*Econometric Theory*, **25**, 2009, 1070–1086. Printed in the United States of America.
doi:10.1017/S0266466608090415

# THE LIMITS OF ECONOMETRICS: NONPARAMETRIC ESTIMATION IN HILBERT SPACES

GRACIELA CHICHILNISKY
*Columbia University*

We extend Bergstrom's 1985 results on nonparametric (NP) estimation in Hilbert spaces to unbounded sample sets. The motivation is to seek the most general possible framework for econometrics, NP estimation with no a priori assumptions on the functional relations nor on the observed data. In seeking the boundaries of the possible, however, we run against a sharp dividing line, which defines a *necessary and sufficient condition* for NP estimation. We identify this condition somewhat surprisingly with a classic statistical assumption on the relative likelihood of bounded and unbounded events (DeGroot, 2004). Other equivalent conditions are found in other fields: decision theory and choice under uncertainty (monotone continuity axiom (Arrow, 1970), insensitivity to rare events (Chichilnisky, 2000), and dynamic growth models (dictatorship of the present; Chichilnisky, 1996). When the crucial condition works, NP estimation can be extended to the sample space $R^+$. Otherwise the estimators, which are based on Fourier coefficients, do not converge: the underlying distributions are shown to have "heavy tails" and to contain purely finitely additive measures. Purely finitely additive measures are not constructible, and their existence has been shown to be equivalent to the axiom of choice in mathematics. Statistics and econometrics involving purely finitely additive measures are still open issues, which suggests the current limits of econometrics.

## DEDICATION

*This paper is in honor of a great man, and attempts to complete a conversation that was left pending for 27 years.*

## 1. INTRODUCTION

In 1985, Rex Bergstrom (1985) constructed a nonparametric (NP) estimator for nonlinear models with bounded sample spaces, using techniques of Hilbert

This paper is based on an invited address at a conference that took place at the University of Essex in May 2006, honoring Rex Bergstrom's memory. I thank the participants for valuable comments and particularly Peter Phillips, the organizer. I am also grateful to William Barnett of the University of Kansas and the participants of a Statistics Department seminar at Columbia University and of the Econometrics Seminar at the Economics Department of Columbia University, particularly Victor De La Pena, Chris Heyde, Marc Henry, and Dennis Kristensen, for helpful comments and suggestions. Address correspondence to Graciela Chichilnisky, 335 Riverside Drive, New York, NY, 10025 USA; email: chichilnisky1@gmail.com.

 © 2009 Cambridge University Press 0266-4666/09 $15.00

https://doi.org/10.1017/S0266466608090415 Published online by Cambridge University Press

spaces,[1] a type of space I introduced in economics (Chichilnisky, 1976, 1977). His article is simple, elegant, and general, but requires an a priori bound on observed data that conflicts with the spirit of NP estimation. This article extends his original results to *unbounded* sample spaces such as the positive real line $R^+$,[2] using earlier work (Chichilnisky, 1976, 1977, 1996, 1997, 1999, 2000, 2002, 2006a, 2006b).

We focus on the statistical assumptions needed for the extension of NP estimation from bounded intervals to the positive real line $R^+$, and find a sharp dividing line, a condition that is both necessary and sufficient for extending NP estimation from bounded intervals to the entire real line. The clue to this condition appeared in the literature as a classic statistical assumption that restricts the asymptotic behavior of the unknown function, and derives from a classical assumption on relative likelihoods. In DeGroot (2004) the condition is denoted $SP_4$, and it compares the likelihood of bounded and unbounded sets. A simple interpretation for this condition is that, no matter how small is a set $B \succ \varnothing$, *it is impossible for every infinite interval* $(n, \infty)$ *to be at least as likely as* $B$.[3] In practical terms, the condition requires that unbounded sets are eventually less likely than *any* bounded set. To check the condition in practice, one examines the relative likelihood of any bounded set $B$ and compares it with infinite sets of the form $(n, \infty)$. Eventually, for large enough $n$, the set $B$ must be more likely than $(n, \infty)$. As a practical example, any continuous integrable density function on the line $f : R^+ \to R$ defines a relative likelihood that satisfies this condition. And as shown below, this condition eliminates "heavy tails."

In order to generalize the NP estimation problem as much as possible, we extend Assumption $SP_4$, which was originally defined only for *density functions*, to *any continuous unknown function* $f : R^+ \to R$. We show that when $SP_4$ is satisfied, the unknown function can be represented by a function in a Hilbert space and the NP estimator can be extended appropriately to $R^+$. But when assumption $SP_4$ fails, the situation is quite different. The estimator does not have appropriate asymptotic behavior at infinity. It appears that a classic statistical assumption holds the cards for extending NP estimation to unbounded sample spaces.

Exploring other areas of the literature, we find other assumptions that we prove to be equivalent to $SP_4$ and in that sense determinant for extending NP estimation to $R^+$. In decision theory one such assumption is the monotone continuity axiom of Arrow (1970); another is the insensitivity to rare events of Chichilnisky (2000, 2006a, 2006b). In optimal growth models it is dictatorship of the present as defined in Chichilnisky (1996). When the key assumptions fail, the estimator does not converge. In all cases, the failure leads to purely finitely additive measures on $R^+$, and to distributions with heavy tails. Econometric results involving purely additive measures are still an open issue, which suggests the current limits of econometrics.

Results extending semi-NP estimation and NP estimation to infinite cases, dealing with separate but related issues, can be found, e.g., in Blundell, Chen, and Kristensen (2006), Stinchcombe (2002), and Chen, Hansen, and Sheinkman (2005),

https://doi.org/10.1017/S0266466608090415 Published online by Cambridge University Press

among others. Other articles in the NP literature include Andrews (1991) and Newey (1997), both of which again assume a compact support for the regressor.

## 2. HILBERT SPACES

The methodology we use here is weighted Hilbert spaces as defined in Chichilnisky (1976, 1977), two publications that introduced Hilbert spaces in economics.

These earlier results suggested the advantages of using Hilbert spaces in econometrics, in particular for NP estimation.[4] The rationale is simple: NP estimation is by nature infinite-dimensional, because when the forms of the functions in the true model are unknown, the most efficient use of the data is to allow the estimated functions (or the number of estimated parameters) to depend on the size of the sample, tending to infinity with the sample size. This provides a natural infinite-dimensional context for NP estimation. In this context, Hilbert spaces are a natural choice, because they are the closest analog to euclidean space in infinite dimensions.

Bergstrom pointed out in personal communication (see Note 1) that there is a natural limitation for the use of Hilbert spaces on the real line $R$. Standard Hilbert spaces such as $L_2(R)$ require that the unknown function approaches zero at infinity, a somewhat unreasonable limitation to impose on the economic model, as it excludes widely used functions such as constant, increasing, and cyclical functions on the line. To overcome his objection, I suggested using weighted Hilbert spaces, since these impose weaker limiting requirements at infinity, as shown below. Bergstrom's article (1985) acknowledged my contribution to NP estimation in Hilbert spaces, but it is restricted to bounded sample spaces: his results apply to $L_2$ spaces of functions defined on a bounded segment of the line, $[a, b] \subset R$.

Below, we extend the original methodology in Bergstrom (1985) to *unbounded* sample spaces by using weighted Hilbert spaces as originally proposed. In exploring the viability of the proofs, we run into an interesting dilemma. When the sample space is the entire positive real line, Hilbert space techniques still require additional conditions on the asymptotic behavior of the unknown function at infinity. In bounded sample spaces such as $[a, b]$, this problem did not arise, because the unknown are continuous, and therefore bounded, and belong to the Hilbert space $L_2[a, b]$. But this is not the case when the sample space is the positive line $R^+$. A continuous real-valued function on $R^+$ may not be bounded and may not be in the space $L_2(R^+)$.[5] Therefore, the Fourier series expansions that are used for defining the estimator may not converge. With unbounded sample spaces, additional statistical assumptions are needed for NP estimation.

Consider the problem of estimating an unknown function $f$ on $R^+$, for example, a capital accumulation path through time or a density function, which are standard nonlinear NP estimation problems. The unknown density function may be continuous, but not a square integrable function on $R^+$, namely, an element of $L_2(R^+)$. Since the NP estimator is defined by approximating values of the Fourier

https://doi.org/10.1017/S0266466608090415 Published online by Cambridge University Press

coefficients of the unknown function (Bergstrom, 1985), when the Fourier coefficients of the estimator do not converge, the estimator itself fails to converge. A similar situation arises in general NP estimation problems where the unknown function may not have the asymptotic behavior needed to ensure the appropriate convergence. This illustrates the difficulties involved in extending NP estimation in Hilbert spaces from bounded to unbounded sample spaces.

The rest of this article focuses on the statistical necessary and sufficient conditions needed for extending the results from bounded intervals to the positive line $R.^{+}$

## 3. STATISTICAL ASSUMPTIONS AND NP ESTIMATION

A brief summary of earlier work follows. Bergstrom's statistical assumptions require that the unknown function $f$ be continuous and bounded a.e. on the sample space $[a, b] \in R.^{6}$ His sample design assumes *separate observations at equidistant points*. The number of parameters increases with the size of the sample space, and disturbances are not necessarily normal.

Bergstrom uses an *orthogonal series* in Hilbert space to derive NP properties and prove convergence theorems. The series is orthonormal in the Hilbert space rather than in the sample space, so the elements remain unchanged as the sample size increases. This series includes polynomials or any dense family of orthonormal functions in the Hilbert space.

An estimator $\hat{f}$ is defined in a simple and natural manner (Bergstrom, 1985): the first $M$ Fourier coefficients of the unknown function $f$ are estimated relative to the orthonormal set. Estimates of the coefficients are obtained from the sample by ordinary least squares regression, setting the rest of the Fourier coefficients to zero. Bergstrom (1985) shows that $\mathrm{E}\left[\int_a^b \{\hat{f}_{MN}(x) - f(x)\}\,dx\right]$ can be made arbitrarily small by a suitable choice of $M$ and $N$. He also defines an estimator $f_N^*(x)$ that is optimal for the sample size $N$, and shows that this converges in a given metric to $f(x)$ as $N \to \infty$. A third theorem shows how an optimal value of $M^*$ is related to the Fourier coefficients and the mean square errors of their estimates obtained from regressions with various values of $M$, and provides the basis for an estimation procedure. A "stopping rule" is also provided for estimating the optimum value of the parameter that, for a given sample, provides the exact number of Fourier coefficients to be estimated. The definition of the estimator and the proofs of these results, require that $f$ be an element of a Hilbert space.

Unbounded sample spaces give rise to a different type of problem. To explain the problem and motivate the results, we first explain why earlier work was restricted to bounded sample spaces.

## 4. WHY BOUNDED SAMPLE SPACES?

When working in Hilbert spaces, there are good technical reasons for requiring that the sample space be bounded. For example, consider the typical Hilbert

https://doi.org/10.1017/S0266466608090415 Published online by Cambridge University Press

space of functions $L_2$, the space of square integrable measurable real-valued functions. Bergstrom (1985) considered the space $L_2[a,b]$ of functions defined on the bounded segment $[a,b] \subset R$, where $[a,b]$ represents the sample space. As he points out, there is no need to assume anything further than continuity for the unknown function.[7] Every continuous unknown function on the segment $[a,b]$ is bounded and belongs to the Hilbert space $L_2[a,b]$.

However, when the sample space is unbounded, such as $R^+$, the square integrability condition of being in $L_2(R^+)$ function imposes significantly more restrictions. For example, for continuous functions of bounded variation, it requires that the functions to be estimated have a well-defined limit at infinity, such as $\lim_{t\to\infty} f(t) = 0$. This is not a reasonable restriction to impose on the unknown function if, for example, the function represents capital accumulation, which typically increases over time. The restriction on $\lim_{t\to\infty} f(t)$ also eliminates other standard cases, such as constant, increasing, or cyclical functions.

In mathematical terms, an appropriate transformation of the line can alleviate the problem. This was the methodology introduced in Chichilnisky (1976, 1977), the first publications to use Hilbert spaces in economics. I defined then a Hilbert space $L_2(R^+)$ with a "weight function" $\overline{\gamma}(t)$ that defines a *finite* density measure for $R^+$.[8] In this case, square integrability requires far less, only that the product of the function times the weight function, $f(t)\overline{\gamma}(t)$, converges to zero at infinity, rather then the function $f(t)$ itself. This is a more reasonable assumption, which is asymptotically satisfied by the solutions in most optimal growth models, where there is a well-defined "discount" factor $\gamma > 1$. The solution I considered was the (weighted) Hilbert space $L_2(R^+, \gamma)$ of all measurable functions $f$, for which the absolute value of the discounted product $f(t)e^{-\gamma t}$ is square integrable (Chichilnisky, 1976, 1977). As already stated, this does not require $\lim_{t\to\infty} f(t) = 0$, and it includes bounded, increasing, and cyclical real-valued functions on $R^+$.[9] It is of course possible to include other weight functions as part of the methodology introduced in Chichilnisky (1976, 1977), provided the weight functions are monotonically decreasing and therefore invertible, but the ones specified in Chichilnisky (1976, 1977) are naturally associated with the models at hand. This solution is an improvement, but the condition that the unknown function belongs to a Hilbert space *still* poses asymptotic restrictions at infinity, which are considered below.

In the case of optimal growth models in Chichilnisky (1977), the methodology of weighted Hilbert spaces is based on a transformation map induced by the model itself, its own discount factor $\overline{\gamma} : R^+ \to [0,1)$, $\overline{\gamma}(t) = e^{-\gamma t}$. Under this transformation, the unbounded sample space $R^+$ is mapped into the bounded sample space $[0,1)$ where the original assumptions and results for bounded sample spaces can be reinterpreted appropriately in a bounded sample space. This is the route followed in this paper.

Before doing so, however, it seems worth discussing briefly a different methodology that has been suggested for NP estimation with unbounded sample spaces,[10] explaining why it may be less suitable.

https://doi.org/10.1017/S0266466608090415 Published online by Cambridge University Press

## 5. COMPACTIFYING THE SAMPLE SPACE

A natural approach to extend NP estimation to unbounded sample spaces would be to compactify the sample space and apply the existing results to the compactified space. For example, the compactification of the positive real line $R^+$ yields a space that is equivalent to a bounded interval $[a, b]$. To proceed with NP estimation, one needs to reinterpret every function $f : R^+ \to R$ as a function defined on the compactified space, $\widetilde{f} : \widetilde{R} \to R$. As we see below, this requires from the onset that the function $f$ on $R$ has a well-defined limiting behavior at infinity, namely, $\lim_{t\to\infty} f(t) < \infty$. Otherwise, $f$ cannot be extended to a function on the compactified space. To lift this constraint, Peter Phillips suggested that one could *estimate* (rather than assume) the behavior of the unknown function at infinity.[11] But in all cases, *some* limit must be assumed for the unknown function, which can be considered an unrealistic requirement. The following example shows why.

Consider the Alexandroff one-point compactification of the real line $R^+$, which consists of "adding" to the real numbers a point of infinity $\{\infty\}$ and defining the corresponding neighborhoods of infinity. This is a frequently-used technique of compactification. A function $f$ on the line $R$ can be extended to a function on the compactified line, but only if $f$ has a well-defined limiting behavior at infinity, namely, if there exists a well-defined $\lim_{t\to\infty} f(t)$. This is not always possible nor a reasonable restriction to impose; for example, this requirement excludes all cyclical functions, for which $\lim_{t\to\infty} f(t)$ does not exist.

One can explore more general forms of compactification, such as the Stone-Cech compactification of the line $\widehat{R}$, the most general possible compactification of the real line.[12] $\widehat{R}$ is a well-behaved Hausdorff space and is a universal compactifier of $R$, which means that every other compactification of $R$ is a subset of it. Any function $f : R \to R$ can be extended to a function on the compactified space, $\widehat{f} : \widehat{R} \to R$. However, it is difficult to interpret Hilbert spaces of functions defined on $\widehat{R}$, since these would be square integrable functions defined on ultrafilters rather than on real numbers. Such spaces do not have a natural interpretation.

To overcome these difficulties, in the following we use weighted Hilbert spaces for NP estimation on unbounded samples spaces.

## 6. NP ESTIMATION ON WEIGHTED HILBERT SPACES

Following Chichilnisky (1976, 1977) consider the sample space $R^+ = [0, \infty)$ with a standard $\sigma$ field and a finite density or "weight function" $\overline{\gamma} : R^+ \to [1, 0)$, $\overline{\gamma}(t) = e^{-\gamma t}$, $\gamma > 1$, $\int_{R^+} e^{-\gamma t} dt < \infty$. Define the weighted Hilbert space $H_\gamma$, also denoted $H$, consisting of all measurable and square integrable functions $g(.) : R^+ \to R$ with the weighted $L_2$ norm $\| \cdot \|$:

$$\| g \|_2 = \left( \int_{R^+} g^2(t) e^{-\gamma t} dt \right)^{1/2}.$$

https://doi.org/10.1017/S0266466608090415 Published online by Cambridge University Press

Observe that the space $H$ contains the space of bounded measurable functions $L_\infty(R^+)$ and includes all periodic and constant functions, as well as many increasing functions.

The weight function $\overline{\gamma}$ induces a homeomorphism, namely, a bicontinuous one to one and onto transformation, between the positive real line and the interval $[1, 0)$, $\overline{\gamma} : R^+ \to [0, 1)$, $\overline{\gamma}(t) = e^{-\gamma t}$. In the following, we use a modified homeomorphism $\delta : R^+ \to [0, 1)$, defined as $\delta(t) = 1 - \overline{\gamma}(t) \in [0, 1)$ to maintain the standard order of the line. The transformation $\delta$ allows us to translate Bergstrom's 1985 methodology, assumptions, and notation, which are valid for $[0, 1)$, to the positive real line $R^+$. The following section interprets the statistical assumptions in Bergstrom (1985) for NP estimation in this new context and introduces new statistical assumptions.

## 7. STATISTICAL ASSUMPTIONS AND RESULTS ON $R^+$

We have a sample of $N$ paired observations $(t_1, y_1), \ldots, (t_N, y_N)$ in which $t_1, \ldots, t_N$ are nonrandom positive real numbers whose values are fixed by the statistician and $y_1, \ldots, y_N$ are random variables whose joint distribution depends on $t_1, \ldots, t_N$. In particular, it is assumed that $\mathrm{E}(y_i) = f(\delta(t_i)) = f(x_i)$, $(i = 1, \ldots, N)$, where $f$ is an unknown function, $\delta : R^+ \to [0, 1)$ is the one to one transformation defined above, and $x_i = \delta(t_i)$.

We are concerned with estimating an unknown function $g : R^+ \to R$ over the sample space $R^+$ or, equivalently, estimating the function $f$ over the bounded interval $[0, 1)$ defined by $f(.) = (g \circ \delta^{-1})(.) = g(\delta^{-1}(.)) : [0, 1) \to R$. The model is precisely described by Assumptions 1 to 4 below, which are a transformed version of the Assumptions in Bergstrom (1985, Sect. 2, p. 11). We also require a new statistical assumption, Assumption 3 below, which is needed due to the unbounded nature of our sample space.[13]

Observe that, given the properties of the transformation map $\delta$, it is statistically equivalent to work with the nonrandom variables $t_1, \ldots, t_N$ or, instead, with the *transformed* nonrandom variables $x_1 = x_1(t_1), \ldots, x_N = x_N(t_N)$. To simplify the comparison with Bergstrom's (1985) results, it seems best to use the latter variables when describing the statistical model.

**Assumption 1** (Sampling assumption). The observable random variables $y_1, \ldots, y_N$ are assumed to be generated by the equations

$$y_i = f(x_i) + u_i = f(\delta(t_i)) + u_i,$$

where

$$x_i = a + \frac{b-a}{2N} \quad \text{and}$$

$$x_{i+1} = x_i + \frac{b-a}{N}, \qquad i = 1, \ldots, N-1,$$

https://doi.org/10.1017/S0266466608090415 Published online by Cambridge University Press

Here, $a, b$ are constants $(1 \geqslant b > a \geqslant 0)$, and $u_1, \ldots, u_N$ are unobservable random variables[14] satisfying the conditions

$$\mathrm{E}(u_i) = 0 \qquad (i = 1, \ldots, N)$$

$$\mathrm{E}(u_i^2) = \sigma^2 \qquad (i = 1, \ldots, N)$$

$$\mathrm{E}(u_i, u_j) = 0 \quad \text{when } i \neq j \text{ and } i, j = 1, \ldots, N.$$

**Assumption 2.** The unknown function $g : R^+ \to R$ is continuous or, equivalently, the "transformed" function $f : [0, 1) \to R$ defined by $f(.) = go\delta^{-1}(.) : [0, 1) \to R$, is continuous.

When the domain of a function $f$—namely the sample space—is the closed bounded interval $[0, 1]$ then, being continuous, $f$ is bounded and $f \in L_2[0, 1]$ as pointed out in Bergstrom (1985, p. 11). One may therefore apply Hilbert spaces techniques for NP estimation.

In our case, the (transformed) function $f$ is defined over the (half open) interval $[0, 1)$. Under appropriate boundary conditions, $f$ can be extended to the closed interval $[0, 1]$. Continuity over the closed bounded interval implies boundedness, and furthermore, it ensures that $f \in L_2[0, 1]$. But this is no longer true when the sample space is the positive real line $R^+$, or, equivalently, the transformed sample space is $\delta(R^+) = [0, 1)$. A continuous function defined on $R^+$ may not be bounded and may not belong to $L_2(R^+)$.[15] For the unbounded sample space $R^+$, we require the following additional statistical assumption on the unknown function:

**Assumption 3.** The unknown function $g : R^+ \to R$ is in the Hilbert space $H$ or, equivalently, the transformed function $f : [0, 1) \to R$ can be extended to a continuous function $f : [0, 1] \to R$.

**Assumption 4.** The countable set of continuous functions $\phi_1(x(t)), \ldots, \phi_N(x(t))$ is a complete orthonormal set in the space $L_2(R^+)$ of square integrable functions on $R^+$ with ordinary Lebesgue measure $\mu$.

This requires that the functions $\phi_j$ be continuous, linearly independent, dense in $L_2(R^+)$, and satisfy the conditions

$$\int_0^1 \phi_j^2(x)\, dx = 1 \qquad (j = 1, 2, \ldots)$$

and

$$\int_0^1 \phi_j(x).\phi_i(x)\, dx = 0 \qquad (j \neq i; j, i = 1, 2, \ldots).$$

Observe that one can consider different orthonormal sets; for example, Bergstrom (1985) considers an orthonormal set consisting of polynomials of increasing order.

https://doi.org/10.1017/S0266466608090415 Published online by Cambridge University Press

On the basis of Assumptions 1, 2, 3, and 4, the following results, which are reproduced from Bergstrom (1985), obtain directly from those of Bergstrom (1985). These results are expressed in the transformed unknown function $f : [0, 1] \to R$ to facilitate comparison with Bergstrom (1985) but can be equivalently expressed on the unknown function $g : R^+ \to R$.

THEOREM 1 (Bergstrom, 1985). *Let* $\hat{f}_{MN}(x)$ *be defined by*

$$\hat{f}_{MN}(x) = \hat{c}_1(M, N)\phi_1(x) + \cdots + \hat{c}_M(M, N)\phi_M(x), \tag{1}$$

*where*

$$\hat{c}_1(M, N), \ldots, \hat{c}_M(M, N)$$

*are the values of*

$$c_1, c_2, \ldots, c_M$$

*that minimize*

$$\sum_{i=1}^{N} \{y_1 - c_1\phi_1(x) - \cdots - c_M\phi_M(x)\}^2;$$

*i.e., there are sample regression coefficients. Then for an arbitrarily small real number* $\varepsilon > 0$, *there is an integer* $M_\varepsilon$ *and a function* $N_\varepsilon(M)$ *such that*

$$\mathrm{E}\left[\int_a^b \hat{f}_{MN}(x) - f(x)\,dx\right] < \varepsilon$$

*if* $M > M_\varepsilon$ *and* $N > N_\varepsilon(M)$.

THEOREM 2 (Bergstrom, 1985). *Let* $M^*$ *be the smallest integer such that*

$$\mathrm{E}\left[\int_0^1 \{\hat{f}_{M^*N}(x) - f(x)\}^2\,dx\right] \le \mathrm{E}\left[\int_0^1 \{\hat{f}_{MN}(x) - f(x)\}^2\,dx\right] \quad (M = 1, \ldots, N),$$

*where* $\hat{f}_{MN}(x)$ *is defined for* $M = 1, \ldots, N$ *by (1), and let* $f_N^*(x)$ *be defined by*

$$f_N^*(x) = \hat{f}_{M^*N}(x).$$

*Then under Assumptions 1–4,*

$$\lim_{N\to\infty}\left[\int_0^1 \{f_N^*(x) - f(x)\}^2\,dx\right] = 0.$$

DEFINITION 1. *Let* $c_1, \ldots, c_n$ *be the Fourier coefficients of* $f(x)$ *relative to the orthonormal set* $\phi_1, \ldots, \phi_n$ *in the transformed set* $[0, 1]$, *namely,*

$$c_j = \int_0^1 f(x)\phi_j(x)\,dx \quad (j = 1, 2, \ldots).$$

https://doi.org/10.1017/S0266466608090415 Published online by Cambridge University Press

Observe that under the conditions the set $\phi_1, \ldots, \phi_n$ is orthonormal and therefore complete in $L_2[0, 1]$ so that

$$\lim_{M\to\infty} \int_0^1 \left\{ f(x) - \sum_{j=1}^{M} c_j \phi_j(x) \right\}^2 dx = 0,$$

and Parseval's equality is satisfied (Kolmogorov and Fomin, 1961, p. 98):

$$\int_0^1 f^2(x)\, dx = \sum_{j=1}^{\infty} c_j^2.$$

THEOREM 3 (Bergstrom, 1985). *Under Assumptions 1–4,*

$$\sum_{j=M+1}^{M^*} c_j^2 \geq \sum_{j=1}^{M^*} \mathrm{E}(\widehat{c}_j(M^*, N) - c_j)^2 - \sum_{j=1}^{M} \mathrm{E}(\widehat{c}_j(M, N) - c_j)^2 \quad (M = 1, \ldots, M^* - 1)$$

$$\sum_{j=M^*+1}^{M} c_j^2 \leqslant \sum_{j=1}^{M} \mathrm{E}(\widehat{c}_j(M, N) - c_j)^2 - \sum_{j=1}^{M^*} \mathrm{E}(\widehat{c}_j(M^*, N) - c_j)^2 \quad (M = M^* + 1, \ldots, N).$$

From the definition of the transformation $\delta$ and Assumptions 1 to 4, the proofs for the theorems above follow directly from Theorems 1, 2, and 3 in Bergstrom (1985).

Theorems 1, 2, and 3 are quite general, but the underlying assumptions (1 to 4) still require interpretation for the case of unbounded sample spaces. The following section tackles this issue.

## 8. STATISTICAL ASSUMPTIONS ON $R^+$

Assumptions 1, 2, and 4 have a ready interpretation in the transformed sample space. Assumption 3 is, however, of a different nature. It requires that the unknown function $g : R^+ \to R$ be an element of a (weighted) Hilbert space or, equivalently, that the transformed unknown function $f : [0, 1) \to R$ can be extended to a continuous function in the Hilbert space $L_2[0, 1]$. This condition is critical: when Assumption 3 is satisfied, Theorems 1, 2, and 3 extend NP estimation to $R^+$, but otherwise these theorems, which depend on the properties of $L_2$ functions and the convergence of their Fourier coefficients, no longer work. What conditions are needed to ensure that Assumption 3 holds?[16] The following provides classical statistical conditions involving *relative likelihoods*, cf. DeGroot (2004, Ch. 6).

THEOREM 4. *If a relative likelihood $\preceq$ satisfies assumptions $SP_1$ to $SP_5$ of DeGroot (2004, Ch. 6), then there exists a probability function $f : R^+ \to R$ representing the relative likelihood $\preceq$, where $f$ is an element of the Hilbert space $L_2(R^+)$ and Assumption 3 above is satisfied.*

https://doi.org/10.1017/S0266466608090415 Published online by Cambridge University Press

**Proof.** Consider the five assumptions $SP_1, \ldots, SP_5$ provided in DeGroot (2004, Ch. 6). Together they imply the existence of a countably additive probability measure on $R^+$ that agrees with the relative likelihood order $\gtrsim$ (cf. DeGroot, 2004, Sect. 6.4, pp. 76–77). Given any countably additive measure $\mu$ on $R$, one can always find a functional representation as a measurable function, $f : R^+ \to R$, that is integrable, $f \in L_1(R^+)$, and satisfying $\mu(A) = \int_A f(x)\,dx$ (Yosida and Hewitt, 1952). In other words, the five assumptions $SP_1, \ldots, SP_5$ guarantee the existence of an absolutely continuous distribution representing the "relative likelihood of events" (DeGroot, 2004).

Since the space of integrable functions on the (positive) real line is contained in the space of square integrable functions on the (positive) real line, $L_1(R^+) \subset L_2(R^+)$ (Yosida and Hewitt, 1952), it follows, under the assumptions, that $f \in L_2(R^+)$, as we wished to prove. Thus the five statistical assumptions of DeGroot (2004) suffice to guarantee Assumption 3, and hence the results of Theorems 1, 2, and 3. ■

Among the five fundamental statistical assumptions of DeGroot (2004), there is one, $SP_4$, that plays a key role: it is necessary and sufficient to extend the NP estimation results to unknown density functions on $R^+$. The next step is to define assumption $SP_4$ and explain its role. The notation $A \succeq B$ indicates that the likelihood of the set or event $A$ is higher than the likelihood of $B$ (see DeGroot, 2004).

DEFINITION 2 (Assumption $SP_4$; DeGroot, 2004). *Let $A_1 \supset A_2 \supset \ldots$ be a decreasing sequence of events, and $B$ some fixed event such that $A_i \gtrsim B$ for $i = 1, 2 \ldots$. Then $\bigcap_{i=1}^{\infty} A_i \gtrsim B_i$.*

To clarify the role of $SP_4$, suppose that each infinite interval of the form $(n, \infty) \subset R$, $n = 1, 2, \ldots$ is regarded as more likely (by the relative likelihood) than some fixed small subset $B$ of $R$. Since the intersection of all these intervals is empty, $B$ must be equivalent to the empty set $\phi$. In other words, if $B$ is more likely than the empty set, $B \succ \phi$, then, regardless of how small $B$ is, it is impossible for every infinite interval $(n, \infty)$ to be as likely as $B$. One way to interpret the role of Assumption $SP_4$ is in averting heavy tails.

DEFINITION 3. *We say that a relative likelihood $\gtrsim$ has heavy tails when, for any given set $B$, there exist an $n > 0$ and a set $C \supset (n, \infty)$, such that $C \gtrsim B$; namely, $C$ is as likely as $B \subset R^+$.*[17]

Intuitively, this definition states that there exist infinite intervals or "tail sets" of the form $(n, \infty)$ with arbitrarily large measure, which may be interpreted as heavy tails.

THEOREM 5. *When assumption $SP_4$ fails, relative likelihoods have heavy tails.*

**Proof.** The logical negation of $SP_4$ implies that there exists a large enough $n$ such that $(n, \infty)$ is as likely as $B$, for any bounded $B$. This implies that the probability

https://doi.org/10.1017/S0266466608090415 Published online by Cambridge University Press

measure of the event $(n, \infty)$ does not go to zero when $n$ goes to infinity. Therefore, one obtains heavy tails as defined above. ■

It is possible to interpret $SP_4$ to apply to any unknown function $f : R^+ \to R$ within the statistical model defined above. For this, one must reinterpret the relationship $\preceq$ that appears in the definition of $SP_4$ as follows.

DEFINITION 4. *Let $f : R^+ \to R$ be a continuous positive valued function. Then the expression $A \preceq B$* means *$\int_A f\,dx < \int_B f\,dx$, where integration is with respect to the standard measure on $R^+$.*[18]

When working in Hilbert spaces, we use a similar definition of the expression $\preceq$ to obtain necessary and sufficient conditions below.

DEFINITION 5. *Let $f : R^+ \to R$ be a continuous function. Then the expression $A \preceq B$ means $\int_A f^2 dx < \int_B f^2 dx$, where integration is with respect to the standard measure on $R^+$.*[19]

The following extends $SP_4$ to any continuous function $f: R^+ \to R$.

DEFINITION 6 (Assumption $SP_4$ in Hilbert spaces). *Let $A_1 \supset A_2 \supset \ldots$ be a decreasing sequence of sets in $R^+$, and $B$ some fixed set such that $A_i \succsim B$, namely, $\int_{A_i} f^2(x)\,dx > \int_B f^2(x)\,dx$ for $i = 1, 2, \ldots$, then $\bigcap_{i=1}^{\infty} A_i \succsim B_i$.*

In other words, if $B$ is any set such that $\int_B f^2(x)\,dx > 0$, then, regardless of how small $B$ is, it is impossible for every infinite interval $(n, \infty)$ to satisfy $\int_{(n,\infty)} f^2(x)\,dx > \int_B f^2(x)\,dx$. This is a reasonable extension of $SP_4$ provided above.

## 9. $SP_4$ IS NECESSARY AND SUFFICIENT FOR EXTENDING NP ESTIMATION TO $R^+$

To obtain specific necessary and sufficient conditions for NP estimation, consider now the statistical model defined above, and assume that all the statistical assumptions of Bergstrom (1985) are satisfied, namely, Assumptions 1, 2, and 4. We study the estimation of an unknown function $g : R^+ \to R$. When the model is restricted to the bounded sample space $[0, 1]$, namely, $g : [0, 1] \to R$, Theorems 1, 2, and 4 of Bergstrom (1985) ensure the existence of an NP estimator in Hilbert spaces with the appropriate asymptotic behavior. The following provides a necessary and sufficient condition for extending the NP estimation results from the sample space $[0, 1]$ to the unbounded sample space $R^+$.

THEOREM 6. *Assumption $SP_4$ of DeGroot (2004), as extended above, is necessary and sufficient for extending NP estimation in Hilbert spaces from the sample space [0,1] to the unbounded sample space $R^+$.*

**Proof.** This follows directly from Theorems 1–5, above. ■

https://doi.org/10.1017/S0266466608090415 Published online by Cambridge University Press

Observe that when $SP_4$ fails, the distribution induced by the density $f$ is not countably additive and cannot be represented by a function in $H$, and the estimator, which is constructed from Fourier coefficients, fails to converge.

## 10. CONNECTION WITH DECISION THEORY

The general applicability of NP estimation, and the central role played by assumption $SP_4$, make it desirable to situate the results in the context of the larger literature. A natural connection that comes to mind is decision theory. There is a logical parallel between the classic assumptions on relative likelihood (DeGroot, 2004) and the classic axioms of decision making under uncertainty (Arrow, 1970). From assumptions on relative likelihood one obtains probability measures that represent the likelihood of events. From the axioms of decision making under uncertainty, one derives subjective probability measures that define expected utility. One would expect to find an axiom in the foundations of choice under uncertainty that corresponds to assumption $SP_4$ on relative utility. Such an axiom exists: it is called Monotone Continuity in Arrow (1970) and, as shown below, it is equivalent to $SP_4$. We use standard definitions for actions and lotteries used in the theory of choice under uncertainty, see, e.g., Arrow (1970) and Chichilnisky (2000).

DEFINITION 7. *A vanishing sequence of sets in the real line $R$ is a family of of sets $\{A_i\}_{i=1,\ldots} \subset R^+$ satisfying $A_1 \supset A_2 \supset \cdots \supset A_i, \ldots,$ and $\bigcap_{i=1}^{\infty} A_i = \phi$.*

DEFINITION 8. *The expression $A \succ B$ is now used to indicate that action $A \subset R$ is preferred to action $B \subset R$.*

DEFINITION 9 (Monotone Continuity Axiom; Arrow, 1970). *Given two actions $A$ and $B$, where $A \succ B$, and a vanishing sequence $\{E_i\}$, suppose that $\{A_i\}$ and $\{B_i\}$ yield the same consequences as $A$ and $B$ on $E_i^c$ and any arbitrary consequence $c$ on $E_i$. Then for all $i$ sufficiently large, $A_i \succ B_i$.*

The following results use the identification (see Yosida and Hewitt, 1952; Yosida, 1974; Chichilnisky, 2000) of a distribution on the line $R$ with a continuous linear real-valued function defined on the space of bounded functions on the line, $L_\infty(R)$.[20]

THEOREM 7. *Assumption $SP_4$ of DeGroot (2004) is equivalent to the Monotone Continuity Axiom (Arrow, 1970).*

**Proof.** The strategy is to show that $SP_4$ and the Monotone Continuity Axiom (MCA) are each necessary and sufficient for the existence of a ranking of events $\preceq$ in $R^+$ (by relative likelihood or by choice, respectively) that is representable by an integrable function on $R^+$.[21] Consider first the Monotone Continuity Axiom. Chichilnisky (2006a, 2006b) showed it is necessary and sufficient for the existence of a choice function that is a continuous linear function on $R$, an element of the dual space $L_\infty^*(R)$, represented by a countably additive measure on $R$ and

https://doi.org/10.1017/S0266466608090415 Published online by Cambridge University Press

thus admitting a representation by an integrable function in $L_1(R)$. The argument is as follows: the dual space $L^*_\infty(R)$ is (by definition) the space of all continuous linear real-valued functions on $L_\infty(R^+)$. It has been shown (Yosida, 1952; Yosida and Hewitt, 1974) that this space consists of both countably additive and purely finitely additive measures on $R$. Chichilnisky (2006a, 2006b) showed that the monotone continuity axiom rules out purely finitely additive linear measures and ensures that the choice criterion is represented by a countably additive measure on $R$ (Thm. 2, Chichilnisky, 2006a, 2006b). Since a countably additive measure on $R$ can always be represented by an integrable function in $L_1(R^+)$ (Yosida, 1952; Yosida and Hewitt, 1974), this completes the first part of the proof. Consider now $SP_4$. DeGroot (2004) showed that Assumption $SP_4$ eliminates distributions that are purely finitely additive, as shown DeGroot (2004, Sect. 6.2, p. 73, par. 3), ensuring that the distribution is represented by a countably additive measure, which completes the proof. ■

## 11. RARE EVENTS AND SUSTAINABILITY

When estimating an unknown path $f$ over time, $SP_4$ can be interpreted as a condition on the behavior of the unknown function on finite and infinite time intervals. A related necessary and sufficient condition has been used in the literature on sustainable development: it is called *dictatorship of the present* (Chichilnisky, 1996). For any order $\preceq$ of continuous bounded paths $f : R^+ \to R$ :

DEFINITION 10. *We say that $\preceq$ is a dictatorship of the present when for any two $f$ and $g$ there exists an $N = N(f, g)$ such that $f \preceq g \Leftrightarrow f' \preceq g'$, for any $f'$ and $g'$ that are identical to $f$ and $g$ on the interval* $[0, N)$.

The condition of dictatorship of the present (Chichilnisky, 1996) is equivalent to the representation of a welfare criterion by countably additive measures, and by an attendant integrable function on the line. The condition is also logically identical to insensitivity to rare events (Chichilnisky, 2000, 2006a, 2006b) when the numbers in the real line $R^+$ represent events rather than time periods.

DEFINITION 11 (Chichilnisky, 2000, 2006a, 2006b). *A ranking of lotteries $W : L \to R$ is called insensitive to rare events when for any two lotteries, $f$ and $g$, there is an $\varepsilon > 0, \varepsilon = \varepsilon(f, g)$ such that $W(f) > W(g) \Leftrightarrow W(f') > W(g')$ for every $f'$ and $g'$ that differ from $f$ and $g$ solely on sets of measure smaller than $\varepsilon$.*

DEFINITION 12 (Chichilnisky, 2000, 2006a, 2006b). *A ranking of lotteries $W : L \to R$ is sensitive to rare events when it is not insensitive to rare events.*

THEOREM 8. *Assumption $SP_4$ is equivalent to Monotone Continuity (Definition 9) and to insensitivity to rare events (Chichilnisky, 2000), and the latter is logically identical to dictatorship of the present. In their appropriate contexts, each of the four conditions ($SP_4$, monotone continuity, insensitivity to rare*

https://doi.org/10.1017/S0266466608090415 Published online by Cambridge University Press

*events, and dictatorship of the present) is necessary and sufficient for extending NP estimation results to* $R^+$.

**Proof.** Chichilnisky (2006a, 2006b) established that insensitivity to rare events is equivalent to the Monotone Continuity Axiom in Arrow (1970; and cf. Thm. 2 in Chichilnisky, 2006). Chichilnisky (1996, 2000) showed that insensitivity to rare events is logically identical to dictatorship of the present. Theorems 6 and 7 complete the proof of Theorem 8. ■

## 12. CONCLUSIONS

We extended Bergstrom's 1985 results on NP estimation in Hilbert spaces to unbounded sample sets, using previous results in Chichilnisky (1976, 1977, 1996, 2000, 2006a, 2006b). The focus was on the statistical assumptions needed for the extension. When estimating an unknown function on the positive line $R^+$, we obtained a necessary and sufficient condition that derives from a classic assumption on relative likelihoods, $SP_4$ in DeGroot (2004). We extended assumption $SP_4$, and therefore the results, to any unknown continuous function $f : R^+ \to R$. We also showed that the $SP_4$ assumption is equivalent to well-known axioms for choice under uncertainty, such as the Monotone Continuity Axiom in Arrow (1970), insensitivity to rare events in Chichilnisky (2000, 2006a, 2006b), and to criteria used for sustainable choice over time, such as dictatorship of the present (Chichilnisky, 1996).

When the key assumptions fail, the estimators on bounded sample spaces that are based on Fourier coefficients do not converge. We showed that this involves heavy tails and purely finitely additive measures, thus suggesting a limit to NP econometrics.

### NOTES

1. In 1980, Rex Bergstrom and I discussed Hilbert spaces at a Colchester pub and he offered me the Keynes Chair of Economics at the University of Essex, which I accepted. Bergstrom was interested in my recent work introducing Hilbert and Sobolev spaces in economics (Chichilnisky, 1976, 1977), and I suggested that Hilbert spaces were a natural space to use in NP econometrics, which apparently inspired his 1985 article, as the article documents. Bergstrom passed away in 2006, and his former student Peter Phillips organized this conference in his honor.

2. In a weighted Hilbert space the real line $R$ is endowed with a finite density function, such as $e^{-t}$, rather than the standard uniform measure on $R$.

3. See DeGroot (2004, Sect. 6.2, p. 73, par. 4) following Assumption $SP_4$.

4. NP estimation in Hilbert spaces was very interesting to me at the time, as I had introduced Hilbert spaces in economic models in a Ph.D. dissertation with Gerard Debreu at U.C. Berkeley (Chichilnisky, 1976) and in a neoclassical optimal growth model (Chichilnisky, 1977), including $L_2$, weighted $L_2$, and Sobolev spaces. A few years later, Gallant (1981) again used Sobolev norms for nonparametric estimation.

5. The same problem arises when the sample space is bounded but not closed, such as $[a, b)$.

6. This implies that the estimator $f$ is in the Hilbert space $L_2$ of square integrable functions denoted $L_2[a, b]$, which is the space of measurable functions $f : [a, b] \to R$ satisfying $\int_a^b | f(x) |^2 dx < \infty$.

https://doi.org/10.1017/S0266466608090415 Published online by Cambridge University Press

7. Bergstrom (1985) required the function to be continuous a.e. on $[a, b]$, which is essentially the same in our case.

8. The weighted Hilbert space $H$ is defined as the space of all square integrable functions on the positive line using the (finite) density function $\delta^{-x}$ : a function $f \in H$ when the weighted integral $\int_{R+} | f(x)\delta^{-x} |^2 dx < \infty$.

9. This weaker assumption always works, but has no natural interpretation when the model lacks a discount factor.

10. Private communication with Peter Phillips.

11. Private communication.

12. This can be described as the space of all ultrafilters of the real line $R$.

13. It is also needed because [0,1) is not compact.

14. To simplify notation, we may assume in the following without loss of generality $a = 0$, $b = 1$.

15. Appropriate boundary conditions are needed for this to be, for example, in the case of continuous functions of bounded variation, $\lim_{x\to\infty} f(x) = 0$.

16. Since we consider weight functions $\overline{\gamma}$, one could interpret the requirement simply as the fact that $g^2$ does not go to infinity too fast. But what does "too fast" mean in a nonparametric context? Compared to what? Imposing a limiting condition at infinity or at a boundary ($x \to 1$) becomes a parametric requirement that conflicts with the intention of nonparametric estimation. Alternatively, one could choose another "weighting" function $\widehat{\gamma}$ on the definition of $H_\gamma$ for which $g \in H_\gamma$, but this becomes an arbitrary parameter and defeats the nonparametric nature of the problem. When estimating a density function $f$ over $R^+$, one may answer the question by reference to the properties of the associated relative likelihood function (DeGroot, 2004, ch. 6), or, when estimating an investment path over time, one may consider the behavior of the associated capital accumulation path. A referee pointed out that an alternative choice of weighting function is the density of the regressor (assuming it has one). In practice, this is not known but could be estimated. If one uses the empirical c.d.f. $\widehat{\Pi}(t) = \sum_i I\{t_i \leqslant t\}/N$, then $x_i = \widehat{\Pi}^{-1}(t_i) = (i-1)/N$. In a strict sense $\widehat{\Pi}$ is not invertible, but this can be handled using a kernel-smoothed version of it. Using the density $\pi(t) = \Pi'(t)$, the central condition becomes $\int_{R_+} g^2(x)\pi(t)\,dt = \mathrm{E}\{g^2(t)\} < \infty$ and highlights that the condition concerns both the regression function and the distribution/transformation of the random variable $t$.

17. Other definitions of the phenomenon known as heavy tails exist, and are not discussed here. Our interpretation is presented as one possible definition of heavy tails that is justified by the fact that it is based only on the "primitives" of the statistical theory, namely, on "relative likelihoods."

18. Observe that this interpretation of the relationship $\preceq$ is identical the definition of relative likelihood when $f$ is a density function. Therefore it agrees with the definition provided in the previous section.

19. Observe that the interpretation of the relationship $\preceq$ is identical to the definition of relative likelihood when $f$ is a density function. Therefore it agrees with the definition provided in the previous section.

20. Namely, an element of the dual space of $L_\infty(R)$, denoted $L^*_\infty(R)$.

21. Namely, a function in $L_1(R^+)$.

## REFERENCES

Andrews, D. (1991) Asymptotic normality of series estimators for nonparametric and semiparametric regression models. *Econometrica* 59, 307–345.

Arrow, K. (1970) *Essays in the Theory of Risk Bearing*. North Holland.

Bergstrom, R. (1985) The estimation of nonparametric functions in a Hilbert space. *Econometric Theory* 1, 7–26.

Blundell, R., X. Chen, & D. Kristensen. (2006) Semi-Nonparametric IV Estimation of Shape Invariant Engel Curves. Working paper, Columbia University.

Chen, X., L.P. Hansen, & J. Sheinkman. (2000) Principal Components and the Long Run. Working paper, revised November 2005.

https://doi.org/10.1017/S0266466608090415 Published online by Cambridge University Press

Chichilnisky, G. (1976) Manifolds of Preferences and Equilibria. Ph.D. dissertation, UC Berkeley.
Chichilnisky, G. (1977) Non linear functional analysis and optimal economic growth. *Journal of Math Analysis and Applications* 61, 504–520.
Chichilnisky, G. (1996) An axiomatic approach to sustainable development. *Social Choice and Welfare* 13, 231–257.
Chichilnisky, G. (1997) What is sustainable development? *Land Economics* 73, 467–491.
Chichilnisky, G. (1999) An axiomatic approach to the theory of choice under uncertainty with catastrophic risks. In G. Chichilnisky (ed.), *Topology and Markets*. American Mathematical Society and The Fields Institute for Mathematical Sciences, Toronto.
Chichilnisky, G. (2002) An axiomatic approach to choice under uncertainty with catastrophic risks. *Resource and Energy Economics* 22, 221–231.
Chichilnisky, G. (2002) Catastrophic risks. In A.H. El Shaarawi, & W. Piergorsch (eds.), *Encyclopedia of Environmetrics*, pp. 274–279. Wiley.
Chichilnisky, G. (2006a) Non Parametric Estimation in Hilbert Spaces. Invited address to the Conference in Honor of Rex Bergstrom, Colchester, Essex, U.K. May.
Chichilnisky, G. (2006b) Topology of Fear. Working paper, Columbia University. Presented at 2006 NBER GE Conference in Honor of Gerard Debreu, U.C. Berkeley, September.
DeGroot, M. (2004) *Optimal Statistical Decisions*. Wiley.
Gallant, A.R. (1981) On the bias of flexible functional forms. *Journal of Econometrics* 15, 211–246.
Kolmogorov, A.N. & S.V. Fomin (1961) *Elements of the Theory of Functions and Functional Analysis*, Vol. 2. Graylock Press.
Newey, W.K. (1997) Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics* 79, 147–168.
Priestley M.B. & M.T. Chao (1972) Non parametric function fitting. *Journal of Royal Statistical Society B* 34, 385–392.
Stinchcombe, M. (2002) Some Genericity Analysis in Nonparametric Statistics. Working paper, December 10.
Yosida, K. (1974) *Funcional Analysis*, 4th ed. Springer Verlag.
Yosida, K. & E. Hewitt (1952) Finitely level independent measures. *Transactions of the American Mathematical Society* 72, 46–66.

https://doi.org/10.1017/S0266466608090415 Published online by Cambridge University Press