1

The Limits of Mathematics and NP Estimation in Hilbert Spaces

Graciela Chichilnisky*
Columbia University, New York
USA

1. Introduction

Non-Parametric (NP) estimation seeks the most general way to find a functional form that fits observed data. Any parametric assumption places a somewhat unnatural *apriori* restriction on the problem. For example linear estimation assumes from the outset that the functional forms described by the data are linear. This is a strong assumption that is often incorrect and prevents us from seeing the essential non-linear nature of economics. Indeed market clearing conditions are typically non-linear and therefore market behavior is not properly described by linear equations. Parametric estimation is therefore a "straight-jacket" that limits or distorts our perception of the world.

But how general is the *NP* philosophy? How far does it go? Are there limits to *NP* econometric estimation, can we assume nothing at all about the parameters of the problem and still obtain successful statistical tests that disclose the funcional forms behind the observed data? This chapter explores the limits to our ability to extend *NP* estimation. Considering the success of *NP* estimation in bounded or compact domains, the chapter seeks ways to identify the limits of extending it from bounded intervals to the entire real line R^+, so as to avoid artificial constraints that contradict the intention of NP estimation. The bounded sample approach is not new. In 1985 Rex Bergstrom (1985) constructed a non - parametric *(NP)* estimator for non-linear models with bounded sample spaces, using techniques of Hilbert spaces[1] a type of space I introduced in Economics (Chichilnisky, 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 chapter extends these original results to *unbounded* sample spaces such as the positive real line R^+,[2] using earlier work (Chichilnisky 1976, 1977, 1996, 2000, 2006, 2009, 2010a, 2010b and 2011). We 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

*UNESCO Professor of Mathematics and Economics, and Director, Columbia Consortium for Risk Management, Columbia University, New York. This chapter was based on an invited address at a conference 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, the participants of a Statistics Department seminar at Columbia University, the Econometrics Seminar at the Economics Department of Columbia University, and particularly Victor De La Pena, Chris Heyde, Marc Henry and Dennis Kristensen, for helpful comments and suggestions.

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 De Groot (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 $B \succ \emptyset$, *it is impossible for every infinite interval* (n, ∞) *to be as least as likely as* B.[3] In practical terms the condition requires that in an increasing sequence of unboudned sets, the sets become 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, ∞). Eventually for large enough n, the set B must be more likely than (n, ∞). 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 *fat 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 proved 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) and another is the *Insensitivity to Rare Events* see Chichilnisky (2000, 2006). In optimal growth models it is *Dictatorship of the Present* as defined in Chichilnisky (1996) and the classic work of Koopmans on "impatience". 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. We show here that they are directly connected with Kurt Godel's work on the *Limits of Mathematics* - which established in the 1940's that any logical system that is sufficiently complex - such as Mathematics - is either ambiguous or leads to contradictions. We are therefore encountering a general phenomenon on the limits of Mathematics.

Results extending *semi NP* estimation and *NP* estimation to infinite cases, dealing with separate but related issues, can be found e.g. in Blundell et al (2006), Stinchcombe (2002), and Chen (2005) 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 (1985) pointed out that there is a natural limitation for the use of Hilbert space 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 they exclude 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 requirement 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 I 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 non - linear *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 coefficients of the unknown function (Berhstrom 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 chapter focuses on 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 is as follows. Bergstrom's statistical assumptions (Bergstrom, 1985) 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 $E\left[\int_a^b \{\hat{f}_{MN}(x) - f(x)\}dx\right]$ can be made arbitrarily small by a suitable choice of M and of 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 which, 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 space of functions L_2, the space of *square integrable* measurable real valued functions. Bergstrom (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, for example, if the function represents capital accumulation, which typically increases over time. The restriction on $\lim_{t\to\infty} f(t)$ eliminates also 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 and 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 well defined a '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 and 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 and 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 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 re-interpreted 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.

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 σ 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_γ, also denoted H, consisting of all measurable and square integrable functions $g(.) : R^+ \to R$ with the weighted L_2 norm $\| \, . \, \|$

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

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 bi-continuous 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 δ 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 $(t_1, y_1) ..., (t_N, y_N)$ in which $t_1, ...t_N$ are non random positive real numbers whose values are fixed by the statistician and $y_1, ...y_N$ are random variables whose joint distribution depends on $t_1, ...t_N$. we may In particular it is assumed that $E(y_i) = f(\delta(t_i)) = f(x_i)$, $(i = 1, ..., N)$ 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), section 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 δ, it is statistically equivalent to work with the non random variables $t_1, ...t_N$ or, instead, with the *transformed* non random variables $x_1 = x_1(t_1) ... x_N = x_N(t_N)$. 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 $y_1, ...y_N$ are assumed to be generated by the equations

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

where

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

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

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

$$E(u_i) = 0$$
$$E(u_i^2) = \sigma^2$$
$$E(u_i, u_j) = 0 \quad i \neq j, i = 1, ..., N.$$

Assumption 2: The unknown function $g : R^+ \rightarrow R$ is continuous or, equivalently, the 'transformed' function $f : [0,1) \rightarrow R$ defined by $f(.) = go\delta^{-1}(.) : [0,1) \rightarrow R$, 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 $g : R^+ \rightarrow R$ is in the Hilbert Space H or, equivalently, the transformed function $f : [0,1) \rightarrow R$ can be extended to a continuous function $f : [0,1] \rightarrow R$.
Assumption 4: The countable set of continuous functions $\phi_1(x(t)), ..., \phi_N(x(t))$ is a complete orthonormal set in the space $L_2(R^+)$ of square integrable functions on R^+ with ordinary Lebesgue measure μ.
This requires that the functions ϕ_j be continuous, linearly independent, dense in $L_2(R^+)$ and satisfy the conditions

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

and

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

Observe that one can consider different orthonormal sets, for example, (Bergstrom, 1985) considers an orthonormal set consisting of polynomials on increasing order.
On the basis of these 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 $f : [0,1] \rightarrow R$ to facilitate comparison with Bergstrom (1985) but can be equivalently expressed on the unknown function $g : R^+ \rightarrow R$:

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

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

where

$$\hat{c}_1(M, N)\phi_1(x), ..., \hat{c}_M(M, N)\phi_M(x)$$

are the values of

$$c_1, c_2, ..., c_M$$

that minimize

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

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

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

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

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

$$E\left[\int_0^1 \{\widehat{f}_{M^*N}(x) - f(x)\}^2\right] \leq E\left[\int_0^1 \{\widehat{f}_{MN}(x) - f(x)\}^2\right] \ (M = 1, ..., N)$$

where $\widehat{f}_{MN}(x)$ *is defined by (1) and let* $f_N^*(x)$ *be defined by*

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

Then under Assumptions 1-4,

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

Definition 3. *Let* $c_1, ..., c_n, ...$ *be the Fourier coefficients of* $f(x)$ *relative to the orthonormal set* $\phi_1, ..., \phi_n$ *in the transformed set* $[0, 1]$*, namely*

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

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

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

and Parseval's inequality is satisfied (Kolmogorov, 1961, p. 98)

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

Theorem 4. *(Bergstrom 1985) Under Assumptions 1-4,*

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

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

From the definition of the transformation δ 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 4 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 4 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. De Groot, 2004, Chapter 6).

Theorem 5. *If a relative likelihood $\preceq$ satisfies assumptions SP_1 to SP_5 of De Groot (2004) Chapter 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.*

Proof: Consider the five assumptions $SP_1, ..., SP_5$ provided in Degroot (2004) Chapter 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), Section 6.4, p. 76-77). Given any countably additive measure μ 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 1952). In other words, the five assumptions $SP_1, ..., SP_5$ 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, 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 our Assumption 3, and hence the results of Theorems 1, 2 and 4.$\blacksquare$
Among the five fundamental statistical assumptions of Degroot (2004) there is one, SP_4, which 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 6. ***Assumption*** *SP_4 (De Groot 2004)*: Let $A_1 \supset A_2 \supset$ ...be a decreasing sequence of events, and B some fixed event such that $A_i \succsim B$ for $i = 1, 2....$, 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 $(n, \infty) \subset R$, $n = 1, 2, ...$ 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 ϕ. In other words, if B is more likely than the empty set, $B \succ \phi$, then regardless of how small is

B is, it is impossible for every infinite interval (n, ∞) to be as likely as B. One way to interpret the role of Assumption SP_4 is in averting 'heavy tails':

Definition 7. *We say that a relative likelihood $\succsim$ has 'heavy tails' when for any given set B, there exist an $N > 0$ such that $n > N$ and $C \supset (n, \infty)$, $C \succsim B$ 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, ∞) with arbitrarily large measure, which may be interpreted as 'heavy tails'.

Theorem 8. *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, ∞) is as likely than B, for any bounded B. This implies that the probability measure of the event (n, ∞) does not go to zero when n goes to infinity. Therefore, one obtains 'heavy tails' as defined above.$\blacksquare$

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 9. 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^+.[1]

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

Definition 10. 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^+.[18]

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

Definition 11. Assumption SP_4 in Hilbert spaces. Let $A_1 \supset A_2 \supset$ ...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 $i = 1, 2....$,then $\bigcap_{i=1}^{\infty} A_i \gtrsim B_i$.

In other words, if B is any set such that $\int_B f^2(x)dx > 0$, then regardless of how small is B, it is impossible for every infinite interval (n, ∞) 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. *SP4* 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^+ :

[1] 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.

Theorem 12. *Assumption SP_4 of De Groot (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, 2, 4, 5 and 8 above. ■
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 an logical parallel between the classic assumptions on relative likelihood (De Groot, 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 13. A vanishing sequence *of sets in the real line R is a family of of sets* $\{A_i\}_{i=1,....} \subset R^+$ *satisfying* $A_1 \supset A_2 \supset ... \supset A_i ...$*,and* $\bigcap_{i=1}^{\infty} A_i = \phi$.

Definition 14. *The expression $A \succ B$ is used to indicate that action $A \subset R$ is 'preferred' to action $B \subset R$.*

Definition 15. *Monotone Continuity Axiom (MCA, 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 (1952, 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, $L_\infty(R)$.[19]

Theorem 16. *Assumption SP_4 of De Groot (2004) is equivalent to the Monotone Continuity Axiom (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^+.[20] Consider first the Monotone Continuity Axiom (MCA). Chichilnisky (2006) 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 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 that this space consists (Yosida, 1952, 1974) of both countably additive and purely finitely additive measures on R. Chichilnisky (2006) showed 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, Theorem 2 of Chichilnisky (2006). Since a countably additive measure on R can always be represented by an integrable function in $L_1(R^+)$ (Yosida, 1952, 1974), this completes the first part of the proof. Consider now SP_4. De Groot De Groot (2004) showed that Assumption SP_4 eliminates distributions that are purely finitely additive, as shown in para. 3, page 73 Section 6.2 of De Groot (2004), 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 17. *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, 2006), when the numbers in the real line R^+ represents events rather than time periods:

Definition 18. *(Chichilnisky (2000, 2006)) 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 ε.*

Definition 19. *(Chichilnisky 2000, 2006) A ranking of lotteries $W : L \to R$ is Sensitive to Rare Events when it is not Insensitive to Rare Events.*

Theorem 20. *Assumption SP_4 is equivalent to Monotone Continuity and to Insensitivity for Rare Events, 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 for Rare Events, and Dictatorship of the Present) is necessary and sufficient for extending NP estimation results to R^+.*

Proof: Chichilnisky (2006) established that *Insensitivity to Rare Events* is equivalent to the *Monotone Continuity Axiom* (*MCA*) in Arrow (1970), cf. Theorem 2 in Chichilnisky (2006). Chichilnisky (1996, 2000) showed that *Insensitivity to Rare Events* is logically identical to *Dictatorship of the Present*. Theorems 12 and 16 complete the proof of the theorem.■

12. K. Godel and the limits of mathematics

The critical condition that allows extending econometric estimation from bounded domains to the entire line is the logical negation of purely finitely additive measures, as was shown in the previous Section. The constructibility of purely finite measures has been shown in turn to be equivalent to the existence of "ultrafilters", to Hahn Banach's theorem and the Axiom of Choice in Mathematics, Chichilnisky (2009, 2010a, 2010b). Furthermore, the Axiom of Choice was established by K. Godel early on to be independent from the other axioms of Mathematics (Godel (1943)), so there is a formulation of Mathematics with the Axiom of Choice and another without it, both are equally valid. This Axiom of Choice itself is therefore an unprovable proposition from the other axioms of Mathematics - thus providing a link with the ambiguity feature of large logical systems that was first identified by K. Godel last century (1943). Equally the two separate and distinct mathematical systems - one with and the other without the Axiom of Choice – are equally valid and they are contradictory with each other. Therefore there exist mathematically and logically correct statements that are contradictory within the classic body of Mathematics. These statements confirm K. Godel's incompleteness theorem for Mathematics.

More precisely, the formal equivalence we are seeking between the results of this chapter and the Limits of Mathematics, has been established in the previous Section of this chapter, where a link was established to Chichilnisky's axiom of "Sensitivity to Rare Events" - which, in Chichilnisky (2009, 2010a, 2010b, 2011), was proven to be identical with the negation of the Axiom of Choice.

The literature on the Limits of Mathematics that was initiated by Godel, is deeply connected therefore with the results on the Limits of Econometrics presented in this chapter.

13. Conclusions

We extended Bergstrom's 1985 results on *NP* estimation in Hilbert spaces to unbounded sample sets, using previous results in Chichilnisky (1976 and 1977). 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* De Groot (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, 2006), and to criteria used for sustainable choice over time, such as *Dictatorship of the Present* (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.

14. Footnotes

1. In 1980 Rex Bergstrom and I discussed Hilbert spaces at a Colchester pub at a time 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 chapter. Bergstrom passed away in 2006, and his former student Peter Phillips organized this conference in his honor. This paper is in honor of a great man, and attempts to complete a conversation between us that was left pending for 27 years.

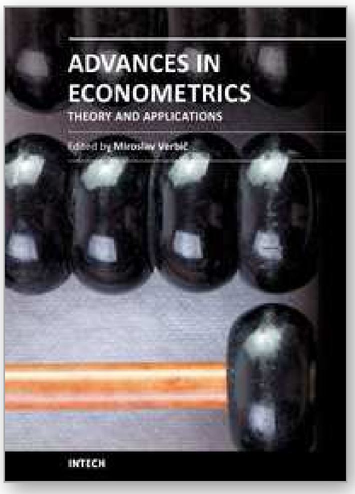
2. In a *weighted* Hilbert space the real line R is endowed with a finite density function, such as e^{-t}, rather than the standard uniform measure on R.
3. See p. 73, Section 6.2 of De Groot (2004), para. 4 following Assumption SP_4.
4. NP estimation in Hilbert Spaces was very interesting to me at the time, as I had introduced Hilbert spaces in economic models in a PhD dissertation with Gerard Debreu at UC 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, in 1981, Gallant (1981) used again Sobolev norms for non - parametric 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 $\int_a^b | f(x) |^2 dx < \infty$.
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 δ^{-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. And the fact that [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 the case, 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 non parametric 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 non-parametric estimation.Alternatively one could choose another "weighting" function $\widehat{\gamma}$ on the definition of H_γ for which $g \in H_\gamma$, but this becomes an arbitrary parameter and defeats the "non parametric" 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 (**?**), 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 cdf, $\widehat{\Pi}(t) = \sum_i I\{t_i \leqslant t\}/N$, then $x_i = \widehat{\Pi}^{-1}(t_i) = (i-1)/N$. $\widehat{\Pi}$ is not invertible in a strict sense, but this can be handled using a kernel-smoothed version of it. Using the density $\pi(t) = \Pi'(t)$, the central condition becomes $\int_{R_+} g^2(x)\pi(t)dt = E\{g^2(t)\} < \infty$ and highlights that the condition concerns both the regression function and the distribution / transformation of the random valiable 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 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.
19. Namely an element of the dual space of $L_\infty(R)$, denoted $L_\infty^*(R)$.
20. Namely a function in $L_1(R^+)$

15. References

[1] Andrews, D. "Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models" *Econometrica*, 59, 307-345, 1991.
[2] Arrow, K. *Essays in the Theory of Risk Bearing*, Amsterdam and Oxford: North Holland (1970).
[3] Bergstrom, R. "The Estimation of Non Parametric Functions in a Hilbert Space", *Econ. Theory*, 1, 7-26, 1985.
[4] Blundell, R. X. Chen and D. Kristensen, "Semi-nonparametric IV estimation of shape invariant Engel curves" Working Paper Columbia University, 2006.
[5] Chen, X. Hansen, L.P. and Sheinkman, J. "Principal Components and the Long Run", Working Paper, October 2000, revised November 2005.
[6] Chichilnisky, G. *Manifolds of Preferences and Equilibria*, PhD dissertation UC Berkeley, 1976
[7] Chichilnisky, G. "Non Linear Functional Analysis and Optimal Economic Growth", *J. Math Analysis and Applications*, Vol 61, No 2, Nov 15, 1977 p. 504-520.
[8] Chichilnisky, G. "An Axiomatic Approach to Sustainable Development" *Social Choice and Welfare*, 3, No 2, p. 231-257, 1996.
[9] Chichilnisky, G. "An Axiomatic Approach to Choice Under Uncertainty with Cataastrophic Risks" in *Topology and Markets*, Fields Institute of Mathematical Sciences, Toronto, Canada, 1996, American Mathematical Society ISSN 1968 5265;22 Providence Rhode Island, 02940 and *Resource and Energy Economics*, 2000. Also *Encyclopedia of Environmetrics*, 2002
[10] Chichilnisky, G. "The Topology of Fear" *Journal of Mathematical Economics*, 2009 - based on Columbia University Working Papaer, presented at 2006 NBER GE Conference in Honor of Gerard Debreu, UC Berkeley, California, September 2006.
[11] Chichilnisky, G. "The Foundation of Probability with Black Swans" *Mathematical Social Sciences*, 2010
[12] Chichilnisky, G. "The Foundation of Statistics with Black Swans" *Journal of Probability and Statistics*, 2011
[13] Chichilnisky, G."Catastrophic Risks with Finite or Infinite States" *International Jounral of Ecological Economics*, March 2011.
[14] Chichilnisky "Sustainable Markets with Short Sales" *Economic Theory*, 2011, in press.
[15] Chichilnisky, G. "Non Parametric Estimation in Hilbert Spaces" *Econometric Theory* 2010, Invited Address at Conference in Honor of Rex Bergstrom, (P. Philips) Colchester, Essex, UK, May 2006.

[16] De Groot, M. *Optimal Statistical Decisions*, John Wiley and Sons, Wiley Classics Library, Hoboken, New Jersey, 2004
[17] Gallant, A.R. “On the bias of flexible functional forms" *J. of Econometrics*, 15, 1981, p. 211-246.
[18] Newey, W.K. "Convergence Rates and Asymptotic Normality for Series Estimators" *Journal of Econometrics*, 79, 147-168, 1997.
[19] Kolmogorov, A.N. and Fomin, S.V. 1961, *Elements of the theory of Functions and Functional Analysis*, 2, New York: Graylock Press.
[20] Priestley M.B. and Chao M.T: "Non parametric Function Fitting", *J. Royal Stat. Society*, 1972.
[21] Stinchcombe, M. “Some Genericity Analysis in Nonparametric Statistics", Working Paper, December 10, 2002.
[22] Yosida, K. *Funcional Analysis* 1974 4th edition, New York Heidelberg: Springer Verlag
[23] Yosida, K. and E. Hewitt, 1952 “Finitely Level Independent Measures", *Transactions of the American Math. Society*, 72, 46-66.



Advances in Econometrics - Theory and Applications

Edited by Dr. Miroslav Verbic

ISBN 978-953-307-503-7
Hard cover, 116 pages
Publisher InTech
Published online 27, July, 2011
Published in print edition July, 2011

Econometrics is becoming a highly developed and highly mathematicized array of its own sub disciplines, as it should be, as economies are becoming increasingly complex, and scientific economic analyses require progressively thorough knowledge of solid quantitative methods. This book thus provides recent insight on some key issues in econometric theory and applications. The volume first focuses on three recent advances in econometric theory: non-parametric estimation, instrument generating functions, and seasonal volatility models. Additionally, three recent econometric applications are presented: continuous time duration analysis, panel data analysis dealing with endogeneity and selectivity biases, and seemingly unrelated regression analysis. Intended as an electronic edition, providing immediate â€œopen accessâ€ to its content, the book is easy to follow and will be of interest to professionals involved in econometrics.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821