\lim_{m\to\infty} Y(\omega, s'_m) = Y(\omega,t)$. Furthermore, since $Y'(\omega, s) = Y(\omega, s)\ \forall s \in \mathcal{J}$, it follows that $\lim_{m\to\infty} Y'(\omega, s'_m) = Y'(\omega, t)$, i.e., the event on the RHS occurs. In the second case: $t \in int(\mathcal{J})$, i.e., $\exists \delta > 0$ such that $\{t' : \|t' - t\| < \delta\} \subseteq \mathcal{J}$. But then, $\exists \delta' = \delta/2 > 0$ such that $\{t' : \|t' - t\| < \delta'\} \subseteq int(\mathcal{J})$. Since $s_m \in S$ and $s_m \to t$, $\exists M$ such that $s_m \in \{S \cap int(\mathcal{J})\}\ \forall m \geq M$. Then, define $\{s'_m\}_{m=1}^{\infty}$ with $s'_m = s_{m+M}$. By construction, then $\{s'_m\}_{m=1}^{\infty} : s'_m \in S'$, $s'_m \to t$, and

$$\liminf_{m\to\infty} Y(\omega, s_m) = \liminf_{m\to\infty} Y(\omega, s'_m) \leq Y(\omega, t) \leq \limsup_{m\to\infty} Y(\omega, s'_m) = \limsup_{m\to\infty} Y(\omega, s_m).$$

Furthermore, since $Y'(\omega, s) = Y(\omega, s)\ \forall s \in \mathcal{J}$, the event on the right hand side (RHS) occurs, as:

$$\liminf_{m\to\infty} Y(\omega, s'_m) = \liminf_{m\to\infty} Y'(\omega, s'_m) \leq Y(\omega, t) = Y'(\omega, t) \leq \limsup_{m\to\infty} Y'(\omega, s'_m) = \limsup_{m\to\infty} Y(\omega, s'_m).$$

Step 3: Conclude the proof. From step 1, we know that S' is a countable dense subset of $\mathcal{J}$. From this, step 2, $\mathcal{J} \subseteq \mathcal{I}$, and that $Y : \Omega \to \mathbb{R}^{\mathcal{I}}$ is a separable stochastic process with separant set S, it is straightforward to conclude that $Y' : \Omega \to \mathbb{R}^{\mathcal{J}}$ is a separable stochastic process with separant set S'. □

Lemma A.3. *Let $Y : \Omega \to \mathbb{R}^{\mathcal{J}}$ be a separable stochastic process in $(\Omega, \mathcal{B}, P)$ with separant set S, and let $x \in C'(\mathcal{J})$. Then, there exists a countable dense subset of $\mathcal{J}$, denoted $D(x)$, such that:*

$$P^*(Y(\omega,t) \leq x(t), \forall t \in \mathcal{J}) = P_*(Y(\omega,t) \leq x(t), \forall t \in \mathcal{J}) = P(Y(\omega,t) \leq x(t), \forall t \in D(x)).$$

As a consequence, $\{Y(\omega,t) \leq x(t), \forall t \in \mathcal{J}\} \in \mathcal{B}$.

Proof. Consider arbitrary $x \in C'(\mathcal{J})$. Denote by S' the countable subset of $\mathcal{J}$ where x is discontinuous and define $D(x) \equiv S \cup S'$. It is easy to verify that $D(x)$ is a countable and dense subset of $\mathcal{J}$ and that Y is a separable stochastic process in $(\Omega, \mathcal{B}, P)$ with separant set $D(x)$. The remainder of the proof is divided into steps.

Step 1: Show that if A is countable, then $\{Y(\omega,t) \leq x(t), \forall t \in A\}$ is measurable. Fix $t \in A$ arbitrarily. Since $Y(\omega, \cdot) : \Omega \to \mathbb{R}^{\mathcal{I}}$ is a stochastic process in $(\Omega, \mathcal{B}, P)$, then for every $t \in A$, $Y(\omega,t) - x(t) : \Omega \to \mathbb{R}$ is a random variable and, so, $\{Y(\omega,t) - x(t) \leq 0\} \in \mathcal{B}$. By repeating this argument $\forall t \in A$, it follows that $\cap_{t\in A} \{Y(\omega,t) - x(t) \leq 0\} = \{Y(\omega,t) \leq x(t), \forall t \in A\} \in \mathcal{B}$.

Step 2: Show that $P^*(Y(\omega,t)\le x(t), \forall t\in\mathcal{J}) = P(Y(\omega,t)\le x(t), \forall t\in D(x))$. Since $D(x)\subseteq\mathcal{J}$, it follows that $\{Y(\omega,t)\le x(t), \forall t\in\mathcal{J}\}\subseteq\{Y(\omega,t)\le x(t), \forall t\in D(x)\}$ then the monotonicity of the outer probability implies that $P^*(Y(\omega,t)\le x(t), \forall t\in\mathcal{J}) \le P^*(Y(\omega,t)\le x(t), \forall t\in D(x))$. Since $D(x)$ is countable, step 1 implies that $\{Y(\omega,t)\le x(t), \forall t\in D(x)\}$ is measurable, i.e., $P(Y(\omega,t)\le x(t), \forall t\in D(x)) = P^*(Y(\omega,t)\le x(t), \forall t\in D(x))$. Thus, $P^*(Y(\omega,t)\le x(t), \forall t\in\mathcal{J}) \le P(Y(\omega,t)\le x(t), \forall t\in D(x))$.

The proof of the step is completed if we show that $P(Y(\omega,t)\le x(t), \forall t\in D(x)) \le P^*(Y(\omega,t)\le x(t), \forall t\in\mathcal{J})$. To show this, consider the following argument:

$$\begin{aligned}
&\{Y(\omega,t)\le x(t), \forall t\in D(x)\}\\
&=\left\{\begin{array}{c}\{Y(\omega,t)\le x(t), \forall t\in\mathcal{J}\}\cup\\ \{\{Y(\omega,t)\le x(t), \forall t\in D(x)\}\cap\{\exists t'\in\mathcal{J}/D(x): Y(\omega,t')>x(t')\}\}\end{array}\right\}\\
&\subseteq\left\{\begin{array}{c}\{Y(\omega,t)\le x(t), \forall t\in\mathcal{J}\}\cup\\ \left\{\exists t'\in\mathcal{J}/D(x):\left\{\begin{array}{c}\forall\{s_m\}_{m=1}^{\infty}: s_m\in D(x),\ s_m\to t',\text{ such that,}\\ \left\{\{Y(\omega,t')-x(t')\}>\limsup_{m\to\infty}\{Y(\omega,s_m)-x(s_m)\}\right\}\end{array}\right\}\right\}\end{array}\right\},
\end{aligned}$$

where we used that $D(x)$ is dense in $\mathcal{J}$ to show that $\{\{Y(\omega,t)\le x(t), \forall t\in D(x)\}\cap\{\exists t'\in\mathcal{J}/D(x): Y(\omega,t')>x(t')\}\}$ implies that there is a sequence $\{t_s\}_{s=1}^{\infty}$ with $t_s\in D(x)\ \forall s\in\mathbb{N}$, and $t_s\to t'$, such that $Y'(\omega,t')-x(t')>0\ge Y'(\omega,t_s)-x(t_s)$ and, thus: $\{Y(\omega,t')-x(t')\}>\limsup_{m\to\infty}\{Y(\omega,s_m)-x(s_m)\}$.

From this, we can conduct the following derivation:

$$\begin{aligned}
&P(Y(\omega,t)\le x(t), \forall t\in D(x)) = P^*(Y(\omega,t)\le x(t), \forall t\in D(x))\\
&\le\left\{\begin{array}{c}P^*(Y(\omega,t)\le x(t), \forall t\in\mathcal{J})+\\ P^*\left(\exists t'\in\mathcal{J}/D(x):\left\{\begin{array}{c}\forall\{s_m\}_{m=1}^{\infty}: s_m\in D(x),\ s_m\to t',\text{ such that,}\\ \left\{\{Y(\omega,t')-x(t')\}>\limsup_{m\to\infty}\{Y(\omega,s_m)-x(s_m)\}\right\}\end{array}\right\}\right)\end{array}\right\}\\
&\le P^*(Y(\omega,t)\le x(t), \forall t\in\mathcal{J}),
\end{aligned}$$

where the first equality follows from step 1, the second inequality follows from the previous argument and the countable subadditivity of the outer probability, and the final inequality follows from the fact that Y is separable on $D(x)$ and Lemma A.1.

Step 3: Show that $P_*(Y(\omega,t)\le x(t), \forall t\in\mathcal{J}) = P(Y(\omega,t)\le x(t), \forall t\in D(x))$. By the same argument used in step 2: $P_*(Y(\omega,t)\le x(t), \forall t\in\mathcal{J}) \le P(Y(\omega,t)\le x(t), \forall t\in D(x))$. The proof of the step is completed if we show that $P_*(Y(\omega,t)\le x(t), \forall t\in\mathcal{J}) \ge P(Y(\omega,t)\le x(t), \forall t\in D(x))$. To show this, consider the following argument:

$$\begin{aligned}
&\{\exists t'\in\mathcal{J}: Y(\omega,t')>x(t')\}\\
&=\{\exists t'\in D(x): Y(\omega,t')>x(t')\}\cup\{\exists t'\in\mathcal{J}/D(x): Y(\omega,t')>x(t')\}\\
&=\left\{\begin{array}{c}\{\exists t'\in D(x): Y(\omega,t')>x(t')\}\cup\\ \{\{Y(\omega,t)\le x(t), \forall t\in D(x)\}\cap\{\exists t'\in\mathcal{J}/D(x): Y(\omega,t')>x(t')\}\}\end{array}\right\}\\
&\subseteq\left\{\begin{array}{c}\{\exists t'\in D(x): Y(\omega,t')>x(t')\}\cup\\ \left\{\exists t'\in\mathcal{J}/D(x):\left\{\begin{array}{c}\forall\{s_m\}_{m=1}^{\infty}: s_m\in D(x),\ s_m\to t',\text{ such that,}\\ \left\{\{Y(\omega,t')-x(t')\}>\limsup_{m\to\infty}\{Y(\omega,s_m)-x(s_m)\}\right\}\end{array}\right\}\right\}\end{array}\right\},
\end{aligned}$$

and from this and arguments in step 2: $P^*(\exists t' \in \mathcal{J} : Y(\omega, t') > x(t')) \leq P^*(\exists t' \in D(x) : Y(\omega, t') > x(t'))$. From here we deduce that:

$$P_*(Y(\omega, t) \leq x(t), \forall t \in \mathcal{J}) \geq P_*(Y(\omega, t) \leq x(t), \forall t \in D(x)) = P(Y(\omega, t) \leq x(t), \forall t \in D(x)),$$

where we have used the basic relationship between inner and outer probability, step 1, and the fact that $D(x)$ is countable. □

A.3 Proofs of Section 3

Proof of Lemma 3.1. Fix $x \in C'(\mathcal{I})$ and $j = 1, \ldots, 2^J$ arbitrarily. Recall that S_j denotes the subset of $\mathcal{I}$ where functions are observed in the j^{th} missing data pattern. By Assumption 1.a, X is a separable stochastic process in $(\Omega, \mathcal{B}, P)$ with a separant set that we denote by S and by Assumption 1.c, S_j is the union of finite intervals of $\mathcal{I}$ and, consequently, $S_j/int(S_j)$ and $(\mathcal{I}/S_j)/int(\mathcal{I}/S_j)$ are both countable subsets of $\mathcal{I}$. Then, by Lemma A.2, $X_{j,1} : \Omega \to \mathbb{R}^{S_j}$ and $X_{j,2} : \Omega \to \mathbb{R}^{\mathcal{I}/S_j}$ defined by $X_{j,1}(t) = X(t)$ $\forall t \in S_j$ and $X_{j,2}(t) = X(t)$ $\forall t \in \mathcal{I}/S_j$ are separable stochastic processes with separant set given by $S_{j,1} = \{S \cap int(S_j)\} \cup \{S_j/int(S_j)\}$ and $S_{j,2} = \{S \cap int(\mathcal{I}/S_j)\} \cup \{(\mathcal{I}/S_j)/int(\mathcal{I}/S_j)\}$, respectively. Then, by Lemma A.3, $\{X(t) \leq x(t), \forall t \in S_j\} \in \mathcal{B}$ and $\{X(t) \leq x(t), \forall t \in \mathcal{I}/S_j\} \in \mathcal{B}$. Since $\{O = S_j\} \in \mathcal{B}$, it follows that $\{\{X(t) \leq x(t), \forall t \in S_j\} \cap \{O = S_j\}\} \in \mathcal{B}$, $\{\{X(t) \leq x(t), \forall t \in \mathcal{I}/S_j\} \cap \{O = S_j\}\} \in \mathcal{B}$, and $\{\{X(t) \leq x(t), \forall t \in \mathcal{I}\} \cap \{O = S_j\}\} \in \mathcal{B}$. These results allow us to make probability statements over these events.

By Law of Total Probability:

$$\begin{aligned}
F_X(x) &= \sum_{j=1}^{2^J} P(\{X(t) \leq x(t), \forall t \in \mathcal{I}\} \cap \{O = S_j\}) \\
&= \left\{ \begin{array}{c} P(\{X(t) \leq x(t), \forall t \in \mathcal{I}\} \cap \{O = \mathcal{I}\}) + \\ \sum_{j=2}^{2^J} P(\{X(t) \leq x(t), \forall t \in S_j\} \cap \{X(t) \leq x(t), \forall t \in \mathcal{I}/S_j\} \cap \{O = S_j\}) \end{array} \right\} \\
&= \left\{ \begin{array}{c} P(\{X'(t) \leq x(t), \forall t \in \mathcal{I}\} \cap \{O = \mathcal{I}\}) + \\ \sum_{j=2}^{2^J} P(\{X'(t) \leq x(t), \forall t \in S_j\} \cap \{X(t) \leq x(t), \forall t \in \mathcal{I}/S_j\} \cap \{O = S_j\}) \end{array} \right\},
\end{aligned}$$

where we have used the convention that $j = 1$ implies $S_j = \mathcal{I}$ and $\mathcal{I}/S_j = \varnothing$ and that $\forall j = 1, \ldots, 2^J$, $X' : \Omega \to \mathbb{R}^{S_j}$ is defined so that $X'(t) = X(t)$ $\forall t \in S_j$. This verifies Eq. (3.1).

For the rest of the proof, choose $j = 2, \ldots, 2^J$ arbitrarily and consider two mutually exclusive and exhaustive cases. In the first case: $P(\{X'(t) \leq x(t), \forall t \in S_j\} \cap \{O = S_j\}) = 0$. Then, the monotonicity of the probability measure implies that $P(\{X'(t) \leq x(t), \forall t \in S_j\} \cap \{X(t) \leq x(t), \forall t \in \mathcal{I}/S_j\} \cap \{O = S_j\}) = 0$. In the second case: $P(\{X'(t) \leq x(t), \forall t \in S_j\} \cap \{O = S_j\}) > 0$. Then, Bayes' rule implies that:

$$\begin{array}{c} P(\{X'(t) \leq x(t), \forall t \in S_j\} \cap \{X(t) \leq x(t), \forall t \in \mathcal{I}/S_j\} \cap \{O = S_j\}) = \\ \left\{ \begin{array}{c} P(\{X(t) \leq x(t), \forall t \in \mathcal{I}/S_j\} \mid \{X'(t) \leq x(t), \forall t \in S_j\} \cap \{O = S_j\}) \\ \times P(\{X'(t) \leq x(t), \forall t \in S_j\} \cap \{O = S_j\}) \end{array} \right\}. \end{array}$$

This completes the proof. □

Proof of Lemma 3.2. The proof is divided into steps. In the first step, we concentrate on the case when

$x \in \mathbb{R}^{\mathcal{I}}$ and show Eq. (3.2). In the second step, we specialize the case to $x \in C'(\mathcal{I})$ and show that the bounds are sharp.

Step 1:Show Eq. (3.2). Fix $x \in \mathbb{R}^{\mathcal{I}}$ arbitrarily. We first show that $F_X(x) \leq F_X^H(x)$. Consider the following derivation:

$$\begin{aligned}
&F_X(x) \equiv P^*(\{X(t) \leq x(t), \forall t \in \mathcal{I}\}) \\
&\leq \textstyle\sum_{j=1}^{2^J} P^*(\{X(t) \leq x(t), \forall t \in \mathcal{I}\} \cap \{O = S_j\}) \\
&\leq \textstyle\sum_{j=1}^{2^J} P^*(\{X(t) \leq x(t), \forall t \in S_j\} \cap \{O = S_j\}) \equiv F_X^H(x),
\end{aligned}$$

where we have used the countable subadditivity and the monotonicity of the outer probability. The fact that $F_X(x) \geq F_X^L(x)$ follows from $\{\{X(t) \leq x(t), \forall t \in \mathcal{I}\} \cap \{O = S_1\}\} \subseteq \{X(t) \leq x(t), \forall t \in \mathcal{I}\}$ and the monotonicity of the outer probability.

Step 2: Fix $x \in C'(\mathcal{I})$ arbitrarily and show that the bounds in Eq. (3.2) are sharp. Fix $j = 1, \ldots, 2^J$ arbitrarily and let $X' : \Omega \to \mathbb{R}^{S_j}$ be defined so that $X'(t) = X(t)\ \forall t \in S_j$. By repeating the argument of Lemma 3.1, it follows that $\{X'(t) \leq x(t), \forall t \in S_j\} \in \mathcal{B}$, $\{X(t) \leq x(t), \forall t \in \mathcal{I}/S_j\} \in \mathcal{B}$, and $\{O = S_j\} \in \mathcal{B}$. These results allow us to make probability statements over these events.

We now show that:

$$\begin{aligned}
&\mathcal{H}(P(\{X'(t) \leq x(t), \forall t \in S_j\} \cap \{X(t) \leq x(t), \forall t \in \mathcal{I}/S_j\} \cap \{O = S_j\})) \\
&= [0, P(\{X'(t) \leq x(t), \forall t \in S_j\} \cap \{O = S_j\})].
\end{aligned} \tag{A.4}$$

There are two possible cases. In the first case: $P(\{X'(t) \leq x(t), \forall t \in S_j\} \cap \{O = S_j\}) = 0$. In this case, elementary arguments imply that Eq. (A.4) holds. In the second case: $P(\{X'(t) \leq x(t), \forall t \in S_j\} \cap \{O = S_j\}) > 0$. By Bayes' rule:

$$\begin{aligned}
&P(\{X'(t) \leq x(t), \forall t \in S_j\} \cap \{X(t) \leq x(t), \forall t \in \mathcal{I}/S_j\} \cap \{O = S_j\}) = \\
&\left\{ \begin{array}{c} P(\{X(t) \leq x(t), \forall t \in \mathcal{I}/S_j\} \mid \{X'(t) \leq x(t), \forall t \in S_j\} \cap \{O = S_j\}) \\ \times P(\{X'(t) \leq x(t), \forall t \in S_j\} \cap \{O = S_j\}) \end{array} \right\}.
\end{aligned}$$

The expression on the RHS is a product of two terms: the first one is not identified and only known to be a number in $[0, 1]$ and the second one is identified. By imposing the logical bounds on the unidentified term, Eq. (A.4) holds.

We now show that:

$$\begin{aligned}
&\mathcal{H}\left(\{P(\{X'(t) \leq x(t), \forall t \in S_j\} \cap \{X(t) \leq x(t), \forall t \in \mathcal{I}/S_j\} \cap \{O = S_j\})\}_{j=2}^{2^J}\right) \\
&= \Pi_{j=2}^{2^J} [0, P(\{X'(t) \leq x(t), \forall t \in S_j\} \cap \{O = S_j\})].
\end{aligned} \tag{A.5}$$

For each coordinate $j = 2, \ldots, 2^J$ such that $P(\{X'(t) \leq x(t), \forall t \in S_j\} \cap \{O = S_j\}) > 0$, the corresponding coordinate of the vector can be decomposed into the product of a point identified probability and non-identified term given by $P(\{X(t) \leq x(t), \forall t \in \mathcal{I}/S_j\} \mid \{X'(t) \leq x(t), \forall t \in S_j\} \cap \{O = S_j\})$. Thus, the vector is not identified due to the non-identification of the vector:

$$\{P(\{X(t) \leq x(t), \forall t \in \mathcal{I}/S_j\} | \{X'(t) \leq x(t), \forall t \in S_j\} \cap \{O = S_j\})\}_{j=2}^{2^J}. \tag{A.6}$$

Moreover, $j_1 \neq j_2$ implies that the coordinates of the non-identified vector in Eq. (A.6) do not impose any restrictions on each other. As a consequence, the identified set for the non-identified vector in Eq. (A.6) is

$\{[0,1]\}_{j=2}^{2^J}$ and, from this, Eq. (A.5) follows.

To conclude, combining Eqs. (3.1) and (A.5) yields:

$$\mathcal{H}\left(F_X\left(x\right)\right)=\left[\begin{array}{l}P\left(\left\{X'\left(t\right)\leq x\left(t\right),\forall t\in\mathcal{I}\right\}\cap\left\{O=\mathcal{I}\right\}\right),\\ \sum_{j=1}^{2^J}P\left(\left\{X'\left(t\right)\leq x\left(t\right),\forall t\in S_j\right\}\cap\left\{O=S_j\right\}\right)\end{array}\right]=\left[F_X^L\left(x\right),F_X^H\left(x\right)\right],$$

and this completes the proof. □

Lemma A.4. *Assume Assumption 1. In general, it is not true that:*

$$\{\Gamma\cap\{G:F_X^L(x)\leq G(x)\leq F_X^H(x),\ \forall x\in\mathbb{R}^{\mathcal{I}}\}\}\subseteq\mathcal{H}\left(F_X\right).$$

Proof. To make the argument transparent, consider a one dimensional example, i.e., $\mathcal{I}=\{t_0\}\in\mathbb{R}$. In this case, Eq. (3.1) yields: $F_X\left(x\right)=P\left(X'\left(t_0\right)\leq x\left(t_0\right)\cap\{O=\mathcal{I}\}\right)+P\left(X\left(t_0\right)\leq x\left(t_0\right)\cap\{O=\varnothing\}\right)$, and also $F_X^L=P\left(X'\left(t_0\right)\leq x\left(t_0\right)\cap\{O=\mathcal{I}\}\right)$ and $F_X^H=P\left(X'\left(t_0\right)\leq x\left(t_0\right)\cap\{O=\mathcal{I}\}\right)+P\left(O=\varnothing\right)$. Consider arbitrary $x_1,x_2:\mathcal{I}=\{t_0\}\to\mathbb{R}$, such that $x_1\left(t_0\right)<x_2\left(t_0\right)$, and let F_X^L, F_X^H, and F be such that:

z	$F_X^L\left(z\right)$	$F_X^H\left(z\right)$	$F\left(z\right)$
x_1	0.1	0.4	0.3
x_2	0.2	0.5	0.3

It is not hard to verify that F_X^H and F_X^L are valid candidates for these functions, and that $F\in\{\Gamma\cap\{G:F_X^L(x)\leq G(x)\leq F_X^H(x),\forall x\in\mathbb{R}^{\mathcal{I}}\}\}$. We now verify that $F\notin\mathcal{H}\left(F_X\right)$. If $F=F_X$, Eq. (3.1) applied to x_1,x_2 would imply that $P\left(X\left(t_0\right)\leq x_2\left(t_0\right)\cap\{O=\varnothing\}\right)<P\left(X\left(t_0\right)\leq x_1\left(t_0\right)\cap\{O=\varnothing\}\right)$. Since $x_1\left(t_0\right)<x_2\left(t_0\right)$, this is a contradiction. □

Proof of Lemma 3.3. This proof is trivial and is therefore omitted. □

Proof of Lemma 3.4. This proof is trivial and is therefore omitted. □

Lemma A.5. *Assume Assumption 1. In general, it is not true that:*

$$T_L'\left(X,Y_{\theta_0}\right)=T_L\left(X,Y_{\theta_0}\right)\ \textit{or}\ T_H\left(X,Y_{\theta_0}\right)=T_H'\left(X,Y_{\theta_0}\right).$$

Proof. For transparency of the argument, consider the one dimensional example used in Lemma A.4, i.e., $\mathcal{I}=\{t_0\}\in\mathbb{R}$. Consider arbitrary $x_1,x_2,x_3:\mathcal{I}=\{t_0\}\to\mathbb{R}$, such that $x_1\left(t_0\right)<x_2\left(t_0\right)<x_3\left(t_0\right)$, and let F_X^L, F_X^H and $F_Y\left(\cdot|\theta_0\right)$ be such that:

z	$F_X^L\left(z\right)$	$F_X^H\left(z\right)$	$F_Y\left(z\vert\theta_0\right)$
x_1	0.1	0.4	0.3
x_2	0.2	0.5	0.3
x_3	0.3	0.6	0.9

It is easy to verify that F_X^H and F_X^L are valid candidates for these functions. We assume that $\mu\left(x_1\right),\mu\left(x_2\right),\mu\left(x_3\right)>0$. As in the main text, $\mathcal{H}'\left(F_X\right)=\{G:F_X^L(x)\leq G(x)\leq F_X^H(x),\forall x\in\mathbb{R}^{\mathcal{I}}\}$.

We begin with lower WCSB. By definition:

$$T'_L(X, Y_{\theta_0}) = \inf_{F \in \mathcal{H}'(F_X)}\{(F(x_1) - 0.3)^2 \mu(x_1) + (F(x_2) - 0.3)^2 \mu(x_2) + (F(x_3) - 0.9)^2 \mu(x_3)\},$$
$$T_L(X, Y_{\theta_0}) = \inf_{F \in \mathcal{H}(F_X)}\{(F(x_1) - 0.3)^2 \mu(x_1) + (F(x_2) - 0.3)^2 \mu(x_2) + (F(x_3) - 0.9)^2 \mu(x_3)\},$$

The solution to the first minimization problem is given by: $F(x_1) = F(x_2) = 0.3$ and $F(x_3) = 0.6$. By the argument used in Lemma A.4, $F \in \left\{\Gamma \cap \left\{G : F_X^L(x) \leq G(x) \leq F_X^H(x), \forall x \in \mathbb{R}^{\mathcal{I}}\right\}\right\}$ but $F \notin \mathcal{H}(F_X)$. The minimizer in the second problem needs to belong to $\mathcal{H}(F_X)$ and so $T'_L(X, Y_{\theta_0}) < T(X, Y_{\theta_0})$.

We now consider the upper WCSB. By definition:

$$T'_H(X, Y_{\theta_0}) = \sup_{F \in \mathcal{H}'(F_X)}\{(F(x_1) - 0.3)^2 \mu(x_1) + (F(x_2) - 0.3)^2 \mu(x_2) + (F(x_3) - 0.9)^2 \mu(x_3)\},$$
$$T_H(X, Y_{\theta_0}) = \sup_{F \in \mathcal{H}(F_X)}\{(F(x_1) - 0.3)^2 \mu(x_1) + (F(x_2) - 0.3)^2 \mu(x_2) + (F(x_3) - 0.9)^2 \mu(x_3)\},$$

The solution to the first maximization problem is given by: $F(x_1) = 0.1$, $F(x_2) = 0.5$, and $F(x_3) = 0.3$. Since F is not weakly increasing, $F \notin \Gamma \supset \mathcal{H}(F_X)$. The minimizer in the second problem needs to belong to $\mathcal{H}(F_X)$ and so $T'_H(X, Y_{\theta_0}) > T_H(X, Y_{\theta_0})$. □

A.4 Proofs of Section 4

Proof of Theorem 4.1. The results in BHHN are derived under the assumption that all of the approximation errors are negligible relative to sampling error. This is formally achieved by taking limits in the following order: $V \to \infty$, $m \to \infty$, and $n \to \infty$. Let $\hat{T}(X, Y_{\hat{\theta}_0})$ denote the (unfeasible) test statistic that we would compute if we were to observe the complete dataset. Then:

$$\liminf_{n\to\infty} \liminf_{m\to\infty} \liminf_{V\to\infty} P\left(n\hat{T}'_L(X, Y_{\hat{\theta}_0}) > t^*_{\hat{\theta}_0}(1-\alpha)\right) \leq \lim_{n\to\infty} \lim_{m\to\infty} \lim_{V\to\infty} P\left(n\hat{T}(X, Y_{\hat{\theta}_0}) > t^*_{\hat{\theta}_0}(1-\alpha)\right) = \alpha,$$

where the inequality follows from $\hat{T}'_L(X, Y_{\hat{\theta}_0}) \leq \hat{T}(X, Y_{\hat{\theta}_0})$ and the equality follows from Theorem 3.2 in BHHN. □

Lemma A.6. *Assume Assumptions 1-5. Then, for all $\varepsilon > 0$:*

$$\lim_{n\to\infty} \lim_{m\to\infty} \lim_{V\to\infty} P\left(\left\|(\hat{T}'_L(X, Y_{\hat{\theta}_0}), \hat{T}'_H(X, Y_{\hat{\theta}_0})) - (T'_L(X, Y_{\theta_0}), T'_H(X, Y_{\theta_0}))\right\| > \varepsilon\right) = 0.$$

Proof. This proof will explicitly consider the argument for the upper WCSB, but the same arguments can be used for the lower WCSB. Also, it is relevant to point out that the following argument is tailored to the order in which the sequences are considered, i.e., $V \to \infty$, $m \to \infty$, and $n \to \infty$. Had the order of the limits been different, then the formal results would still be true, but the formal argument would be slightly different.

Let Φ denote space of pairs of functions (F_1, F_2) such that (i) $F_1, F_2 : \mathbb{R}^{\mathcal{I}} \to \mathbb{R}$ and (ii) $\forall x \in \mathbb{R}^{\mathcal{I}}$,

$0 \leq F_1(x) \leq F_2(x) \leq 1$. Define $\Psi_H, \check{\Psi}_H, \tilde{\Psi}_H, \hat{\Psi}_H : \Phi \to [0,1]$ as follows:

$$\begin{aligned}
\Psi_H(F_1,F_2) &\equiv \sup_{F \in \{G: F_1(x) \leq G(x) \leq F_1(x),\ \forall x \in \mathbb{R}^{\mathcal{I}}\}} \int (F(x) - F_Y(x|\theta_0))^2 d\mu(x), \\
\check{\Psi}_H(F_1,F_2) &\equiv \sup_{F \in \{G: F_1(x) \leq G(x) \leq F_1(x),\ \forall x \in \mathbb{R}^{\mathcal{I}}\}} \int (F(x) - F_Y(x|\hat{\theta}_0))^2 d\mu(x), \\
\tilde{\Psi}_H(F_1,F_2) &\equiv \sup_{F \in \{G: F_1(x) \leq G(x) \leq F_1(x),\ \forall x \in \mathbb{R}^{\mathcal{I}}\}} \int (F(x) - \hat{F}_Y(x|\hat{\theta}_0))^2 d\mu(x), \\
\hat{\Psi}_H(F_1,F_2) &\equiv \sup_{F \in \{G: F_1(x) \leq G(x) \leq F_1(x),\ \forall x \in \mathbb{R}^{\mathcal{I}}\}} \tfrac{1}{V} \sum_{j=1}^{V} (F(Z_j) - \hat{F}_Y(Z_j|\hat{\theta}_0))^2,
\end{aligned}$$

where $\hat{\theta}_0$, $\hat{F}_Y$, and $\{Z_j\}_{j=1}^V$ are as in Assumptions 1 and 4. By definition: $T'_H(X, Y_{\theta_0}) = \Psi_H(F_X^L, F_X^H)$ and $\hat{T}'_H(X, Y_{\hat{\theta}_0}) = \hat{\Psi}_H(\hat{F}_X^L, \hat{F}_X^H)$ and, thus, by the triangular inequality:

$$\left|\hat{T}'_H\left(X, Y_{\hat{\theta}_0}\right) - T'_H(X, Y_{\theta_0})\right| \leq \left\{ \begin{array}{c} \left|\hat{\Psi}_H(\hat{F}_X^L, \hat{F}_X^H) - \tilde{\Psi}_H(\hat{F}_X^L, \hat{F}_X^H)\right| + \left|\tilde{\Psi}_H(\hat{F}_X^L, \hat{F}_X^H) - \check{\Psi}_H(\hat{F}_X^L, \hat{F}_X^H)\right| \\ + \left|\check{\Psi}_H(\hat{F}_X^L, \hat{F}_X^H) - \Psi_H(\hat{F}_X^L, \hat{F}_X^H)\right| + \left|\Psi_H(\hat{F}_X^L, \hat{F}_X^H) - \Psi_H(F_X^L, F_X^H)\right| \end{array} \right\}.$$

The objective of the rest of the proof is to show that each of the terms in the RHS is $o_p(1)$, as $V \to \infty$, $m \to \infty$, and $n \to \infty$. The remainder of the proof is divided into steps.

Step 1: Show that $\Psi_H(\hat{F}_X^L, \hat{F}_X^H) - \Psi_H(F_X^L, F_X^H) = o_p(1)$, as $V \to \infty$, $m \to \infty$, and $n \to \infty$. We show something stronger, namely that $\Psi_H(\hat{F}_X^L, \hat{F}_X^H) - \Psi_H(F_X^L, F_X^H) = o_p(1)$, as $(V, m, n) \to (\infty, \infty, \infty)$. For any $n, m, V \in \mathbb{N}$, denote by $P_{n,m,V}$ the probability measure with respect to the randomness in $\mathcal{X}_n$, $\mathcal{Y}_{\theta,m}$, and $\mathcal{Z}_V$. Consider the following derivation for any $n, m, m', V, V' \in \mathbb{N}$:

$$\begin{aligned}
&P_{n,m,V}\left(\left|\Psi_H(\hat{F}_X^L, \hat{F}_X^H) - \Psi_H(F_X^L, F_X^H)\right| > \varepsilon\right) \\
&= \int_{\mathcal{X}_n, \mathcal{Y}_{\theta,m}, \mathcal{Z}_V} dP\left(\left|\Psi_H(\hat{F}_X^L, \hat{F}_X^H) - \Psi_H(F_X^L, F_X^H)\right| > \varepsilon | \mathcal{X}_n, \mathcal{Y}_{\theta,m}, \mathcal{Z}_V\right) dP(\mathcal{X}_n | \mathcal{Y}_{\theta,m}, \mathcal{Z}_V)\, dP(\mathcal{Y}_{\theta,m}, \mathcal{Z}_V) \\
&= \int_{\mathcal{X}_n} dP\left(\left|\Psi_H(\hat{F}_X^L, \hat{F}_X^H) - \Psi_H(F_X^L, F_X^H)\right| > \varepsilon | \mathcal{X}_n\right) dP(\mathcal{X}_n) \\
&= P_{n,m',V'}\left(\left|\Psi_H(\hat{F}_X^L, \hat{F}_X^H) - \Psi_H(F_X^L, F_X^H)\right| > \varepsilon\right),
\end{aligned}$$

where the first equality holds by definition, the second equality holds because $\Psi_H(\hat{F}_X^L, \hat{F}_X^H) - \Psi_H(F_X^L, F_X^H)$ is exclusively a function of $\mathcal{X}_n$ and $\mathcal{X}_n$ is independent of $\mathcal{Y}_{\theta,m}$ and $\mathcal{Z}_V$ (Assumption 5.c), and the final equality holds by undoing the previous steps for a sample of $\mathcal{X}_n$, $\mathcal{Y}_{\theta,m'}$, and $\mathcal{Z}_{V'}$. Since $n, m, m', V, V' \in \mathbb{N}$ were arbitrarily chosen, this reveals that the distribution of the event $\{|\Psi_H(\hat{F}_X^L, \hat{F}_X^H) - \Psi_H(F_X^L, F_X^H)| > \varepsilon\}$ does not depend on m and V. As a consequence, $\Psi_H(\hat{F}_X^L, \hat{F}_X^H) - \Psi_H(F_X^L, F_X^H) = o_p(1)$, as $(V, m, n) \to (\infty, \infty, \infty)$ if and only if $\Psi_H(\hat{F}_X^L, \hat{F}_X^H) - \Psi_H(F_X^L, F_X^H) = o_p(1)$, as $n \to \infty$. Thus, the step is complete if we show that $\Psi_H(\hat{F}_X^L, \hat{F}_X^H) - \Psi_H(F_X^L, F_X^H) = o_p(1)$, as $n \to \infty$. For convenience, we further divide this into steps.

Step 1.1: Show that if $(F_1, F_2) \in \Phi$, then $\forall \varepsilon > 0$, $\exists \delta = \delta(\varepsilon, F_1, F_2) > 0$ such that $(F'_1, F'_2) \in \Phi$ and $||F'_1 - F_1||_\mu + ||F'_2 - F_2||_\mu \leq \delta$ implies $|\Psi_H(F'_1, F'_2) - \Psi_H(F_1, F_2)| \leq \varepsilon$. Fix $\varepsilon > 0$ and set $\varepsilon_0 \in (0, \varepsilon)$. Let $H : \mathbb{R}_+ \to \mathbb{R}_+$ be defined by:

$$H(y) \equiv 2y^{1/2}(\varepsilon + \Psi_H(F_1, F_2))^{1/2} + y.$$

Note that H is continuous, $H(0) = 0$, and $\lim_{y \to \infty} H(y) = \infty$. By the intermediate value theorem $\exists \delta > 0$

such that:

$$2\delta^{1/2}\left(\varepsilon+\Psi_H\left(F_1,F_2\right)\right)^{1/2}+\delta=\varepsilon-\varepsilon_0>0. \tag{A.7}$$

Consider any $(F_1',F_2')\in\Phi$ and $||F_1'-F_1||_\mu+||F_2'-F_2||_\mu\leq\delta$. We now show that $|\Psi_H\left(F_1',F_2'\right)-\Psi_H\left(F_1,F_2\right)|\leq\varepsilon$.

By definition of supremum, $\exists G'\in\{G:F_1'(x)\leq G(x)\leq F_2'(x),\ \forall x\in\mathbb{R}^{\mathcal{I}}\}$ such that $\Psi_H(F_1',F_2')-\varepsilon_0\leq||G'-F_Y\left(\cdot|\theta_0\right)||_\mu$. Furthermore, if we define $G'':\mathbb{R}^{\mathcal{I}}\to\mathbb{R}$ as follows: $G''\left(x\right)\equiv\min\{\max\{G'\left(x\right),F_1\left(x\right)\},F_2\left(x\right)\}$, we have that $G''\in\{G:F_1(x)\leq G(x)\leq F_2(x),\ \forall x\in\mathbb{R}^{\mathcal{I}}\}$. We now show that $||G''-G'||_\mu\leq\delta$. To see this, consider the following derivation.

$$\begin{aligned}
&\|G''-G'\|_\mu=\int\left(G''\left(x\right)-G'\left(x\right)\right)^2d\mu\left(x\right)\\
&\leq\int\left(F_1\left(x\right)-F_1'\left(x\right)\right)^2 1\left[G'\left(x\right)<F_1\left(x\right)\right]d\mu\left(x\right)+\int\left(F_2'\left(x\right)-F_2\left(x\right)\right)^2 1\left[G'\left(x\right)>F_2\left(x\right)\right]d\mu\left(x\right)\\
&\leq\|F_1'-F_1\|_\mu+\|F_2'-F_2\|_\mu\leq\delta,
\end{aligned}$$

where the first inequality follows from the definition of G'' and that $G'\in\{G:F_1'(x)\leq G(x)\leq F_2'(x),\forall x\in\mathbb{R}^{\mathcal{I}}\}$. Based on this, consider the following derivation:

$$\begin{aligned}
&\Psi_H\left(F_1',F_2'\right)-\varepsilon_0\leq||G'-F_Y\left(\cdot|\theta_0\right)||_\mu\\
&=\int\left(G'\left(x\right)-G''\left(x\right)+G''\left(x\right)-F_Y\left(x|\theta_0\right)\right)^2d\mu\left(x\right)\\
&\leq||G''-G'||_\mu+2(||G''-G'||_\mu)^{1/2}(||G''-F_Y\left(\cdot|\theta_0\right)||)^{1/2}+||G''-F_Y\left(\cdot|\theta_0\right)||_\mu\\
&\leq\delta+2\delta^{1/2}\left(\Psi_H\left(F_1,F_2\right)\right)^{1/2}+\Psi_H\left(F_1,F_2\right)\\
&\leq\delta+2\delta^{1/2}\left(\varepsilon+\Psi_H\left(F_1,F_2\right)\right)^{1/2}+\Psi_H\left(F_1,F_2\right)=\varepsilon-\varepsilon_0+\Psi_H\left(F_1,F_2\right),
\end{aligned}$$

where we have used that $G''\in\{G:F_1(x)\leq G(x)\leq F_2(x),\forall x\in\mathbb{R}^{\mathcal{I}}\}$ and, thus, $\|G''-F_Y\left(\cdot|\theta_0\right)\|_\mu\leq\Psi_H\left(F_1,F_2\right)$, $||G''-G'||_\mu\leq\delta$, $\varepsilon>0$, and the definition of δ in Eq. (A.7). From the previous argument, it follows immediately that $\Psi_H(F_1',F_2')-\Psi_H\left(F_1,F_2\right)\leq\varepsilon$.

We now repeat the previous argument reversing the roles of $\Psi_H\left(F_1,F_2\right)$ and $\Psi_H\left(F_1',F_2'\right)$. By doing this, we define $\tilde{G}'$ such that $\Psi_H(F_1',F_2')-\varepsilon_0\leq||\tilde{G}'-F_Y\left(\cdot|\theta_0\right)||_\mu$ and $\tilde{G}''$ such that $\tilde{G}''\left(x\right)\equiv\min\{\max\{\tilde{G}'\left(x\right),F_1'\left(x\right)\},F_2'\left(x\right)\}$, and that satisfies $||\tilde{G}''-\tilde{G}'||_\mu\leq\delta$. Then, we deduce that:

$$\begin{aligned}
&\Psi_H\left(F_1,F_2\right)-\varepsilon_0\leq||\tilde{G}'-F_Y\left(\cdot|\theta_0\right)||_\mu\\
&=\int(\tilde{G}'\left(x\right)-\tilde{G}''\left(x\right)+\tilde{G}''\left(x\right)-F_Y\left(x|\theta_0\right))^2d\mu\left(x\right)\\
&\leq\delta+2\delta^{1/2}(\Psi_H(F_1',F_2'))^{1/2}+\Psi_H(F_1',F_2')\\
&\leq\delta+2\delta^{1/2}\left(\varepsilon+\Psi_H\left(F_1,F_2\right)\right)^{1/2}+\Psi_H(F_1',F_2')\leq\varepsilon-\varepsilon_0+\Psi_H(F_1',F_2'),
\end{aligned}$$

where we have repeated previous arguments and used that that $\Psi_H(F_1',F_2')-\Psi_H\left(F_1,F_2\right)\leq\varepsilon$. From this, it follows immediately that $\Psi_H\left(F_1,F_2\right)-\Psi_H(F_1',F_2')\leq\varepsilon$, which gives: $|\Psi_H\left(F_1',F_2'\right)-\Psi_H\left(F_1,F_2\right)|\leq\varepsilon$.

Step 1.2: Show that $(\hat{F}_X^L,\hat{F}_X^H)\in\Phi$, and $||\hat{F}_X^L-F_X^L||_\mu+||\hat{F}_X^H-F_X^H||_\mu=o_p\left(1\right)$ as $n\to\infty$. By definition, it is trivial that $(\hat{F}_X^L,\hat{F}_X^H)\in\Phi$. We concentrate the rest of the proof in showing that $||\hat{F}_X^L-F_X^L||_\mu+||\hat{F}_X^H-F_X^H||_\mu=$

$o_p(1)$. For any $j = 1, \ldots, 2^J$, define $\varpi_{n,j} = ||\hat{F}_j(\cdot) - F_j(\cdot)||_\mu$, where $\forall x \in L_2(\mathcal{I})$:

$$\hat{F}_j(x) \equiv \frac{1}{n}\sum_{i=1}^n 1(X_i'(t) \leq x(t), \forall t \in S_j)\, 1(O_i = S_j),$$
$$F_j(x) \equiv P(\{X'(t) \leq x(t), \forall t \in S_j\} \cap \{O = S_j\}).$$

By the same arguments as in the proof of Theorem 3.1 in BHHN, $n^{1/2}(\hat{F}_j(\cdot) - F_j(\cdot)) : \Omega \to L_2(\mathcal{I})$ is a stochastic process that converges weakly to a Gaussian process in $L_2(\mathcal{I})$ and, thus, $\varpi_{n,j} = o_p(1)$ as $n \to \infty$. On the one hand, $\hat{F}_X^L(\cdot) - F_X^L(\cdot) = \hat{F}_1(\cdot) - F_1(\cdot)$ and so: $||\hat{F}_X^L - F_X^L||_\mu = ||\hat{F}_1 - F_1||_\mu = \varpi_{n,1} = o_p$, as $n \to \infty$. On the other hand, $\hat{F}_X^H(\cdot) - F_X^H(\cdot) = \sum_{j=1}^{2^J} \hat{F}_j(\cdot) - F_j(\cdot)$ and so, by Hölder's inequality:

$$\begin{aligned}
&||\hat{F}_X^H - F_X^H||_\mu = \int (\hat{F}_X^H(x) - F_X^H(x))^2 d\mu(x) \\
&= \int |[\textstyle\sum_{a=1}^{2^J} (\hat{F}_a(x) - F_a(x))][\sum_{b=1}^{2^J} (\hat{F}_b(x) - F_b(x))]| d\mu(x) \\
&\leq \textstyle\sum_{j=1}^{2^J} \int (\hat{F}_j(x) - F_j(x))^2 d\mu(x) = \sum_{j=1}^{2^J} \varpi_{n,j} = o_p(1), \text{ as } n \to \infty.
\end{aligned}$$

Step 1.3: Complete the argument. Fix $\varepsilon > 0$ arbitrarily. Since $(F_X^L, F_X^H) \in \Phi$, then step 1.1 implies that $\exists \delta = \delta(\varepsilon, F_X^L, F_X^H) > 0$ such that for all $(F_1', F_2') : \{(F_1', F_2') \in \Phi \cap ||F_1' - F_X^L||_\mu + ||F_2' - F_X^H||_\mu \leq \delta\} \subseteq \{|\Psi_H(F_1', F_2') - \Psi_H(F_X^L, F_X^H)| \leq \varepsilon\}$. As a consequence:

$$P(\{(\hat{F}_X^H, \hat{F}_X^H) \in \Phi \cap ||\hat{F}_X^H - F_X^L||_\mu + ||\hat{F}_X^H - F_X^H||_\mu \leq \delta\}) \leq P(|\Psi_H(F_1', F_2') - \Psi_H(F_X^L, F_X^H)| \leq \varepsilon).$$

By step 1.2, the LHS converges to one as $n \to \infty$, which implies that the RHS also converges to one as $n \to \infty$.

Step 2: Show that $\check{\Psi}_H(\hat{F}_X^L, \hat{F}_X^H) - \Psi_H(\hat{F}_X^L, \hat{F}_X^H) = o_p(1)$, as $V \to \infty$, $m \to \infty$, and $n \to \infty$. We show something stronger, namely that $\check{\Psi}_H(\hat{F}_X^L, \hat{F}_X^H) - \Psi_H(\hat{F}_X^L, \hat{F}_X^H) = o_p(1)$, as $(V, m, n) \to (\infty, \infty, \infty)$. To show this, we can use the same argument used in step 1 to argue that, for any $\varepsilon > 0$, $\{|\check{\Psi}_H(\hat{F}_X^L, \hat{F}_X^H) - \Psi_H(\hat{F}_X^L, \hat{F}_X^H)| > \varepsilon\}$ does not depend on m and V. To achieve this result, we rely on the fact that $\hat{\theta}_0$ is a function of $\mathcal{X}_n$ (Assumption 2.b) and that $\mathcal{X}_n$ is independent of $\mathcal{Y}_{\theta,m}$ and $\mathcal{Z}_V$ (Assumption 5.c). As a consequence, $\check{\Psi}_H(\hat{F}_X^L, \hat{F}_X^H) - \Psi_H(\hat{F}_X^L, \hat{F}_X^H) = o_p(1)$, as $(V, m, n) \to (\infty, \infty, \infty)$ if and only if $\check{\Psi}_H(\hat{F}_X^L, \hat{F}_X^H) - \Psi_H(\hat{F}_X^L, \hat{F}_X^H) = o_p(1)$, as $n \to \infty$. Thus, the step is complete if we show that $\check{\Psi}_H(\hat{F}_X^L, \hat{F}_X^H) - \Psi_H(\hat{F}_X^L, \hat{F}_X^H) = o_p(1)$, as $n \to \infty$. We begin with the following derivation:

$$\begin{aligned}
||F_Y(\cdot|\hat{\theta}_0) - F_Y(\cdot|\theta_0)||_\mu &= \int (F_Y(x|\hat{\theta}_0) - F_Y(x|\theta_0))^2 d\mu(x) \\
&= \int \left|F_Y(x|\hat{\theta}_0) - F_Y(x|\theta_0)\right| \left|\sum_{i=1}^p (\partial F_Y(x|\tilde{\theta})/\partial\theta_i)(\hat{\theta}_{0,i} - \theta_{0,i})\right| d\mu(x) \\
&\leq \int |\sum_{i=1}^p (\partial F_Y(x|\tilde{\theta})/\partial\theta_i)(\hat{\theta}_{0,i} - \theta_{0,i})| d\mu(x) \\
&\leq ||\hat{\theta}_0 - \theta_0|| \left\{\sup_{\theta \in \mathcal{O}} \int [(\partial F_Y(x|\theta)/\partial\theta')(\partial F_Y(x|\theta)/\partial\theta)] d\mu(x)\right\}^{1/2} \\
&= O_p(n^{-1/2}) O(1) = o_p(1), \text{ as } n \to \infty.
\end{aligned}$$

where the first equality holds by definition, the second equality holds by the mean value theorem ($\tilde{\theta} \in \mathbb{R}^p$ is between θ_0 and $\hat{\theta}_0$), the first inequality holds because $\forall (\theta, x) \in \Theta \times \mathbb{R}^{\mathcal{I}}$, $F_Y(x|\theta) \in [0, 1]$, the second inequality holds by Hölder's inequality, and the final inequality holds by Assumptions 2.b and 3.

Next, for a fixed function F, consider the following derivation:

$$
\begin{aligned}
&||F - F_Y(\cdot|\hat{\theta}_0)||_\mu = \int (F(x) - F_Y(x|\hat{\theta}_0))^2 d\mu(x) \\
&= \int (F(x) - F_Y(x|\theta_0) + F_Y(x|\theta_0) - F_Y(x|\hat{\theta}_0))^2 d\mu(x) \\
&\leq \left\{ \begin{array}{c} \int (F(x) - F_Y(x|\theta_0))^2 d\mu(x) + \int (F_Y(x|\hat{\theta}_0) - F_Y(x|\theta_0))^2 d\mu(x) + \\ 2[(\int (F(x) - F_Y(x|\theta_0))^2 d\mu(x))(\int (F_Y(x|\hat{\theta}_0) - F_Y(x|\theta_0))^2 d\mu(x))]^{1/2} \end{array} \right\} \\
&\leq ||F - F_Y(\cdot|\theta_0)||_\mu + ||F_Y(\cdot|\hat{\theta}_0) - F_Y(\cdot|\theta_0)||_\mu + 2||F_Y(\cdot|\hat{\theta}_0) - F_Y(\cdot|\theta_0)||_\mu^{1/2},
\end{aligned}
$$

where we have used Hölder's inequality and the fact that $\forall(\theta, x) \in \Theta \times \mathbb{R}^{\mathcal{I}}$, $F(x), F_Y(x|\theta_0) \in [0,1]$. Taking supremum on both sides of the inequality with respect to $F \in \{G : \hat{F}_X^L(x) \leq G(x) \leq \hat{F}_X^H(x), \forall x \in \mathbb{R}^{\mathcal{I}}\}$, and using that $||F_Y(\cdot|\hat{\theta}_0) - F_Y(\cdot|\theta_0)||_\mu = o_p(1)$, as $n \to \infty$, it then follows that $\check{\Psi}_H(\hat{F}_X^L, \hat{F}_X^H) - \Psi_H(\hat{F}_X^L, \hat{F}_X^H) \leq o_p(1)$, as $n \to \infty$. By reversing the roles of $F_Y(\cdot|\hat{\theta}_0)$ and $F_Y(\cdot|\theta_0)$ and repeating the argument, it follows that $\Psi_H(\hat{F}_X^L, \hat{F}_X^H) - \check{\Psi}_H(\hat{F}_X^L, \hat{F}_X^H) \leq o_p(1)$, as $n \to \infty$.

Step 3: We show that $\tilde{\Psi}_H(\hat{F}_X^L, \hat{F}_X^H) - \check{\Psi}_H(\hat{F}_X^L, \hat{F}_X^H) = o_p(1)$, as $V \to \infty$, $m \to \infty$, and $n \to \infty$. We first use an argument similar to the one used in step 1 to argue that, for any $\varepsilon > 0$, $\{|\tilde{\Psi}_H(\hat{F}_X^L, \hat{F}_X^H) - \check{\Psi}_H(\hat{F}_X^L, \hat{F}_X^H)| > \varepsilon\}$ does not depend on V. To achieve this result, we rely on the fact that $\hat{\theta}_0$ is a function of $\mathcal{X}_n$ (Assumption 2.b), which implies that $\tilde{\Psi}_H(\hat{F}_X^L, \hat{F}_X^H) - \check{\Psi}_H(\hat{F}_X^L, \hat{F}_X^H)$ is only a function of $\mathcal{X}_n$ and $\mathcal{Y}_{\theta,m}$, and that $\mathcal{X}_n$ and $\mathcal{Y}_{\theta,m}$ are independent of $\mathcal{Z}_V$ (Assumption 5.c). Thus, the step is complete if we show that $\Psi_H(\hat{F}_X^L, \hat{F}_X^H) - \Psi_H(F_X^L, F_X^H) = o_p(1)$, as $m \to \infty$ and $n \to \infty$.

We condition on $\mathcal{X}_n$ and consider the following argument. By Assumption 1.b, $\hat{\theta}_0$ is only a function of $\mathcal{X}_n$ and, consequently, it is conditioned upon. Consider now the stochastic behavior of $||\hat{F}_Y(\cdot|\hat{\theta}_0) - F_Y(\cdot|\hat{\theta}_0)||_\mu$, where $x \in \mathbb{R}^{\mathcal{I}}$:

$$
\begin{aligned}
\hat{F}_Y(x|\hat{\theta}_0) &= m^{-1} \sum_{i=1}^{m} 1(Y_{\hat{\theta}_0,i}(t) \leq x(t), \forall t \in \mathcal{I}), \\
F_Y(x|\hat{\theta}_0) &= P(Y_{\hat{\theta}_0}(\omega, t) \leq x(t), \forall t \in \mathcal{I})),
\end{aligned}
$$

By Assumption 5.a and the arguments used in the proof of Theorem 3.1 in BHHN, $m^{1/2}(\hat{F}_Y(\cdot|\hat{\theta}_0) - F_Y(\cdot|\hat{\theta}_0)) : \Omega \to L_2(\mathcal{I})$ is a stochastic process that converges weakly to a Gaussian process in $L_2(\mathcal{I})$ and, thus, $\{||\hat{F}_Y(\cdot|\hat{\theta}_0) - F_Y(\cdot|\hat{\theta}_0)||_\mu|\mathcal{X}_n\} = o_p(1)$ as $m \to \infty$. Therefore, for any $\varepsilon > 0$:

$$
\begin{aligned}
\lim_{n\to\infty} \lim_{m\to\infty} P\left(||\hat{F}_Y(\cdot|\hat{\theta}_0) - F_Y(\cdot|\hat{\theta}_0)||_\mu > \varepsilon\right) &= \lim_{n\to\infty} \lim_{m\to\infty} \int P\left(||\hat{F}_Y(\cdot|\hat{\theta}_0) - F_Y(\cdot|\hat{\theta}_0)||_\mu > \varepsilon|\mathcal{X}_n\right) dP(\mathcal{X}_n) \\
&= \lim_{n\to\infty} \int \lim_{m\to\infty} P\left(||\hat{F}_Y(\cdot|\hat{\theta}_0) - F_Y(\cdot|\hat{\theta}_0)||_\mu > \varepsilon|\mathcal{X}_n\right) dP(\mathcal{X}_n) = 0,
\end{aligned}
$$

where we have used the previous finding and the dominated convergence theorem. As a consequence, $||\hat{F}_Y(\cdot|\hat{\theta}_0) - F_Y(\cdot|\hat{\theta}_0)||_\mu = o_p(1)$ as $m \to \infty$ and $n \to \infty$.

Next, consider the following derivation for a fixed distribution F:

$$\begin{aligned}
&\left\|F-\hat{F}_Y(\cdot|\hat{\theta}_0)\right\|_\mu=\int\left(F\left(x\right)-\hat{F}_Y(x|\hat{\theta}_0)\right)^2d\mu\left(x\right)\\
&=\int\left(F\left(x\right)-\hat{F}_Y(x|\hat{\theta}_0)+\hat{F}_Y(x|\hat{\theta}_0)-F_Y(x|\hat{\theta}_0)\right)^2d\mu\left(x\right)\\
&\leq\left\{\begin{array}{c}\int\left(F\left(x\right)-\hat{F}_Y(x|\hat{\theta}_0)\right)^2d\mu\left(x\right)+\int\left(\hat{F}_Y(x|\hat{\theta}_0)-F_Y(x|\hat{\theta}_0)\right)^2d\mu\left(x\right)+\\ 2\left[\left(\int\left(F\left(x\right)-\hat{F}_Y(x|\hat{\theta}_0)\right)^2d\mu\left(x\right)\right)\left(\int\left(\hat{F}_Y(x|\hat{\theta}_0)-F_Y(x|\hat{\theta}_0)\right)^2d\mu\left(x\right)\right)\right]^{1/2}\end{array}\right\}\\
&\leq\left\|F-F_Y(\cdot|\hat{\theta}_0)\right\|_\mu+\left\|\hat{F}_Y(\cdot|\hat{\theta}_0)-F_Y(\cdot|\hat{\theta}_0)\right\|_\mu+2\left\|\hat{F}_Y(\cdot|\hat{\theta}_0)-F_Y(\cdot|\hat{\theta}_0)\right\|_\mu^{1/2},
\end{aligned}$$

where we have used Hölder's inequality and the fact that $\forall\left(\theta,x\right)\in\Theta\times\mathbb{R}^{\mathcal{I}}$, $F\left(x\right),F_Y\left(x|\theta_0\right)\in[0,1]$. Taking supremum on both sides of the inequality with respect to $F\in\{G:\hat{F}_X^L(x)\leq G(x)\leq\hat{F}_X^H(x)\ \forall x\in\mathbb{R}^{\mathcal{I}}\}$, and using that $||\hat{F}_Y(\cdot|\hat{\theta}_0)-F_Y(\cdot|\hat{\theta}_0)||_\mu=o_p\left(1\right)$ as $m\to\infty$ and $n\to\infty$ it follows that $\tilde{\Psi}_H(\hat{F}_X^L,\hat{F}_X^H)-\check{\Psi}_H(\hat{F}_X^L,\hat{F}_X^H)\leq o_p\left(1\right)$, as $m\to\infty$ and $n\to\infty$. By reversing the roles of $F_Y(\cdot|\hat{\theta}_0)$ and $F_Y\left(\cdot|\theta_0\right)$ and repeating the argument, it follows that $\check{\Psi}_H(\hat{F}_X^L,\hat{F}_X^H)-\tilde{\Psi}_H(\hat{F}_X^L,\hat{F}_X^H)\leq o_p\left(1\right)$, as $m\to\infty$ and $n\to\infty$.

Step 4: In this step, we show that $\hat{\Psi}_H(\hat{F}_X^L,\hat{F}_X^H)-\tilde{\Psi}_H(\hat{F}_X^L,\hat{F}_X^H)=o_p\left(1\right)$, as $V\to\infty$, $m\to\infty$, and $n\to\infty$. We condition on $\mathcal{X}_n$ and $\mathcal{Y}_{\theta,m}$, and consider the following argument. By Assumption 1.b, $\hat{\theta}_0$ is only a function of $\mathcal{X}_n$ and, consequently, it is conditioned upon. In this setting, consider:

$$\begin{aligned}
&\hat{\Psi}_H(\hat{F}_X^L,\hat{F}_X^H)-\tilde{\Psi}_H(\hat{F}_X^L,\hat{F}_X^H)\\
&=\left\{\begin{array}{c}\sup_{F\in\{G:\hat{F}_X^L(x)\leq G(x)\leq\hat{F}_X^H(x),\ \forall x\in\mathbb{R}^{\mathcal{I}}\}}V^{-1}\sum_{j=1}^V(F\left(Z_j\right)-\hat{F}_Y(Z_j|\hat{\theta}_0))^2\\ -\sup_{F\in\{G:\hat{F}_X^L(x)\leq G(x)\leq\hat{F}_X^H(x),\ \forall x\in\mathbb{R}^{\mathcal{I}}\}}\int(F\left(x\right)-\hat{F}_Y(x|\hat{\theta}_0))^2d\mu\left(x\right)\end{array}\right\}\\
&=\left\{\begin{array}{c}\frac{1}{V}\sum_{j=1}^V\max\{(\hat{F}_X^L\left(Z_j\right)-\hat{F}_Y(Z_j|\hat{\theta}_0))^2,(\hat{F}_X^H\left(Z_j\right)-\hat{F}_Y(Z_j|\hat{\theta}_0))^2\}\\ -\int\max\{((\hat{F}_X^L\left(x\right)-\hat{F}_Y(x|\hat{\theta}_0))^2,(\hat{F}_X^H\left(x\right)-\hat{F}_Y(x|\hat{\theta}_0))^2\}d\mu\left(x\right)\end{array}\right\},
\end{aligned}$$

Assumption 5.b implies that, conditional on $\mathcal{X}_n$ and $\mathcal{Y}_{\theta,m}$, $\{\max\{(\hat{F}_X^L\left(Z_j\right)-\hat{F}_Y(Z_j|\hat{\theta}_0))^2,(\hat{F}_X^H\left(Z_j\right)-\hat{F}_Y(Z_j|\hat{\theta}_0))^2\}\}_{j=1}^V$ is an i.i.d. sample and, thus, the weak law of large numbers implies that $\{\hat{\Psi}_H(\hat{F}_X^L,\hat{F}_X^H)-\tilde{\Psi}_H(\hat{F}_X^L,\hat{F}_X^H)|\mathcal{X}_n,\mathcal{Y}_{\theta,m}\}=o_p\left(1\right)$ as $V\to\infty$. Therefore, for any $\varepsilon>0$:

$$\begin{aligned}
&\lim_{n\to\infty}\lim_{m\to\infty}\lim_{V\to\infty}P(|\hat{\Psi}_H(\hat{F}_X^L,\hat{F}_X^H)-\tilde{\Psi}_H(\hat{F}_X^L,\hat{F}_X^H)|>\varepsilon)\\
&=\lim_{n\to\infty}\lim_{m\to\infty}\lim_{V\to\infty}\int P(|\hat{\Psi}_H(\hat{F}_X^L,\hat{F}_X^H)-\tilde{\Psi}_H(\hat{F}_X^L,\hat{F}_X^H)|>\varepsilon|\mathcal{X}_n,\mathcal{Y}_{\theta,m})dP\left(\mathcal{X}_n,\mathcal{Y}_{\theta,m}\right)\\
&=\lim_{n\to\infty}\lim_{m\to\infty}\int\lim_{V\to\infty}P(|\hat{\Psi}_H(\hat{F}_X^L,\hat{F}_X^H)-\tilde{\Psi}_H(\hat{F}_X^L,\hat{F}_X^H)|>\varepsilon|\mathcal{X}_n,\mathcal{Y}_{\theta,m})dP\left(\mathcal{X}_n,\mathcal{Y}_{\theta,m}\right)=0,
\end{aligned}$$

where we have used the previous finding and the dominated convergence theorem. □

Proof of Theorem 4.2. By Theorem 3.2 in BHHN, $t_{\hat{\theta}_0}^*\left(1-\alpha\right)=t_{\theta_0}\left(1-\alpha\right)+\gamma_n$, where $\gamma_n=o_p\left(1\right)$, as $V\to\infty$, $m\to\infty$, and $n\to\infty$, and $t_{\theta_0}\left(1-\alpha\right)$ is the $\left(1-\alpha\right)$ quantile of a non-degenerate and non-negative distribution (of a random variable denoted by V_0 in BHHN). From this, if follows that $\forall\alpha\in(0,1)$, $t_{\theta_0}\left(1-\alpha\right)\in\left(0,\infty\right)$. By definition, $D=T_H'\left(X,Y_{\theta_0}\right)\in(0,1]$. Fix $\varepsilon\equiv D/2$ and consider the following derivation:

$$
\begin{aligned}
&P(n\hat{T}'_H(X,Y_{\hat{\theta}_0}) \le t^*_{\hat{\theta}_0}(1-\alpha)) \\
&= P(n\hat{T}'_H(X,Y_{\hat{\theta}_0}) \le t_{\theta_0}(1-\alpha)+\gamma_n \cap |\gamma_n|>\varepsilon) + P(n\hat{T}'_H(X,Y_{\hat{\theta}_0}) \le t_{\theta_0}(1-\alpha)+\gamma_n \cap |\gamma_n|\le\varepsilon) \\
&\le P(|\gamma_n|>\varepsilon) + P(\hat{T}'_H(X,Y_{\hat{\theta}_0}) \le (t_{\theta_0}(1-\alpha)+\varepsilon)/n).
\end{aligned}
$$

Since $(t_{\theta_0}(1-\alpha)+\varepsilon) < \infty$, $\exists N \in \mathbb{N}$ such that $\forall n \ge N$: $(t_{\theta_0}(1-\alpha)+\varepsilon)/n - T'_H(X,Y_{\theta_0}) < -T'_H(X,Y_{\theta_0})/2 = -\varepsilon$. Thus, $\forall n \ge N$:

$$
P(n\hat{T}'_H(X,Y_{\hat{\theta}_0}) \le t^*_{\hat{\theta}_0}(1-\alpha)) \le P(|\gamma_n|>\varepsilon) + P(|\hat{T}'_H(X,Y_{\hat{\theta}_0}) - T'_H(X,Y_{\theta_0})|>\varepsilon).
$$

By $\gamma_n = o_p(1)$, as $V\to\infty$, $m\to\infty$, and $n\to\infty$ and by Lemma A.6, the right hand side of the previous equation is $o(1)$, as $V\to\infty$, $m\to\infty$, and $n\to\infty$. Thus, the required result follows from taking limits as $V\to\infty$, $m\to\infty$, and $n\to\infty$ in the previous equation. □

Proof of Theorem 4.3. The proof follows from arguments similar to those used in the proof of Theorem 4.2. □

Lemma A.7. *Assume Assumptions 1-5. Consider the following particularly simple setup. Suppose that $\mathcal{I}=\{t_0\}$, $\Theta=\{\theta_0\}$, $X(t_0)\sim N(0,1)$, and $Y_{\theta_0}(t_0)\sim N(\delta_n,1)$, where $\{\delta_n\}_{n=1}^{\infty}$ is a non-stochastic sequence that depends on the sample size. Suppose that there are only two missing data patterns: $O=\mathcal{I}$ (i.e. the function is completely observed), or $O=\varnothing$ (i.e. the function is completely unobserved). Consider the following cases:*

1. *Suppose that $\delta_n=\delta\neq 0$, $P(O=\varnothing)=0.5$, and the data are missing at random. To simplify our computations, also assume that $\mu(x)$ is degenerate at $x(t_0)=0$. In this case, $T(X,Y_{\theta_0})=D>0$ and $T'_L(X,Y_{\theta_0})=0$ (i.e. case 3 occurs), and*

$$
\lim_{n\to\infty}\lim_{m\to\infty}\lim_{V\to\infty} P\left(t^*_{\hat{\theta}_0}(1-\alpha) < n\hat{T}'_L(X,Y_{\hat{\theta}_0}) \le n\hat{T}'_H(X,Y_{\hat{\theta}_0})\right) > 0. \tag{A.8}
$$

2. *Suppose that $\delta_n=\delta\neq 0$ and $P(O=\varnothing)=1$. In this case, $T(X,Y_{\theta_0})=D>0$ and $T'_L(X,Y_{\theta_0})=0$ (i.e. case 3 occurs), and*

$$
\lim_{n\to\infty}\lim_{m\to\infty}\lim_{V\to\infty} P\left(t^*_{\hat{\theta}_0}(1-\alpha) < n\hat{T}'_L(X,Y_{\hat{\theta}_0}) \le n\hat{T}'_H(X,Y_{\hat{\theta}_0})\right) = 0. \tag{A.9}
$$

3. *Suppose that $\delta_n\neq 0$, $\delta_n\to 0$ and $P(O=\varnothing)=1$. In this case, $T(X,Y_{\theta_0})=D_n>0$ and $D_n\to 0$ (i.e., we are considering local alternative hypotheses), $T'_L(X,Y_{\theta_0})=0$, and Eq. (A.9) occurs.*

Proof. <u>Part 1:</u> Since $\mu(x)$ is degenerate at $x(t_0)=0$ and $P(O=\varnothing)=0.5$, then $F_L(x_0)=F_Y(x_0|\theta_0)=0.25$ and $F_H(x_0)=0.75$, which implies that $T'_L(X,Y_{\theta_0})=0$. On the other hand: $T(X,Y_{\theta_0})=\int(\Phi(x-\delta)-\Phi(x))^2\,d\mu(x)=D>0$.

We now consider the performance of the test. We begin by deriving the asymptotic distribution of $n\hat{T}'_L(X,Y_{\hat{\theta}_0})$. Since $\mu(x)$ is degenerate at $x(t_0)=0$, then there is no Monte Carlo integration error and, thus, V does not affect the behavior of the test statistic. By considering $m\to\infty$, the error in approximating the distribution of $Y_{\theta_0}(t_0)$ vanishes and, consequently, we can consider the value of $F_Y(x_0|\theta_0)$ to be known. Finally, we consider the behavior as $n\to\infty$. By our assumptions, $\{(X'_i,O_i)\}_{i=1}^n$ is i.i.d. and since the data are missing at random, $1\left(X_i' \leq 0\right) 1\left(O_i = \mathcal{I}\right) \sim Be\left(0.25\right)$, $\{1\left(X_i' \leq 0\right) 1\left(O_i = \mathcal{I}\right) + 1\left(O_i = \varnothing\right)\} \sim Be\left(0.75\right)$, and

$$\begin{aligned}
\hat{F}_L(x_0) &= n^{-1}\textstyle\sum_{i=1}^n 1\left(X_i' \leq 0\right) 1\left(O_i = \mathcal{I}\right), \\
\hat{F}_H(x_0) &= n^{-1}\textstyle\sum_{i=1}^n \{1\left(X_i' \leq 0\right) 1\left(O_i = \mathcal{I}\right) + 1\left(O_i = \varnothing\right)\}.
\end{aligned}$$

By the strong law of large numbers if follows that, for any $\varepsilon > 0$, $P(\liminf\{|\hat{F}_H(x_0) - F_H(x_0)| < \varepsilon\}) = 1$ (see, e.g., page 70 in Billingsley (1995)), which implies that $\liminf\{1(\hat{F}_H(x_0) \geq F_Y(x_0|\theta_0)) = 0\}$, a.s. By the Central Limit Theorem $\sqrt{n}(\hat{F}_L(x_0) - F_Y(x_0|\theta_0)) = \sqrt{n}(\hat{F}_L(x_0) - F_L(x_0)) \xrightarrow{d} \zeta$, where $\zeta \sim N(0, 0.1875)$. It then follows that:

$$\begin{aligned}
&n\hat{T}_L'(X, Y_{\hat{\theta}_0}) = n\hat{T}_L'(X, Y_{\theta_0}) \\
&= (\sqrt{n}(\hat{F}_L(x_0) - F_Y(x_0|\theta_0))^2 1(\hat{F}_L(x_0) \geq F_Y(x_0|\theta_0)) + (\sqrt{n}(\hat{F}_H(x_0) - F_Y(x_0|\theta_0))^2 1(\hat{F}_H(x_0) \geq F_Y(x_0|\theta_0)) \\
&= [\sqrt{n}(\hat{F}_L(x_0) - F_Y(x_0|\theta_0))]_+^2 + o_p\left(1\right) \xrightarrow{d} \zeta_+^2, \text{ as } V \to \infty,\ m \to \infty, \text{ and } n \to \infty.
\end{aligned}$$

We next consider the asymptotic distribution of $t^*_{\hat{\theta}_0}\left(1-\alpha\right)$. Theorem 3.2 in BHHN implies that $t^*_{\hat{\theta}_0}\left(1-\alpha\right) = t_{\theta_0}\left(1-\alpha\right) + \gamma_n$, where $t_{\theta_0}\left(1-\alpha\right) \in (0, \infty)$ and $\gamma_n = o_p\left(1\right)$, as $V \to \infty$, $m \to \infty$, and $n \to \infty$. Fix $\varepsilon > 0$ arbitrarily. By combining the previous findings:

$$\begin{aligned}
&P(t^*_{\hat{\theta}_0}(1-\alpha) < n\hat{T}_L'(X, Y_{\hat{\theta}_0}) \leq n\hat{T}_H'(X, Y_{\hat{\theta}_0})) \\
&\leq P(t_{\theta_0}(1-\alpha) + \gamma_n < n\hat{T}_L'(X, Y_{\hat{\theta}_0}) \cap \gamma_n \geq -\varepsilon) \\
&\leq P(t_{\theta_0}(1-\alpha) - \varepsilon < n\hat{T}_L'(X, Y_{\hat{\theta}_0})) \\
&\to P(\zeta_+^2 \geq (t_{\theta_0}(1-\alpha) - \varepsilon)), \text{ as } V \to \infty,\ m \to \infty, \text{ and } n \to \infty.
\end{aligned}$$

where $\zeta \sim N(0, 0.1875)$ and $t_{\theta_0}(1-\alpha) \in (0, \infty)$. This provides an asymptotic upper bound for the object of interest. By applying the same argument with $P(t^*_{\hat{\theta}_0}(1-\alpha) \geq n\hat{T}_L'(X, Y_{\hat{\theta}_0}))$, we obtain a lower bound for the object of interest. By combining these bounds and by letting $\varepsilon \downarrow 0$:

$$\lim_{n\to\infty}\lim_{m\to\infty}\lim_{V\to\infty} P(t^*_{\hat{\theta}_0}(1-\alpha) < n\hat{T}_L'(X, Y_{\hat{\theta}_0}) \leq n\hat{T}_H'(X, Y_{\hat{\theta}_0})) = P(\zeta_+^2 \geq t_{\theta_0}(1-\alpha)) > 0,$$

where the strict inequality follows from $\zeta \sim N(0, 0.1875)$ and $t_{\theta_0}(1-\alpha) < \infty$. This verifies Eq. (A.8).

Part 2: Since $P\left(O = \varnothing\right) = 1$, then $\forall x \in \mathbb{R}^I$, $F_L\left(x\right) = 0$ and $F_H\left(x\right) = 1$, and, so: $T_L'\left(X, Y_{\theta_0}\right) = 0$. On the other hand: $T\left(X, Y_{\theta_0}\right) = \int \left(\Phi\left(x - \delta\right) - \Phi\left(x\right)\right)^2 d\mu\left(x\right) = D > 0$.

We now consider the performance of the test. First, $P\left(O = \varnothing\right) = 1$ implies that the sample is composed of completely missing observations (a.s.) and, hence, $\hat{T}_L'(X, Y_{\hat{\theta}_0}) = 0$ (a.s.). Second, by the same argument as in part 1, $t^*_{\hat{\theta}_0}\left(1-\alpha\right) = t_{\theta_0}\left(1-\alpha\right) + \gamma_n$, where $t_{\theta_0}\left(1-\alpha\right) \in (0, \infty)$ and $\gamma_n = o_p\left(1\right)$, as $V \to \infty$, $m \to \infty$, and $n \to \infty$. Fix $\varepsilon = t_{\theta_0}\left(1-\alpha\right)/2 > 0$ and consider the following derivation:

$$\begin{aligned}
&P(t^*_{\hat{\theta}_0}\left(1-\alpha\right) \geq n\hat{T}_L'(X, Y_{\hat{\theta}_0})) \geq P(t^*_{\hat{\theta}_0}\left(1-\alpha\right) \geq \varepsilon \cap \hat{T}_L'(X, Y_{\hat{\theta}_0}) = 0) \\
&\geq P(t^*_{\hat{\theta}_0}\left(1-\alpha\right) - t_{\theta_0}\left(1-\alpha\right) \geq \varepsilon - t_{\theta_0}\left(1-\alpha\right)) + P(\hat{T}_L'(X, Y_{\hat{\theta}_0}) = 0) - 1 \\
&\geq P(|t^*_{\hat{\theta}_0}\left(1-\alpha\right) - t_{\theta_0}\left(1-\alpha\right)| \leq t_{\theta_0}\left(1-\alpha\right)/2) + P(\hat{T}_L'(X, Y_{\hat{\theta}_0}) = 0) - 1 \\
&\to 1 \text{ as } V \to \infty,\ m \to \infty, \text{ and } n \to \infty.
\end{aligned}$$

This result implies Eq. (A.9).

Part 3: All of the results in part 2 can also be used in this case. The only result that remains to be shown is that $T\left(X, Y_{\theta_0}\right) = D_n \to 0$, which follows from $\delta_n \to 0$ and the dominated convergence theorem. $\square$