cemmap

centre for microdata methods and practice

Permutation tests for equality of distributions of functional data

Federico A. Bugni
Joel L. Horowitz

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP18/18

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

An ESRC Research Centre

PERMUTATION TESTS FOR EQUALITY OF DISTRIBUTIONS OF FUNCTIONAL DATA

by

Federico A. Bugni
Department of Economics
Duke University
Durham, NC 27708 USA

and

Joel L. Horowitz
Department of Economics
Northwestern University
Evanston, IL 60208 USA

March 2018

Abstract

Economic data are often generated by stochastic processes that take place in continuous time, though observations may occur only at discrete times. For example, electricity and gas consumption take place in continuous time. Data generated by a continuous time stochastic process are called functional data. This paper is concerned with comparing two or more stochastic processes that generate functional data. The data may be produced by a randomized experiment in which there are multiple treatments. The paper presents a test of the hypothesis that the same stochastic process generates all the functional data. In contrast to existing methods, the test described here applies to both functional data and multiple treatments. The test is presented as a permutation test, which ensures that in a finite sample, the true and nominal probabilities of rejecting a correct null hypothesis are equal. The paper also presents the asymptotic distribution of the test statistic under alternative hypotheses. The results of Monte Carlo experiments and an application to an experiment on billing and pricing of natural gas illustrate the usefulness of the test.

Key words: Functional data, permutation test, randomized experiment, hypothesis test

JEL Listing: C12, C14

Part of this research was carried out while Joel L. Horowitz was a visitor at the Department of Economics, University College London, and the Centre for Microdata Methods and Practice. The research of Federico Bugni was supported in part by NIH Grant 40-4153-00-0-85-399 and NSF Grant SES-1729280.

PERMUTATION TESTS FOR EQUALITY OF DISTRIBUTIONS OF FUNCTIONAL DATA

1. INTRODUCTION

Economic data are often generated by stochastic processes that can be viewed as taking place in continuous time, though observations may occur only at discrete times. Examples are gas and electricity consumption by households, asset prices or returns, and wages. Data generated from a continuous time stochastic process are random functions and are called functional data. The analysis of functional data is a well-established research area in statistics that has generated a vast literature. See, for example, Hall and Hossein-Nasab (2006); Jank and Shmueli (2006); Ramsay and Silverman (2002, 2005); Yao, Müller, and Wang (2005); and the references therein.

In this paper, we are concerned with comparing two or more stochastic processes that generate functional data. These processes are produced by a randomized experiment in which there may be multiple treatments. There are one or more treatment groups and one control group. Our objective is to test the hypothesis that the same stochastic process generates the functional data in all the groups. More precisely, the null hypothesis is that the functional data (random functions) generated by the stochastic processes for the (possibly multiple) treatment groups and the control group have the same probability distribution. Our interest in this hypothesis is motivated by experiments in billing and pricing of gas and pricing of electricity that have been conducted in several countries, including the US and Ireland. In a typical experiment, households are assigned randomly to treatment and control groups. The treatment groups have one or more experimental billing or price schedules, and the control group has regular billing and pricing. Consumption of gas or electricity by households in the treatment and control groups is measured at frequent time intervals for several months. For example, in the Irish experiment on gas billing and pricing that we analyze later in this paper, consumption was measured every 30 minutes for twelve months. Gas consumption takes place in continuous time, though it is measured only at discrete times. The consumption path of a household is a random function of continuous time. The consumption paths of all households in the treatment groups (control group) are random samples of functions generated by the treatment (control) consumption processes. The hypothesis tested in this paper is that the consumption processes of the treatment and control groups are the same. The alternative hypothesis is that the treatment and control processes differ on a set of time intervals with non-zero Lebesgue measure.

If the hypothesis to be tested pertained to the distributions of finite-dimensional random variables, then testing could be carried out using the Cramér-von Mises or Kolmogorov-Smirnov two-sample tests, among others (Schilling 1986, Henze 1988) or multi-sample generalizations of these tests. But the Cramér-von Mises and Kolmogorov-Smirnov tests do not apply to random functions, which are infinite-dimensional random variables. Methods are also available for testing the hypothesis that continuous time data or, equivalently, random functions are generated by a known stochastic process or a process that is known up to a finite-dimensional parameter (Bugni, Hall, Horowitz, and Neumann 2009; Cuesta-Albertos, del Barrio, Fraiman, and Matrán 2007; Cuesta-Albertos, Fraiman and Ransford 2006; Hall and Tajvidi 2002; Kim and Wang 2006). Methods of parametric time-series analysis can also be used in this setting. However, the method described here is nonparametric. It does not assume that the stochastic processes generating the data have known parametric or semiparametric forms.

Another possibility is to carry out nonparametric tests of hypotheses of equality of specific features (e.g., moments) of the processes generated by the various treatment groups. For example, Harding and Lamarche (2014) compared moments of the distributions of electricity consumption in the treatment and control groups in a time-of-day pricing experiment. However, a test of equality of specific moments does not reveal whether the processes generated by the various groups differ in other ways. The method described in this paper facilitates such an investigation.

There are several existing methods for carrying out non-parametric two-sample distributional tests. Székely and Rizzo (2004) describe a two-sample test for data that may be high-dimensional but not functional. The test of Székely and Rizzo (2004) is not applicable to functional data and, apart from consistency, its asymptotic power properties are unknown. Schilling (1986) and Henze (1988) describe two-sample nearest neighbor tests for multivariate (not functional) data. These tests are not applicable to multiple treatments, and their asymptotic power properties are unknown. Hall and Tajvidi (2002) describe a two-sample permutation test for functional data. The test of Hall and Tajvidi (2002) applies to functional data and is an alternative to the test developed in this paper when there is a single treatment group in addition to the control group. The test of Hall and Tajvidi (2002) is not applicable to experiments with multiple treatments.

The test described in this paper is applicable to experiments with multiple treatments as well as experiments with one treatment group and a control group. This is an important advantage of our test compared to others. Experiments with multiple treatments are common in many fields (see, for example, Chong, Cohen, Field, Nakasone, and Torero (2016); Ashraf, Field, and Lee (2014); and Field, Jayachandran, Pande, and Rigol (2016), among many others). The experiment on gas billing and pricing analyzed later in this paper has multiple treatments. The tests Schiling (1986), Henze (1988), and Hall and Tajvidi (2002) are not applicable to experiments with multiple treatments.

The test described in this paper is motivated by Bugni, Hall, Horowitz, and Neumann (2009) (hereinafter BHHN), who describe a Cramér-von Mises-type test of the hypothesis that a sample of random functions was generated by a continuous time stochastic process that is known up to a finite-dimensional parameter. BHHN give a bootstrap method for estimating the test's critical value. This paper presents a Cramér-von Mises type test of the hypothesis that two or more samples of random functions were generated by the same unknown stochastic process. The alternative hypothesis is that the samples were generated by different stochastic processes. In contrast to the test of BHHN, the test presented here is implemented as a permutation test, which ensures that in a finite sample, the true and nominal probabilities of rejecting a correct null hypothesis are equal. A test based on the bootstrap does not have this property. Nor does any other test based on an asymptotic approximation. The test proposed here has non-trivial power against alternative hypotheses that differ from the null hypothesis by $O(n^{-1/2})$, where n is the number of observations in the largest sample. "Non-trivial" means that the power of the test exceeds the probability with which the test rejects a correct null hypothesis. The asymptotic local power of the permutation test is the same as it would be if the critical value of the test were based on the asymptotic distribution of the test statistic under the null hypothesis. Thus, there is no penalty in terms of asymptotic power for the permutation test's elimination of the finite-sample error in the probability of rejecting a correct null hypothesis.

Section 2 of this paper presents the proposed test statistic for the case of a single treatment group and a control group. Section 2 explains how the critical values are obtained, and describes the procedure for implementing the test. Section 3 presents the properties of two-sample version of the test under the null and alternative hypotheses. Section 4 extends the results of Sections 2 and 3 to experiments in which there are several treatment groups and a control group. Section 5 discusses methods for selecting a user-chosen measure that is used in the test. Section 6 applies the test to data from a multiple-treatment experiment on the pricing of gas. Section 7 presents the results of simulation studies of the test's behavior using a design that mimics the experiment analyzed in Section 6. Section 8 presents concluding comments. The proofs of theorems are in the appendix, which is Section 9.

2. DESCRIPTION OF THE TEST IN THE SINGLE TREATMENT CASE

2.1 The Test Statistic

Let $\mathcal{I} = [0,T]$ be a closed interval, and let $L_2(\mathcal{I})$ denote the set of real-valued, square-integrable functions on $\mathcal{I}$. We consider two stochastic processes (or random functions) on $\mathcal{I}$: $X(t) \in L_2(\mathcal{I})$ and $Y(t) \in L_2(\mathcal{I})$. For example, $X(t)$ may correspond to the treatment group and $Y(t)$ to the control group. In the gas pricing experiment, $\mathcal{I}$ is the period of time over which gas consumption is observed. $X(t)$ and $Y(t)$, respectively, are gas consumption at time t by individuals in the treatment and control groups. Let F_X and F_Y respectively be the probability distribution functions of $X(t)$ and $Y(t)$. That is, for any non-stochastic function $z \in L_2(\mathcal{I})$,

(2.1) $$F_X(z) = P[X(t) \le z(t) \text{ for all } t \in \mathcal{I}]$$

and

(2.2) $F_Y(z) = P[Y(t) \le z(t) \text{ for all } t \in \mathcal{I}]$.

The null hypothesis to be tested is

(2.3) $H_0: F_X(z) = F_Y(z)$

for all $z \in L_2(\mathcal{I})$. The alternative hypothesis is

(2.4) $H_1: P_\mu[F_X(Z) \ne F_Y(Z)] > 0$,

where μ is a probability measure on $L_2(\mathcal{I})$ and Z is a random function with probability distribution μ. H_1 is equivalent to the hypothesis that $F_X(z) \ne F_Y(z)$ on a set of z's with non-zero μ measure. The measure μ is analogous to a weight function in tests of the Cramer-von Mises type, among others. Like the weight function in other tests, μ in the test presented here influences the directions of departure from H_0 in which the test has high power. The choice of μ is discussed in Section 4.

Now define

$$\tau = \int [F_X(z) - F_Y(z)]^2 d\mu(z).$$

Then $\tau = 0$ under H_0 and $\tau > 0$ under H_1. A test of H_0 can be based on a sample analog of τ that is scaled to have a non-degenerate limiting distribution. To obtain the analog, let $\{X_i(t): i = 1,...,n\}$ and $\{Y_i(t): i = 1,...,m\}$ denote random samples (sample paths) of n and m realizations of $X(t)$ and $Y(t)$, respectively. Make

<u>Assumption 1</u>: (i) $X(t)$ and $Y(t)$ are separable, μ-measurable stochastic processes. (ii) $\{X_i(t): i = 1,...,n\}$ is an independent random sample of $X(t)$. $\{Y_i(t): i = 1,...,m\}$ is an independent random sample of $Y(t)$ and is independent of $\{X_i(t): i = 1,...,n\}$.

Also assume for the moment that $X_i(t)$ and $Y_i(t)$ are observed for all $t \in \mathcal{I}$. The more realistic setting in which $X_i(t)$ and $Y_i(t)$ are observed only at a discrete set of points $t \in \mathcal{I}$ is treated in the next paragraph.[1] Define the empirical distribution functions

(2.5) $\hat{F}_X(z) = n^{-1} \sum_{i=1}^{n} I[X_i(t) \le z(t) \text{ for all } t \in \mathcal{I}]$

[1] $X(t)$ and $Y(t)$ are stochastic processes, such as gas consumption, that take place in continuous time but can be observed (measured) only at discrete time points, say $t_1, t_2, ..., t_J$. A test of a hypothesis about the discrete-time processes $X(t_j)$ and $Y(t_j)$ ($j = 1,...,J$) is an approximation to a test of about the continuous time processes $X(t)$ and $Y(t)$ ($t \in \mathcal{I}$). The power of a test of a hypothesis about the discrete-time processes may decrease as the number of time points J increases. Therefore, we develop a test that has desirable properties in the continuous time setting but can be used with discrete time.

and

(2.6) $\hat{F}_Y(z) = m^{-1}\sum_{i=1}^{m} I[Y_i(t) \le z(t) \text{ for all } t \in \mathcal{I}]$.

Define the test statistic

(2.7) $\tau_{nm} = (n+m)\int [\hat{F}_X(z) - \hat{F}_Y(z)]^2 d\mu(z)$.

H_0 is rejected if τ_{nm} is larger than can be explained by random sampling error. The integral in (2.7) may not have a closed analytic form. In that case, τ_{mn} can be replaced with a simulation estimator that is obtained by randomly sampling μ. Let $\{Z_\ell: \ell = 1,..., L\}$ be such a sample. Then the simulation version of τ_{nm} is

(2.8) $\hat{\tau}_{nm} = (n+m)L^{-1}\sum_{\ell=1}^{L}[\hat{F}_X(Z_\ell) - \hat{F}_Y(Z_\ell)]^2$.

Arguments like those used to prove Theorem 3.3 of BHHN can be used to show that $\hat{\tau}_{nm} \to^{a.s.} \tau_{nm}$ with respect to the probability measure μ as $L \to \infty$. However, the α-level permutation test based on $\hat{\tau}_{nm}$ rejects a correct H_0 with probability exactly α, even if L is finite. See Theorem 3.1.

Now suppose that $X_i(t)$ and $Y_i(t)$ are observed only at the discrete times $\{t_j: j = 1,..., J;\ 0 \le t_j \le T\}$. Then the empirical distribution functions $\hat{F}_X$ and $\hat{F}_Y$ are replaced by

$$\tilde{F}_X[z(t_1),..., z(t_J)] = n^{-1}\sum_{i=1}^{n} I[X_i(t_j) \le z(t_j) \text{ for all } j = 1,..., J]$$

and

$$\tilde{F}_Y[z(t_1),..., z(t_J)] = m^{-1}\sum_{i=1}^{m} I[Y_i(t_j) \le z(t_j) \text{ for all } j = 1,..., J].$$

The test statistic remains as in (2.7), except the arguments of the empirical distribution functions are the finite-dimensional vector $[z(t_1),...., z(t_J)]'$. The test statistic is

$$\tau_{nm} = (n+m)\int \{\tilde{F}_X[z(t_1),..., z(t_J)] - \tilde{F}_Y[z(t_1),..., z(t_J)]\}^2 d\mu(z).$$

Define $\zeta_j = z(t_j)$ ($j = 1,..., J$) Then τ_{nm} is equivalent to

(2.9) $\tau_{nm} = (n+m)\int [\tilde{F}_X(\zeta_1,...,\zeta_J) - \tilde{F}_Y(\zeta_1,...,\zeta_J)]^2 f_J(\zeta_1,...,\zeta_J) d\zeta_1 ... d\zeta_J$,

where f_J is the probability density function on $\mathbb{R}^J$ induced by μ.

2.2 The Critical Value and Test Procedure

Under H_0 and mild regularity conditions, the empirical process $(n+m)^{1/2}[\hat{F}_X(z)-\hat{F}_Y(z)]$ converges weakly to a Gaussian process, and $(n+m)^{1/2}[\tilde{F}_X(\zeta_1,...,\zeta_J)-\tilde{F}_Y(\zeta_1,...,\zeta_J)]$ is asymptotically normal. These results can be used to derive the asymptotic distribution of τ_{nm} under H_0 with either continuous-time or discrete-time observations of $X(t)$ and $Y(t)$. The asymptotic distribution can be used in the usual way to obtain asymptotic critical values of τ_{nm}. It is likely that the bootstrap can be used to estimate the asymptotic critical values if, as usually happens, the analytic asymptotic distribution is intractable. However, asymptotic approximations can be inaccurate and misleading in finite samples. We avoid this problem by carrying out a permutation test based on τ_{nm}. The critical value of a permutation test does not depend on asymptotic approximations. The true and nominal probabilities of rejecting a correct null hypothesis with a permutation test are equal in finite samples. Moreover, the asymptotic power of the permutation test is the same as the power the test based on the asymptotic critical value. This section explains the permutation test procedure and how to obtain critical values for permutation tests based on τ_{nm}. As is explained in Section 2.1, the same results apply to the simulation version of τ_{mn}.

Let $\alpha \in (0,1)$ be the nominal level of the test. The α-level critical value is computed by evaluating the test statistic for permutations of the combined sample of $n+m$ observations of $\{X_i: i=1,...,n;\ Y_i: i=1,...,m\}$. There are $Q=(m+n)!$ ways of dividing the $(n+m)$ observations in the combined sample into one set of m observations and another of n observations. Let $q=1,...,Q$ index these divisions or permutations, and let τ_{nmq} denote the test statistic based on the q'th permutation. The α-level critical value of τ_{mn} is the $(1-\alpha)$ quantile of τ_{nmq} over $q=1,...,Q$. Denote this by $t^*_{nm}(1-\alpha)$. Then,

$$t^*_{nm}(1-\alpha)=\inf\left\{t\in\mathbb{R}:\ Q^{-1}\sum_{q=1}^{Q} I(\tau_{nmq}\le t)\ge 1-\alpha\right\}.$$

If Q is large, then $t^*_{nm}(1-\alpha)$ can be estimated with arbitrary accuracy by replacing the sums over all Q permutations of the observations with sums over a random sample of $\tilde{Q}$ permutations. The α-level test rejects a correct H_0 with probability exactly α, even if $t^*_{nm}(1-\alpha)$ is estimated by this random sampling method (Lehmann and Romano 2005, p. 636).

Among the $(n+m)!$ permutations of the data, only the $(n+m)!/(n!m!)$ combinations consisting of one group of n observations and another of m observations yield distinct values of τ_{nmq}. Therefore, the permutation test can be defined in terms of combinations of the data, rather than permutations. The critical value and properties of the test are the same, regardless of whether τ_{nmq} is defined using permutations or combinations.

To carry out the permutation test based on τ_{nm}, define

$$\phi_{nm} = \begin{cases} 1 \text{ if } \tau_{nm} > t^*_{nm}(1-\alpha) \\ a \text{ if } \tau_{nm} = t^*_{nm}(1-\alpha) \\ 0 \text{ if } \tau_{nm} < t^*_{nm}(1-\alpha) \end{cases}$$

where

$$a = \frac{Q\alpha - Q^+}{Q^0},$$

$$Q^+ = \sum_{q=1}^{Q} I[\tau_{nmq} > t^*_{nm}(1-\alpha)],$$

and

$$Q^0 = \sum_{q=1}^{Q} I[\tau_{nmq} = t^*_{nm}(1-\alpha)].$$

The permutation test based on τ_{nm} rejects H_0 with probability ϕ_{nm}. That is, the test rejects H_0 if $\phi_{nm} = 1$ and rejects H_0 with probability a if $\phi_{nm} = a$. The outcome of the permutation test is random if $\tau_{nm} = t^*_{nm}(1-\alpha)$. The test rejects a correct H_0 with probability exactly α. A non-stochastic level α test can be obtained by replacing a above with 0.

3. PROPERTIES OF THE TEST IN THE SINGLE TREATMENT CASE

3.1 Finite Sample Properties under H_0

The proposed test is an example of a randomization test. Lehmann and Romano (2005, Ch. 15) provide a general discussion of randomization tests. Let $\boldsymbol{G}_{nm}$ denote the group of $Q = (m+n)!$ permutations of the $m+n$ observations $\{X_i: i = 1,...,n;\ Y_i: i = 1,...,m\}$ that produce one set of n observations and another of m observations. Let $(\mathcal{X}_n, \mathcal{Y}_m) = \{X_i: i = 1,...,n;\ Y_i: i = 1,...,m\}$ denote the original sample and $(\mathcal{X}_{nq}, \mathcal{Y}_{mq})$ denote the q'th permutation. Then

$$(\mathcal{X}_{nq}, \mathcal{Y}_{mq}) = g(\mathcal{X}_n, \mathcal{Y}_m)$$

for some function $g \in \boldsymbol{G}_{nm}$.

The following theorem gives the finite-sample behavior of τ_{nm} under H_0 with the critical value $t^*_{nm}(1-\alpha)$.

Theorem 3.1: Let assumption 1 hold. For any distribution P that satisfies H_0 and any $\alpha \in (0,1)$,

$$E_P(\phi_{nm}) = \alpha . \quad \blacksquare$$

Theorem 3.1 implies that the true and nominal rejection probabilities of the tests proposed in this paper are equal regardless of:

1. The measure μ or probability density function f_J that is used to define the test statistic.
2. Whether $X_i(t)$ and $Y_i(t)$ are observed in continuous time or only at discrete points in time.
3. Whether the integrals in (2.7) and (2.9) are calculated in closed form or estimated by simulation as in (2.8).
4. Whether t^*_{nm} is computed using all Q possible permutations of the data or only an independent random sample of $\tilde{Q} < Q$ permutations.

3.2 Asymptotic Properties under H_1

This section presents asymptotic properties under H_1 of the permutation test based on τ_{nm}. These include consistency and power under local alternative hypotheses. Define the randomization distribution function of τ_{nm} as

$$\text{(3.1)} \qquad \hat{R}_{nm}(t) = Q^{-1}\sum_{q=1}^{Q} I(\tau_{nmq} \le t) .$$

The critical value $t^*_{nm}(1-\alpha)$ is the $1-\alpha$ quantile of this distribution.

The following theorem is a extension of a result of Hoeffding (1952). See, also, Lehmann and Romano (2005, Theorem 15.2.3). In contrast to Hoeffding (1952), our theorem below does not require continuity of the limiting distribution of τ_{nm}. The theorem is the starting point for investigating the asymptotic behavior of the randomization test under H_1.

Theorem 3.2: Define $\mathcal{W}_{nm} = (\mathcal{X}_n, \mathcal{Y}_m)$, and let P_{nm} denote the probability distribution of $\mathcal{W}_{nm}$. Let G_{nm} and G'_{nm} be random variables that are uniformly distributed on $\boldsymbol{G}_{nm}$ independently of $\mathcal{W}_{nm}$ and each other. Let $\tau_{nm}(G_{nm}\mathcal{W}_{nm})$ denote the test statistic τ_{nm} evaluated using the transformed observations $G_{nm}\mathcal{W}_{nm}$. Suppose that under the sequence of probability measures $\{P_{nm} : n, m = 1,...,\infty\}$ and as $n, m \to \infty$, (3.2) $[\tau_{nm}(G_{nm}\mathcal{W}_{nm}), \tau_{nm}(G'_{nm}\mathcal{W}_{nm})] \to^d (\tau, \tau')$,

where τ and τ' are independently and identically distributed random variables with cumulative distribution function $R(\cdot)$. Define

$$r(1-\alpha) = \inf\{t \in \mathbb{R}: R(t) \geq 1-\alpha\}.$$

Then,

1. As $n, m \to \infty$, $\hat{R}_{nm}(t) \to^p R(t)$ for every t that is a continuity point of R.
2. If $R(t)$ is continuous and strictly increasing at $t = r(1-\alpha)$, then

$$t^*_{nm}(1-\alpha) \to^p r(1-\alpha)$$

as $n, m \to \infty$.

3. Let $\tau_{nm} \to^d Z$ as $n, m \to \infty$, where Z is a random variable with cumulative distribution function $\mathcal{J}$. Then

(a) $$\lim_{s \to r(1-\alpha)^-} \mathcal{J}(s) \leq \liminf_{n,m\to\infty} P_{nm}[\tau_{nm} \leq t^*_{nm}(1-\alpha)]$$

$$\leq \limsup_{n,m\to\infty} P_{nm}[\tau_{nm} \leq t^*_{nm}(1-\alpha)] \leq \mathcal{J}[r(1-\alpha)]$$

(b) If $\mathcal{J}(t)$ is continuous at $t = r(1-\alpha)$, then

$$\lim_{n,m\to\infty} P_{nm}[\tau_{nm} \leq t^*_{nm}(1-\alpha)] = \mathcal{J}[r(1-\alpha)].$$ ■

The main result of Theorem 3.2 is part 3, which describes the asymptotic behavior of the permutation test based on τ_{nm}. Parts 1 and 2 are intermediate results. Under H_0, part 3 is a straightforward consequence of Lemma 9.1 in the appendix and Theorem 3.1.

Using part 3 to calculate the power of the test against a specific alternative requires showing that condition (3.2) holds and determining $\mathcal{J}$. Lemma 9.2 in the appendix derives the distribution R in the statement of Theorem 3.2 and shows that (3.2) holds under the null hypothesis or a sequence of local alternative hypotheses.[2] Theorem 3.3 below establishes consistency of the permutation test against fixed alternatives. Theorems 3.4 and 3.5 derive $\mathcal{J}$ under a sequence of local alternative hypotheses and obtain the permutation test's asymptotic local power. To obtain the asymptotic properties of the permutation test under alternative hypotheses, we make

Assumption 2: As $n \to \infty$, $m = m(n) \to \infty$ and $m/n \to \lambda$ for some finite $\lambda > 0$.

---

[2] Lemma 9.2 is a central intermediate result in this paper. A noteworthy aspect of the proof of Lemma 9.2 is that it obtains the asymptotic distribution of a function of the empirical processes $n^{1/2}(\hat{F}_X - F_X)$ and $n^{1/2}(\hat{F}_Y - F_Y)$ without an assumption of stochastic equicontinuity.

If $X(t)$ and $Y(t)$ are observed at a fixed, finite set of points $t_1,...,t_J$, let $\zeta = (\zeta_1,....,\zeta_J)'$ be a $J \times 1$ vector. Define the cumulative distribution functions

$$F_X(\zeta) = P[X(t_j) \le \zeta_j \text{ for all } j = 1,..., J]$$

and

$$F_Y(\zeta) = P[Y(t_j) \le \zeta_j \text{ for all } j = 1,..., J].$$

Let μ_J be the measure induced on $\mathbb{R}^J$ by μ. If $X(t)$ and $Y(t)$ are observed in continuous time, define the cumulative distribution functions $F_X(z)$ and $F_Y(z)$ as in (2.1) and (2.2). Let μ be the measure in (2.4), and let $z(t) \in L_2(\mu)$ be a function.

The following theorem establishes consistency of the permutation test against a fixed alternative when $X(t)$ and $Y(t)$ are observed at the discrete times $(t_1,...,t_J)$ or in continuous time.

Theorem 3.3: Let assumptions 1 and 2 hold. If

$$\int [F_X(z) - F_Y(z)]^2 d\mu(z) > 0,$$

then

$$\lim_{n\to\infty} P[\tau_{nm} > t^*_{nm}(1-\alpha)] = 1. \quad \blacksquare$$

We now consider the asymptotic local power of the permutation test when $X(t)$ and $Y(t)$ are observed at a the finite set of points $(t_1,...,t_J)$. Let $\tilde{\Upsilon}(\zeta)$ be a Gaussian process indexed by $\zeta \in \mathbb{R}^J$ that has mean zero and covariance function

$$\text{cov}[\tilde{\Upsilon}(\zeta), \tilde{\Upsilon}(\tilde{\zeta})] = [(1+\lambda)^2 / \lambda]\{F_Y[\min(\zeta, \tilde{\zeta})] - F_Y(\zeta) F_Y(\tilde{\zeta})\}, \tag{3.3}$$

where $\min(\zeta, \tilde{\zeta})$ is the $J \times 1$ vector whose j'th component ($j = 1,..., J$) is $\min(\zeta_j, \tilde{\zeta}_j)$. Define a sequence of local alternatives by

$$F_{nX}(\zeta) = F_Y(\zeta) + (n+m)^{-1/2} \tilde{D}(\zeta)$$

for every $\zeta \in \mathbb{R}^J$ and some function $\tilde{D}$ such that

$$\int \tilde{D}(\zeta)^2 d\mu_J < \infty.$$

F_X is now indexed by the sample size n because, under a sequence of local alternatives, F_X changes as n increases. F_Y can also be indexed by m. We do not index F_Y this way because doing so adds complexity to the notation without changing the result. Define $r(1-\alpha)$ as the $1-\alpha$ quantile of the distribution of the random variable

$$\int [\tilde{\Upsilon}(\zeta)]^2 d\mu_J.$$

The following theorem gives the asymptotic power of the permutation test against sequences of local alternatives when $X(t)$ and $Y(t)$ are observed at a finite set of points.

<u>Theorem 3.4</u>: Let assumptions 1 and 2 hold. Then,

$$P\left\{\int[\tilde{\Upsilon}(\zeta)+\tilde{D}(\zeta)]^2 d\mu_J > r(1-\alpha)\right\} \le \liminf_{n\to\infty} P[\tau_{nm} > t^*_{nm}(1-\alpha)]$$

$$\le \limsup_{n\to\infty} P[\tau_{nm} > t^*_{nm}(1-\alpha)]$$

$$\le \lim_{\delta\to 0^+} P\left\{\int[\tilde{\Upsilon}(\zeta)+\tilde{D}(\zeta)]^2 d\mu_J > r(1-\alpha)-\delta\right\}. \blacksquare$$

It follows from Theorem 3.4 that the α-level permutation test based on τ_{nm} has asymptotic local power exceeding α whenever $\int[\tilde{D}(\zeta)]^2 d\mu_J > 0$.

We now consider the asymptotic local power of the permutation test when $X(t)$ and $Y(t)$ are observed in continuous time. Define $\Upsilon(z)$ as a Gaussian process indexed by $z \in L_2(\mu)$ with mean zero and covariance function

(3.4) $\quad \mathrm{cov}[\Upsilon(z),\Upsilon(\tilde{z})] = [(1+\lambda)^2/\lambda]\{F_Y[\min(z,\tilde{z})] - F_Y(z)F_Y(\tilde{z})\}$,

where $\min(z,\tilde{z})$ is the function of $t \in \mathcal{I}$ defined by $\min(z,\tilde{z}) = \{\min[z(t),\tilde{z}(t)]:\ t\in\mathcal{I}\}$. Define a sequence of local alternatives by

$$F_{nX}(z) = F_Y(z) + (n+m)^{-1/2} D(z)$$

for every $z \in L_2(\mu)$ and some function D such that

$$\int D(z)^2 d\mu < \infty .$$

As in the discrete case, F_X is indexed by n because, under a sequence of local alternatives, F_X changes as n increases. Define $r(1-\alpha)$ as the $1-\alpha$ quantile of the distribution of the random variable

$$\int \Upsilon(z)^2 d\mu .$$

The following theorem gives the asymptotic power of the permutation test against sequences of local alternatives when $X(t)$ and $Y(t)$ are observed in continuous time.

<u>Theorem 3.5</u>: Let assumptions, 1 and 2 hold. Then,

$$P\left\{\int[\Upsilon(z)+D(z)]^2 d\mu > r(1-\alpha)\right\} \le \liminf_{n\to\infty} P[\tau_{nm} > t_{nm}^*(1-\alpha)]$$

$$\le \limsup_{n\to\infty} P[\tau_{nm} > t_{nm}^*(1-\alpha)]$$

$$\le \lim_{\delta\to 0^+} P\left\{\int[\Upsilon(z)+D(z)]^2 d\mu > r(1-\alpha)-\delta\right\}. \blacksquare$$

As in the discrete-t case, the α-level permutation test based on τ_{nm} for continuous t has asymptotic local power exceeding α whenever $\int[\tilde{D}(\zeta)|^2 \, d\mu > 0$.

It follows from Theorems 3.2(2), 3.4, and 3.5 that if the limiting distribution of τ_{nm} is continuous, then the permutation test statistic and a non-permutation test based on τ_{nm} in (2.7) have the same asymptotic distribution under local alternative hypotheses.

4. EXTENSION TO MULTIPLE TREATMENTS

This section outlines the extension of the results of Sections 2 and 3 to the case in which there are two or more treatment groups and a single control group. We assume that the outcomes of all treatment groups are continuously observed. As in the previous sections, results for discretely observed outcomes can be obtained by replacing the measure μ for continuously observed outcomes with a measure that concentrates on the observed times $\{t_j : j = 1,...,J\}$. Let $s = 0,1,...,S$ index treatment groups with the control group labelled $s = 0$. Let $X_s(t)$ denote the outcome process in treatment group s. For each $s = 0,...,S$ define the cumulative distribution function

$$F_s(z) = P[X_s(t) \le z(t) \text{ for all } t \in \mathcal{I}].$$

The null hypothesis is

$$H_0: \ F_s = F_0 \ \text{ for all } s = 1,...,S.$$

The alternative hypothesis is

$$H_1: \ P_\mu[F_s(Z) \ne F_0(Z) \text{ for some } s = 1,...,S] > 0.$$

Let $\{X_{is}(t): i = 1,...,n_s\}$ denote a random sample (sample paths) of n_s realizations of $X_s(t)$. Define $n = \sum_{s=0}^{S} n_s$. The following assumptions extend assumptions 1 and 2 to the case of multiple treatments.

Assumption 1′: (i) $X_s(t)$ ($s = 0,...,S$) is a separable, μ-measurable stochastic process. (ii) $\{X_{is}(t): i = 1,...,n_s\}$ is an independent random sample of $X_s(t)$.

Assumption 2′: For each s there is a constant $\pi_s > 0$ such that $n_s / n \to \pi_s$ as $n \to \infty$.

For each $s = 0,...,S$ define the empirical distribution function

$$\hat{F}_s(z) = n_s^{-1}\sum_{i=1}^{n} I[X_{is}(t) \le z(t) \text{ for all } t \in \mathcal{I}].$$

Let μ be the measure defined in Section 2.1, and define $\boldsymbol{n} = (n_0, n_1,..., n_S)'$. The test statistic is

$$\tau_{\boldsymbol{n}} = \sum_{s=1}^{S}(n_0 + n_s)\int [\hat{F}_0(z) - \hat{F}_s(z)]^2 d\mu(z).$$

The multiple-treatment test is implemented by permuting the observed sample paths so that there are n_s permuted observations in treatment group s. Let $\tau_{\boldsymbol{n}q}$ denote the statistic obtained from permutation q. The critical value of the multiple-treatment test statistic $\tau_{\boldsymbol{n}}$ is obtained using the method described in Section 2.2 with τ_{nmq} replaced by $\tau_{\boldsymbol{n}q}$. Denote the α-level critical value by $t_{\boldsymbol{n}}^*(1-\alpha)$. As in the single-treatment case, the α-level multiple-treatment test rejects a correct H_0 with probability exactly α.

To obtain the multiple-treatment analogs of Theorems 3.3 and 3.5, define a sequence of local alternative hypotheses by

$$F_{ns}(z) = F_0(z) + n^{-1/2}D_s(z)$$

for each $s = 1,...,S$, every $z \in L_2(\mu)$ and functions D_s such that

$$\int D_s(z)^2 d\mu(z) < \infty.$$

For each $s = 1,...,S$, define $\Upsilon_s(z)$ as the Gaussian processes with means of zero, covariance functions

$$\text{cov}[\Upsilon_s(z), \Upsilon_s(\tilde{z})] = (\pi_s^{-1} + \pi_0^{-1})\{F_0[\min(z, \tilde{z})] - F_0(z)F_0(\tilde{z})\}$$

and cross-covariance functions

$$\text{cov}[\Upsilon_s(z), \Upsilon_{\tilde{s}}(\tilde{z})] = \pi_0^{-1}\{F_0[\min(z, \tilde{z})] - F_0(z)F_0(\tilde{z})\}; \quad s \ne \tilde{s},$$

where $\min(z, \tilde{z})$ is as defined in (3.4). Define $r(1-\alpha)$ as the $1-\alpha$ quantile of the distribution of the random variable

$$\int \sum_{s=1}^{S} [\Upsilon_s(z)]^2 d\mu(z).$$

The multiple-treatment analog of Theorem 3.3 is:

Theorem 4.1: Let assumptions 1′ and 2′ hold. If

$$\int \sum_{s=1}^{S} [F_s(z) - F_0(z)]^2 \, d\mu(z) > 0 \,,$$

then

$$\lim_{n\to\infty} P[\tau_{\boldsymbol{n}} > t^*_{\boldsymbol{n}}(1-\alpha)] = 1 \,. \quad \blacksquare$$

The multiple-treatment analog of Theorem 3.5 is:

Theorem 4.2: Let assumptions 1′ and 2′ hold. Then,

$$P\left\{ \int \sum_{s=1}^{S} [\Upsilon_s(z) + D_s(z)]^2 d\mu(z) > r(1-\alpha) \right\} \le \liminf_{n\to\infty} P[\tau_{\boldsymbol{n}} > t^*_{\boldsymbol{n}}(1-\alpha)]$$

$$\le \limsup_{n\to\infty} P[\tau_{\boldsymbol{n}} > t^*_{\boldsymbol{n}}(1-\alpha)]$$

$$\le \lim_{\delta\to 0^+} P\left\{ \int \sum_{s=1}^{S} [\Upsilon_s(z) + D_s(z)]^2 d\mu(z) > r(1-\alpha) - \delta \right\}. \quad \blacksquare$$

5. THE MEASURE μ

As was stated in Section 2.1, the measure μ influences the directions of departure from H_0 in which the test presented here has high power. This section presents informal suggestions about how μ can be constructed. We emphasize that regardless of the choice of μ, the probability that the α-level permutation test rejects a correct null hypothesis is exactly α. A more formal approach to constructing μ is outlined at the end of this section.

To obtain a flexible class of measures, let $\{\psi_k : i = 1,2,...\}$ be a complete, orthonormal basis for $L_2[\mathcal{I}]$. For example, we use a basis of trigonometric functions in Sections 6 and 7. Let μ be the probability measure generated by the random function

$$(5.1) \qquad Z(t) = \sum_{k=1}^{\infty} b_k \psi_k(t) \,,$$

where the Fourier coefficients $\{b_k\}$ are random variables satisfying

$$(5.2) \qquad \sum_{k=1}^{\infty} b_k^2 < \infty$$

with probability 1. Sample paths $Z_j(t)$ are generated randomly by sampling the b_k's randomly from some distribution such that (5.2) holds with probability 1. The distribution of the b_k's implies the measure μ. Therefore, μ can be specified by specifying the distribution of the b_k 's and the basis functions $\{\psi_k\}$, which ensures that μ is a probability distribution on $L_2[0,1]$. The test statistic can be computed using (2.8) by truncating the infinite sum in (5.1) at some integer K, randomly sampling the b_k 's and computing $Z_i(t)$'s as

$$Z_i(t) = \sum_{k=1}^{K} b_{ki}\psi_k(t),$$

where b_{ki} is the i 'th realization of the random variable b_k.

The mean of $Z(t)$ is

$$E[Z(t)] = \sum_{k=1}^{K} E(b_k)\psi_k(t).$$

An investigator who expects $|F_X[z(t)] - F_Y[z(t)]|$ to be relatively large in certain ranges of t can choose $E[Z(t)]$ to be a function, say $w(t)$, that is large in those ranges and set

$$E(b_k) = \int_0^T w(t)\psi_k(t)dt.$$

An investigator who has no such expectations might choose $w(t)$ to be a constant. Given a choice of $w(t)$ and the resulting mean Fourier coefficients $E(b_k)$, the b_k 's can be specified as

$$b_k = E(b_k) + \rho_k U_k,$$

where the U_k 's are random variables that are independently and identically distributed across values of k with $E(U_k) = 0$ and $Var(U_k) = 1$, and the ρ_k 's are non-stochastic constants satisfying

$$\sum_{k=1}^{\infty} \rho_k^2 < \infty.$$

The distributions of the U_k 's can set equal to $U[-3^{1/3}, 3^{1/3}]$ or $N(0,1)$ if the distributions of the processes $X(t)$ and $Y(t)$ have thin tails. If $X(t)$ and $Y(t)$ have heavy-tailed distributions, then one might consider taking the variables U_k to have heavy-tailed distributions such as Student-t with a low number of degrees of freedom.

A more formal approach to choosing μ is to specify an alternative hypothesis, specify the distributions of the Fourier coefficients b_k up to finitely many parameters, and choose the parameters through Monte Carlo simulation to maximize power or asymptotic local power (given in Theorems 3.4, 3.5, and 4.2) against the alternative. The computation required to implement this approach is difficult and time-consuming, because the objective function of the optimization problem is non-convex and must be evaluated through high-dimensional numerical integration. We carried out the power-optimization approach with several of the Monte Carlo designs described in Section 7 and found that it produced little increase in power over the informal choice of μ described in Section 6.

6. AN EMPIRICAL APPLICATION

This section reports the application of our test to data produced by the smart metering consumer behavior trial (CBT) for gas conducted by the Commission for Energy Regulation (CER) of Ireland. The CER is Ireland's independent regulator of electricity and natural gas. The goal of the CBT was to investigate the effects of several different billing and pricing treatments on residential customers' consumption of gas. The gas consumption of each customer in the CBT was measured every half hour by a smart meter. The CER kindly provided the data produced by the CBT and related documentation (Commission for Energy Regulation 2011).

The CBT was divided into two periods, a baseline period that took place from December 2009 through May 2010 and an experimental period that took place from June 2010 through May 2011. During the baseline period, all customers participating in the CBT were charged the standard rate for gas and were billed bimonthly in the usual way. During the experimental period, customers were assigned randomly to a control group or one of four treatment groups. Customers then received different treatments depending on their assignments. Customers in the control group continued to be charged the standard rate and billed bimonthly. Customers in the first treatment group were charged at the standard rate and billed bimonthly but also received a detailed report on their energy usage with recommendations about how to reduce consumption. Customers in the second treatment group were charged the standard rate but billed monthly instead of bimonthly. Customers in the third treatment group were charged at the standard rate and billed bimonthly but also received an in-home electronic device that displayed their instantaneous gas consumption and its cost. Customers in the fourth treatment group, like those in the third group, were billed bimonthly and received the in-home device. In addition, these customers were charged a variable rate according to the seasonal wholesale cost of procuring gas. Depending on the season, the rate these customers were charged was between 16 percent below the standard rate (in June through September 2010) and 17 percent above the standard rate (in December 2010 and January 2011).

The analysis in this section is concerned with gas consumption during the experimental period, when customers received different treatments depending on their assignment. We test the null hypothesis that the distributions of gas consumption by customers in the four treatment groups and the control group were the same in each month from June-December 2010. The data consist of observations of the gas consumption of 1492 customers at half-hour intervals. The numbers of customers in the treatment and control groups are shown in Table 1.

Figures 1-3 provide an informal illustration of the differences between the distributions of gas consumption in the five groups. Figure 1 shows average monthly gas consumption by customers in the control and four treatment groups; Figure 2 shows the average standard deviation of customers' consumption; and Figure 3 shows the average correlation coefficient of consumption in consecutive half-hour periods. It can be seen that the differences among the means and standard deviations of consumption in the different treatment groups are small, but there are larger differences among the correlation coefficients. Thus, the main effect of the experimental treatments appears to be a shift in the dependence structure of gas consumption.

We applied our test and the test of Székely and Rizzo (2004) to consumption in each of the months from June through December. As is explained in Section 1, Székely and Rizzo (2004) provide the only existing test that accommodates multiple treatments. In our test, we used a trigonometric basis in (5.1) and a truncated series expansion. Thus, (5.1) became

$$Z(t) = b_1 + \sum_{k=1}^{(K-1)/2} \sqrt{2} b_{2k} \cos[k\pi(2t-T)/T] + \sum_{k=1}^{(K-1)/2} \sqrt{2} b_{2k+1} \sin[k\pi(2t-T)/T],$$

where K is an integer and T is the number of half hours in a month. The Fourier coefficients were

$$b_1 \sim N(\mu_1, 1/K),$$

where

$$\mu_1 = \text{median}_i \max_t \{X_i(t): \ i = 1,...,1492; \ t = 1,...,T\}$$

and

$$b_k \sim N(0, 1/K); \ k > 1.$$

The parameter μ_1 is the mean of $Z(t)$ and is set near the center of the support of the data. Our test would have low power if μ_1 were outside of or too close to the boundaries of the support. We computed p-values for our test with for $K = 3,5,...,15$ and found little variation over this range. Therefore, we report only p-values for $K = 15$. The integrals in the definition of τ_n are population averages of functionals of $Z(t)$. We used $L = 4000$ draws of $Z(t)$ to approximate these integrals. Equation (2.8) shows the approximation for the single-treatment case. The approximation for multiple treatments, as in the CBT, is similar. We used 500 permutations of the data to compute critical values for our test and the test of Székely and Rizzo (2004).

The results of the tests are shown in Table 2. The first row of Table 2 shows the p-values obtained using our test, and the second row shows the p-values obtained using the test of Székely and Rizzo (2004). Our test rejects the null hypothesis of no treatment effect at the 0.05 level in July and at the 0.10 level in August. It does not reject the null hypothesis in June ($p > 0.50$) or September-

December ($p \geq 0.7$ in each month). The test of Székely and Rizzo (2004) does not reject the null hypothesis in any of the months June-December ($p > 0.26$ in each month). Our test and the test of Székely and Rizzo (2004) are permutation tests, so both have correct finite-sample sizes. Therefore, the results shown in Table 2 indicate that our test detects a treatment effect that is not detected by the test of Székely and Rizzo (2004).

7. MONTE CARLO EXPERIMENTS

This section reports the results of Monte Carlo experiments that explore the finite-sample properties of our test. The designs of the experiments are based on the empirical illustration of Section 6. We simulate observations of half-hour gas consumption during a 30-day month. Thus, $\mathcal{I} = \{1,...,T\}$ with $T = 1440$ half hours. Each simulated dataset consists of $n = 150$ individuals who are distributed evenly among a control group and two treatment groups. Thus, $s = 0,1,2$, $n_0 = n_1 = n_2 = 50$, and $n = \sum_{s=0}^{2} n_s = 150$. Each simulated dataset $\{X_{is}(t): t \in \mathcal{I};\ i = 1,...,n_s;\ s = 0,1,2\}$ was generated as follows:

1. Draw random variables $\{\xi_i(t): (i,t) \in \{1,...,n\} \times \mathcal{I}\}$ independently from the $N(0,1)$ distribution.
2. For all $i = 1,...,n_s$ and $s = 0,1,2$; set $\tilde{X}_{is}(0) = \xi_i(0)$.
3. For all $i = 1,...,n_s$; $s = 0,1,2$; and $t \in \mathcal{I}$, set $\tilde{X}_{is}(t) = \rho_s(t)\tilde{X}_{is}(t-1) + \xi_i(t)\sqrt{1-\rho_s^2(t)}$, where $\rho_s(t)$ is a parameter defined below.
4. For all $i = 1,...,n_s$; $s = 0,1,2$; and $t \in \mathcal{I}$, set $X_{is}(y) = \mu_s(t) + \sigma_s(t)\tilde{X}_{is}(t)$, where $\mu_s(t)$ and $\sigma_s(t)$ are parameters defined below.

The resulting random variables $\{X_{is}(t): t \in \mathcal{I};\ i = 1,...,n_s;\ s = 0,1,2\}$ are normally distributed with

1. $E[X_{is}(t)] = \mu_s(t)$.
2. $Var[X_{is}(t)] = \sigma_s^2(t)$.
3. $Corr[X_{is}(t), X_{is}(t-1)] = \rho_s(t)$ for all $t \in \mathcal{I}$ with $t > 1$.

In addition, $X_{i_1 s_1}(t_1)$ is independent of $X_{i_2 s_2}(t_2)$ if $i_1 \neq i_2$ or $s_1 \neq s_2$.

The specification of the experimental design is completed by defining the parameters $\mu_s(t)$, $\sigma_s(t)$, and $\rho_s(t)$. We chose the parameters of the control group ($s = 0$) to correspond to the CBT data in June 2010. For values of t corresponding the first half hour of the day ($t = 1, 49, 97, ...$) we set

$[\mu_0(t), \sigma_0(t), \rho_0(t)]$ equal to the averages of those parameters in the CBT data over the first half hours of days in June 2010. For values of t corresponding to the second half hour of each day ($t = 2, 50, 98, ...$) we set $[\mu_0(t), \sigma_0(t), \rho_0(t)]$ equal to the averages of those parameters in the CBT data over the second half hours of days in June 2010. The values of $[\mu_0(t), \sigma_0(t), \rho_0(t)]$ for the remaining half hours were set similarly. The values of $[\mu_s(t), \sigma_s(t), \rho_s(t)]$ $(s = 1, 2)$ for the two treatment groups varied according to the experiment. We did experiments with 10 different sets of values of $[\mu_s(t), \sigma_s(t), \rho_s(t)]$, which we call parameter designs. The 10 parameter designs are:

1. No treatment effect: $[\mu_s(t), \sigma_s(t), \rho_s(t)] = [\mu_0(t), \sigma_0(t), \rho_0(t)]$ for all t and $s = 1, 2$.

2. Mean shift for treatment group 1: $[\mu_1(t), \sigma_1(t), \rho_1(t)] = [\mu_0(t) + 0.05, \sigma_0(t), \rho_0(t)]$ and $[\mu_2(t), \sigma_2(t), \rho_2(t)] = [\mu_0(t), \sigma_0(t), \rho_0(t)]$.

3. Mean shift for both treatment groups: $[\mu_1(t), \sigma_1(t), \rho_1(t)] = [\mu_2(t), \sigma_2(t), \rho_2(t)] = [\mu_0(t) + 0.05, \sigma_0(t), \rho_0(t)]$.

4. Mean shift for treatment group 1 and variance shift for treatment group 2: $[\mu_1(t), \sigma_1(t), \rho_1(t)] = [\mu_0(t) + 0.05, \sigma_0(t), \rho_0(t)]$ and $[\mu_2(t), \sigma_2(t), \rho_2(t)] = [\mu_0(t), \sigma_0(t) + 0.05, \rho_0(t)]$.

5. Mean shift for treatment group 1 and correlation shift for treatment group 2: $[\mu_1(t), \sigma_1(t), \rho_1(t)] = [\mu_0(t) + 0.05, \sigma_0(t), \rho_0(t)]$ and $[\mu_2(t), \sigma_2(t), \rho_2(t)] = [\mu_0(t), \sigma_0(t), \rho_0(t) + 0.2]$.

6. Variance shift for treatment group 1: $[\mu_1(t), \sigma_1(t), \rho_1(t)] = [\mu_0(t), \sigma_0(t) + 0.05, \rho_0(t)]$ and $[\mu_2(t), \sigma_2(t), \rho_2(t)] = [\mu_0(t), \sigma_0(t), \rho_0(t)]$.

7. Variance shifts for both treatment groups: $[\mu_1(t), \sigma_1(t), \rho_1(t)] = [\mu_2(t), \sigma_2(t), \rho_2(t)] = [\mu_0(t), \sigma_0(t) + 0.05, \rho_0(t)]$.

8. Correlation shift for treatment group 1: $[\mu_1(t), \sigma_1(t), \rho_1(t)] = [\mu_0(t), \sigma_0(t), \rho_0(t) + 0.2]$ and $[\mu_2(t), \sigma_2(t), \rho_2(t)] = [\mu_0(t), \sigma_0(t), \rho_0(t)]$.

9. Correlation shift for treatment group 1 and variance shift for treatment group 2: $[\mu_1(t), \sigma_1(t), \rho_1(t)] = [\mu_0(t), \sigma_0(t), \rho_0(t) + 0.2]$ and $[\mu_2(t), \sigma_2(t), \rho_2(t)] = [\mu_0(t), \sigma_0(t) + 0.05, \rho_0(t)]$.

10. Correlation shift for both treatment groups: $[\mu_1(t), \sigma_1(t), \rho_1(t)] = [\mu_2(t), \sigma_2(t), \rho_2(t)] = [\mu_0(t), \sigma_0(t), \rho_0(t) + 0.2]$.

There were 1000 Monte Carlo replications in each experiment. Each experiment consists of computing the empirical probability that the null hypothesis of no treatment effect is rejected at the nominal 0.05 level. We compare the rejection probabilities of our test with those of the test of Székely and Rizzo (2004), which is the only existing test that applies to multiple treatments. The power of our test depends on K. Accordingly, we carried out experiments with $K = 3,5,...,15$. The highest power occurs with $K = 15$. All experiments used $L = 4000$.

The results of the experiments are shown in Table 3. The results with design 1 indicate that our test and the test of Székely and Rizzo (2004) both have empirical probabilities of rejecting a correct null hypothesis that are close to the nominal probability. With our test, it is not possible to reject at the 0.05 level the hypothesis that the empirical and nominal rejection probabilities are equal. Our test is a permutation test, so this result is expected. The test of Székely and Rizzo (2004) is more powerful than our test in parameter designs 2-4, which include a mean shift, though our test with a large K has substantial power in design 4, which has a variance shift in addition to a mean shift. In parameter designs 8-10, which include a correlation shift without a mean shift, our test is more powerful than that of Székely and Rizzo (2004). The latter test has very low power, whereas our test has substantial power. Design 5 includes mean and correlation shifts. In this design, our test with $K > 9$ is more powerful than that of Székely and Rizzo (2004). The results for the two parameter designs with variance shifts without correlation shifts are mixed. Our test is more powerful than that of Székely and Rizzo (2004) in design 7, but the test of Székely and Rizzo (2004) is slightly more powerful than ours in design 6.

We summarize the Monte Carlo results as follows. Our test and the test of Székely and Rizzo (2004) both reject a correct null hypothesis with the correct (nominal) probability. However, the two tests have different abilities to detect departures from the null hypothesis. The test of Székely and Rizzo (2004) is particularly good at detecting mean shifts but has very low power against correlation shifts. In contrast, our test has high power against correlation shifts. We believe that this explains the empirical results of Section 6, as the CBT experimental treatment changed the correlation structure of gas consumption but had little effect on the mean or variance of consumption.

8. CONCLUSIONS

Economic data are often generated by stochastic processes that take place in continuous time, though observations may occur only at discrete times. Data generated by a continuous time stochastic process are called functional data. This paper has been concerned with comparing two or more stochastic processes that generate functional data. The data may be produced by a randomized experiment in which there are multiple treatments. The paper has presented a permutation test of the hypothesis that the same stochastic process generates all the functional data. In contrast to existing methods, the test described here applies to both functional data and multiple treatments. The results of Monte Carlo experiments and an application to an experiment on billing and pricing of natural gas have illustrated the usefulness of the test.

9. MATHEMATICAL APPENDIX: PROOFS OF THEOREMS

9.1 Proofs of Theorems 3.1-3.6, and Theorems 4.1-4.2

This section provides proofs of the results stated in Sections 3-4. Section 9.2 provides auxiliary lemmas that are used in the proofs.

Define the randomization hypothesis as

Definition: Under H_0, $(\mathcal{X}_n, \mathcal{Y}_m) \sim g(\mathcal{X}_n, \mathcal{Y}_m)$ for every $g \in \boldsymbol{G}_{nm}$.

We now have

Lemma 9.1: Let assumption 1 hold. Then the randomization hypothesis holds for $(\mathcal{X}_n, \mathcal{Y}_m)$. ■

Proof: Let $\mathcal{W}_{nm} = (\mathcal{X}_n, \mathcal{Y}_m)$. For any permutation $q = 1,...,Q$, define $\mathcal{W}_{nmq} = (\mathcal{X}_{nq}, \mathcal{Y}_{mq})$. For each $g \in \boldsymbol{G}$ there is a permutation q such that $\mathcal{W}_{nmq} = g(\mathcal{W}_{nm})$. Under H_0, $\mathcal{W}_{nm}$ is an independently and identically distributed (iid) sample of size $n+m$ with cumulative distribution function (CDF) $F_X = F_Y \equiv F$. Therefore, $\mathcal{W}_{nmq}$ is an iid sample with CDF F, and $g(\mathcal{W}_{nm}) = \mathcal{W}_{nmq} \sim \mathcal{W}_{nm}$. Q.E.D.

Proof of Theorem 3.1: For any $w \in \text{supp}(\mathcal{W}_{nm})$, the α-level permutation test defined in Section 2.2 can be written

$$\phi(w) = \begin{cases} 1 \ \text{ if } \hat{T}(w) > \hat{T}^{(k)}(w) \\ a(w) \ \text{ if } \hat{T}(w) = \hat{T}^{(k)}(w) \\ 0 \ \text{ if } \hat{T}(w) < \hat{T}^{(k)}(w), \end{cases} \tag{9.1}$$

where $\hat{T}(w)$ denotes the test statistic τ_{nm} when $\mathcal{W}_{nm} = w$, $\hat{T}^{(k)}(w)$ denotes the k'th largest value of $\{\hat{T}(gw)\}_{g \in \boldsymbol{G}}$,

$$k = Q - \sup\{\gamma \in \mathbb{N}: \ \gamma \le Q\alpha\},$$

$$Q^0(w) = \sum_{g \in \boldsymbol{G}} I[\hat{T}(gw) = \hat{T}^{(k)}(w)],$$

$$Q^+(w) = \sum_{g \in \boldsymbol{G}} I[\hat{T}(gw) > \hat{T}^{(k)}(w)] ,$$

and

$$a(w) = [Q\alpha - Q^+(w)]/Q^0(w).$$

Let $\hat{T}^{(k)}(gw)$ denote the k'th largest value of $\hat{T}(gw)$. For each $g \in \boldsymbol{G}$, $\hat{T}^{(k)}(w) = \hat{T}^{(k)}(gw)$, $Q^0(gw) = Q^0(w)$, and $Q^+(gw) = Q^+(w)$. Consequently, $a(gw) = a(w)$. Moreover,

$$\phi(gw)=\begin{cases}1 \text{ if } \hat{T}(gw)>\hat{T}^{(k)}(w)\\ a(w) \text{ if } \hat{T}(gw)=\hat{T}^{(k)}(w)\\ 0 \text{ if } \hat{T}(gw)<\hat{T}^{(k)}(w)\end{cases}$$

and

$$\sum_{g\in\boldsymbol{G}}\phi(gw)=Q^{+}(w)+a(w)Q^{0}(w)$$

$$=Q^{+}(w)+\frac{Q\alpha-Q^{+}(w)}{Q^{0}(w)}Q^{0}(w)=Q\alpha.$$

Therefore, if $\mathcal{W}_{nm}\sim P$ for some distribution P supported on $\mathrm{supp}(\mathcal{W}_{nm})$, then

(9.2) $\quad Q^{-1}\sum_{g\in\boldsymbol{G}}E_P\{\phi[g(\mathcal{W}_{nm})]\}=\alpha$.

By the randomization hypothesis, $\mathcal{W}_{nm}\sim g(\mathcal{W}_{nm})$, so

(9.3) $\quad E_P[\phi(\mathcal{W}_{nm})=E_P\{\phi[g(\mathcal{W}_{nm})]\}$.

The theorem follows by combining (2), (3), and $Q=|\boldsymbol{G}|$. Q.E.D.

Proof of Theorem 3.2: Parts (a) and (b) are proved by Lehmann and Romano (2005, Theorem 15.2.3). Part (c) is similar to Lemma 5 of Andrews and Guggenberger (2010). Part (d) is a corollary of part (c). Q.E.D.

The following notation is used in Lemma 9.2, which is stated in the next paragraph. Let $\mathcal{O}$ denote the fixed subset of $\mathcal{I}=[0,T]$ on which $X(t)$ and $Y(t)$ are observed. $\mathcal{O}=\{t_1,...,t_J\}$ if $X(t)$ and $Y(t)$ are observed only at the discrete times $t_1,...,t_J$. $\mathcal{O}=[0,T]$ if $X(t)$ and $Y(t)$ are observed in continuous time. Let $F_X(z,\mathcal{O})=P[X(t)\le z(t)\ \ \forall\, t\in\mathcal{O}]$ and $F_Y(z,\mathcal{O})=P[Y(t)\le z(t)\ \ \forall\, t\in\mathcal{O}]$.

For any function $D(z)$ satisfying

$$\int D(z)^2 d\mu<\infty,$$

define

$$F_{nX}(z,\mathcal{O})=F_Y(z,\mathcal{O})+(n+m)^{1/2}D(z).$$

Let $\{\psi_k:\ k=1,2,...\}$ be a complete orthonormal basis for $L_2(\mu)$ with the properties that are specified after (9.4). Let Υ^* be the process defined at (3.3) if $\mathcal{O}=\{t_1,...,t_J\}$ and the process defined at (3.4) if $\mathcal{O}=[0,T]$. Then

$$\int \Upsilon^*(z)\{\psi_k(z)\}_{k=1}^K d\mu(z) \sim N(\mathbf{0}_{K\times 1}, \Sigma_K)$$

for any positive integer K, where Σ_K is the $K \times K$ matrix whose $(k, \tilde{k})$ component is

(9.4) $$(\Sigma_K)_{k,\tilde{k}} = [(\lambda+1)^2/\lambda]\int_z \int_{\tilde{z}} \{F_Y[\min(z,\tilde{z});\mathcal{O}] - F_Y(z;\mathcal{O})F_Y(\tilde{z};\mathcal{O})\}\psi_k(z)\psi_{\tilde{k}}(\tilde{z})d\mu(z)d\mu(\tilde{z}).$$

The basis $\{\psi_k\}$ can always be chosen so that

$$\sum_{k=1}^{\infty}(\Sigma_K)_{k,k} < \infty .$$

Define $N = n + m$, and let $\mathcal{W} = \{W_i :\ i = 1,...,N\}$ denote the combined samples of observations of X and Y with $W_i = X_i$ if $1 \le i \le n$ and $W_i = Y_i$ if $n+1 \le i \le n+m$.

Lemma 9.2: Let assumptions 1 and 2 hold. Let q_N and $\tilde{q}_N$ be two permutations of $\{1,...,N\}$ that are sample independently from the uniform distribution on $\{1,2,...,N\}$. Then

$$(\tau_{nmq_N}, \tilde{\tau}_{nmq_N}) \to^d (\tau, \tilde{\tau}),$$

where τ and $\tilde{\tau}$ are independently distributed as $\int \Upsilon^*(z)^2 d\mu(z)$.

Proof: For any permutation $q \in \{1,...,Q\}$ of $\{1,2,...,N\}$, let i_q denote the position in the permutation of observation i of $\mathcal{W}$. For any function $z(t)$ ($t \in \mathcal{O}$) define

$$H_{Nq}(z) = N^{-1/2}\sum_{i=1}^{N} U_{iq} I[W_i(t) \le z(t)\ \ \forall\, t \in \mathcal{O}],$$

where

$$U_{iq} = (N/n)I(i_q \le n) - (N/m)I(i_q > n).$$

Then

$$(\tau_{nmq_N}, \tau_{nm\tilde{q}_N}) = \left[\int H_{Nq_N}(z)^2 d\mu(z), \int H_{N\tilde{q}_N}(z)^2 d\mu(z)\right]$$

By the Cramér-Wold device, it suffices to show that

$$\alpha\tau_{Nq_N} + \beta\tau_{N\tilde{q}_N} \to^d \alpha\tau + \beta\tilde{\tau}.$$

for any constants α and β. For any positive integer K and any $q \in \{q_N, \tilde{q}_N\}$,

$$H_{Nq}(z) = \sum_{k=1}^{\infty} c_{Nqk}\psi_k(z)$$

$$= H_{NqK1}(z) + H_{NqK2}(z),$$

where

$$c_{Nqk} = \int H_{Nq}(z)\psi_k(z)d\mu(z),$$

$$H_{NqK1}(z) = \sum_{k=1}^{K} c_{Nqk}\psi_k(z),$$

and

$$H_{NqK2}(z) = \sum_{k=K+1}^{\infty} c_{Nqk}\psi_k(z).$$

Also define

$$\tau_{NqK1} = \int H_{NqK1}(z)^2 d\mu(z) = \sum_{k=1}^{K} c_{Nqk}^2$$

and

$$\tau_{NqK2} = \int H_{NqK2}(z)^2 d\mu(z) = \sum_{k=K+1}^{\infty} c_{Nqk}^2 ,$$

where the second equality in the both lines follows from orthonormality of $\{\psi_k\}$. Similarly,

(9.5) $$\Upsilon^*(z) = \sum_{k=1}^{\infty} b_k\psi_k(z) = \Upsilon_{K1}^*(z) + \Upsilon_{K2}^*(z),$$

where

(9.6) $$b_k = \int \Upsilon^*(z)\psi_k(z)d\mu(z),$$

(9.7) $$\Upsilon_{K1}^*(z) = \sum_{k=1}^{K} b_k\psi_k(z),$$

and

(9.8) $$\Upsilon_{K2}^*(z) = \sum_{k=K+1}^{\infty} b_k\psi_k(z).$$

Also define

(9.9) $$\tau_{K1} = \int \Upsilon_{K1}^*(z)^2 d\mu(z) = \sum_{k=1}^{K} b_k^2$$

and

(9.10) $$\tau_{K2} = \int \Upsilon_{K2}^*(z)^2 d\mu(z) = \sum_{k=K+1}^{\infty} b_k^2 .$$

Let $\tilde{\Upsilon}^*(z)$ be a process that is independent of but has the same distribution as $\Upsilon^*(z)$. Define $\tilde{\Upsilon}^*_{K1}(z)$, $\tilde{\Upsilon}^*_{K2}(z)$, $\tilde{b}_k$, $\tilde{\tau}_{K1}$, and $\tilde{\tau}_{K2}$ by replacing $\Upsilon^*(z)$ with $\tilde{\Upsilon}^*(z)$ in (9.5)-(9.10). To prove the theorem, it suffices to show that

(9.11) $\alpha\tau_{K1} + \beta\tilde{\tau}_{K1} \to^d \alpha\tau + \beta\tilde{\tau}$

as $K \to \infty$,

(9.12) $\alpha\tau_{Nq_N K1} + \beta\tau_{N\tilde{q}_N K1} \to^d \alpha\tau_{K1} + \beta\tilde{\tau}_{K1}$

as $N \to \infty$ for any positive integer K, and

(9.13) $\alpha(\tau_{Nq_N} - \tau_{Nq_N K1}) + \beta(\tau_{N\tilde{q}_N} - \tau_{N\tilde{q}_N K1}) = \alpha\tau_{Nq_N K2} + \beta\tau_{N\tilde{q}_N K2} \to^p 0$

as $N \to \infty$ followed by $K \to \infty$.

We begin with (9.11). It suffices to show that $\tau_{K1} \to^p \tau$ and $\tilde{\tau}_{K1} \to^p \tilde{\tau}$ as $K \to \infty$. We show that $\tau_{K1} \to^p \tau$. The same argument shows that $\tilde{\tau}_{K1} \to^p \tilde{\tau}$. Now $\tau - \tau_{K1} = \tau_{K2}$, so (9.7) follows from

$$E(\tau_{K2}) = \sum_{k=K+1}^{\infty} E(d_k^2) \to 0$$

as $K \to \infty$ because $E[\Upsilon^*(z)] \in L_2(\mu)$.

Next we show that (9.12) holds. For any positive integer K define

$$C_{NK} = \left(\{c_{Nq_N k}\}_{k=1}^K, \{c_{N\tilde{q}_N k}\}_{k=1}^K\right)$$

and

$$B_K = \left(\{b_k\}_{k=1}^K, \{\tilde{b}_k\}_{k=1}^K\right).$$

Let $\breve{\Sigma}_K$ be the $2K \times 2K$ matrix

$$\breve{\Sigma}_K = \begin{pmatrix} \Sigma_K & 0_{K\times K} \\ 0_{K\times K} & \Sigma_K \end{pmatrix},$$

where Σ_K is defined in (9.4). Part 2 of Lemma 9.4 implies that $C_{NK} \to^d N(0, \breve{\Sigma}_K) \sim B_K$ as $N \to \infty$. Result (9.12) now follows from the continuous mapping theorem.

To prove (9.13), it suffices to show $\tau_{Nq_N K2} \to^p 0$ as $N \to \infty$ followed by $K \to \infty$. The same argument shows that $\tau_{N\tilde{q}_N K2} \to^p 0$ as $N \to \infty$ followed by $K \to \infty$. By Lemma 9.3,

$$E(\tau_{Nq_NK2}) = \sum_{k=K+1}^{\infty} E(c_{Nq_Nk}^2)$$

$$= \sum_{k=K+1}^{\infty} \left\{ N^{-1/2}\left(1+\frac{m}{n}\right)\int_{z_1}\int_{z_2} D[(\min(z_1,z_2)]\psi_k(z_1)\psi_k(z_2)d\mu(z_1)d\mu(z_2) \right.$$

$$-2N^{-1/2}\left(1+\frac{m}{n}\right)\left[\int_z D(z)\psi_k(z)d\mu(z)\right]\left[\int_z F_Y(z;\mathcal{O})\psi_k(z)d\mu(z)\right]+(\Sigma_K)_{k,k}\left(2+\frac{n}{m}+\frac{m}{n}\right)\frac{\lambda}{(\lambda+1)^2}$$

$$\left. + N^{-2}\left[\int_z D(z)\psi_k(z)d\mu(z)\right]^2 \sum_{i=1}^{n}\sum_{j=1}^{n} E(U_{iq_N}U_{jq_N}) \right\}.$$

The last expression is bounded as $N \to \infty$ for every positive integer K, which implies that

$$\lim_{K\to\infty}\lim_{N\to\infty} E(\tau_{Nq_NK2}) = 0 .$$

The result (9.13) follows from this and Markov's inequality. Q.E.D.

Proof of Theorem 3.3: Arguments like those used to prove Lemma 9.2 show that $(\tau_{nmq_N}, \tau_{nm\tilde{q}_N})/N \to^p (0,0)$. Theorem 3.3 follows from this result. Q.E.D.

Proofs of Theorems 3.4 and 3.5: These theorems follow from Theorem 3.2 and Lemma 9.2. Q.E.D.

Proofs of Theorem 4.1 and 4.2: These theorems follow from arguments similar to those used to prove Theorems 3.3-3.6. Q.E.D.

9.2 Auxiliary Lemmas

Define D and N as in the paragraph preceding Lemma 9.2.

Lemma 9.3: Let assumption 2 hold, and let q_N and $\tilde{q}_N$ be two permutations of $\{1,...,N\}$ that are sampled independently from the uniform distribution on $\{1,2,...,N\}$. Let i_q denote the position of observation i ($i = 1,...,N$) in permutation q of the original sample. Define

$$U_{iq_N} = (N/n)I(i_{q_N} \le n) - (N/m)I(i_{q_N} > n) .$$

Define U_{iq_N} similarly with $\tilde{q}_N$ in place of q_N. Then as $N \to \infty$,

$$(9.14) \quad N^{-1/2}\sum_{i=1}^{n} U_{iq_N} = O_p(1) ,$$

$$(9.15) \quad N^{-1}\sum_{i=1}^{n} U_{iq_N}^2 \to^p (\lambda+1)/\lambda ,$$

(9.16) $$N^{-1}\sum_{i=n+1}^{n} U_{iq_N}^2 \to^p \lambda+1,$$

(9.17) $$N^{-1}\sum_{i=1}^{n} U_{iq_N} U_{i\tilde{q}_N} \to^P 0,$$

(9.18) $$N^{-1}\sum_{i=n+1}^{N} U_{iq_N} U_{i\tilde{q}_N} \to^P 0,$$

(9.19) $$E\left(\sum_{i=1}^{n} U_{iq_N}\right) = E\left(\sum_{i=n+1}^{N} U_{iq_N}\right) = 0,$$

(9.20) $$N^{-1}E\left|\sum_{i=1}^{n} q_{iq_N}^2\right| = 1 + n/m,$$

and

(9.21) $$N^{-1}E\left|\sum_{i=n+1}^{N} q_{iq_N}^2\right| = 1 + m/n.$$

Proof: We begin by obtaining preliminary results that are used to prove (9.14)-(9.21). The quantity

$$n^{-1}\sum_{i=1}^{n} I(iq_N \le N)$$

has a hypergeometric distribution for n draws from a population of size N that has n "successes." Therefore,

$$En^{-1}\sum_{i=1}^{n} I(q_N \le n) = n/N \to (1+\lambda)^{-1}$$

and

$$Var\left[n^{-1}\sum_{i=1}^{n} I(q_N \le n)\right] = \frac{m^2}{N^2(N-1)} \to 0$$

as $N \to \infty$. It follows that

$$E\left[n^{-1}\sum_{i=1}^{n} I(q_N \le n)\right] \to^p (1+\lambda)^{-1}.$$

By a similar argument,

$$E\left[m^{-1}\sum_{i=n+1}^{N} I(q_N \le n)\right] \to^p (1+\lambda)^{-1}.$$

In addition, Theorem 1 of (Lahiri, Chatterjee, and Matti 2007) implies that

$$N^{1/2}\left[N^{-1}\sum_{i=1}^{n} I(q_N \le n) - (n/N)^2\right] = O_p(1)\,.$$

Now consider the limiting behavior of

$$n^{-1}\sum_{i=1}^{n} I(i_{q_N} \le n) I(i_{\tilde{q}_N} \le n)$$

and

$$m^{-1}\sum_{i=n+1}^{N} I(i_{q_N} \le n) I(i_{\tilde{q}_N} \le n)\,.$$

Fix $\overline{i}_2 \in \{0,\ldots,n\}$ arbitrarily. Consider the even that out of the observations indexed by $i = 1,\ldots,n$, there are exactly $\overline{i}_2$ such that $I(i_{\tilde{q}} \le n) = 1$. By the hypergeometric distribution, the probability of this event is

$$\binom{n}{\overline{i}_2}\binom{m}{n-\overline{i}_2}\binom{N}{n}^{-1}.$$

In addition, because the permutations q_N and $\tilde{q}_N$ are independent, $I(i_{q_N} \le n) I(i_{\tilde{q}_N} \le n)$ has the hypergeometric distribution,

$$E_{\overline{i}_2}\left[n^{-1}\sum_{i=1}^{n} I(i_{q_N} \le n) I(i_{\tilde{q}_N} \le n)\right] = \frac{n\overline{i}_2}{N}\,,$$

and

$$Var_{\overline{i}_2}\left[n^{-1}\sum_{i=1}^{n} I(i_{q_N} \le n) I(i_{\tilde{q}_N} \le n)\right] = n\frac{\overline{i}_2}{N}\frac{N-\overline{i}_2}{N}\frac{m}{N-1}\,,$$

where $E_{\overline{i}_2}$ and $Var_{\overline{i}_2}$, respectively, denote the mean and variance conditional on

$$\sum_{i=1}^{n} I(i_{\tilde{q}} \le n) = \overline{i}_2.$$

The unconditional mean is

$$E\left[n^{-1}\sum_{i=1}^{n} I(i_{q_N} \le n) I(i_{q_{\tilde{N}}} \le n)\right] = (n/N)^2 \to (1+\lambda)^{-1}.$$

The unconditional variance satisfies

$$Var\left[n^{-1}\sum_{i=1}^{n} I(i_{q_N} \le n) I(i_{\tilde{q}_N} \le n)\right] \to 0\,.$$

Therefore,

$$n^{-1}\sum_{i=1}^{n} I(i_{q_N} \le n) I(i_{q_{\tilde{N}}} \le n) \to^p (1+\lambda)^{-2}.$$

By an analogous argument,

$$m^{-1}\sum_{i=n+1}^{N} I(i_{q_N} \le n) I(i_{q_{\tilde{N}}} \le n) \to^p (1+\lambda)^{-2}.$$

We now use the foregoing results to prove (9.14)-(9.21). Result (9.14) now follows from

$$N^{-1/2}\sum_{i=1}^{n} U_{iq_N} = (N/n + N/m)N^{1/2}\left[N^{-1}\sum_{i=1}^{N} I(i_{q_N} \le n) - (n/N)^2\right] = O_p(1).$$

Result (9.15) follows from

$$N^{-1}\sum_{i=1}^{N} U_{iq_N}^2 = (N/n^2)\sum_{i=1}^{n} I(i_{q_N} \le n) + (N/m)^2 \sum_{i=1}^{n} I(i_{q_N} > n) \to^p (\lambda+1)/\lambda.$$

A similar argument gives (9.16). Result (9.17) follows from

$$N^{-1}\sum_{i=1}^{N} U_{iq_N} U_{i\tilde{q}_N} = N^{-1}\sum_{i=1}^{n}\left[\frac{N}{n} I(i_{q_N} \le n) + \frac{N}{m} I(i_{q_N} > n)\right]\left[\frac{N}{n} I(i_{\tilde{q}_N} \le n) + \frac{N}{m} I(i_{\tilde{q}_N} > n)\right]$$

$$\to^p \left[\frac{1+\lambda}{\lambda^2} - 2\frac{1+\lambda}{\lambda^2} + \frac{1+\lambda}{\lambda^2}\right] = 0.$$

A similar argument yields (9.18).

To obtain (9.19) observe that

$$\sum_{i=1}^{n} U_{iq_N} = \frac{N}{n}\sum_{i=1}^{n} I(i_{q_N} \le n) - \frac{nN}{m}\sum_{i=1}^{N} I(i_{q_N} > n).$$

This and the preliminary results imply that

$$E\left(\sum_{i=1}^{n} U_{iq_N}\right) = 0.$$

This and

$$\sum_{i=1}^{N} U_{iq_N} = 0$$

imply that

$$E\left(\sum_{i+n+1}^{N} U_{iq_N}\right) = 0,$$

which establishes (9.19).

$$E\left(\sum_{i=n+1}^{n} U_{iq_N}\right) = 0\,,$$

which establishes (9.19).

To prove (9.20), observe that

$$N^{-1}\sum_{i=1}^{N} U_{iq_N}^2 = \frac{N}{n^2}\sum_{i=1}^{n} I(i_{q_N} \le n) + \frac{N}{m^2}\sum_{i=1}^{n} I(i_{q_N} > n)\,.$$

This result and the preliminary results imply that

$$E\left[N^{-1}\sum_{i=1}^{n} U_{iq_N}^2\right] = 1 + n/m\,.$$

In addition,

To prove (9.21), observe that

$$N^{-1}\sum_{i=1}^{N} U_{iq_N}^2 = \frac{N}{n^2}\sum_{i=1}^{n} I(i_{q_N} \le n) + \frac{N}{m^2}\sum_{i=1}^{n} I(i_{q_N} > n)\,.$$

This result and the preliminary results imply that

$$E\left[N^{-1}\sum_{i=1}^{n} U_{iq_N}^2\right] = 1 + n/m\,,$$

which establishes (9.21). Q.E.D.

<u>Lemma 9.4</u>: Let assumption 2 hold, q_N and $\tilde{q}_N$ be two permutations of $\{1,...,N\}$ that are sampled independently from the uniform distribution on $\{1,2,...,N\}$, $\hat{F}_{Xq_N}$ ($\hat{F}_{X\tilde{q}_N}$) be the empirical distribution function of the first n observations in permutation q_N ($\tilde{q}_N$), and $\hat{F}_{Yq_N}$ ($\hat{F}_{Y\tilde{q}_N}$) be the empirical distribution function of observations $n+1,...,N$. Then

(9.22) $$N^{1/2}\int [\hat{F}_X(z;\mathcal{O}) - \hat{F}_Y(z;\mathcal{O})]\{\psi_k(z)\}_{k=1}^{K} d\mu(z) \to^d N(\Xi,\Sigma_K)$$

and

(9.23) $$N^{1/2}\begin{bmatrix}\int [\hat{F}_{Xq_N}(z;\mathcal{O}) - \hat{F}_{Yq_N}(z;\mathcal{O})]\{\psi_k(z)\}_{k=1}^{K} d\mu(z) \\ \int [\hat{F}_{X\tilde{q}_N}(z;\mathcal{O}) - \hat{F}_{Y\tilde{q}_N}(z;\mathcal{O})]\{\psi_k(z)\}_{k=1}^{K} d\mu(z)\end{bmatrix} \to^d N\left(\mathbf{0}_{2K\times 1}, \begin{bmatrix}\Sigma_K & \mathbf{0}_{K\times K} \\ \mathbf{0}_{K\times K} & \Sigma_K\end{bmatrix}\right),$$

where

$$\Xi = \int D(z)\{\psi_k(z)\}_{k=1}^{K} d\mu(z)$$

and Σ_K is the $K \times K$ matrix defined in (9.4).

<u>Proof</u>: Let $\{W_i : i = 1,...,N\}$ denote the combined sample of observations of X and Y.

Proof of (9.23): Let i_q denote the position of observation i ($i=1,\ldots,N$) in permutation q of the original sample. Then for any permutation q,

(9.24) $$N^{1/2}[\hat{F}_{Xq}(z;\mathcal{O})-\hat{F}_{Yq}(z;\mathcal{O})]=N^{1/2}\sum_{i=1}^{n}U_{iq}I[W_i(t)\le z(t)\ \ \forall\, t\in\mathcal{O}],$$

where

$$U_{iq}=(N/n)I(i_q\le n)-(N/m)I(i_q>n).$$

Step 1: We show that

(9.25) $$N^{-1/2}\sum_{i=1}^{N}\begin{pmatrix}U_{iq_N}\\ U_{i\tilde{q}_N}\end{pmatrix}\left\{\int I[W_i(t)\le z(t)\ \ \forall\, t\in\mathcal{O}]\psi_k(z)d\mu(z)-\mu_{iNk}\right\}_{k=1}^{K}\to^d N\left(\boldsymbol{0}_{2K\times 1},\begin{bmatrix}\Sigma_K & \boldsymbol{0}_{K\times K}\\ \boldsymbol{0}_{K\times K} & \Sigma_K\end{bmatrix}\right),$$

where

$$\mu_{iNk}=I(i\le n)\int F_{nX}(z;\mathcal{O})\psi_k(z)d\mu(z)+I(i>n)\int F_Y(z;\mathcal{O})\psi_k(z)d\mu(z).$$

Let $\alpha,\beta\in\mathbb{R}$ and $\gamma\in\mathbb{R}^K$ be arbitrary constants. By the Cramér-Wold device, it suffices to show that

(9.26) $$N^{-1/2}\sum_{i=1}^{n}\Upsilon_i\to^d N(0,\sigma^2),$$

where

$$\Upsilon_i=\sum_{k=1}^{K}\gamma_k(\alpha U_{iq_N}+\beta U_{i\tilde{q}_N})\left\{\int I[W_i(t)\le z(t)\ \ \forall\, t\in\mathcal{O}]\psi_k(z)d\mu(z)-\mu_{iNk}\right\}$$

and

$$\sigma^2=(\alpha^2+\beta^2)\sum_{k,\tilde{k}=1}^{K}\gamma_k\tilde{\gamma}_{\tilde{k}}\Sigma_{K,k\tilde{k}}.$$

To establish (9.26), observe that conditional on $(q_N,\tilde{q}_N)$, $\{\Upsilon_i\}_{i=1}^{N}$ is a sequence of independent mean-0 random variables with variances

$$\sigma_{iN}^2=\sum_{k,\tilde{k}=1}^{K}\gamma_k\gamma_{\tilde{k}}(\alpha U_{iq_N}+\beta U_{i\tilde{q}_N})^2\Big\{I(i\le n)\int_z\int_{\tilde{z}}\{F_{nX}[\min(z,\tilde{z});\mathcal{O}]-F_{nX}(z;\mathcal{O})F_{nX}(\tilde{z};\mathcal{O})\}\psi_k(z)\psi_{\tilde{k}}(\tilde{z})d\mu(z)d\mu(\tilde{z})$$

$$+I(i>n)\int_z\int_{\tilde{z}}\{F_Y[\min(z,\tilde{z});\mathcal{O}]-F_Y(z;\mathcal{O})F_Y(\tilde{z};\mathcal{O})\}\psi_k(z)y_{\tilde{k}}(\tilde{z})d\mu(z)d\mu(\tilde{z})\Big\}.$$

By Lemma 9.3,

(9.27) $$\sigma_N^2\equiv N^{-1}\sum_{i=1}^{N}\sigma_{iN}^2\to\sigma^2>0$$

with probability 1 relative to the distribution of $(q_N, \tilde{q}_N)$. Moreover, for any sufficiently small $\delta > 0$ and as $N \to \infty$,

$$E(|\Upsilon_i|^{2+\delta} | q_N, \tilde{q}_N) = K^{2+\delta} \max_{k \le K} |\gamma_k|^{2+\delta} [\max(|\alpha|, |\beta|)]^{2+\delta} \left[\frac{N}{\min(n,m)}\right]^{2+\delta} \max_{\tilde{k} \le K} \left[\int |\psi_{\tilde{k}}(z)| \, d\mu(z)\right]^{2+\delta}$$

(9.28) $$\to K^{2+\delta} \max_{k \le K} |\gamma_k|^{2+\delta} [\max(|\alpha|, |\beta|)]^{2+\delta} \left[\frac{\lambda+1}{\min(1,\lambda)}\right]^{2+\delta} \max_{\tilde{k} \le K} \left[\int |\psi_{\tilde{k}}(z)| \, d\mu(z)\right]^{2+\delta} < \infty.$$

Result (9.26) and, therefore, (9.25), now follows from (9.27), (9.28), and a triangular array central limit theorem (Serfling 1980, p. 30).

Step 2 For any $k = 1, \ldots, K$

(9.29) $$N^{-1/2} \sum_{i=1}^{N} (U_{iq_N}, U_{i\tilde{q}_N}) \mu_{1Nk} = N^{-1/2} \sum_{i=1}^{n} (U_{iq_N}, U_{i\tilde{q}_N}) \int [F_{nX}(z; \mathcal{O}) - F_Y(z; \mathcal{O})] \psi_k(z) d\mu(z).$$

Lemma 9.4 implies that

$$N^{-1/2} \sum_{i=1}^{N} (U_{iq_N}, U_{i\tilde{q}_N}) = O_p(1).$$

Therefore, the right-hand side of (9.29) is $O_p(1)$. Combining this result, (9.24), and (9.25) yields (9.23).

Proof of (9.22): For any $k = 1, \ldots, K$

(9.30) $$N^{1/2}[\hat{F}_X(z; \mathcal{O}) - \hat{F}_Y(z; \mathcal{O})] = N^{1/2} \sum_{i=1}^{n} U_i \int I[W_i(t) \le z(t) \;\; \forall \, t \in \mathcal{O}] \psi_k(z) d\mu(z),$$

where

$$U_i = (N/n) I(i \le n) - (N/m) I(i > n).$$

By an argument similar to that used in the proof of (9.23),

(9.31) $$N^{1/2} \sum_{i=1}^{n} U_i \left\{\int I[W_i(t) \le z(t) \;\; \forall \, t \in \mathcal{O}] \psi_k(z) d\mu(z) - \mu_{iNk}\right\}_{k=1}^{K} \to^d N(\mathbf{0}_{K \times 1}, \Sigma).$$

Therefore,

(9.32) $$N^{-1/2} \sum_{i=1}^{N} U_i \mu_{iNk} = N^{1/2} \int [\hat{F}_X(z; \mathcal{O}) - \hat{F}_Y(z; \mathcal{O})] \psi_k(z) d\mu(z).$$

Result (9.22) follows from (9.30)-(9.32). Q.E.D.

REFERENCES

Andrews, D.W.K. and P. Guggenberger (2010). Asymptotic Size and a Problem with Subsampling with the m Out of n Bootstrap. *Econometric Theory*, 26, 426–468.

Bugni, F.A., P. Hall, J.L. Horowitz, and G.R. Neumann (2009). Goodness-of-fit tests for functional data. *The Econometrics Journal*, 12. S1-S18.

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

Cuesta-Albertos, 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, 4814-4831.

Hall, P. and M. Hosseini-Nasab (2006). On properties of functional principal components analysis. *Journal of the Royal Statistical Society*, Series B., 68, 109-126.

Hall, P. and N. Tajvidi (2002). Permutation tests for equality of distributions in high-dimensional settings. *Biometrika*, 89, 359-374.

Harding, M. and C. Lamarche (2014). Empowering consumers through smart technology: experimental evidence on the consequences of time-of-use electricity pricing. Working paper, Department of Economics, Stanford University.

Henze, N. (1988) A multivariate two-sample test based on the number of nearest neighbor type coincidences. *Annals of Statistics*, 16, 772-783.

Hoeffding, W. (1952). The large-sample power of tests based on permutations of observations. *Annals of Mathematical Statistics*, 3, 169-192.

Jank, W. and G. Shmueli (2006). Functional data analysis in electronic commerce research. *Statistical* Science, 21, 155-166.

Kim, M.S. and S. Wang (2006). Sizes of two bootstrap-based nonparametric specification tests for the drift function in continuous time models. *Computational Statistics and Data Analysis*, 50, 1793-1806.

Lahiri, S.N., A. Chatterjee, and T. Matti (2007). Normal approximation to the hypergeometric distribution in nonstandard cases and a sub-Gaussian Berry-Esséen theorem. *Journal of Statistical Planning and Inference*, 137, 3570-3590.

Lehmann, E.L. and J.P. Romano (2005). *Testing Statistical Hypotheses*, 3rd edition, New York: Springer.

Ramsay, J.O. and B.J. Silverman (2002). *Applied Functional Data Analysis: Methods and Case Studies.* New York: Springer.

Ramsay, J.O. and B.W. Silverman (2005). *Functional Data Analysis*. New York: Springer.

Romano, J.P. (1989). Bootstrap and randomization tests of some nonparametric hypotheses. *Annals of Statistics*, 17, 141-159.

Schilling, M. (1986). Multivariate two-sample tests based on nearest neighbors. *Journal of the American Statistical Association*, 81, 799-806.

Serfling, R.J. (1980). *Approximation Theorems in Mathematical Statistics*. New York: Wiley.

Székely, G. and M. Rizzo (2004). Testing for equal distributions in high-dimension. Working paper, Bowling Green State University and Ohio University.

Yao, F., H.-G. Müller, and J.-L. Wang (2005). Functional linear regression analysis for longitudinal data. *Annals of Statistics*, 33, 2873-2903.

TABLE 1: DISTRIBUTION OF CUSTOMERS AMONG GROUPS

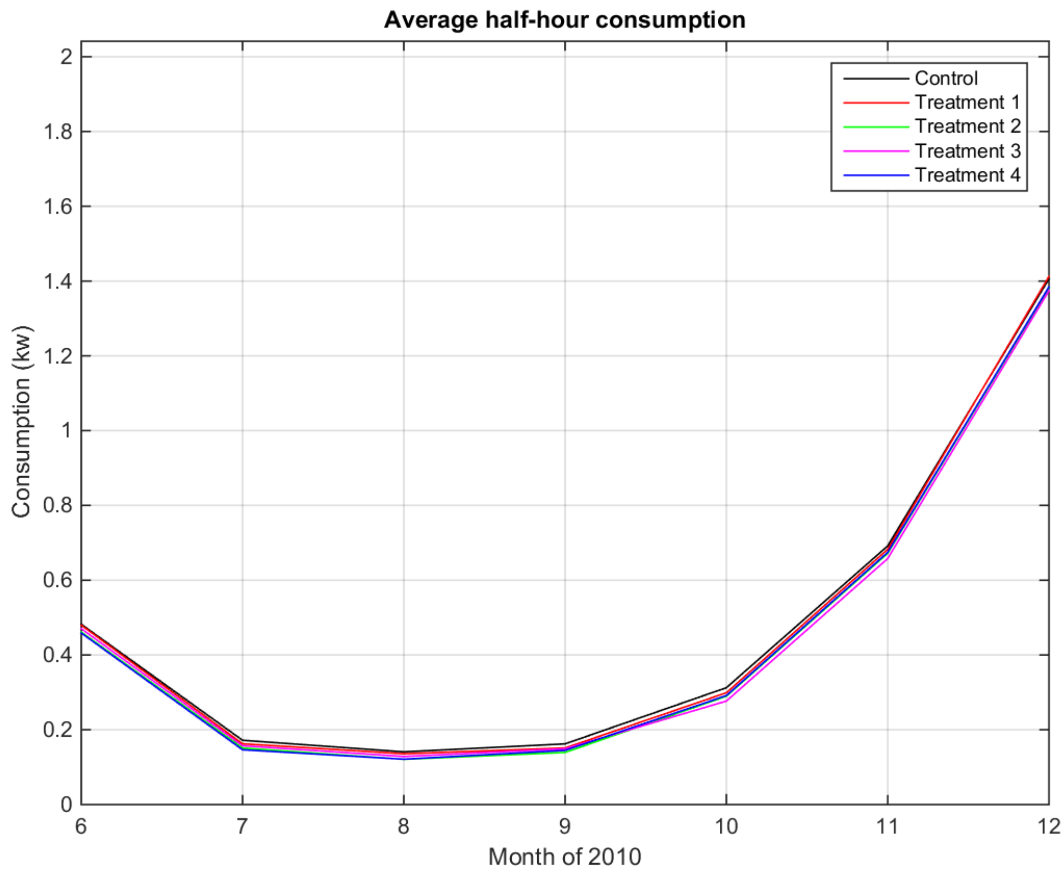
	Control	Treatment 1	Treatment 2	Treatment 3	Treatment 4	Total
Number of Customers	524	236	227	251	254	1492
Percentage of Customers	35.1	15.8	15.2	16.8	17.0	100

TABLE 2: *P*-VALUES OF THE TESTS

Test	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
Ours	0.500	0.035	0.095	0.840	0.864	0.705	0.955
Székely and Rizzo (2004)	0.455	0.385	0.570	0.265	0.824	0.975	0.990

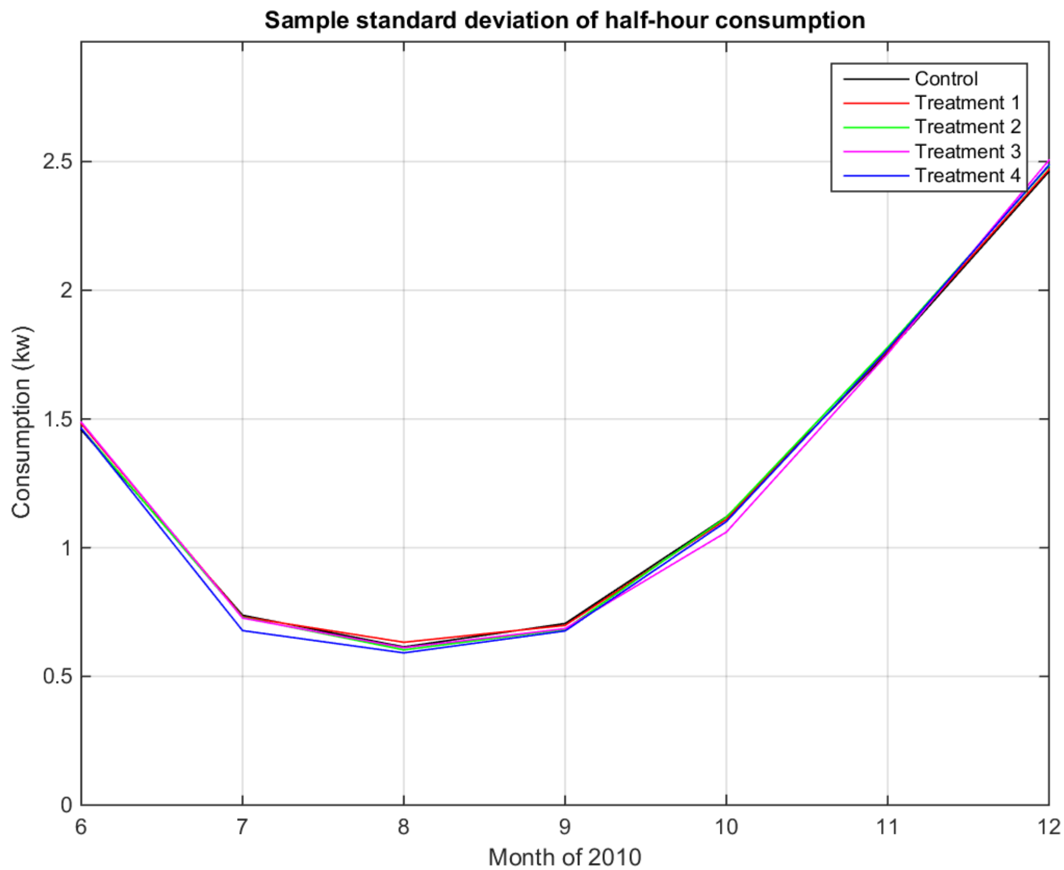
TABLE 3: EMPIRICAL REJECTION PROBABILITIES IN THE MONTE CARLO EXPERIMENTS

Test	K	Design 1	Design 2	Design 3	Design 4	Design 5	Design 6	Design 7	Design 8	Design 9	Design 10
Ours	3	0.048	0.095	0.159	0.462	0.185	0.460	0.857	0.131	0.544	0.251
	5	0.044	0.109	0.202	0.548	0.280	0.553	0.923	0.184	0.697	0.387
	7	0.047	0.128	0.244	0.597	0.402	0.625	0.956	0.251	0.803	0.510
	9	0.044	0.139	0.260	0.656	0.502	0.669	0.971	0.330	0.868	0.598
	11	0.046	0.144	0.303	0.689	0.577	0.698	0.980	0.413	0.918	0.655
	13	0.047	0.160	0.319	0.727	0.641	0.730	0.985	0.472	0.940	0.721
	15	0.046	0.166	0.365	0.741	0.702	0.752	0.986	0.541	0.950	0.774
Székely-Rizzo		0.067	0.592	0.608	0.827	0.551	0.833	0.806	0.074	0.199	0.070



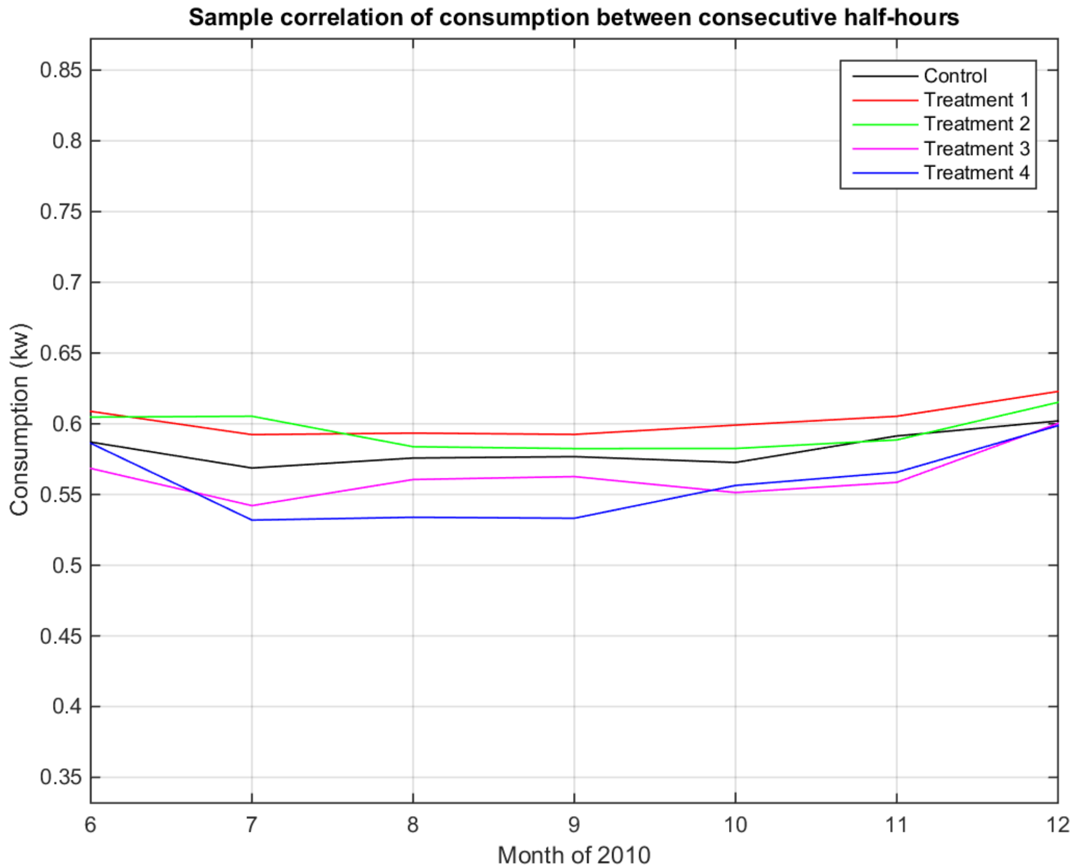
Average half-hour consumption
Control
Treatment 1
Treatment 2
Treatment 3
Treatment 4
Consumption (kw)
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
6
7
8
9
10
11
12
Month of 2010

Figure 1



Sample standard deviation of half-hour consumption
Control
Treatment 1
Treatment 2
Treatment 3
Treatment 4
Consumption (kw)
0
0.5
1
1.5
2
2.5
6
7
8
9
10
11
12
Month of 2010

Figure 2



Sample correlation of consumption between consecutive half-hours
Control
Treatment 1
Treatment 2
Treatment 3
Treatment 4
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
0.45
0.4
0.35
Consumption (kw)
6
7
8
9
10
11
12
Month of 2010

Figure 3