QUARTERLY OF APPLIED MATHEMATICS
VOLUME LIII, NUMBER 1
MARCH 1995, PAGES 105-128

STATISTICAL TESTS OF FIT IN ESTIMATION PROBLEMS FOR STRUCTURED POPULATION MODELING

By

BEN G. FITZPATRICK

North Carolina State University, Raleigh, North Carolina

Abstract. We present a statistical model of the type of data used in many applications of least squares estimation techniques to structured population problems. By viewing the data as a random sample from the size density, we may apply various statistical goodness of fit tests assess dynamic models for the data. We examine in particular classical χ^2 tests and a Cramér-von Mises test, based on empirical probability distributions.

1. Introduction. In this paper we examine some sample-theoretic questions associated with the inverse problem of fitting population data with a parameterized model. The specific type of problem that motivates this work is in the estimation of growth, mortality, and fecundity rates, g, μ, and k, respectively, in a size structured model such as the Sinko-Streifer model

$$v_t + (gv)_x = -\mu v, \qquad x_0 < x < x_1, \ t > 0,$$

with boundary conditions

$$\begin{aligned} g(t, x_1)v(t, x_1) &= 0, \\ g(t, x_0)v(t, x_0) &= \int_{x_0}^{x_1} k(t, y)v(t, y)\, dy. \end{aligned}$$

The function v represents the size density of the population at time t; that is, the number of individuals having size between a and b at time t is given by

$$N = \int_a^b v(t, x)\, dx.$$

Many inverse problems involving this model center on determining the parameters g, μ, and k from "histogram" data given by functions $H(t_i, \cdot)$ at various times $\{t_i\}_{i=1}^n$. The notation indicates the fact that the data is treated as being known as a function of x; we have referred to the data as a histogram because it is typically given as piecewise constant functions constructed from a finite collection of measurements

Received September 23, 1992.
1991 *Mathematics Subject Classification.* Primary 62E20; Secondary 92A15.
Research supported in part under grant AFOSR-91-0021.

of individuals' sizes. An example of such data, for mosquitofish observed at the University of California, Davis, in an experimental rice field is studied in [BBKW1], [BF1], and [BVWLKRC]. In this setup, the rice field is drained; then the fish fall into traps and are measured and counted. One can expect a reasonably accurate observation of the sizes of all the individuals in the population in this case.

In other experimental settings, however, one may not obtain such a complete collection of data. A typical example is that of observing sizes of fish caught in a river, such as the California Department of Fish and Game's tracking of larval striped bass in the Sacramento-San Joaquin Estuary [BBKW1], [BBKW2], [L]. One may still construct size histograms from the data, but it is no longer clear that the entire population has been counted and measured. The question that arises, then, is this: how do we compare the model and the observed histogram? The model gives numbers for the population as a whole. While it is certainly possible to estimate the number of individuals that were not observed, this approach would seem to introduce a difficult-to-quantify level of uncertainty into the process of determining model parameters.

It is our goal in the present paper to discuss a statistical approach to the modeling of the observations that incorporates histogram-type data into parameter estimation in a way that is robust with respect to total population size. By viewing the model as a probability density (when properly normalized), we may consider the data as being sampled from the model. We thus combine standard least squares techniques with nonparametric statistical methods, which may be used for model goodness-of-fit testing.

The paper is organized as follows. In Sec. 2, we set up our model problem, and we discuss our modeling of the observations. In Sec. 3, we set up the notation and assumptions needed for analysis of the statistical tests. We devote Sec. 4 to finding the asymptotic distribution of statistics related to Pearson's χ^2 test, which are based on a fixed histogram "bin size". We study in Sec. 5 questions related to asymptotics of the parameter estimates as the number of measurements increases while the bin size is decreased. These asymptotics lead us to a test of fit based on empirical distributions. The behavior of the empirical distribution and the associated tests in the subject of Sec. 6.

2. Statistical interpretation of size density data. We take the following as basic ingredients in our model problem. First, we have a parameterized model function $v(t, x; q)$, which denotes the time-dependent size density of the population. Sizes are assumed to occur only in a known range $[x_0, x_1]$. The function v is referred to as the model function because in many cases it arises as the solution of a differential equation that models the population. The parameter q is to be determined: all that is known *a priori* is that q belongs to a known set Q of possible parameters. To determine q, we observe sizes of members of the population at chosen times $t_1, \dots, t_m$. The measurements taken at time t_i are denoted by $\{\xi_{i,j}\}_{j=1}^{N_i}$. The number $\xi_{i,j}$ represents the size of the jth individual measured at time t_i.

As mentioned above, the measurements are often reported in terms of histograms $H(t_i, x)$, which are constructed as follows. We determine a "bin size" h, and we divide the interval $[x_0, x_1]$ into subintervals $J_k = [y_k, y_{k+1})$ of length h. The value of $H(t_i, \cdot)$ on the subinterval J_k is given by the number of sizes from the collection $\{\xi_{ij}\}_{j=1}^{N_i}$, divided by the interval length, h. Thus, when we integrate $H(t_i, x)$ over $x \in J_k$, we get the number of individuals which lie in that subinterval.

The next step in the usual estimation approach (see, for example, [Ba], [BBKW1], [BF1], [BK]) is to fit the model function to this histogram data, using the following least squares criterion:

$$J(q) = \sum_{i=1}^{m} \int_{x_0}^{x_1} |H(t_i, x) - v(t_i, x; q)|^2 \, dx.$$

This method presumes that H gives not only the structure but also the total count of the population.

We take here a somewhat different approach. We begin by defining the normalized size density

$$u(t, x; q) = \frac{v(t, x; q)}{\int_{x_0}^{x_1} v(t, y; q) \, dy}.$$

Our next step is to take a statistical model for the data. We assume that the histogram $H(t_i, \cdot)$ has been derived from a sample $\{\xi_{i,j}\}_{j=1}^{N_i}$, which is a random sample from a probability density function $f(t_i, \cdot)$. If the model is a good one, we should expect that $f(t, \cdot) \approx u(t, \cdot; q^*)$, for some value q^* of the parameter. That is, the normalized density u represents the likelihood of obtaining an observation of a given size. The actual observations at time t_i can then be interpreted as values taken by random variables $X_{i,1}, \dots, X_{i,N_i}$, which are independent and identically distributed having probability density $f(t_i, \cdot)$. We may then take this statistical information into account when fitting the model to the data. Moreover, we now have the possibility of using statistical tests to assess goodness of fit quantitatively.

This statistical modeling of the data suggests at least two possible approaches to deriving estimates of the parameter q, the most obvious choice being maximum likelihood. If, additionally, we assume the samples at each time are (stochastically) independent, we may write the likelihood function $\ell(q; \xi)$ of the sample $\xi = \{\xi_{i,j}\}$:

$$\ell(q; \xi) = \prod_{i=1}^{m} \prod_{j=1}^{N_i} u(t_i, \xi_{i,j}; q).$$

The maximum likelihood estimator is obtained, of course, by maximizing this expression with respect to q. Combining maximum likelihood with differential equation models for densities is an idea that has been considered in many other works (see [S] for a pharmacokinetics example).

Another approach, which is closely related to the above-mentioned least squares criterion, involves nonparametric density estimation. Here we scale the histogram $H(t_i, \cdot)$ by the number of individuals observed; that is, we set $\hat{u}(t_i, x) = H(t_i, x)/N_i$. One can view this function as an estimator for the density $f(t_i, \cdot)$. We may then incorporate the density estimate into a "normalized" least squares cost functional, such as

$$J(q) = \sum_{i=1}^{m} \int_{x_0}^{x_1} |\hat{u}(t_i, x) - u(t_i, x; q)|^2 \, dx .$$

This cost functional avoids the scaling problem associated with lack of knowledge of the total count for the population.

A major benefit of the statistical approach is the availability of tests of fit. We discuss below variants of Pearson's χ^2 test, which are based on approximations of the probability distribution of scaled versions of the cost J evaluated at its minimizer. With such a probability distribution in hand, one may test whether or not a computed residual is statistically significant. We also develop asymptotic distributions for Cramér-von Mises tests, which use empirical distributions rather than densities.

In the next section, we set up some notation, and we lay out the assumptions we need in order to analyze the parameter estimates and model fit.

3. Notation and assumptions. We begin by constructing the histogram estimator of the density. Given a positive number h, we partition the interval $[x_0, x_1]$ into M_h disjoint subintervals $J_k = [y_{k-1}, y_k)$, with $y_k = x_0 + kh$, with $hM_h = x_1 - x_0$ (and $J_{M_h} = [y_{M_h - 1}, y_{M_h}]$). The histogram density estimator $\hat{u}(t_i, x)$ is defined by

$$\hat{u}(t_i, x) = \frac{1}{hN_i} \sum_{k=1}^{M_h} I_{J_k}(x) \sum_{j=1}^{N_i} I_{J_k}(X_{i,j}) .$$

Here I_A denotes the indicator function for the set A: $I_A(x) = 1$ if $x \in A$, and $I_A(x) = 0$ if $x \notin A$. The above definition just gives the proportion of the sample at time t_i that lies in J_k, divided by the subinterval length. As mentioned in the previous section, division by h scales the histogram so that integration yields probability. In order to proceed, we shall need some assumptions concerning the data and the model. These assumptions are given below.

(D) The families of random variables $\{X_{i,1}, \dots, X_{i,N_i}\}_{i=1}^m$ are independent in i. For each i, the random variables $X_{i,1}, \dots, X_{i,N_i}$ are independent and identically distributed, with density $f(t_i, \cdot)$, which is bounded uniformly in x and t_i. The observed data $\{\{\xi_{i,j}\}_{j=1}^{N_i}\}_{i=1}^m$ are values taken by the above random variables.

(M) The model function $u(t, x; q)$ is, for each $t \geq 0$ and $q \in Q$, a probability density function (of x) which is supported on the finite interval $[x_0, x_1]$. The parameter set Q is a compact subset of $\mathbb{R}^p$. The mapping $q \mapsto u(t, \cdot; q)$ is twice continuously differentiable from Q to $L^2(x_0, x_1)$ and is uniformly bounded over $(x_0, x_1) \times Q$, for each t.

(H) There is a unique parameter $q^* \in \mathrm{Int}(Q)$ such that $f(t_i, \cdot) = u(t_i, \cdot; q^*)$, $1 \leq i \leq m$.

We use throughout the notation $\mathcal{N} = (N_1, \dots, N_m)$ to denote the number of observations at the fixed times $t_1, \dots, t_m$. We also set $N = N_1 + \cdots + N_m$.

The assumptions (D) and (M) are of course the properties we require of the data and the model, respectively. The underlying probability space for the random variables will be denoted $(\Omega, \mathcal{F}, P)$. The assumption (H) represents the hypothesis that we wish to test; that is, we are interested in determining whether or not the model generated that data.

To facilitate the analysis, we need some notation. We put $H = L^2(x_0, x_1)$, and we denote by $\langle\cdot, \cdot\rangle$ and $\|\cdot\|$ the standard inner product and norm in H. We set $\varphi_k^{(h)}(x) = I_{J_k}(x)/\sqrt{h}$. Note that these functions form an orthonormal set in H, for each fixed value of h. We denote by P_h the orthogonal projection from H onto the subspace spanned by $\{\varphi_k^{(h)}\}_{k=1}^{M_h}$, which we denote by H_h. We write $f(t_i)$, $u(t_i; q)$, and $\hat{u}(t_i)$ when we are referring to $f(t_i, \cdot)$, $u(t_i, \cdot; q)$, and $\hat{u}(t_i, \cdot)$ as functions belonging to H. We remark that, if $g \in H$, we have $P_h g \to g$ in H, and that $P_h g(x) \to g(x)$ for almost every $x \in [x_0, x_1]$, with respect to Lebesgue measure (see [DS, pp. 214 and 297]). We do not distinguish linear operators on H_h from their matrix representations, which are given in terms of the basis. For example, if A is a linear operator on H_h, we write $A_{kl} = \langle\varphi_k^{(h)}, A\varphi_l^{(h)}\rangle$.

It will be convenient for our purposes to write the histogram in terms of the orthonormal functions $\varphi^{(h)}$:

$$\begin{aligned}\hat{u}_h(t_i, x) &= \sum_{k=1}^{M_h}\left(\frac{1}{\sqrt{h}N_i}\sum_{j=1}^{N_i} I_{J_k}(X_{i,j})\right)\varphi_k^{(h)}(x) \\ &= \frac{1}{N_i}\sum_{j=1}^{N_i}\left(\frac{1}{\sqrt{h}}\sum_{k=1}^{M_h}\varphi_k^{(h)}(x) I_{J_k}(X_{i,j})\right).\end{aligned}$$

With this form, we may examine $\hat{u}_h(t_i)$ as a random vector (taking values in $H_h = \mathbb{R}^{M_h}$) which is an average of independent, identically distributed random vectors. A straightforward calculation shows that the mean of this random vector is $P_h f(t_i)$:

$$E\left[\frac{1}{\sqrt{h}}\sum_{k=1}^{M_h}\varphi_k^{(h)} I_{J_k}(X_{i,j})\right] = \frac{1}{\sqrt{h}}\sum_{k=1}^{M_h}\varphi_k^{(h)}\int_{J_k} f(t_i, y)\,dy = P_h f(t_i),$$

$$E[\hat{u}_h(t_i)] = \frac{1}{N_i}\sum_{j=1}^{N_i}\left(E\frac{1}{\sqrt{h}}\sum_{k=1}^{M_h}\varphi_k^{(h)} I_{J_k}(X_{i,j})\right) = P_h f(t_i).$$

We shall also need the covariance:

$$\begin{aligned}V_{kl}^{i,h} &= E\left[\left(\frac{1}{h} I_{J_k}(X_{i,j})\right)\left(\frac{1}{h} I_{J_l}(X_{i,j})\right)\right] \\ &\quad - \frac{1}{h}\left(\int_{J_k} f(t_i, y)\,dy\right)\left(\int_{J_l} f(t_i, y)\,dy\right) \\ &= \delta_{kl}\frac{1}{h}E\left[I_{J_k}(X_{i,j})\right] - \frac{1}{h}\left(\int_{J_k} f(t_i, y)\,dy\right)\left(\int_{J_l} f(t_i, y)\,dy\right) \\ &= \frac{1}{h}\delta_{kl}\int_{J_k} f(t_i, y)\,dy - \frac{1}{h}\left(\int_{J_k} f(t_i, y)\,dy\right)\left(\int_{J_l} f(t_i, y)\,dy\right), \\ \Sigma_{kl}^{i,h} &= \operatorname{Cov}(\hat{u}(t_i)) = \frac{1}{N_i} V_{kl}^{i,h}.\end{aligned}$$

We begin the analysis of the least squares problem by studying the behavior of the cost functional

$$J^h_{\mathcal{N}}(q) = \sum_{i=1}^{m} \int_{x_0}^{x_1} |\hat{u}(t_i, x) - P_h u(t_i, x; q)|^2 \, dx$$
$$= \sum_{i=1}^{m} \|\hat{u}(t_i) - P_h u(t_i; q)\|^2 .$$

This cost functional is slightly different from the more commonly used

$$\widehat{J}^h_{\mathcal{N}}(q) = \sum_{i=1}^{m} \|\hat{u}(t_i) - u(t_i; q)\|^2 ;$$

however, the analysis for $\widehat{J}$ requires extra smoothness (with respect to x) on the model function.

We denote by $\hat{q} = \hat{q}^h_{\mathcal{N}}$ a minimizer of $J^h_{\mathcal{N}}$ over Q. We denote by J^* the deterministic least squares cost:

$$J^*(q) = \sum_{i=1}^{m} \|f(t_i) - u(t_i; q)\|^2 \, dx .$$

In some cases below we shall be interested in the deterministic cost for the projected problem:

$$J^*_h(q) = \sum_{i=1}^{m} \|P_h f(t_i) - P_h u(t_i; q)\|^2 \, dx .$$

Below, we use the notation "$\xrightarrow{\mathscr{D}}$" to denote weak convergence of probability measures (that is, convergence in distribution). We denote by $N(\mu, C)$ a normally distributed random vector with mean μ and covariance C.

4. Fixed bin size and χ^2 tests. In this section we examine a goodness-of-fit test which resembles the classical χ^2 test of Pearson (see Chapter 30 of [KS] and the references therein). This test, which is relatively simple to apply, is based on asymptotic arguments, letting the number of measurements increase while the bin size h remains constant. The purpose of the test is to determine statistically if the model fits the observed data. The test statistics we discuss below can be thought of as measures of the "distance from the model to the data".

Under the assumption (M), we have the existence of a minimizer $\hat{q} = \hat{q}^h_{\mathcal{N}}$ of $J^h_{\mathcal{N}}$ over Q. The classical approach to the χ^2 test is to form the statistic

$$X^2 = \sum_{i=1}^{m} \sum_{k=0}^{M_h - 1} \left(\frac{N_i(\hat{u}(t_i, y_k) - P_h u(t_i, y_k; \hat{q})}{\sqrt{N_i P_h u(t_i, y_k; \hat{q})}} \right)^2 .$$

This random variable is asymptotically χ^2 if $\hat{q}$ is replaced by q^*. Our goal in this section is to find the asymptotic distribution under the assumptions laid out above. We begin with the following lemma.

LEMMA 4.1. Assume that (D), (M), and (H) hold, and assume that, for h sufficiently small, $J_h^*(q)$ has a unique minimizer over Q at q^*. Then, for fixed h sufficiently small, $\hat{q}_{\mathscr{N}}^h \to q^*$, with probability one, as $\mathscr{N} \to \infty$.

Proof. First, we recall that

$$\hat{u}(t_i) = \frac{1}{N_i}\sum_{j=1}^{N_i}\left(\frac{1}{\sqrt{h}}\sum_{k=1}^{M_h}\varphi_k^{(h)} I_{J_k}(X_{i,j})\right) \to P_h f(t_i),$$

as $N_i \to \infty$, with probability one by Khintchine's law of large numbers (see [Bi1, p. 250]). Thus, $J_{\mathscr{N}}^h(q) \to J_h^*(q)$, uniformly in q. Hence, the minimizers must converge, and we have $\hat{q} \to q^*$, with probability one, as desired. □

To obtain an asymptotic distribution for X^2 and related statistics, we first use the central limit theorem on $\hat{u}(t_i)$. We begin by defining the random vectors

$$Y_i = \hat{u}(t_i) - P_h u(t_i, \hat{q}) \quad \text{and} \quad Z_i = \hat{u}(t_i) - P_h u(t_i, q^*).$$

We "stack" these vectors of length M_h into vectors $\vec{Y}$ and $\vec{Z}$ of length mM_h. Given positive numbers $\alpha_1, \ldots, \alpha_m$, we define the matrix V_α^h as a block diagonal $mM_h \times mM_h$ matrix, having blocks $\alpha_i V^{i,h}$. Let V^h denote a block diagonal matrix with blocks $V^{i,h}$. Next, we set

$$B_k^i = \left\langle \varphi_k^{(h)}, \frac{\partial P_h u(t_i; q^*)^{\mathrm{T}}}{\partial q}\right\rangle \in \mathbb{R}^p,$$

and we define the $mM_h \times mM_h$ matrix A by

$$(A\vec{\psi})_i = -2\frac{\partial P_h u(t_i; q^*)}{\partial q}\left(\frac{\partial^2 J_h^*(q^*)}{\partial q^2}\right)^{-1}\sum_{j=1}^m B_j \psi_j.$$

LEMMA 4.2. Assume that (D), (M), and (H) hold, and assume that, for h sufficiently small, $J_h^*(q)$ has a unique minimizer over Q (which by (H) must be at q^*). Then, for fixed h sufficiently small,

$$\vec{Y} \xrightarrow{\mathscr{D}} N(0, (I+A)V_\alpha^h(I+A)^{\mathrm{T}}),$$

when $\mathscr{N} \to \infty$ in such a way that $N_i/N = N_i/(N_1 + N_2 + \cdots + N_m) \to \alpha_i > 0$.

Proof. We first note that

$$\begin{aligned}
\hat{u}(t_i) - P_h u(t_i, \hat{q}) &= \hat{u}(t_i) - P_h u(t_i, q^*) + P_h u(t_i; q^*) - P_h u(t_i, \hat{q}) \\
&= \hat{u}(t_i) - P_h u(t_i, q^*) + \frac{\partial P_h u(t_i; \overline{q})}{\partial q}(q^* - \hat{q}) \\
&= \hat{u}(t_i) - P_h u(t_i, q^*) + \frac{\partial P_h u(t_i; \overline{q})}{\partial q}\left(\frac{\partial^2 J_{\mathscr{N}}^h(\tilde{q})}{\partial q^2}\right)^{-1}\frac{\partial J_{\mathscr{N}}^h(q^*)}{\partial q} \\
&= \hat{u}(t_i) - P_h u(t_i, q^*) + \frac{\partial P_h u(t_i; \overline{q})}{\partial q}\left(\frac{\partial^2 J_{\mathscr{N}}^h(\tilde{q})}{\partial q^2}\right)^{-1} \\
&\quad * \left(-2\sum_{j=1}^m \left\langle \hat{u}(t_j) - P_h u(t_j, q^*), \frac{\partial P_h u(t_j; q^*)^{\mathrm{T}}}{\partial q}\right\rangle\right),
\end{aligned} \tag{4.1}$$

where $\bar{q}$ and $\tilde{q}$ are vectors having components between those of $\hat{q}$ and q^*. Next, we note that the argument of Lemma 4.1 gives us that

$$\frac{\partial P_h u(t_i;\bar{q})}{\partial q} \to \frac{\partial P_h u(t_i;q^*)}{\partial q}$$

and that

$$\frac{\partial^2 J^h_{\mathscr{N}}(\tilde{q})}{\partial q^2} \to \frac{\partial^2 J^*_h(q^*)}{\partial q^2}$$

as $\mathscr{N} \to \infty$. We now define random vectors

$$Y_i = \hat{u}(t_i) - P_h u(t_i, \hat{q}) \quad \text{and} \quad Z_i = \hat{u}(t_i) - P_h u(t_i, q^*).$$

The calculation (4.1) reduces to the following identity:

$$Y_i = Z_i - 2\frac{\partial P_h u(t_i;\bar{q})}{\partial q}\left(\frac{\partial^2 J^h_{\mathscr{N}}(\tilde{q})}{\partial q^2}\right)^{-1}\sum_{j=1}^{m}\left\langle Z_i, \frac{\partial P_h u(t_j;q^*)^{\mathrm{T}}}{\partial q}\right\rangle; \tag{4.2}$$

hence, we may write $\vec{Y} = (I + A^h_{\mathscr{N}})\vec{Z}$, where the matrix $A^h_{\mathscr{N}}$ is defined by the last term of (4.2).

The asymptotic normality then stems from the central limit theorem, applied to Z: note that

$$Z_i = \frac{1}{hN_i}\left(\sum_{k=1}^{M_h} I_{J_k}(x)\sum_{j=1}^{N_i} I_{J_k}(X_{i,j}) - \int_{J_k} u(t_i, x; q^*)\,dx\right),$$

which is a sum of independent, identically distributed random variables with covariance $V^{i,h}/N_i$. Moreover, $\{Z_i\}_{i=1}^m$ are independent. Hence, $\sqrt{N}\vec{Z} \xrightarrow{\mathscr{D}} N(0, V^h_\alpha)$. Since also $A^h_{\mathscr{N}} \to A$ with probability one, Slutsky's theorem then gives us that

$$\sqrt{N}\vec{Y} \xrightarrow{\mathscr{D}} N(0, (I+A)V^h_\alpha(I+A)^{\mathrm{T}}),$$

as desired. □

Next, we define C_i to be a diagonal matrix, whose diagonal entries are

$$1/\sqrt{P_h u(t_i, y_k; q^*)},$$

and let C be the block diagonal matrix with blocks C_i. The following result gives us the distribution of two possible test statistics to use for goodness-of-fit questions.

Theorem 4.3. *Assume that (D), (M), and (H) hold, and assume that, for h sufficiently small, $J^*_h(q)$ has a unique minimizer over Q. Then, for fixed h sufficiently small,*

$$N|Y|^2 \xrightarrow{\mathscr{D}} S^{\mathrm{T}}S \quad \text{and} \quad X^2 \xrightarrow{\mathscr{D}} W^{\mathrm{T}}W,$$

where S is distributed $N(0, (I+A)V^h_\alpha(I+A)^{\mathrm{T}})$, and W is distributed $N(0, C^{1/2}(I+A)V^h(I+A)^{\mathrm{T}}C^{1/2})$, when $\mathscr{N} \to \infty$ in such a way that $N_i/N \to \alpha_i > 0$.

Proof. The proof follows directly from the above lemmas. Note that $N|Y|^2$ tends (by Theorem 5.1 of [Bi2, p. 30]) in distribution to $S^{\mathrm{T}}S$. For the statistic X^2, we note that

$$X^2 = \sum_{i=1}^{m} \frac{1}{N_i} Y_i^{\mathrm{T}} \widehat{C}_i Y_i,$$

where $\widehat{C}_i$ is a diagonal matrix with entries $1/\sqrt{P_h u(t_i, y_k; \hat{q})}$. Note that $\widehat{C}_i \to C_i$ almost surely, as $\mathscr{N} \to \infty$. Slutsky's theorem with the previous lemma gives us that $X^2 \xrightarrow{\mathscr{D}} W^{\mathrm{T}}W$. □

To use either of these tests requires the following steps.

1. Compute the eigenvalues of the covariance matrix of the normal random vector associated with the asymptotic quadratic form of interest. We denote this quadratic form by F.

2. Compute the characteristic function

$$\phi(t) \equiv \mathrm{E}(e^{itF}) = \prod_{l=1}^{mM_h} (1 - 2i\lambda_l t)^{-1/2}$$

of the limiting quadratic form F. Here $\{\lambda_l\}$ is the set of eigenvalues of the covariance.

3. Compute the inverse Fourier transform of $\phi(-t)$ to obtain the density of the form F.

Once we have the density, we compute P-values and thresholds associated with testing the hypothesis. The idea, of course, is that if the quadratic form is large, then the model is not fitting the data well. The difference between the two quadratic forms is in the scaling matrix C. The statistic $|Y|^2$ is a simple measure of the L^2 distance between the histogram and the model density, while the statistic X^2 is a weighted L^2 distance, and the weight is chosen to provide a relative difference. We remark that X^2 is more commonly used in general statistical settings. In future studies we plan to undertake a comprehensive comparison of these two statistics together with those discussed in Sec. 6 below.

5. Large sample theory for parameter estimates. In this section it is our goal to study consistency and asymptotic normality results for estimators determined by least squares. These results give us specific convergence information on parameter estimators as we increase the number of observations while decreasing the bin size. The following theorem gives us a consistency result for least squares estimators.

THEOREM 5.1. Assume that (D) and (M) hold, and suppose that $h = h_{\mathscr{N}} \to 0$, as $\mathscr{N} \to \infty$, slowly enough that $\log(N_i)/hN_i \to 0$. Then $J^h_{\mathscr{N}}(q) \to J^*(q)$, uniformly in q, with probability one. If (H) also holds, then $\hat{q} \to q^*$ with probability one.

Proof. We note first that Lemma 2.3.3 of [PR] gives us that

$$\mathrm{E}[\exp(\lambda^i_{N_i}(\hat{u}(t_i, x) - P_h f(t_i, x)))] \le \exp \frac{(\lambda^i_{N_i})^2 M}{N_i h},$$

where $\lambda^i_{N_i} = (N_i h \log N_i)^{1/2}$, and M is a bound for $f(t_i, x)$. Next, we set $\mu_{N_i} = 4\log(N_i)$, and we note that, as in Lemma 2.3.4 of [PR], if $\varepsilon > 0$, we have from Chebyshev's inequality

$$\begin{aligned} &\mathrm{P}[\hat{u}(t_i, x) - P_h f(t_i, x) \geq (\varepsilon + \mu_{N_i})/\lambda^i_{N_i}] \\ &\quad \leq \exp\left(-\varepsilon - \mu_{N_i} + \frac{(\lambda^i_{N_i})^2 M}{N_i h}\right) = \frac{1}{N_i^3 e^{\varepsilon}}. \end{aligned}$$

Next, we note that, since $\hat{u}(t_i)$ and $P_h f(t_i)$ are piecewise constant on the M_h intervals $\{J_k\}$, we have

$$\begin{aligned} &\mathrm{P}\left[\sup_{[x_0, x_1]} (\hat{u}(t_i, x) - P_h f(t_i, x)) \geq (\varepsilon + \mu_{N_i})/\lambda^i_{N_i}\right] \\ &\quad \leq M_h/(N_i^3 e^{\varepsilon}) \leq C/N_i^2, \end{aligned}$$

for some constant C, since $hN_i \to \infty$.

We now note that $(\varepsilon + \mu_{N_i})/\lambda^i_{N_i} \to 0$, as $N_i \to \infty$, so that, given $\delta > 0$, we have that

$$\sum_{N_i=1}^{\infty} \mathrm{P}\left[\sup_{[x_0, x_1]} (\hat{u}(t_i, x) - P_h f(t_i, x)) \geq \delta\right] < \infty.$$

Using a similar argument, we obtain

$$\sum_{N_i=1}^{\infty} \mathrm{P}\left[\sup_{[x_0, x_1]} (\hat{u}(t_i, x) - P_h f(t_i, x)) \leq -\delta\right] < \infty.$$

The Borel-Cantelli lemma gives us that

$$\mathrm{P}\left[\lim_{N_i \to \infty} \sup_{[x_0, x_1]} |\hat{u}(t_i, x) - P_h f(t_i, x)| = 0\right] = 1,$$

since $f(t_i)$ is bounded and supported on $[x_0, x_1]$.

To connect this result to the least squares cost, we note that $P_h f(t_i) \to f(t_i)$ in $H = L^2(x_0, x_1)$. Thus, we have

$$\mathrm{P}\left[\lim_{N_i \to \infty} \|\hat{u}(t_i) - f(t_i)\| = 0\right] = 1$$

by the triangle inequality, and the fact that uniform convergence of bounded functions on a set of finite measure implies L^2 convergence. Since $\|\hat{u}(t_i) - f(t_i)\| \to 0$ with probability one, independently of q, the cost functional result follows.

Assuming that (H) holds, we have that

$$J^*(q) = \sum_{i=1}^{m} \int_{x_0}^{x_1} |u(t_i, x; q^*) - u(t_i, x; q)|^2 \, dx,$$

and that this functional has a unique minimizer, which is q^*. Since Q is a compact set, any subsequence of minimizers $\{\hat{q}\}$ has a further subsequence that is convergent.

The uniform convergence of $J_{\mathcal{N}}^h$ to J^* guarantees that any convergent subsequence of $\{\hat{q}\}$ must in fact converge to the minimizer q^* of J^*. Hence, every subsequence has a subsequence that converges to q^*, which implies $\hat{q} \to q^*$. This convergence holds with probability one, from the convergence of the cost functionals. $\square$

We remark here that under (D) and (M), we may also obtain convergence results for the derivatives of the cost functional. In particular, we have that

$$\begin{aligned}
&\frac{\partial}{\partial q}\|\hat{u}(t_i) - P_h u(t_i\,;\,q)\|^2 \\
&\quad = -2\int_{x_0}^{x_1}(\hat{u}(t_i\,,x) - P_h u(t_i\,,x\,;\,q))\frac{\partial P_h u(t_i\,,x\,;\,q)}{\partial q}\,dx \\
&\quad \to -2\int_{x_0}^{x_1}(f(t_i\,,x) - u(t_i\,,x\,;\,q))\frac{\partial u(t_i\,,x\,;\,q)}{\partial q}\,dx
\end{aligned}$$

uniformly in q, with probability one. Likewise, for the second derivative we have

$$\begin{aligned}
&\frac{\partial^2}{\partial q^2}\|\hat{u}(t_i\,,x) - P_h u(t_i\,,x\,;\,q)\|^2 \\
&\quad = 2\int_{x_0}^{x_1}\frac{\partial P_h u(t_i\,;\,q)^{\mathrm{T}}}{\partial q}\frac{\partial P_h u(t_i\,;\,q)}{\partial q}\,dx \\
&\qquad - 2\int_{x_0}^{x_1}(\hat{u}(t_i\,,x) - P_h u(t_i\,,x\,;\,q))\frac{\partial^2 P_h u(t_i\,,x\,;\,q)}{\partial q^2}\,dx \\
&\quad \to 2\int_{x_0}^{x_1}\frac{\partial u(t_i\,;\,q)^{\mathrm{T}}}{\partial q}\frac{\partial u(t_i\,;\,q)}{\partial q}\,dx \\
&\quad \to -2\int_{x_0}^{x_1}(f(t_i\,,x) - u(t_i\,,x\,;\,q))\frac{\partial^2 u(t_i\,,x\,;\,q)}{\partial q^2}\,dx
\end{aligned}$$

uniformly, with probability one. We will need these facts later.

Our next problem will be to obtain an asymptotic distribution for the estimator $\hat{q} = \hat{q}_{\mathcal{N}}^h$. The difficulties here arise from letting $h \to 0$. To proceed, we shall need asymptotic information on the variance, which will allow us to use the central limit theorem. We begin by extending the matrix $V^{i,h}$ to an operator on $\mathscr{V}^{i,h}$ on H:

$$\begin{aligned}
\mathscr{V}^{i,h}\psi &= \sum_{k=1}^{M_h}\sum_{l=1}^{M_h} V_{kl}^{i,h}\langle\psi\,,\varphi_l^{(h)}\rangle\varphi_k^{(h)} \\
&= \sum_{k=1}^{M_h}\sum_{l=1}^{M_h}\left(\delta_{kl}\frac{1}{\sqrt{h}}\langle f(t_i)\,,\varphi_k^{(h)}\rangle - \langle f(t_i)\,,\varphi_k^{(h)}\rangle\langle f(t_i)\,,\varphi_l^{(h)}\rangle\right)\langle\varphi_l^{(h)}\,,\psi\rangle\varphi_k^{(h)}\,.
\end{aligned}$$

Note, of course, that the range of $\mathscr{V}^{i,h}$ lies in H_h.

The behavior of these operators, as $h \to 0$, is at the heart of our central limit theorem. We define $\mathscr{V}^i\psi(x) = f(t_i\,,x)\psi(x)$, which, under (D), is a bounded linear operator on $H = L^2(x_0\,,x_1)$. Our next step is the following lemma.

Lemma 5.2. Assume that (D) holds. Then $\mathscr{V}^i$ is the strong limit of $\mathscr{V}^{i,h}$, as $h \to 0$.

Proof. We first define the operators $\mathscr{V}_1^{i,h}$ and $\mathscr{V}_2^{i,h}$ by

$$\mathscr{V}_1^{i,h}\psi = \frac{1}{\sqrt{h}}\sum_{k=1}^{M_h}\langle f(t_i), \varphi_k^{(h)}\rangle\langle\varphi_k^{(h)}, \psi\rangle\varphi_k^{(h)},$$

$$\mathscr{V}_2^{i,h}\psi = -\sum_{k=1}^{M_h}\sum_{l=1}^{M_h}\langle f(t_i), \varphi_k^{(h)}\rangle\langle f(t_i), \varphi_l^{(h)}\rangle\langle\varphi_l^{(h)}, \psi\rangle\varphi_k^{(h)}.$$

We have that $\mathscr{V}^{i,h} = \mathscr{V}_1^{i,h} + \mathscr{V}_2^{i,h}$. We will show that $\mathscr{V}_1^{i,h} \to \mathscr{V}^i$ strongly, and that $\mathscr{V}_2^{i,h} \to 0$ strongly. We first note that

$$\begin{aligned} P_h f(t_i)(x)P_h\psi(x) &= \sum_{k=1}^{M_h}\sum_{l=1}^{M_h}\langle\varphi_k^{(h)}, f(t_i)\rangle\langle\varphi_l^{(h)}, \psi\rangle\varphi_k^{(h)}(x)\varphi_l^{(h)}(x) \\ &= \sum_{k=1}^{M_h}\sum_{l=1}^{M_h}\langle\varphi_k^{(h)}, f(t_i)\rangle\langle\varphi_l^{(h)}, \psi\rangle\frac{1}{\sqrt{h}}\delta_{kl}\varphi_k^{(h)}(x) \\ &= \sum_{k=1}^{M_h}\langle\varphi_k^{(h)}, f(t_i)\rangle\langle\varphi_k^{(h)}, \psi\rangle\frac{1}{\sqrt{h}}\varphi_k^{(h)}(x) = \mathscr{V}_1^{i,h}\psi(x). \end{aligned}$$

Since $P_h f(t_i) \to f(t_i)$ for almost every x, and since $\sup_x |P_h f(t_i)(x)| \leq \sup_x |f(t_i, x)|$, we have by the dominated convergence theorem that $P_h f(t_i)\psi \to f(t_i)\psi$ in H, for every $\psi \in H$. Moreover, $P_h\psi \to \psi$, for every $\psi \in H$, so that

$$\mathscr{V}_1^{i,h}\psi = P_h f(t_i)P_h\psi \to f(t_i)\psi = \mathscr{V}^i\psi$$

in H. It remains to show that $\mathscr{V}_2^{i,h}\psi \to 0$. We have

$$\begin{aligned} \mathscr{V}_2^{i,h}\psi &= -\sum_{k=1}^{M_h}\sum_{l=1}^{M_h}\langle f(t_i), \varphi_k^{(h)}\rangle\langle f(t_i), \varphi_l^{(h)}\rangle\langle\varphi_l^{(h)}, \psi\rangle\varphi_k^{(h)} \\ &= -\sum_{l=1}^{M_h}\langle f(t_i), \varphi_l^{(h)}\rangle\langle\varphi_l^{(h)}, \psi\rangle\sum_{k=1}^{M_h}\langle f(t_i), \varphi_k^{(h)}\rangle\varphi_k^{(h)} \\ &= \sum_{l=1}^{M_h}\langle f(t_i), \varphi_l^{(h)}\rangle\langle\varphi_l^{(h)}, \psi\rangle\sqrt{h}\mathscr{V}^i P_h\psi. \end{aligned}$$

Note that, by the Cauchy-Schwarz and Bessel inequalities, we have

$$\sum_{l=1}^{M_h}|\langle f(t_i), \varphi_l^{(h)}\rangle\langle\varphi_l^{(h)}, \psi\rangle| \leq \|f(t_i)\|\,\|\psi\|,$$

and since $\sqrt{h}\mathscr{V}^i P_h\psi \to 0$, we have

$$\mathscr{V}_2^{i,h}\psi \to 0$$

for each $\psi \in H$, and the lemma is proved. □

Note that the operator $\mathscr{V}^i$ is a "diagonal" operator, implying that if it were the covariance operator of a random process in $L^2(x_0, x_1)$, that process would essentially be a white noise process. Thus we cannot expect convergence in distribution of the processes in $H = L^2$. However, we can obtain a weaker result that will be enough for our purposes. Suppose $\psi \in H$. If X_h is an H_h-valued random vector distributed $N(0, \mathscr{V}^{i,h})$, then $\langle \psi, X_h \rangle$ is a real-valued random variable distributed $N(0, \langle \psi, \mathscr{V}^{i,h}\psi \rangle)$. Note that Lemma 5.2 gives us that $\langle \psi, \mathscr{V}^{i,h}\psi \rangle \to \langle \psi, \mathscr{V}^i \psi \rangle$, so that $\langle \psi, X_h \rangle \xrightarrow{\mathscr{D}} Y$, as $h \to 0$, where Y is distributed $N(0, \langle \psi, \mathscr{V}^i \psi \rangle)$. Our next step is to develop a central limit theorem for such inner products involving the histogram estimator.

LEMMA 5.3. Suppose that (D), (M), and (H) hold, and that $1 \le i \le m$. Let $g \in H$. Then,

$$\sqrt{N_i}\langle \hat{u}(t_i, x) - P_h u(t_i, x; q^*), g \rangle \xrightarrow{\mathscr{D}} N(0, \langle g, \mathscr{V}^i g \rangle),$$

when $N_i \to \infty$ and $h \to 0$ in such a way that $hN_i \to \infty$.

Proof. We first note that

$$\begin{aligned} N_i\langle \hat{u}(t_i, x) &- P_h u(t_i, x; q^*), g \rangle \\ &= \left\langle \sum_{k=1}^{M_h} \left(\frac{1}{\sqrt{h}N_i} \sum_{j=1}^{N_i} I_{J_k}(X_{i,j}) \right) \varphi_k^{(h)} - P_h u(t_i; q^*), g \right\rangle \\ &= \left\langle \sum_{j=1}^{N_i} \sum_{k=1}^{M_h} \left(\frac{1}{\sqrt{h}} I_{J_k}(X_{i,j}) - u_{i,k}^* \right) \varphi_k^{(h)}, P_h g \right\rangle \\ &= \sum_{j=1}^{N_i} \left\langle \sum_{k=1}^{M_h} \left(\frac{1}{\sqrt{h}} I_{J_k}(X_{i,j}) - u_{i,k}^* \right) \varphi_k^{(h)}, P_h g \right\rangle, \end{aligned}$$

where $u_{i,k}^* = \langle \varphi_k^{(h)}, u(t_i; q^*) \rangle$. We have now a sum of independent, identically distributed, zero-mean random variables. Since the distribution of the random variables depends on h, which is tending to 0 as $N_i \to \infty$, we will need to verify Lindeberg's condition for a central limit result (see, for example, [Bi1, p. 310]). Toward that end, we define

$$Y_{N_{i,j}} = \left\langle \sum_{k=1}^{M_h} \left(\frac{1}{\sqrt{h}} I_{J_k}(X_{i,j}) - u_{i,k}^* \right) \varphi_k^{(h)}, P_h g \right\rangle.$$

As we remarked above, for fixed N_i the random variables are independent and identically distributed, with mean 0. We set $\rho_{N_i}^2 = \mathrm{E}[Y_{N_i,1}^2]$ and $s_{N_i}^2 = N_i \rho_{N_i}^2$. Note that

$$\rho_{N_i}^2 = \langle P_h g, \mathscr{V}^{i,h} P_h g \rangle \to \langle g, \mathscr{V}^i g \rangle,$$

as $N_i \to \infty$. The Lindeberg condition, then, is that

$$\lim_{N_i \to \infty} \frac{1}{\rho_{N_i}^2} \int_{[|Y_{N_i,1}| \ge \varepsilon s_{N_i}]} Y_{N_i,1}^2 \, dP = 0$$

for every $\varepsilon > 0$. To proceed, we suppose that $\langle g, \mathscr{V}^i g\rangle > 0$. To verify the Lindeberg condition, we suppose that we are given $\varepsilon > 0$, and we examine $P[|Y_{N_i,1}| \geq \varepsilon s_{N_i}]$. We first note that, for each $\omega \in \Omega$, there exists $k_0 = k_0(\omega)$ such that $X_{i,1}(\omega) \in J_{k_0}$. Then we notice that

$$\begin{aligned}
Y_{N_i,1}^2 &= \left|\left\langle \sum_{k=1}^{M_h} \left(\frac{1}{\sqrt{h}} I_{J_k}(X_{i,1}) - u_{i,k}^*\right) \varphi_k^{(h)}, P_h g\right\rangle\right|^2 \\
&\leq \left\|\sum_{k=1}^{M_h} \left(\frac{1}{\sqrt{h}} I_{J_k}(X_{i,1}) - u_{i,k}^*\right) \varphi_k^{(h)}\right\|^2 \|g\|^2 \\
&= \sum_{k=1}^{M_h} \left|\frac{1}{\sqrt{h}} I_{J_k}(X_{i,1}) - u_{i,k}^*\right|^2 \|g\|^2 \\
&= \left(\left|\frac{1}{\sqrt{h}} - u_{i,k_0}^*\right|^2 + \sum_{k\neq k_0} |u_{i,k}^*|^2\right) \|g\|^2, \\
&\leq \left(\frac{1}{h} + \sum_{k=1}^{M_h} |u_{i,k}^*|^2\right) \|g\|^2,
\end{aligned}$$

which gives us a uniform bound on $Y_{N_i,1}^2$. Moreover, we see that

$$\begin{aligned}
&\left(\frac{1}{h} + \sum_{k=1}^{M_h} |u_{i,k}^*|^2\right) \|g\|^2 \geq \varepsilon^2 s_{N_i}^2 \\
&\qquad \Leftrightarrow 1 + h\sum_{k=1}^{M_h} |u_{i,k}^*|^2 \geq hN_i \frac{\varepsilon^2 \rho_{N_i}^2}{\|g\|^2}.
\end{aligned}$$

Since $\rho_{N_i}^2 \to \langle g, \mathscr{V}^i g\rangle > 0$ and $\sum_{k=1}^{M_h} |u_{i,k}^*|^2 \leq \|u(t_i; q^*)\|^2 < \infty$, the second inequality must be reversed, for N_i sufficiently large. Thus, we have that

$$\mathrm{P}[|Y_{N_i,1}| \geq \varepsilon s N_i] = 0$$

for N_i sufficiently large, so that the Lindeberg condition is satisfied.

In case $\langle g, \mathscr{V}^i g\rangle = 0$, we note that

$$\operatorname{Var}\left(\frac{1}{\sqrt{N_i}} \sum_{j=1}^{N_i} Y_{N_i,j}\right) = \rho_{N_i}^2 \to 0,$$

and we have the desired result. □

THEOREM 5.4. Suppose that (D), (M), and (H) hold, and that we take a sequence of observations of size $\mathscr{N}^s = (N_1^s, N_2^s, \dots, N_m^s)$ with $N^s = \sum_i N_i^s \to \infty$ as $s \to \infty$ in such a way that $N_i^s/N^s \to \alpha_i > 0$. We also take a sequence of bin sizes $h^s \to 0$ in such a way that $h^s N_i^s \to \infty$ and $\log(N_i^s)/(h^s N_i^s) \to 0$, and suppose that the quantity

$$\mathscr{I} = \sum_{i=1}^{m} \alpha_i \int_{x_0}^{x_1} \frac{\partial u(t_i, x; q^*)^{\mathrm{T}}}{\partial q} \frac{\partial u(t_i, x; q^*)}{\partial q} \, dx$$

is positive definite. Then,

$$\sqrt{N^s}(\hat{q} - q^*) \xrightarrow{\mathscr{D}} N(0, \sigma^2)$$

as $s \to \infty$, where

$$\sigma^2 = \sum_{i=1}^{m} \alpha_i \mathscr{I}^{-1} W^i \mathscr{I}^{-1},$$

and where $W^i = 4\langle \partial u(t_i; q^*)/\partial q, \mathscr{V}^i \partial u(t_i; q^*)/\partial q \rangle$.

Proof. We begin by differentiating the cost functional:

$$\begin{aligned}\frac{\partial J_{\mathscr{N}}^h(q^*)}{\partial q} &= -2 \sum_{i=1}^{m} \int_{x_0}^{x_1} (\hat{u}(t_i, x) - P_h u(t_i, x; q^*)) \frac{\partial P_h u(t_i, x, q^*)}{\partial q} \, dx \\ &= -2 \sum_{i=1}^{m} \left\langle \hat{u}(t_i) - P_h u(t_i; q^*), P_h \frac{\partial u(t_i; q^*)}{\partial q} \right\rangle .\end{aligned}$$

We note that from Lemma 5.3 and our assumptions we have

$$-2\sqrt{N_i} \left\langle \hat{u}(t_i) - P_h u(t_i; q^*), P_h \frac{\partial u(t_i; q^*)}{\partial q} \right\rangle \xrightarrow{\mathscr{D}} N(0, W^i).$$

Our only obstacle in obtaining asymptotic normality for the differentiated cost is that the terms need to be multiplied by different N_i's. From our hypotheses, we have the sequence $\{N_i^s\}_{s=1}^{\infty}$, which satisfies $N_i^s \to \infty$ and $N_i^s/N^s \to \alpha_i > 0$ as $s \to \infty$. Then we have that

$$\begin{aligned}\sqrt{N^s} \frac{\partial J_{\mathscr{N}}^h(q^*)}{\partial q} &= -2\sqrt{N^s} \sum_{i=1}^{m} \left\langle \hat{u}(t_i) - P_h u(t_i; q^*), P_h \frac{\partial u(t_i; q^*)}{\partial q} \right\rangle \\ &= -2 \sum_{i=1}^{m} \frac{\sqrt{N_i^s}}{\sqrt{N^s}} \sqrt{N_i^s} \left\langle \hat{u}(t_i) - P_h u(t_i; q^*), P_h \frac{\partial u(t_i; q^*)}{\partial q} \right\rangle \\ &\xrightarrow{\mathscr{D}} N\left(0, \sum_{i=1}^{m} \alpha_i W^i\right) .\end{aligned}$$

We remark that the interchange of " $\xrightarrow{\mathscr{D}}$ " with the sum is allowed because the observations at each time are assumed in (D) to be independent of those at other times.

Now we have from Taylor's theorem that

$$\sqrt{N^s} \frac{\partial J_{\mathscr{N}}^h(\hat{q})}{\partial q} = \sqrt{N^s} \frac{\partial J_{\mathscr{N}}^h(q^*)}{\partial q} + \sqrt{N^s} \frac{\partial^2 J_{\mathscr{N}}^h(\overline{q})}{\partial q^2} (\hat{q} - q^*),$$

where the components of the vector $\overline{q}$ lie between the corresponding components of $\hat{q}$ and q^*. Since $\hat{q} \to q^*$ with probability one, we also have $\overline{q} \to q^*$ with probability one. Next, we note that, because q^* is in the interior of Q, we have that $\hat{q}$ is "eventually" in the interior; that is, $\mathrm{P}[\hat{q} \in \mathrm{Int}(Q)] \to 1$ as $s \to \infty$. Since $\hat{q}$ is a minimizer of $J^h_{\mathscr{N}^s}$, we have

$$\mathrm{P}\left[\sqrt{N^s}\frac{\partial J^h_{\mathscr{N}}(\hat{q})}{\partial q} = 0\right] \geq \mathrm{P}[\hat{q} \in \mathrm{Int}(Q)] \to 1$$

as $s \to \infty$. Recall that

$$\begin{aligned}
&\frac{\partial^2}{\partial q^2}\|\hat{u}(t_i, x) - P_h u(t_i, x; q)\|^2 \\
&= 2\int_{x_0}^{x_1} \frac{\partial P_h u(t_i; q)}{\partial q}^{\mathrm{T}} \frac{\partial P_h u(t_i; q)}{\partial q}\, dx \\
&\quad - 2\int_{x_0}^{x_1} (\hat{u}(t_i, x) - P_h u(t_i, x; q))\frac{\partial^2 u(t_i, x; q)}{\partial q^2}\, dx \\
&\to 2\int_{x_0}^{x_1} \frac{\partial u(t_i; q)}{\partial q}^{\mathrm{T}} \frac{\partial u(t_i; q)}{\partial q}\, dx \\
&\quad - 2\int_{x_0}^{x_1} (f(t_i, x) - u(t_i, x; q))\frac{\partial^2 u(t_i, x; q)}{\partial q^2}\, dx \\
&= 2\int_{x_0}^{x_1} \frac{\partial u(t_i; q)}{\partial q}^{\mathrm{T}} \frac{\partial u(t_i; q)}{\partial q}\, dx \\
&\quad - 2\int_{x_0}^{x_1} (u(t_i, x; q^*) - u(t_i, x; q))\frac{\partial^2 u(t_i, x; q)}{\partial q^2}\, dx,
\end{aligned}$$

with probability one, with the last equality holding under (H). Hence, $\partial^2 J^h_{\mathscr{N}^s}/\partial q^2 \to \partial^2 J^*/\partial q^2$ uniformly. Thus, $\partial^2 J^h_{\mathscr{N}^s}(\overline{q})/'\partial q^2$ converges to $\partial^2 J^*(q^*)/\partial q^2$ with probability one. Note that, under (H), we have $\partial^2 J^*(q^*)/\partial q^2 = \mathscr{I}$. Also, since $\partial^2 J^*(q^*)/\partial q^2$ is positive definite, $(\partial^2 J^h_{\mathscr{N}^s}(\overline{q})/\partial q^2)^{-1}$ exists and converges to $(\partial^2 J^*(q^*)/\partial q^2)^{-1}$, almost surely. Thus we have

$$(\partial^2 J^h_{\mathscr{N}^s}(\overline{q})/\partial q^2)^{-1} \to (\partial^2 J^*(q^*)/\partial q^2)^{-1}$$

with probability one. By Slutsky's theorem (see for example [Bi2, p. 31]), we then obtain

$$\sqrt{N^s}(\hat{q} - q^*) = -\sqrt{N^s}\left(\frac{\partial^2 J^h_{\mathscr{N}^s}(\overline{q})}{\partial q^2}\right)^{-1} \frac{\partial J^h_{\mathscr{N}}(q^*)}{\partial q} \xrightarrow{\mathscr{D}} N\left(0, \sum_{i=1}^{m} \alpha_i \mathscr{I}^{-1} W^i \mathscr{I}^{-1}\right),$$

as desired. □

6. The empirical process and tests of fit. In this section, we are interested in deriving a different goodness-of-fit test for our model. The test we develop, which is closely related to the Cramér-von Mises test, depends on empirical distributions, rather than densities, constructed from the data. These tests are generally appropriate when the full collection of data $\{\xi_{i,j}\}$ is available.

We begin with some notation. We let U denote the parametric distribution function for the data: $U(t, x; q) = \int_{x_0}^{x} u(t, \xi; q)\, d\xi$. Under the hypothesis (H), we have that there is a unique q^* such that $U(t_i, x; q^*)$ is the distribution function for the data at time t_i. We define the transformed data $\widehat{U}$ by

$$\widehat{U}_{i,j} = U(t_i, X_{i,j}, \hat{q}) = \int_{x_0}^{X_{i,j}} u(t, \xi; q)\, d\xi. \tag{6.1}$$

Note that, under (D) and (H), we have that $\{U(t_i, X_{i,j}; q^*) : 1 \le i \le m,\ 1 \le j \le N_i\}$ are independent random variables that are uniformly distributed on $[0, 1]$. We denote by $\widehat{F}_{N_i}(t_i, y)$ the proportion of the time t_i data satisfying $\widehat{U}_{i,j} \le y$, and we let $\widehat{F}(y)$ denote the proportion of the combined transformed data that is no greater than y. These functions are given by the following:

$$\widehat{F}_{N_i}(t_i, y) = \frac{1}{N_i} \sum_{j=1}^{N_i} I_{[\widehat{U}_{i,j} \le y]},$$

$$\widehat{F}_{\mathcal{N}}(y) = \frac{1}{N} \sum_{i,j} I_{[\widehat{U}_{i,j} \le y]} = \sum_{i=1}^{m} \frac{N_i}{N} \widehat{F}_{N_i}(t_i, y).$$

Recall that $\mathcal{N} = (N_1, N_2, \dots, N_m)$ and $N = N_1 + N_2 + \cdots + N_m$. We next scale these processes to obtain our goodness-of-fit statistics. We put

$$\begin{aligned} \widehat{Z}_{N_i}(t_i, y) &= \sqrt{N_i}(\widehat{F}_{N_i}(t_i, y) - y), \\ \widehat{Z}_{\mathcal{N}}(y) &= \sqrt{N}(\widehat{F}_{\mathcal{N}}(y) - y) = \sum_{i=1}^{m} \sqrt{\frac{N_i}{N}} \widehat{Z}_{N_i}(t_i, y). \end{aligned} \tag{6.2}$$

These processes compare the empirical distribution of the transformed data $\widehat{U}_{i,j}$ to the uniform distribution function. Before we proceed to our main theorem for the asymptotic distribution of these processes, we need a little more notation. We set

$$F^*_{N_i}(t_i, y) = \frac{1}{N_i} \sum_{j=1}^{N_i} I_{[U^*_{i,j} \le y]},$$

$$F^*_{\mathcal{N}}(y) = \frac{1}{N} \sum_{i,j} I_{[U^*_{i,j} \le y]} = \sum_{i=1}^{m} \frac{N_i}{N} F^*_{N_i}(t_i, y),$$

where $U^*_{i,j} = U(t_i, X_{i,j}; q^*)$. The processes $Z^*_{N_i}(t_i, y)$ and $Z^*_{\mathcal{N}}(y)$ are defined in an analogous manner. Next, we set

$$x_i(y, q) = \inf\{x : U(t_i, x; q) = y\},$$

$$\hat{y}_i(y) = U(t_i, x_i(y, \hat{q}); q^*).$$

The regularity of these processes is important in the arguments below. We shall make use of the following model regularity assumption.

(MR) The functions

$$(y, q) \mapsto U(t_i, x_i(y, q), q)$$

and

$$(y, q) \mapsto \partial U(t_i, x_i(y, q), q)/\partial q$$

are continuous functions of y and q at $q = q^*$.

Note that all of the functions defined above are functions of $y \in [0, 1]$. We view them as random elements of the metric space $D[0, 1]$ of right-continuous functions having left-hand limits. The metric here is the Skorohod metric. The theory of random functions in D can be found in [Bi2] or [EK], and it has been applied successfully to empirical distribution problems by many authors (see in particular [SW] and [D]). Our first result is a minor adjustment to Lemma 1 to [D].

LEMMA 6.1. Suppose that (D), (M), (H), and (MR) hold. Then

$$\sum_{i=1}^{m} \left| \sqrt{\frac{N_i}{N}} (Z_{N_i}^*(t_i, \cdot) - Z_{N_i}^*(t_i, \hat{y}_i(\cdot))) \right| \to 0$$

in $D[0, 1]$, in probability, as $\mathcal{N} \to \infty$ and $h \to 0$ in such a way that $hN_i \to \infty$, $\log(N_i^s)/h^s N_i^s \to 0$, $N_i/N \to \alpha_i > 0$.

Proof. First, we note that $y = U(t_i, x_i(y, q); q)$, for every $q \in Q$. Then, we have

$$\begin{aligned} |\hat{y}_i(y) - y| &= |U(t_i, x_i(y, \hat{q}); q^*) - U(t_i, x_i(y, \hat{q}); \hat{q})| \\ &\le \int_{x_0}^{x_i(y, \hat{q})} |u(t_i, x; \hat{q}) - u(t_i, x; q^*)| \, dx \\ &\le \int_{x_0}^{x_1} |u(t_i, x; \hat{q}) - u(t_i, x; q^*)| \, dx \\ &\le \sqrt{x_1 - x_0} \, \|u(t_i; \hat{q}) - u(t_i; q^*)\|, \end{aligned}$$

so that $\hat{y}_i(y) \to y$ uniformly on $[0, 1]$, in probability. Next, we note that the processes $\{Z_{N_i}^*(t_i, \cdot)\}_{i=1}^m$ are independent, by (D).

The argument in Lemma 1 of [D] then gives us that

$$(Z_{N_1}^*(t_1, \cdot), \dots, Z_{N_m}^*(t_m, \cdot), \hat{y}_1, \dots, \hat{y}_m) \xrightarrow{\mathcal{D}} (W_1^0, \dots, W_m^0, Y, \dots, Y),$$

where Y denotes the deterministic "identity" process $Y(y) = y$, and $(W_1^0, \dots, W_m^0)$ is a vector process of m independent Brownian bridges. As in [D], we use a result in [Bi2] (see Corollary 1, p. 31), to obtain

$$(Z_{N_1}^*(t_1, \hat{y}_1) - Z_{N_1}^*(t_1, Y), \dots, Z_{N_m}^*(t_m, \hat{y}_m) - Z_{N_m}^*(t_m, Y)) \xrightarrow{\mathcal{D}} 0.$$

Appealing once more to the same result of [Bi2], we have that

$$\sum_{i=1}^{m} \left| \sqrt{\frac{N_i}{N}} (Z_{N_i}^*(t_i, \cdot) - Z_{N_i}^*(t_i, \hat{y}_i(\cdot))) \right| \xrightarrow{\mathcal{D}} 0,$$

and convergence in distribution to a deterministic limit implies convergence in probability to that limit; so we have the stated result. □

Our next step toward a limiting distribution for the statistic $\widehat{Z}_{\mathcal{N}}$ is to relate it to $Z^*_{\mathcal{N}}$. We begin by noting that $U^*_{i,j} \le \hat{y}_i(y)$ if and only if $X_{i,j} \le x_i(y, q^*)$, which holds if and only if

$$\widehat{U}_{i,j} = U(t_i, X_{i,j}; \hat{q}) \le U(t_i, x_i(y, \hat{q}); \hat{q}) = y,$$

so that

$$F^*_{N_i}(t_i, \hat{y}_i(y)) = \widehat{F}_{N_i}(t_i, y).$$

Thus, we have

$$\begin{aligned}
\widehat{Z}_{\mathcal{N}}(y) &= \sum_{i=1}^{m} \sqrt{\frac{N_i}{N}} \sqrt{N_i}(\widehat{F}_{N_i}(t_i, y) - y) \\
&= \sum_{i=1}^{m} \sqrt{\frac{N_i}{N}} \sqrt{N_i}(F^*_{N_i}(t_i, \hat{y}_i(y)) - y) \\
&= \sum_{i=1}^{m} \sqrt{\frac{N_i}{N}} \left(Z^*_{N_i}(t_i, \hat{y}_i(y)) + \sqrt{N_i}(\hat{y}_i(y) - y) \right);
\end{aligned}$$

so to identify the limit we must find the joint distribution of the two processes in this sum. The behavior of the first term by itself is handled by the previous lemma. We next consider the $\hat{y}_i$ term

$$\begin{aligned}
\frac{N}{\sqrt{N_i}}(\hat{y}_i(y) - y) &= \frac{N}{\sqrt{N_i}} U(t_i, x_i(y, \hat{q}), q^*) - U(t_i, x_i(y, \hat{q}), \hat{q}) \\
&= \frac{\partial U(t_i, x_i(y, \hat{q}), \overline{q})}{\partial q} \frac{N}{\sqrt{N_i}}(q^* - \hat{q}),
\end{aligned}$$

where the components of $\overline{q}$ lie between those of q^* and $\hat{q}$.

Under the assumption (MR), we have that

$$\frac{\partial U(t_i, x_i(y, \hat{q}), \overline{q})}{\partial q} \to \frac{\partial U(t_i, x_i(y, q^*), q^*)}{\partial q},$$

uniformly in y, with probability one, so that the asymptotic distribution of $\hat{y}_i$ is tied to that of the parameter estimate $\hat{q}$. Below we shall use the notation $\varepsilon_{\mathcal{N}}(y)$ to denote a generic process in $D[0, 1]$ that tends to 0 in probability, as $\mathcal{N} \to \infty$. Its actual value may change from equation to equation. We have

$$\begin{aligned}
\frac{N}{\sqrt{N_i}}(\hat{y}_i(y) - y) &= \frac{N}{\sqrt{N_i}} \frac{\partial U(t_i, x_i(y, q^*), q^*)}{\partial q}(q^* - \hat{q}) + \varepsilon_{\mathcal{N}}(y) \\
&= \frac{\partial U(t_i, x_i(y, q^*), q^*)}{\partial q} \frac{N}{\sqrt{N_i}} \left(\frac{\partial^2 J^h_{\mathcal{N}}(q^*)}{\partial q^2} \right)^{-1} \frac{\partial J^h_{\mathcal{N}}(q^*)}{\partial q} + \varepsilon_{\mathcal{N}}(y) \\
&= \frac{\partial U(t_i, x_i(y, q^*), q^*)}{\partial q} \frac{N}{\sqrt{N_i}} \left(\frac{\partial^2 J^h_{\mathcal{N}}(q^*)}{\partial q^2} \right)^{-1} \\
&\quad * \sum_{j=1}^{m} \left\langle \hat{u}(t_j) - P_h u(t_j; q^*), P_h \frac{\partial u(t_j; q^*)}{\partial q} \right\rangle + \varepsilon_{\mathcal{N}}(y).
\end{aligned}$$

We next set

$$g_j^h(x) = \sum_{k=1}^{M_h} \left(\frac{1}{\sqrt{h}} I_{J_k}(x) - u_{jk}^* \right) \left\langle \varphi_k^{(h)}, P_h \frac{\partial u(t_j, q^*)}{\partial q} \right\rangle ,$$

where $u_{jk}^* = P_h u(t_j; q^*)(x)$, $x \in J_k$, and

$$A(y) = -2 \sum_{i=1}^{m} \alpha_i \frac{\partial U(t_i, x_i(y, q^*); q^*)^{\mathrm{T}}}{\partial q} \mathscr{I}^{-1} .$$

Then, we recall that

$$\begin{aligned}
&\sum_{j=1}^{m} \left\langle \hat{u}(t_j) - P_h u(t_j; q^*), P_h \frac{\partial u(t_j, q^*)}{\partial q} \right\rangle \\
&\quad = \sum_{j=1}^{m} \frac{1}{N_j} \sum_{l=1}^{N_j} \left(\sum_{k=1}^{M_h} \left\langle \varphi_k^{(h)}, P_h \frac{\partial u(t_j, q^*)}{\partial q} \right\rangle \left(I_{[X_{j,l} \in J_k]} - \int_{J_k} u(t_j, x; q)\, dx \right) \right) \\
&\quad = \sum_{j=1}^{m} \frac{1}{N_j} \sum_{l=1}^{N_j} g_j^h(X_{j,l}) .
\end{aligned}$$

Thus, we have

$$\begin{aligned}
\widehat{Z}_n(y) &= \sum_{i=1}^{m} \sqrt{\frac{N_i}{N}} \left(Z_{N_i}^*(t_i, \hat{y}_i(y)) + \sqrt{N}_i (\hat{y}_i(y) - y) \right) \\
&= \sum_{i=1}^{m} \sqrt{\frac{N_i}{N}} \left(Z_{N_i}^*(t_i, y) + \sqrt{N_i} (\widehat{y}_i(y) - y) \right) + \varepsilon_{\mathscr{N}}(y) \\
&= \sum_{i=1}^{m} \sqrt{\alpha_i} Z_{N_i}^*(t_i, y) + \alpha_i \sqrt{N} (\hat{y}_i(y) - y) + \varepsilon_{\mathscr{N}}(y) \\
&= \sum_{i=1}^{m} \sqrt{\alpha_i} Z_{N_i}^*(t_i, y) - 2\alpha_i \left[\sum_{j=1}^{m} \alpha_j \frac{\partial U(t_j, x_j(y, q^*); q^*)^{\mathrm{T}}}{\partial q} \mathscr{I}^{-1} \right] \\
&\qquad * \sum_{i=1}^{m} \frac{1}{\sqrt{\alpha_j N_j}} \sum_{l=1}^{N_j} g_j^h(X_{j,l}) + \varepsilon_{\mathscr{N}}(y) ,
\end{aligned}$$

so that

$$\widehat{Z}_N(y) = \sum_{i=1}^{m} \left(\sqrt{\alpha_i} Z_{N_i}^*(t_i, y) + A(y) \frac{1}{\sqrt{\alpha_i N_i}} \sum_{l=1}^{N_j} g_i^h(X_{i,l}) \right) + \varepsilon_{\mathscr{N}}(y) . \qquad (\dagger)$$

To determine the asymptotic behavior of this process, we may disregard the term $\varepsilon_{\mathscr{N}}(y)$, and concentrate on the sum. Moreover, the sum is a sum of independent processes, under (D), so that we may examine each term separately. We then study weak convergence of these processes, following the approach of [D]. Our problem has some subtle differences from the problem discussed in [D], which we shall point out in the proof below. Before we get to the theorem, we need some more notation. We set

$$R_i = \langle \partial u(t_i ; q^*)/\partial q , \mathscr{V}^i \partial u(t_i ; q^*)^{\mathrm{T}}/\partial q \rangle ,$$

and

$$H_i(y) = \int_{x_0}^{x_i(y, q^*)} \frac{\partial u(t_i , x ; q^*)}{\partial q} u(t_i , x ; q^*)\, dx - x_i(y , q^*) \int_{x_0}^{x_1} \frac{\partial u(t_i , x ; q^*)}{\partial q} u(t_i , x ; q^*)\, dx .$$

THEOREM 6.2. Assume that (D), (M), (H), and (MR) hold. Then, for each i,

$$\widehat{Z}_{\mathscr{N}}(\cdot) \xrightarrow{\mathscr{D}} W$$

in $D[0, 1]$, where $\mathscr{N} \to \infty$ and $h \to 0$ in such a way that $hN_i \to \infty$ and $N_i/N \to \alpha_i > 0$, and where W is a zero mean Gaussian process whose covariance is given by

$$\begin{aligned} \mathrm{E}[W(u)W(v)] = \sum_{i=1}^{m} (\alpha_i(\min(u , v) - uv) &+ A(u)H_i(v) \\ &+ A(v)H_i(u) \frac{1}{\alpha_i} A(u)R_i A(v)^{\mathrm{T}}) . \end{aligned} \tag{6.3}$$

Proof. From Theorem 15.5 of [Bi2], we will have the stated convergence if we show that the finite-dimensional distributions converge and that the sequence of processes is tight. The tightness argument is almost identical to that of [D], and so we omit it. We begin by examining finite-dimensional distributions. In a sense, this is the crucial step, because it is at this point that we find the appropriate limiting covariance. This step is also similar to that of [D], but it does have some differences. In particular, the second term in the last line of (†) must be examined in a different way because of the nature of the function g_i^h.

Let K be a positive integer, and let $0 < y_1 < y_2 < \cdots < y_K < 1$ be given. For each i, we will show that the finite-dimensional distributions converge; that is,

$$\begin{pmatrix} \sqrt{\alpha_i} Z^*_{N_i}(t_i , y_1) + A(y_1) \frac{1}{\sqrt{\alpha_i N_i}} \sum_{l=1}^{N_j} g_i^h(X_{i,l}) \\ \vdots \\ \sqrt{\alpha_i} Z^*_{N_i}(t_i , y_K) + A(y_K) \frac{1}{\sqrt{\alpha_i N_j}} \sum_{l=1}^{N_j} g_i^h(X_{i,l}) \end{pmatrix} \xrightarrow{\mathscr{D}} \begin{pmatrix} W_i(y_1) \\ \vdots \\ W_i(y_K) \end{pmatrix} .$$

First, we define

$$d_k^i(x) = \begin{cases} 1 - y_k , & \text{if } U(t_i , x ; q^*) \le y_k , \\ y_k , & \text{if } U(t_i , x ; q^*) > y_k , \end{cases}$$

and we set $d_{kl}^i = d_k^i(X_{il})$ We note that $\mathrm{E}[d_{kl}^i] = 0$, $\mathrm{Cov}(d_{kl}^i , d_{k'l}^i) = \min(y_k , y_{k'}) - y_k y_{k'}$, and that

$$Z^*_{N_i}(t_i , y_k) = \frac{1}{\sqrt{N_i}} \sum_{l=1}^{N_i} d_{kl}^i .$$

Now, we set

$$c_{kl}^{i} = \sqrt{\alpha_i} d_{kl}^{i} + A(y_k) g_i^h (X_{il}) / \sqrt{\alpha_i},$$

and we note that

$$\widehat{Z}_{N_i}(t_i, y_k) = \frac{1}{\sqrt{N_i}} \sum_{l=1}^{N_i} c_{kl}^{i} + \varepsilon_{N_i}(y_k).$$

Again, we have that $\mathrm{E}[c_{kl}^i] = 0$, and direct computations yield

$$\begin{aligned} C_{kk'}^{i,h} &\equiv \operatorname{Cov}(c_{kl}^i, c_{k'l}^i) \\ &= \alpha_i(\min(y_k, y_{k'}) - y_k y_{k'}) + A(y_k)\mathrm{E}[d_{k'l}^i g_i^h(X_{il})] \\ &\quad + A(y_{k'})\mathrm{E}[d_{kl}^i g_i^h(X_{il})] + \frac{1}{\alpha_i} A(y_{k'}) \operatorname{Cov}(g_i^h(X_{il})) A(y_k)^{\mathrm{T}} \\ &= \alpha_i(\min(y_k, y_{k'}) - y_k y_{K'}) + A(y_k) H_i^h(y_{k'}) \\ &\quad + A(y_{k'}) H_i^h(y_k) + \frac{1}{\alpha_i} A(y_{k'}) \operatorname{Cov}(g_i^h(X_{il})) A(y_k)^{\mathrm{T}}, \end{aligned}$$

where

$$H_i^h(y) = \int_{x_0}^{x_i(y;q^*)} g_i^h(x) u(t_i, x; q^*)\, dx.$$

To obtain the stated result, we show first that the covariances converge to the desired limit, and second that the limit is a normal random vector. We have that

$$\operatorname{Cov}(g_i^h(X_{il})) = \left\langle P_h \frac{\partial u(t_i, x; q^*)}{\partial q}, \mathscr{V}^{i,h} \frac{\partial u(t_i, x; q^*)}{\partial q} \right\rangle;$$

so by Lemma 3.2 we have the correct limit for the last term. Next, we note that

$$\begin{aligned} \langle g_i^h, \psi \rangle &= \left\langle \sum_{l=1}^{M_h} \left(\frac{1}{\sqrt{h}} I_{J_l}(\cdot) - u_{il}^* \right), \psi \right\rangle \left\langle \varphi_l^{(h)}, P_h \frac{\partial u(t_i, q^*)}{\partial q} \right\rangle \\ &= \left\langle \sum_{l=1}^{M_h} (\varphi_l^{(h)} - u_{il}^*), \psi \right\rangle \left\langle \varphi_l^{(h)}, P_h \frac{\partial u(t_j, q^*)}{\partial q} \right\rangle \\ &= \sum_{l=1}^{M_h} \left(\langle \varphi_l^{(h)}, \psi \rangle - u_{jk}^* \langle 1, \psi \rangle \right) \left\langle \varphi_k^{(h)}, P_h \frac{\partial u(t_j, q^*)}{\partial q} \right\rangle \\ &= \left\langle P_h \psi, P_h \frac{\partial u(t_j, q^*)}{\partial q} \right\rangle - \left\langle P_h u(t_i; q^*), P_h \frac{\partial u(t_j, q^*)}{\partial q} \right\rangle \langle 1, \psi \rangle \\ &\to C_{kk'}^i \equiv \left\langle \psi, \frac{\partial u(t_j, q^*)}{\partial q} \right\rangle - \left\langle u(t_i; q^*), \frac{\partial u(t_j, q^*)}{\partial q} \right\rangle \langle 1, \psi \rangle, \end{aligned}$$

as $h \to 0$. Setting

$$\psi(x) = I_{[x_0, x_i(y_k; q^*)]}(x) u(t_i, x; q^*)$$

gives the correct limit covariance for the H_i^h terms.

To establish that the limit random vector is normally distributed, we note that the random vectors $\vec{c}_l^i = (c_{1l}^i, \dots, c_{Kl}^i)$, $1 \le l \le N_i$, are independent, zero mean random vectors with covariance $C^{i,h}$, which converges to C^i as $h \to 0$. A straightforward application of Lindeberg's central limit theorem gives us the desired result. □

Finally, we consider the goodness-of-fit statistic

$$W_{\mathcal{N}}^2 = \int_0^1 |\widehat{Z}_{\mathcal{N}}(y)|^2 \, dy,$$

which is based on the classical Cramér-von Mises statistic (see, for example [SW]). As in Theorem 4.3, we appeal to Theorem 5.1 of [Bi2] to obtain the following result.

THEOREM 6.3. Assume that the conditions of Theorem 4.2 hold. Then

$$W_{\mathcal{N}}^2 \xrightarrow{\mathcal{D}} \int_0^1 W^2(y) \, dy,$$

where W is a zero mean Gaussian process with covariance given by (6.3). □

We close this section with a description of how one might use Theorem 6.3 to test for goodness of fit of a structured model.

1. Transform the data as in (6.1).
2. Compute $\widehat{Z}_{\mathcal{N}}$ as in (6.2).
3. Compute $W_{\mathcal{N}}^2$.
4. Compute the eigenvalues λ_l of the covariance kernel given in (6.3).
5. Compute the inverse Fourier transform of the characteristic function

$$\psi(t) = \prod_{l=1}^{\infty} (1 - 2i\lambda_l t)^{-1/2},$$

to obtain the probability density of $\int_0^1 W^2(y) \, dy$.

This approach has been applied in similar situations (see [SW, p. 212]). We remark that several approximations have to be made in this procedure. First, the covariance depends on the unknown parameter. One might plug in the least squares estimator. Second, the covariance will have an infinite number of eigenvalues. Third, these eigenvalues must be found numerically. The density obtained, then, from the above procedure must be regarded as approximate.

7. Conclusions. We have presented in this paper an alternate approach to least squares estimation in structured population applications. The usual histogram data is scaled to obtain a nonparametric estimate of the density of the population. Scaling the structured model leads to a postulated parametric density, which can then be compared to the nonparametric estimate through adaptations of classical statistical goodness-of-fit tests.

It is important to note that we have not considered the impact of approximating the parametric model density. Such approximations are necessary in the many applications in which the model density arises from a differential equation. We plan in future studies to tackle this problem. We would expect the results of the present paper to remain valid for many of the usual approximation schemes (see [BF2] for similar approximation results in nonlinear regression situations).

We should also point out that the above statistical theory could be adapted to other forms of data representation. In particular, a commonly used nonparametric density estimation approach, based on kernels (see, e.g., [PR]), would be employed, and we would expect to obtain theorems of a very similar nature.

REFERENCES

[Ba] H. T. Banks, *On a variational approach to some parameter estimation problems*, Distributed Parameter Systems, Lecture Notes in Control and Information Sciences, vol. 75 (F. Kappel, K. Kunisch, and W. Schappacher, eds.), Springer-Verlag, Berlin, 1985, pp. 1–23

[BBKW1] H. T. Banks, L. Botsford, F. Kappel, and C. Wang, *Modeling and estimation in size structured population models*, Mathematical Ecology (Proc., Trieste, 1986), World Sci. Publ., Singapore, 1988, pp. 521–541

[BBKW2] H. T. Banks, L. Botsford, F. Kappel, and C. Wang, *Estimation of growth and survival in size-structured cohort data: An application to larval striped bass (Morone Saxatilis)*, J. Math. Biology **30**, 125–150 (1991)

[BF1] H. T. Banks and B. G. Fitzpatrick, *Estimation of growth rate distributions in size structured population models*, Quart. Appl. Math. **49**, 215–235 (1991)

[BF2] H. T. Banks and B. G. Fitzpatrick, *Statistical tests for model comparison in parameter estimation problems for distributed parameter systems*, J. Math. Biology **28**, 501–527 (1990)

[BK] H. T. Banks and K. Kunisch, *Estimation Techniques for Distributed Parameter Systems*, Birkhäuser, Boston, MA, 1989

[Bi1] P. Billingsley, *Probability and Measure*, Wiley, New York, 1979

[Bi2] P. Billingsley, *Convergence of Probability Measures*, Wiley, New York, 1968

[BVWLKRC] L. Botsford, B. Vandracek, T. Wainwright, A. Linden, R. Kope, D. Reed, and J. J. Cech, *Population development of the mosquitofish, Gambusia Affinis, in rice fields*, Environmental Biology of Fishes **20**, 143–154 (1987)

[B] N. Dunford and J. T. Schwartz, *Linear Operators*, I: *General Theory*, Wiley-Interscience, New York, 1958

[D] J. Durbin, *Weak convergence of the sample distribution function when parameters are estimated*, Ann. Statist. **1**, 279–290 (1973)

[EK] S. N. Ethier and T. G. Kurtz, *Markov Processes: Characterization and Convergence*, Wiley, New York, 1986

[KS] M. Kendall and A. Stuart, *The Advanced Theory of Statistics*, vol. 2, 4th ed., Griffin, London, 1979

[L] A. F. Low, 1986 *Striped bass egg and larva survey in the Sacramento-San Joaquin Estuary*, California Department of Fish and Game, Sacramento, CA, 1986

[PR] B. L. S. Prakasa Rao, *Nonparametric functional estimation*, Academic Press, Orlando, FL, 1983

[S] A. N. Schumitzky, *Nonparametric EM algorithms for estimating prior distributions*, Appl. Math. Comput. **45**, 143–157 (1991)

[SW] G. R. Shorack and J. A. Wellner, *Empirical processes with applications to statistics*, Wiley, New York, 1986