DESCRIPTIVE ECONOMETRICS FOR NON-STATIONARY TIME SERIES WITH EMPIRICAL ILLUSTRATIONS

BY

PETER C. B. PHILLIPS

COWLES FOUNDATION PAPER NO. 1023

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

2001

http://cowles.econ.yale.edu/

JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **16**: 389–413 (2001)
DOI: 10.1002/jae.600

DESCRIPTIVE ECONOMETRICS FOR NON-STATIONARY TIME SERIES WITH EMPIRICAL ILLUSTRATIONS

PETER C. B. PHILLIPS*

Cowles Foundation for Research in Economics, Yale University, PO Box 208281, New Haven, CT 06520-8281, USA

SUMMARY

Recent work by the author on methods of spatial density analysis for time series data with stochastic trends is reviewed. The methods are extended to include processes with deterministic trends, formulae for the mean spatial density are given, and the limits of sample moments of non-stationary data are shown to take the form of moments with respect to the underlying spatial density, analogous to population moments of a stationary process. The methods are illustrated in some empirical applications and simulations. The empirical applications include macroeconomic data on inflation, financial data on exchange rates and political opinion poll data. It is shown how the methods can be used to measure empirical hazard rates for inflation and deflation. Empirical estimates based on historical US data over the last 60 years indicate that the predominant inflation risks are at low levels (2–6%) and low two-digit levels (10–12%), and that there is also a significant risk of deflation around the −1% level. Copyright © 2001 John Wiley & Sons, Ltd.

1. INTRODUCTION

A guiding principle in much of Denis Sargan's research was the 'marriage' of time series and simultaneity. In one of his earliest published papers, Sargan (1953) studied the properties of the correlogram and periodogram. Subsequently, his work (1958, 1959) on instrumental variables provided a methodology of estimation that was well suited to the joint treatment of simultaneity and time series dynamics in parametric models. Later on, his work with Espasa (1977) sought solutions to much more general time series regression problems in the presence of simultaneity. Inspired by the work of Hannan (1963), these solutions utilized spectral methods which permit an investigator to be agnostic with respect to the time series properties of the errors in an econometric model. These methods now fall into the category of semiparametric approaches to modelling, treating the errors, as they do, in a non-parametric fashion while leaving the systematic components of the model in parametric form. Since this early work, non-parametric and semiparametric approaches to econometric modelling have grown in popularity in both time series and microeconometric applications. In view of the common lack of prior information from economic theory models about dynamic formulations, the use of these methods in time series contexts has seemed highly desirable to many researchers including those who have argued for the use of unrestricted vector autoregressions (Sims, 1980). Spectral methods, like unrestricted VARs, can be regarded as tools for describing the data and it is this rather more general subject that concerns us in the present paper.

* Correspondence to: Professor P. C. B. Phillips, Cowles Foundation for Research in Economics, Yale University, P.O. Box 208281, New Haven, CT 06520-8281, USA. E-mail: peter.phillips@yale.edu
Contract/grant sponsor: NSF; Contract/grant number: SES 94-22922; SBR-9730295.

Received 19 January 1999
Revised 9 August 2000

Description is the starting point in much empirical work. In almost all econometric applications it is common to look at the data that has been collected, and find ways of revealing what seem to be its principal features. Graphical representations taking the form of time series plots and charts give us particularly useful ways of envisioning information and conducting *c*omparative analyses over time. Tellingly these methods appear both in formal applied econometric analysis and in the popular press, where they figure prominently in business page discussions of the behaviour of economic time series like asset prices, exchange rates, business confidence and production. Of course, while it is relatively easy to point informally to characteristics that such data seem to display in visual plots and charts (and this is, indeed, what is done daily in the popular business press), it is rather a different matter to formalize this process of description.

In describing the characteristics of economic time series, we are greatly assisted by a presumption of stationarity. For then we can utilize time invariant parameters like the mean, the variance, and the autocorrelogram to build a groundwork of descriptive statistics which involve the sample analogues of these quantities. These parameters provide points of bearing that are useful in summarizing a particular series and in comparing different series. Extending the groundwork further, there are the underlying time invariant marginal and finite dimensional probability densities of the data as well as its spectral density and higher-order spectra, all functional quantities that can be estimated from the observed data in terms of sample analogue functions. While inferential procedures can and, indeed, have been developed for all these quantities, they are first and foremost descriptive tools.

Unfortunately, this entire groundwork of descriptive statistics gets lost when the presumption of stationarity is removed because the underlying time invariant quantities no longer exist. The sample analogues are still, of course, computable in the same way, but their interpretation is not the same and they typically no longer converge without restandardization as the sample size increases. Instead, in many cases of interest like time series that have random wandering characteristics, these sample analogues end up having random rather than non-random limits, as shown in Phillips (1986, 1987, 1988). Notwithstanding this random limiting behaviour, it remains of interest whether any of these sample analogues continue to be useful as descriptive tools for non-stationary data and, if they are, how they should be interpreted.

It is little exaggeration to say that there are presently in use no tools of descriptive statistics for non-stationary data. In recent work, the Phillips (1998a) suggested some methods of spatial density analysis that apply in a fairly natural way to non-stationary data with stochastic trends and made a start in addressing the questions just raised in the last paragraph. This paper reviews those methods, shows how they may be extended to more general processes and gives some illustrations of the methods in practical empirical work and simulations. The empirical applications chosen for this paper include macroeconomic data on inflation, financial data on exchange rates and political opinion poll data. In the inflation application, it is shown how the methods can be used to measure historical hazards for inflation and deflation, the former being of interest in view of recent inflation-targeting policies by monetary authorities in several countries, and the latter being of interest in view of recent experience in the Asian economies.

The paper is organized as follows. Section 2 outlines the spatial density techniques given in Phillips (1998a) and the backbone of limit theory that justifies these descriptive methods. Guidance on the practical interpretation of these results in applications is provided and some extensions of these methods to processes with deterministic trends and long-memory components are discussed. Formulae for the mean spatial density are derived, and the theory is used to show how sample moments of non-stationary data have limits that can be expressed as moments with respect to an

underlying spatial density. These moments therefore continue to provide descriptive characteristics of the data in the limit, in spite of non-stationarity, and thereby retain their use as summary statistics. Section 3 gives some illustrations of the methods with simulated data. Section 4 reports empirical applications to inflation, exchange rates and opinion poll data. Section 5 makes some concluding comments in relation to more general issues of econometric methodology. Notation is summarized in Section 6.

2. CONCEPTS, BACKGROUND THEORY, AND ASYMPTOTICS

2.1. Preliminaries

To fix ideas and facilitate the development, we suppose that the data are well modelled by a stochastic trend with a single unit root. The theory discussed here was initiated in Phillips (1998a) and Phillips and Park (1998) and subsections 2.2–2.4 and 2.6–2.7 overview the particular methods from those papers that we will use here. These ideas are shown in sub-section 2.8 to be relevant in interpreting the sample moments of non-stationary data. The distribution of the spatial density is discussed and its mean functional is calculated in sub-section 2.9. Extensions of the ideas to processes with deterministic trends are given in sub-section 2.10. Further extensions to more general cases, like time series with fractional roots or near unit roots, are presently under development and are briefly mentioned also. These are likely to be important in empirical work where there is evidence of long-range dependence or fractional stochastic trends. The methods we are about to discuss continue to be applicable in such cases in their present form, but with some changes in the technical details and limit theory that are in the process of being worked out.

Accordingly, this paper mainly concentrates on a unit root time series $y_t = \Sigma_1^t u_s$, whose increments u_t form a stationary time series with zero mean and finite absolute moments to order $p > 2$, and which satisfies the functional law:

$$Y_n(\cdot) = \frac{y_{[n\cdot]}}{\sqrt{n}} \Rightarrow B(\cdot) \equiv BM(\sigma^2) \tag{1}$$

Here, σ^2 is the long-run variance of u_t, or $2\pi f_{uu}(0)$, where f_{uu} is the spectrum of u_t. Primitive conditions for equation (1) are well known (e.g. see Phillips and Solo, 1992) and we do not pursue them here. In fact, we need to go a little further than equation (1) and arrange the probability space in such a way that the processes Y_n and B lie in the same space. This can be accomplished by embedding arguments. In particular, we can use the Hungarian strong approximation (e.g. Csörgõ and Horváth, 1993) to y_t, which enables the construction of an expanded probability space that includes a Brownian motion $B(\cdot)$ for which:

$$\sup_{0\le k\le n} |y_k - B(k)| = o_{a.s.}(n^{1/p})$$

so that

$$\sup_{0\le k\le n} \left| \frac{y_k}{\sqrt{n}} - B\left(\frac{k}{n}\right) \right| = o_{a.s.}(1) \tag{2}$$

Since $Y_n(r) = (y_{t-1}/\sqrt{n})$, for $(t-1)/n \le r < t/n$, $t > 1$, we get the direct representation:

$$Y_n(r) = B\left(\frac{[nr]}{n}\right) + o_{a.s.}(1) \tag{3}$$

which effectively embeds the process Y_n in the Brownian motion B. Nothing will be lost if we proceed as if the space has already been set up to ensure equation (3).

When dealing with more general time series than a unit root, we need to establish (or appeal to) appropriate embeddings that extend equation (3) and adjust the probability space correspondingly. Thus, if $y_t = \beta t + \Sigma_1^t u_s = \beta t + y_t^0$ is a unit root process with drift, we can arrange the space so that:

$$Y_n^0(r) = \frac{y^0_{[nr]}}{\sqrt{n}} = B^0\left(\frac{[nr]}{n}\right) + o_{a.s.}(1)$$

and then:

$$Y_n(r) = y_{[nr]} = \beta nr + B^0([nr]) + o_{a.s.}(\sqrt{n})$$

in place of equation (3). Similarly, if $y_t = \Sigma_1^t e^{(t-j)c/n} u_j$ is a near unit root process with local to unity coefficient c, then y_t can be embedded in a diffusion process $J_c(r) = \int_0^r e^{(r-s)c} dB(s)$ and the following strong approximation (see lemma A3 of Phillips, 1998b) holds:

$$\sup_{r\in[0,1]} \left| \frac{y_{[nr]}}{\sqrt{n}} - J_c(r) \right| = o_{a.s.}(1)$$

giving:

$$Y_n(r) = \frac{y_{[nr]}}{\sqrt{n}} = J_c\left(\frac{[nr]}{n}\right) + o_{a.s.}(1)$$

in place of equation (3).

2.2. Sojourn Time and Spatial Density

A central idea in the spatial density analysis of Phillips (1998a) is to replace the notion of a time invariant marginal probability density with that of a time dependent stochastic process that measures spatial density as it has occurred over a given temporal trajectory of the time series. However, instead of measuring probability density at spatial points, this alternate quantity measures the density contributed at different spatial points towards the total quadratic variation of the process over a given interval of time. We explain this difference as follows.

If X_t is a strictly stationary time series with absolutely continuous distribution and probability measure P, then its time invariant probability density can be written in terms of the formula:

$$pdf_X(s) = \lim_{\varepsilon\to 0} \frac{1}{2\varepsilon} \int \mathbf{1}(|x - s| < \varepsilon) dP \tag{4}$$

Accordingly, as this definition makes clear, the quantity $pdf_X(s)$ measures the contribution to the overall probability P that comes from around the spatial point s. Inverting equation (4), we get for a measurable set A:

$$P(X_t \in A) = \int \mathbf{1}(A) dP = \int_A pdf_X(s) ds \tag{5}$$

For a non-stationary series like y_t, there is no time invariant probability measure P and it is no longer sensible to think of decomposing probability into densities from different spatial regions. Moreover, each trajectory $(y_t\colon t = 1, \ldots, n)$ has its own special characteristics, as does that of the limit Brownian motion process B. In view of equations (1) and (3), it is simplest to replace

the time index t with the index r representing the fraction of the sample included in the series by time t. Then, the corresponding trajectories of the series and the limit process are given by $(Y_n(r)\colon r \in [0, 1])$ and $(B(r)\colon r \in [0, 1])$. The quadratic variation of the limit process B is given by the square bracket process:

$$[B]_r = \int_0^r (dB)^2 = r\sigma^2 \tag{6}$$

which is a simple linear deterministic function of r in this Brownian motion case. Over the full sample (i.e. when $r = 1$), the total quadratic variation of the limit process B is simply σ^2. We may therefore contemplate decomposing this variation into densities from different spatial regions according to:

$$L_B(r, s) = \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_0^r \mathbf{1}(|B(t) - s| < \varepsilon) d[B]_t \tag{7}$$

The limit in equation (7) is known to exist almost surely and the limit function $L_B(r, s)$ is called the local time of the Brownian motion B at the spatial point s. $L_B(r, s)$ is a continuous stochastic process in both its temporal and spatial arguments and it measures the sojourn time that the process B spends in the vicinity of s over the time interval $[0, r]$. Obviously, it is increasing in r (i.e. as more time elapses, the number of visits to s, and hence the density there, can only increase). These properties of the local time process of Brownian motion are rather well known (e.g. Revuz and Yor, 1994).

Just as we can invert equation (4) to form (5), we may invert (7) to deliver the occupation time (in variational units) that B spends in the set A over the time interval $[0, r]$, i.e.:

$$\int_0^r \mathbf{1}(B(t) \in A)\, d[B]_t = \int_A L_B(r, s)\, ds$$

When A is the whole real line, this formula produces the decomposition:

$$[B]_r = \int_{-\infty}^{\infty} L_B(r, s)\, ds$$

which, in view of equation (6), leads quite simply to:

$$\sigma^2 = \int_{-\infty}^{\infty} L_B(1, s)\, ds \tag{8}$$

an expression that breaks down the variance of $B(1)$ (i.e. the limiting variance of $Y_n(1)$) into the contributions associated with each spatial point s visited by B (respectively, Y_n). Standardizing equation (8) by σ^{-2} we have:

$$1 = \int_{-\infty}^{\infty} \overline{L}_B(1, s)\, ds$$

where:

$$\overline{L}_B(1, s) = \sigma^{-2} L_B(1, s) = \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_0^1 \mathbf{1}(|B(t) - s| < \varepsilon)\, dt \tag{9}$$

is chronological local time in the sense that it measures sojourn time at s purely in chronological units (rather than variance based units). $\overline{L}_B(1, s)$ is normalized so that the total amount of time spent by the process in the vicinity of all points that it could possibly visit is unity

(i.e. the length of the time interval [0, 1] over which the process has been active). Similarly, $\overline{L}_B(r, s) = \sigma^{-2} L_B(r, s) = \lim_{\varepsilon \to 0} (1/2\varepsilon) \int_0^r \mathbf{1}(|B(t) - s| < \varepsilon)\, dt$ and $r = \int_{-\infty}^{\infty} \overline{L}_B(r, s)\, ds$.

Roughly speaking, we can think of $\overline{L}_B(1, s)$ as the proportion of time over the unit interval [0, 1] that B spends in the vicinity of the spatial point s. Similarly, $\overline{L}_B(r, s)$ can be interpreted as the proportion of time over the interval [0, r] that B spends in the vicinity of the spatial point s.

2.3. Estimating a Spatial Density

We can construct spatial measures for the observed time series y_t that are similar to the measure $\overline{L}_B(1, s)$ for the sojourn time of Brownian motion. In particular, the quantity:

$$\sum_{t=1}^{n} \mathbf{1}\left(\left|\frac{y_t}{\sqrt{n}} - s\right| < \varepsilon_n\right)$$

simply counts the number of times $(y_t/\sqrt{n})$ lies within ε_n of s, and so the quantity:

$$\frac{1}{2\varepsilon_n} \frac{1}{n} \sum_{t=1}^{n} \mathbf{1}\left(\left|\frac{y_t}{\sqrt{n}} - s\right| < \varepsilon_n\right) = \frac{1}{2\varepsilon_n} \int_0^1 \mathbf{1}(|Y_n(r) - s| < \varepsilon_n) dr \tag{10}$$

measures the relative amount of time that $(y_t/\sqrt{n})$ lies within ε_n of s, given the total number of observations (n) and the length of the interval $(2\varepsilon_n)$. In view of the embedding equation (3), (10) is approximately:

$$\frac{1}{2\varepsilon_n} \int_0^1 \mathbf{1}(|B(r) - s| < \varepsilon_n) dr \tag{11}$$

whose limit as $\varepsilon_n \to 0$, according to equation (9), is $\overline{L}_B(1, s)$. Thus, equation (10) is an empirical estimate of the chronological time, $\overline{L}_B(1, s)$, spent by the limit process B in the vicinity of the spatial point s. We can think of (10) as a spatial density estimate constructed with a uniform kernel and with bandwidth parameter ε_n. Using a general kernel function $K(\cdot)$ in place of a uniform kernel, we would have the estimate:

$$\widehat{L}_B(1, s) = \frac{1}{n} \frac{1}{\varepsilon_n} \sum_{t=1}^{n} K\left(\frac{1}{\varepsilon_n}\left(s - \frac{y_t}{\sqrt{n}}\right)\right)$$

where $K(\cdot)$ is a symmetric, non-negative function that integrates to unity. Setting $K_\varepsilon(\cdot) = (1/\varepsilon) K(\cdot/\varepsilon)$, we can rewrite this formula as:

$$\begin{aligned} \widehat{L}_B(1, s) &= \frac{1}{\sqrt{n}} \frac{1}{\sqrt{n}\varepsilon_n} \sum_{t=1}^{n} K\left(\frac{1}{\sqrt{n}\varepsilon_n}(\sqrt{n}s - y_t)\right) \\ &= \frac{1}{\sqrt{n}} \frac{1}{h_n} \sum_{t=1}^{n} K\left(\frac{1}{h_n}(\sqrt{n}s - y_t)\right) \\ &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} K_{h_n}(\sqrt{n}s - y_t) \end{aligned} \tag{12}$$

or as

$$\widehat{L}_B\left(1, \frac{a}{\sqrt{n}}\right) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} K_{h_n}(a - y_t) \tag{13}$$

both of which are expressed as a kernel function in terms of the original series y_t.

The bandwidth parameter in equation (12) is $h_n = \sqrt{n}\varepsilon_n$ and the spatial position, $a = \sqrt{n}s$, is now measured in units of $\sqrt{n}$. As is apparent from equation (11), it is important that $\varepsilon_n \to 0$ if we are to achieve a consistent estimate of the local time $\overline{L}_B$ (1, s). However, it is no longer vital that $h_n \to 0$, and, indeed, $\widehat{L}_B$ (1, s) may still be consistent for $\overline{L}_B$ (1, s) even when h_n is constant or slowly increasing. The reason is that in conventional kernel density estimation for stationary processes a shrinking bandwidth serves the purpose of focusing attention on a particular spatial point, whereas in the present situation y_t tends naturally on its own to drift away from any given spatial point at the rate $\sqrt{n}$. So, on the one hand, there is less reason to have to focus attention on given spatial points, because this focus occurs in a natural way. But, on the other hand, there are necessarily a reduced number of relevant observations in estimating the spatial density $\overline{L}_B$ (1, s). In fact, there are only $O(\sqrt{n})$ relevant observations to each spatial point, compared with the usual *O(n)* relevant observations in stationary density estimation. In effect, this is because the number of returns that a unit root process, like a random walk, makes to the origin or some other fixed point is of the order of $\sqrt{n}$ (cf. Feller, 1957, p. 83), whereas all n observations of a stationary times series are relevant in fitting its probability density at any point in the support. This difference in the order of magnitude explains the standardizing factor $1/\sqrt{n}$ that appears in equation (12). By contrast, if X_t is a strictly stationary time series with probability density (4), then the corresponding kernel density estimate is given by:

$$\widehat{pdf}_X(s) = \frac{1}{n} \sum_{t=1}^{n} K_{h_n}(s - X_t) = \frac{1}{n} \frac{1}{h_n} \sum_{t=1}^{n} K\left(\frac{1}{h_n}(s - X_t)\right) \tag{14}$$

where the standardizing factor is $1/n$ and the bandwidth parameter $h_n \to 0$.

2.4. Limit Theory

Under some rather weak regularity conditions that are discussed in Phillips (1998a) it is shown that:

$$\widehat{L}_B(r, s) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} K_{h_n}(\sqrt{n}s - y_t) \to_{a.s.} \overline{L}_B(r, s - B_0(\kappa)) \tag{15}$$

This result allows for non-trivial initial conditions in the process $y_t = \Sigma_1^t u_s + y_0$, showing that the point $(s - B_0(\kappa))$ at which the spatial density is estimated may be affected by initial conditions if these are not $O_p(1)$. In particular, if the initial conditions extend into the distant past as in:

$$y_0 = \sum_{j=0}^{[n\kappa]} u_{-j}, \qquad \text{for some } \kappa \geq 0$$

for which the following functional law holds:

$$n^{-\frac{1}{2}} \sum_{j=0}^{[n\kappa]} u_{-j} \Rightarrow B_0(\kappa)$$

then in place of equation (1) we have:

$$n^{-\frac{1}{2}} y_{[nr]} \Rightarrow B(r) + B_0(\kappa)$$

with the Brownian motions B and B_0 being independent. Again, matters can be arranged so that the probability space of y_t includes B_0.

A limit distribution theory for the estimate $\hat{L}_B(r, s)$ is also available. Indeed, using the normalization factor $c_n = 1/\varepsilon_n = \sqrt{n}/h_n$, we have the following mixed normal limit theory (Phillips, 1998a):

$$\sqrt{c_n}[\hat{L}_B(r, s) - \overline{L}_B(r, s - B_0(\kappa))] \Rightarrow MN(0, 16K_2\overline{L}_B(r, s - B_0(\kappa))) \tag{16}$$

where the constant is:

$$K_2 = \int_0^\infty \int_0^\infty K(p)(p \wedge t)K(t)dpdt$$

a quadratic functional of the kernel K and the covariance function of Brownian motion. When K is a normal kernel, $K_2 = \pi^{-1/2}(2^{1/2} - 1)$. The convergence rate in equation (16) is $\sqrt{c_n} = n^{1/4}/\sqrt{h_n}$, which is slower than the conventional rate of convergence of kernel density estimates in the stationary case (i.e. $\sqrt{nh_n}$) for all bandwidths $h_n = c/n^b$ with $b < 1/4$.

Using equation (16), asymptotic confidence intervals for $\overline{L}_B(r, s - B_0(\kappa))$ can be constructed at each spatial point, giving e.g.:

$$\hat{L}_B(r, s) \pm 1.96 \left(\frac{16K_2\hat{L}_B(r, s)}{c_n} \right)^{\frac{1}{2}}$$

as a 95% confidence interval for $\overline{L}_B(r, a - B_0(\kappa))$. By virtue of the standardization of y_t, the quantities measure spatial departures from the origin in units of $\sqrt{n}$ and the spatial points are centred on a standardized initial condition. Hence, for a confidence interval directly at the point s^0, we would compute:

$$\hat{L}_B\left(r, s^0 + n^{-\frac{1}{2}} y_0\right) \pm 1.96 \left(\frac{16K_2\hat{L}_B\left(r, s^0 + n^{-\frac{1}{2}} y_0\right)}{c_n} \right)^{\frac{1}{2}} \tag{17}$$

These intervals can be compared with the confidence intervals of kernel estimates of a stationary probability density like equation (15). By traditional theory (Silverman, 1986) a 95% confidence interval for the probability density $pdf_X(s)$ is:

$$\widehat{pdf}_X(s) \pm 1.96 \left(\frac{k_2\widehat{pdf}_X(s)}{nh} \right)^{\frac{1}{2}} \tag{18}$$

where:

$$k_2 = \int_{-\infty}^\infty K(r)^2 dr \tag{19}$$

Leaving aside initial conditions, the only differences between equations (17) and (18) are in the scale factors and the convergence rate. The former arises because of persistent temporal dependence in the trajectory of y_t, and the latter because of the effective number of observations. In other respects, (17) simply extends the theory of non-parametric density estimation to spatial density estimation for stochastic processes.

2.5. Interpreting Spatial Density Estimates

The interpretation of spatial density estimates like $\widehat{L}_B(r, s)$ is similar to that of a conventional probability density estimate. As indicated earlier, in rough terms $\overline{L}_B(1, s)$ is the proportion of time over the unit interval [0, 1] that the limit process B spends in the vicinity of the spatial point s. Similarly, the quantity $\widehat{L}_B(r, s)$ is an estimate of the proportion of time over [0, r] that the standardized series $y_t/\sqrt{n}$ spends in the vicinity of s. The quantity $\widehat{L}_B(r, a/\sqrt{n})$ is then an estimate of the proportion of time over [0, r] that the series y_t spends in the vicinity of a. Note that:

$$\int_{-\infty}^{\infty} \widehat{L}_B\left(1, \frac{a}{\sqrt{n}}\right) da = \int_{-\infty}^{\infty} \widehat{L}_B(1, s) ds \sqrt{n} = \sqrt{n}$$

so that on this adjusted scale of spatial measurement the total amount of sojourn time is set to $\sqrt{n}$ rather than unity. With this modification, we can think of spatial density estimates just like conventional density estimates, the difference being that in the latter we are distributing probability across space whereas in the former we are distributing sojourn time across space. As is apparent from equations (13) and (14), the same formula applies in each case up to normalization by $\sqrt{n}$. Thus if we use the spatial density estimate (13) in the case of stationary rather than non-stationary data we will obtain the usual kernel density estimate scaled by $\sqrt{n}$. In that case, we are effectively distributing a non-unitary $\sqrt{n}$ probability across spatial points, rather than the $\sqrt{n}$ sojourn time.

2.6. Hazard Functions

Following Phillips (1998a), we define the spatial hazard function $H_B(r, a)$ associated with a given spatial density $L_B(r, s)$ as follows:

$$H_B(r, a) = \frac{L_B(r, a)}{\int_a^{\infty} L_B(r, s) ds} \tag{20}$$

This definition is entirely analogous to the hazard rate $\lambda(x)$ associated with a continuous probability density $pdf_X(x)$, which has the form:

$$\lambda(x) = \frac{pdf_X(x)}{\int_x^{\infty} pdf_X(s) ds} = \frac{pdf_X(x)}{1 - cdf_X(x)} = \frac{pdf_X(x)}{\overline{F}(x)}$$

where $cdf_X(x)$ is the cdf of the distribution and $\overline{F}(x) = 1 - cdf_X(x)$ is the survival function. It is also useful to define the left-sided hazard:

$$h_B(r, a) = \frac{L_B(r, a)}{\int_{-\infty}^{a} L_B(r, s) ds} \tag{21}$$

The functions (20) and (21), like the spatial density $L_B(r, a)$ itself, have two arguments; time and space. Note that H_B and h_B are invariant to whether $L_B(r, a)$ or $\overline{L}_B(r, a)$ are used in definitions (20) and (21).

The spatial hazard functions can be interpreted as follows. $H_B(r, a)$ measures the conditional likelihood computed over the sample path to time r that B takes the value a, given that it takes values at least as big as a. Similarly, $h_B(r, a)$ measures the conditional risk computed over the same sample path to time r that B takes the value a, given that it takes values at least as small as a. To be more concrete, suppose the time series y_t is inflation and that $B(r)$ is the weak limit of $n^{-1/2} y_{[nr]}$. The spatial hazard $H_B(r, a = s/\sqrt{n})$ then measures the conditional likelihood computed over the period $[0, r]$ (which is expressed in standardized units of fractions of the overall sample, so that it corresponds to an observation period of $t = \sqrt{n}r$ for y_t) of an inflation rate of s, given that inflation is at least as great as s. Similarly, $h_B(r, a = s/\sqrt{n})$ measures the conditional risk computed over the period $[0,r]$ (respectively, an observation period of $t = \sqrt{n}r$ for y_t) of an inflation rate of s, given that inflation is at least as small as s. In the latter case, when s is negative we can think of $h_B(r, a = s/\sqrt{n})$ as measuring the hazard of deflation rather than inflation. That is, if there is deflation, $h_B(r, a = s/\sqrt{n})$ measures the relative likelihood of a deflation rate of s.

For a given time period r, the shape of the spatial hazard rate functions $H_B(r, a = s/\sqrt{n})$ and $h_B(r, a = s/\sqrt{n})$ can be studied in just the same way as we look at hazard rates in the analysis of independent (or strictly stationary) data. In such studies it is usually of interest to find out whether the hazard declines, increases or stays constant as we increase s, whether there are multiple peaks in the hazard and so on. Since our hazard rates also depend on time r, we can consider what happens to the hazard rate functions as the time period changes or as new data is introduced. In the case of inflation, this means that we look at whether the hazard of a certain rate of inflation of deflation rises or falls over time.

2.7. Hazard Functions Asymptotics

The spatial hazard functions $H_B(r, a)$ and $h_B(r, a)$ are empirically estimable using the non-parametric spatial density estimate $\widehat{L}_B(r, s/\sqrt{n})$. In particular, we may construct:

$$\widehat{H}_B\left(r, \frac{s}{\sqrt{n}}\right) = \frac{\widehat{L}_B(r, s/\sqrt{n})}{\int_a^\infty \widehat{L}_B(r,\ p/\sqrt{n})(d\,p/\sqrt{n})}, \quad \widehat{h}_B\left(r, \frac{s}{\sqrt{n}}\right) = \frac{\widehat{L}_B(r, s/\sqrt{n})}{\int_{-\infty}^a \widehat{L}_B(r,\ p/\sqrt{n}) d\,p/\sqrt{n}}$$

These estimators are consistent and have mixed normal distributions under the same regularity conditions as the limit theory for the spatial density estimate $\widehat{L}_B(r, s/\sqrt{n})$. In particular, we have:

$$\widehat{H}_B\left(r, \frac{s}{\sqrt{n}}\right) \to_{a.s.} H_B(r, a - B_0(\kappa)), \quad \widehat{h}_B\left(r, \frac{s}{\sqrt{n}}\right) \to_{a.s.} h_B(r, a - B_0(\kappa))$$

and:

$$\sqrt{c_n}\left[\widehat{H}_B\left(r, n^{-\frac{1}{2}}s\right) - H_B(r, a - B_0(\kappa))\right] \Rightarrow 4MN\left(0, K_2 \frac{H_B(r, a - B_0(\kappa))^2}{\overline{L}_B(r, a - B_0(\kappa))}\right)$$

$$\sqrt{c_n}\left[\widehat{h}_B\left(r, n^{-\frac{1}{2}}s\right) - h_B(r, a - B_0(\kappa))\right] \Rightarrow 4MN\left(0, K_2 \frac{h_B(r, a - B_0(\kappa))^2}{\overline{L}_B(r, a - B_0(\kappa))}\right)$$

The results for $\widehat{H}_B(r, n^{-1/2}s)$ are proved in Phillips (1998a) and those for $\widehat{h}_B(r, n^{-1/2}s)$ follow in a similar way.

2.8. Interpreting Moments

In stationary cases, the ergodic theorem and the existence of moments are all that are required to ensure good limiting behaviour of sample moments. Thus, if X_t is strictly stationary and ergodic, and X_t^k is integrable then:

$$n^{-1}\sum_{t=1}^{n} X_t^k \to_{a.s.} EX_t^k \tag{22}$$

a result that allows us to interpret sample moments of integrable functions of the data in terms of population moments.

In non-stationary cases, the absence of ergodicity means that quite different limiting behaviour can be expected of sample moments than what occurs in equation (22). Nevertheless, sample moments still play a role as descriptive statistics for the sample data, because they continue to summarize key features of the data in much the same way as in the stationary case. In many cases, empirical investigators want to characterize the data by some data-reduction techniques without having to resolve potentially complicated issues like choosing between stationary and non-stationary generating mechanisms. Thus, it is of interest to be able to interpret sample moment characteristics for non-stationary data.

For a time series y_t satisfying equation (1), it is now a very well-known application of functional limit theory and continuous mappings (Phillips, 1986, 1987, 1988) that standardized moments of y_t have the following asymptotic behaviour:

$$\frac{1}{n^{1+\frac{k}{2}}}\sum_{t=1}^{n} y_t^k \Rightarrow \int_0^1 B(r)^k dr \tag{23}$$

There is another way of writing this result. We use the fact that if g is a measurable and locally integrable map then:

$$\int_0^1 g(B(r))\, dr = \int_{-\infty}^{\infty} g(s)\overline{L}_B(1, s)\, ds \tag{24}$$

This result, known as the occupation time formula (Revuz and Yor, 1994), converts temporal integrals into spatial integrals involving local time. We have written formula (24) here in terms of chronological local time $\overline{L}_B(1, s)$. Applying formula (24) to (23) reveals that we have the alternate representation:

$$\frac{1}{n^{1+\frac{k}{2}}}\sum_{t=1}^{n} y_t^k = \frac{1}{n}\sum_{t=1}^{n}\left(\frac{y_t}{\sqrt{n}}\right)^k \Rightarrow \int_{-\infty}^{\infty} s^k \overline{L}_B(1, s)\, ds \tag{25}$$

which is the kth moment of the spatial distribution $\overline{L}_B(1, s)$. In effect, sample moments of y_t converge weakly to corresponding sample moments of the spatial distribution of the limit process. In this form, (25) is a very natural analogue of (22). Moreover, it makes it clear the sense in which sample moments continue to be useful descriptive characteristics of

the data. Whether the data are stationary or non-stationary, sample moments provide summary information about the observed data that reflect characteristics of an underlying population. This population is just the stationary distribution in the case of a stationary, ergodic time series, whereas it is the sojourn distribution of the limit process in the non-stationary case.

2.9. The Distribution of the Spatial Density Process

Since the spatial density is a random process, it helps in understanding its behaviour to know more about its probability distribution and some of its characteristics. If $W(t)$ is standard Brownian motion, then the probability density of the local time $L_W(r, s)$ changes according to the values of the arguments r and s. The density is also influenced by initialization of the process. For $W(0) = 0$ and for $s = a > 0$ the density of $L_W(r, s)$ is:

$$f(y) = \sqrt{\frac{2}{\pi r}} e^{-\frac{(y+a)^2}{2r}}$$

Perhaps, the most important single characteristic of this distribution is the mean spatial density $E(L_W(r, a))$. At the spatial value $s = a > 0$, the mean is calculated as follows:

$$\begin{aligned}
E(L_W(r,a)) &= \sqrt{\frac{2}{\pi r}} \int_0^\infty y e^{-\frac{(y+a)^2}{2r}} dy \\
&= \sqrt{\frac{2}{\pi r}} \int_a^\infty (z-a) e^{-\frac{z^2}{2r}} dz \\
&= \frac{1}{2}\sqrt{\frac{2}{\pi r}} \int_{a^2}^\infty \left(q^{\frac{1}{2}} - a\right) e^{-\frac{q}{2r}} q^{-\frac{1}{2}} dq \\
&= \frac{1}{2}\sqrt{\frac{2}{\pi r}} \left[\int_{a^2}^\infty e^{-\frac{q}{2r}} dq - a\int_{a^2}^\infty e^{-\frac{q}{2r}} q^{-\frac{1}{2}} dq\right] \\
&= \sqrt{\frac{2r}{\pi}} \left[\int_{a^2/2r}^\infty e^{-p} dp - a(2r)^{-\frac{1}{2}} \int_{a^2/2r}^\infty e^{-p} p^{-\frac{1}{2}} dp\right] \\
&= \sqrt{\frac{2r}{\pi}} e^{-\frac{a^2}{2r}} - \frac{a}{\sqrt{\pi}} \Gamma\left(\frac{1}{2}, \frac{a^2}{2r}\right)
\end{aligned} \tag{26}$$

where $\Gamma(b, x) = \int_x^\infty e^{-p} p^{b-1} dp$ is an incomplete gamma function. Combining equation (26) with a similar calculation for $s = a < 0$, we end up with the general formula:

$$E(L_W(r,a)) = \sqrt{\frac{2r}{\pi}} e^{-\frac{a^2}{2r}} - \frac{|a|}{\sqrt{\pi}} \Gamma\left(\frac{1}{2}, \frac{a^2}{2r}\right) \tag{27}$$

and the mean spatial density function is seen to be symmetric about the origin $a = 0$. The function is graphed in Figure 1.

More complex calculations can be performed to produce a related result for the mean local time of a Brownian motion with drift and these will be reported elsewhere.

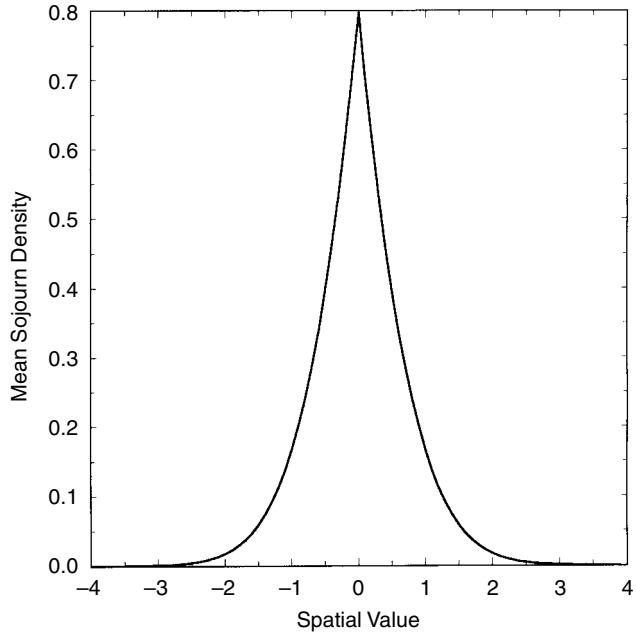

Figure 1. Mean local time of Brownian motion

2.10. Extensions of Local Time to General Processes

For the spatial density of a general continuous process $X(t)$ to be defined over $[0, r]$, we need the limit:

$$L_X(r, s) = \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_0^r 1(|X(t) - s| < \varepsilon)\, dt \tag{28}$$

to exist, thereby giving the analogue for $X(t)$ of the chronological local time of Brownian motion. More formally, local time $L_X(r, s)$ is the spatial Lebesgue density of the occupation measure:

$$\mu_r(A) = \int_0^r \mathbf{1}(X(t) \in A)\, dt \quad \forall A \in \mathcal{B}$$

where $\mathcal{B}$ is the Borel algebra on the real line $\mathcal{R}$, and it will exist when μ is absolutely continuous with respect to Lebesgue measure on $\mathcal{R}$. A necessary and sufficient condition for the existence of (28) is the condition:

$$\liminf_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_{[0,r]} P(|X(s) - X(t)| \leq \varepsilon) ds < \infty \quad a.e.\ t \in [0, r] \tag{29}$$

A multidimensional version of this result is proved in Geman and Horowitz (1980, theorem 21.12), together with some related results on existence and the path properties of $L_X(r, s)$ —see also Bosq (1998). Condition (29) is satisfied for continuous processes like Brownian motion with drift, where $X(t) = \beta t + B(t)$, and for fractional processes like the fractional Brownian motion:

$$B_H(r) = \frac{1}{\Gamma\left(H + \frac{1}{2}\right)} \int_0^r (r - s)^{H - \frac{1}{2}}\, dB(s) \tag{30}$$

With these extensions, the concept of a spatial density becomes applicable to a wide range of non-stationary time series whose limits after standardization take the form of continuous stochastic processes. A theory for local time in the case of the fractional Brownian motion (30) has recently been developed in Tyurin and Phillips (1999) and is applied to analyse non-linear functionals of fractional processes and to provide a basis for descriptive statistical inference about such processes.

3. SIMULATIONS

This section illustrates the spatial density estimates in some commonly arising cases with simulated data. We use single replicates in these simulations and explain why these are useful here. When the data is non-stationary, it is of the nature of a spatial density that it is trajectory dependent even in the limit. In such cases, there are not functional invariants like a marginal probability density (a relevant invariant in the stationary case) to be estimated. The spatial density is itself a random process that alters with the observed trajectory and with its two arguments —the spatial coordinate and the period of observation. Performing multiple replications would lead to the estimation of some functional, such as an average, of this random process, which is an altogether different object. The distinction will become clearer in Figure 3 (a) below, where the effects are highlighted by plotting both the spatial density and the mean functional in the random walk case. The following simulations are provided to give some idea of the 'typical' shape of these densities in some prototypical cases, to contrast them with stationary cases and thereby to assist readers in understanding the use of these methods in practice.

The model for y_t is taken to be a Gaussian first-order autoregression with and without a deterministic trend component. The generating mechanisms are as follows:

Model 1: Gaussian AR(1) with no trend

$$y_t = \rho y_{t-1} + u_t, \quad u_t \equiv \text{iid } N(0, 1)$$

Model 2: Gaussian AR(1) with trend

$$y_t = bt + y_t^0,\ y_t^0 = \rho y_{t-1}^0 + u_t, \quad u_t \equiv \text{iid } N(0, 1)$$

The AR coefficients selected were $\rho = 0.5,\ 1.0$, covering stationary and unit root cases, and the deterministic trend coefficient $b = 0.05$ was chosen for model 2. When the model is stationary, the initialization y_0 is drawn from the stationary distribution. When the models have a unit root, the initialization is set to zero. The sample size is set at $n = 500$. Figures 2–5 display the data (Figures X(b)) and the resulting spatial density estimates (Figures X(a)). The bandwidth in these exercises, and in our empirical illustrations in Section 4, is chosen using the rule $h_n = n^{-1/5}$. Pointwise 95% confidence bands for the spatial density are given by the broken lines in the figures.

For model 1, y_t is stationary when $|\rho| < 1$ and the invariant distribution is $N(0, \sigma_y^2)$ with $\sigma_y^2 = 1/(1 - \rho^2)$. It is this invariant distribution that is being estimated by the spatial density estimate in the $\rho = 0.5$ case (Figure 2(a)), subject to rescaling the probability by $\sqrt{n}$, as mentioned earlier. In the unit root case (Figure 3(a)), the spatial density estimates the sojourn time of the series at each point that it visits. The differences between the two cases are very apparent in the figures.

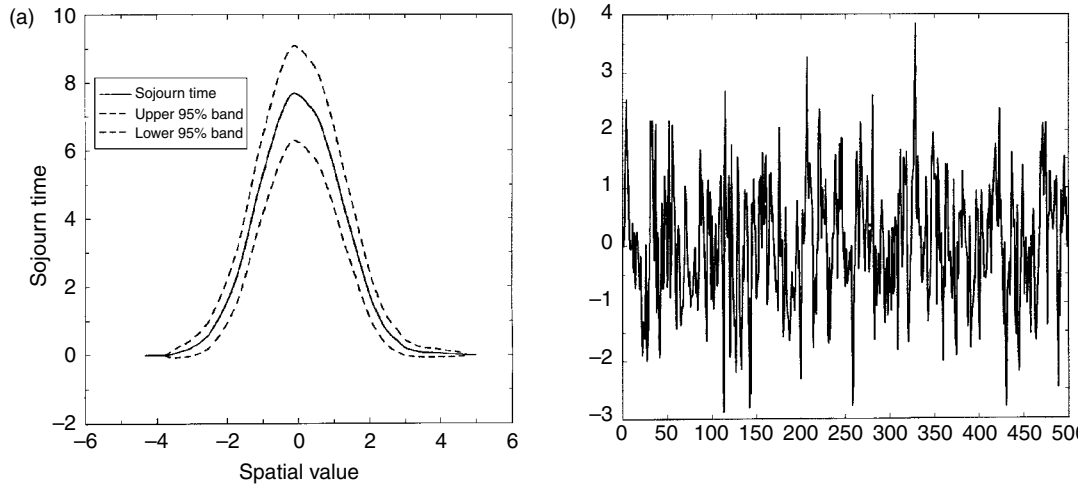

Figure 2. (a) Spatial density of AR(1): $\rho = 0.5$; (b) AR(1) data: $\rho = 0.5$

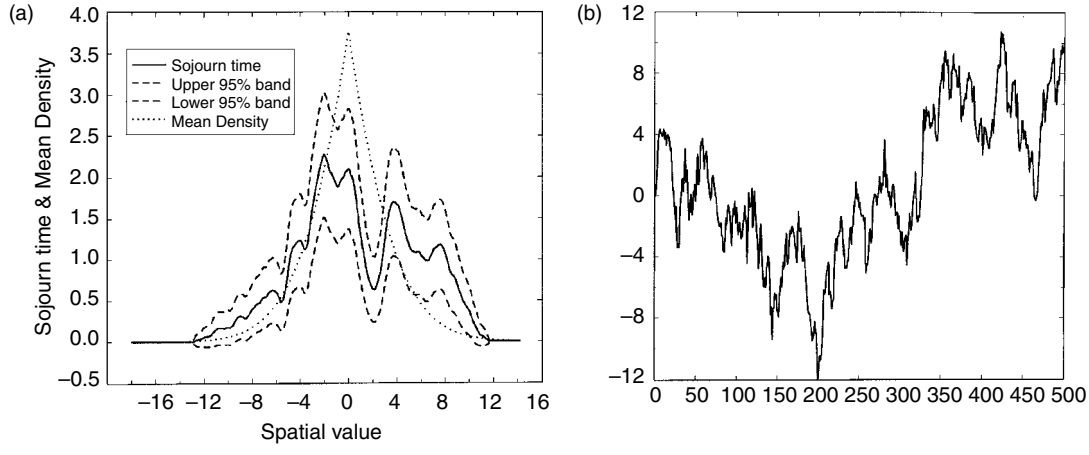

Figure 3. (a) Spatial density of random walk; (b) Random walk

In the stationary case the fitted curve is smooth and a good approximation to a rescaled normal density. In the unit root case, the curve is irregular and shows substantial variation in sojourn levels over quite a large range of spatial values. This irregular curve is our point estimate of the stochastic process $L_B(1, s)$, for which the Brownian motion B is the limit process of $n^{-1/2} y_{[n\cdot]}$. It gives us summary information about the spatial points y_t has visited and the relative proportion of those visits to the full sample. Apparently, the regions $[-4, 1]$ and $[3.5, 4.5]$ are the most frequently visited in this sample path. The spatial support is also far wider in the unit root case.

Figure 3(a) also shows the mean spatial density (given by the dotted line in the figure). The mean density is calculated using formula (27), but with spatial coordinates given by $a/\sqrt{n}$ to allow for the fact that the spatial density of the data rather than the normalized data is being

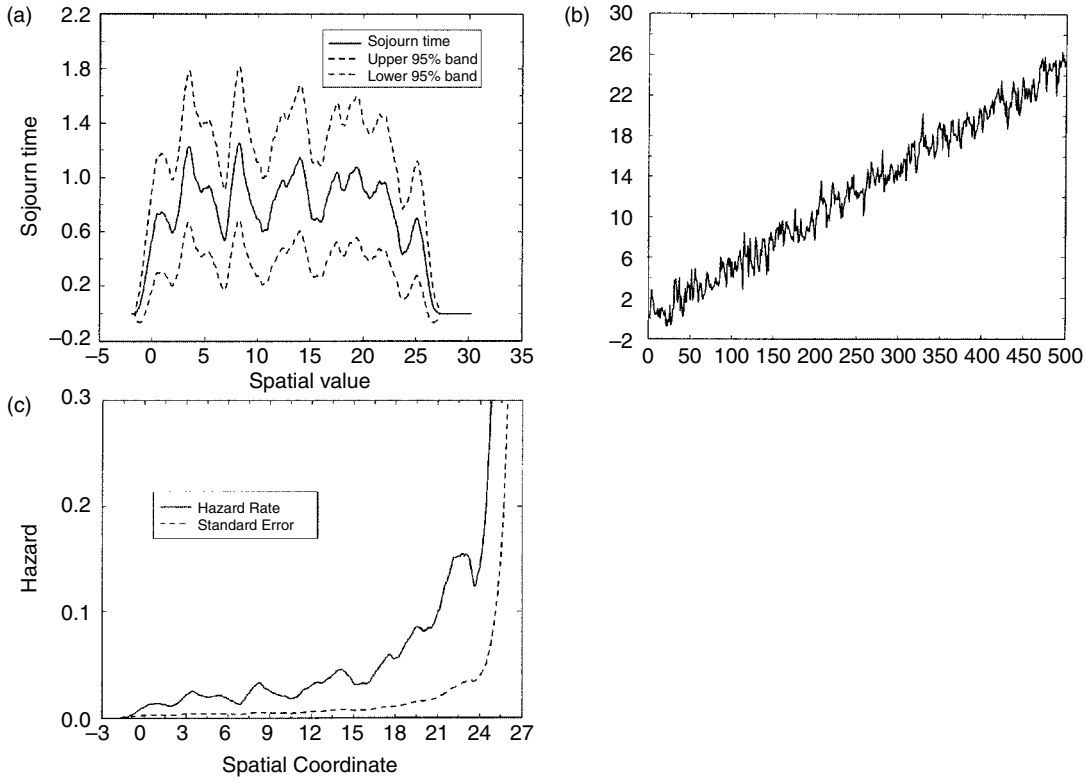

Figure 4. (a) Spatial density of AR(1) + trend (5%): $\rho = 0.5$; (b) Data for AR(1) + trend (5%): $\rho = 0.5$; (c) Hazard function: AR(1) + trend (5%)

considered (cf. the discussion in Section 2.5). As is apparent from the graph, the density estimate follows the general shape of the mean functional well for negative spatial values, but is below the mean functional at the origin and above it in the far-right tail. Other realizations of the random walk process will, of course, lead to different configurations in relation to the mean functional.

Model 2 allows for trend stationary and random walk with drift realizations. The same innovations were used for generating the data in this case as for Model 1. Figure 4 gives the trend stationary results. With 5% trend growth, the data is now distributed over a much wider region than for Model 1. The spatial density estimate is no longer a smooth curve (as in Figure 2) and has wide confidence bands. From these bands it is apparent that the spatial density outcome is compatible with a uniform spatial distribution over the support, barring the very ends of the range. This marries with the notion of the data being stationary about the fixed trend line $y = bt$, and will be discussed further below. This behaviour would be accentuated were the data generated more frequently about this line (i.e. infill observations keeping the span of the data fixed).

The stationary component has distribution $y_t^0 =_d N(0, \sigma_y^2)$, so then $y_t =_d N(bt, \sigma_y^2)$ and the density of the data is invariant up to the mean. Some calculations reveal that the spatial density

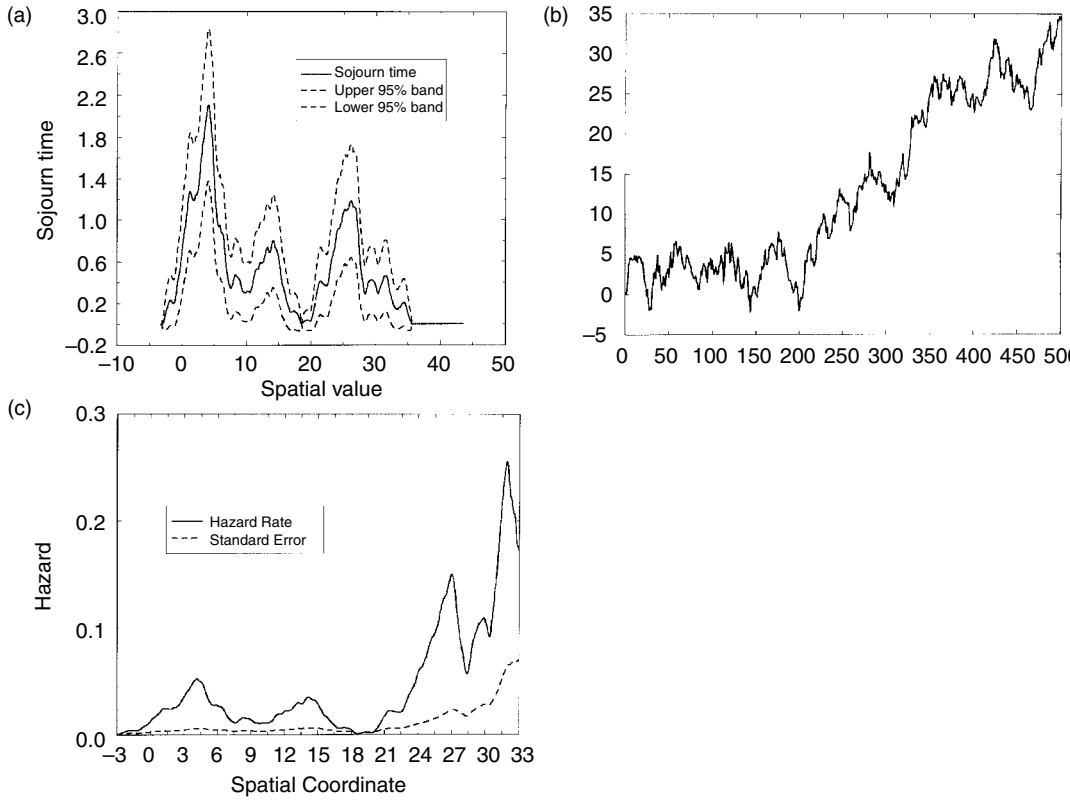

Figure 5. (a) Spatial density: random walk + drift (5%); (b) Random walk + drift (5%); (c) Hazard function: Random walk + drift (5%)

estimate in this case has mean value:

$$E\left(\widehat{L}_B\left(1, \frac{a}{\sqrt{n}}\right)\right) = E\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} K_{h_n}(a - y_t)\right)$$

$$= \frac{1}{\sqrt{n}}\sum_{t=1}^{n} E(K_{h_n}(a - bt - y_t^0))$$

$$= \frac{1}{\sqrt{2\pi n}(\sigma_y^2 + h_n^2)^{\frac{1}{2}}}\sum_{t=1}^{n} e^{-\frac{(a-bt)^2}{2(\sigma_y^2+h_n^2)}[1+O(h_n^4)]} \tag{31}$$

When $b = 0$, equation (31) is:

$$\frac{\sqrt{n}}{\sqrt{2\pi}(\sigma_y^2 + h_n^2)^{\frac{1}{2}}} e^{-\frac{a^2}{2(\sigma_y^2+h_n^2)}}[1 + O(h_n^4)]$$

which is the density of $N(0, \sigma_y^2 + h_n^2)$ evaluated at the spatial point a and scaled by $\sqrt{n}$. (The variance term, $\sigma_y^2 + h_n^2$, is well known from conventional kernel density asymptotics—e.g. Härdle

and Linton, 1994). When $b \neq 0$, equation (31) can be approximated for large n by writing $bt \sim bn_0(t/n)$ where $n_0 < n$ is large and fixed. Then, (31) is approximately:

$$\frac{1}{\sqrt{2\pi n}(\sigma_y^2 + h_n^2)^{\frac{1}{2}}} \sum_{t=1}^{n} e^{-\frac{(a-bn_0\frac{t}{n})^2}{2(\sigma_y^2+h_n^2)}[1+O(h_n^4)]} \sim \frac{\sqrt{n}}{\sqrt{2\pi}(\sigma_y^2 + h_n^2)^{\frac{1}{2}}} \int_0^1 e^{-\frac{(a-bn_0 r)^2}{2(\sigma_y^2+h_n^2)}[1+O(h_n^4)]} dr$$

$$= \frac{\sqrt{n}}{|b|n_0} \tag{32}$$

which we note to be independent of the spatial point a. Thus, in the trend stationary case, the spatial density estimate has approximately constant mean, corroborating the form of the empirical estimate found in Figure 4. Thus, one of the characteristics of spatial densities of trend stationary series is that that the mean density is approximately constant over the support. As we will see, this behaviour is quite different from the spatial density of a random walk with drift.

Figure 5 gives the spatial density of a random walk with a 5% drift, so that the deterministic component here is the same as it is in the trend stationary case. It is apparent from Figure 5(a) that the regions [3, 5] and [23, 27] are the most frequently visited in this sample path. Clearly, the spatial support is far wider than in the case of no drift (Figure 2), but the curve is just as irregular.

This can be explained by noting that if $M(r) = br + B(r)$ is Brownian motion with drift, then $dM = bdr + dB$, so that $(dM)^2 = (dB)^2 = dt$ and the quadratic variation process of M is the same as that of B, i.e. $[M]_t = [B]_t = \int_0^t (dB)^2$. Then:

$$L_M(r, s) = \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_0^r \mathbf{1}(|M(t) - s| < \varepsilon)\, d[M]_t$$

$$= \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_0^r \mathbf{1}(|B(t) - (s - bt)| < \varepsilon)\, dt$$

and the sojourn time of M at s can be regarded as a version of the sojourn time of Brownian motion recentred around the drift.

Finally, we compute the hazard functions for the trend stationary and difference stationary data shown in Figure 4(b) and 5(b). The hazard function for the trend stationary data is shown in Figure 4(c). According to (32), we would expect the hazard rate for (what is on average, at least) a uniform density to be a steadily increasing function, and Figure 4(c) shows this to be approximately so. On the other hand, for the difference stationary data, as is apparent from the spatial density estimate given in Figure 5(b), there are regions of greater spatial concentration around 3–5, 13–15 and 25–27. These regions show up as peaks in the hazard function shown in Figure 5(c). All these peaks appear significant. In both Figures 4(c) and 5(c) the standard errors rise with the spatial coordinate and ultimately the hazard estimates become unreliable as the data thins out in the right tail.

4. EMPIRICAL APPLICATIONS

4.1. Exchange Rate Data and Target Zones

Exchange rate data under floating regimes typically behave as if they have no fixed mean and are usually well represented by unit root processes. Intermediate between fixed and flexible exchange

rate regimes are target zone systems where exchange rates are permitted to float within bands. This exchange rate mechanism was operated by the European Monetary System (EMS) over the 1980s and 1990s. For most countries in the EMS, the bands were set at $\pm 2.25\%$ around a central parity that was occasionally adjusted by currency realignments. Such exchange rate target zones have been the subject of considerable research. A theory model for exchange rate target zones was developed by Krugman (1991, 1992) and has formed the starting point in much of the subsequent research on this topic. This model allows for the determination of the exchange rate within the band in terms of a non-linear function of economic fundamentals which are represented by Brownian motion. The non-linear function arises because of the presence of monetary policy intervention designed to keep the exchange rate in the band. It has an S-curve shape and approaches the edge of the band in a smooth tangential fashion that is characterized as 'smooth pasting' (Dixit, 1993). One of the empirical implications of this theory is that the distribution of exchange rates within the band should be U-shaped, so that there is, in effect, a greater concentration of observations closer to the edge of the band than in the center of the band.

The empirical evidence in support of this implication of the target zone model can be assessed by using our spatial density approach. Figure 6(b) shows the French Franc/German mark exchange rate from March 1979 through to March 1992 in terms of deviations from the central parity at the *beginning* of the period. This data was used by Svensson (1992) in his review of the target zone literature. The $\pm 2.25\%$ bands around the central parity are shown by broken lines in this figure. Successive realignments in the central parity are also shown, occurring in September 1979, October 1981, June 1982, March 1983, April 1986 and January 1987. All these realignments devalued the franc against the mark.

Combining these periods together by subtracting the central parity in each period, we may focus on the behaviour of the exchange rates within the bands, which was Krugman's (1991) concern, and ignore issues of jumps in the bands themselves, which is a different issue that has attracted subsequent attention (e.g. Pesaran and Ruge-Murcia, 1999). Figure 6(c) shows this transformed data and Figure 6(a) gives the spatial density for this data. The figure shows strong evidence of bimodality in the data, revealing a clear peak in spatial density at the lower edge of the band around -2%. The remainder of the data appears to be spread out fairly evenly over the positive region of the band [0, 2.25]. These results provide partial support for the conclusion from the target zone model that exchange rates should tend to cluster near the edges of the band. However, the spatial density is clearly not U-shaped, and there appears to be a strong tendency in these data for the exchange rate to spend a good deal of time in the centre of the band as well as near the edges, a feature that is also fairly evident in the time plot of the data. What the spatial density in Figure 6(a) adds to a close examination of the time plot of the data is a quantitative evaluation of the relative importance of different spatial locations, precisely the matter that needs to be addressed in considering the empirical implications of the target zone theory model. Thus, while the target zone model is rejected, it appears that there are some implications of the model that do find support in the data.

4.2. Inflation Data: Measuring Inflation and Deflation Hazards

Figure 7(a) shows annual CPI inflation rates for the USA based on monthly data for the CPI over the period 1934:1–1997:12. The time plot shows several periods of two-digit inflation, periods of deflation and substantial volatility in inflation, especially at higher rates. Figure 7(b) gives the

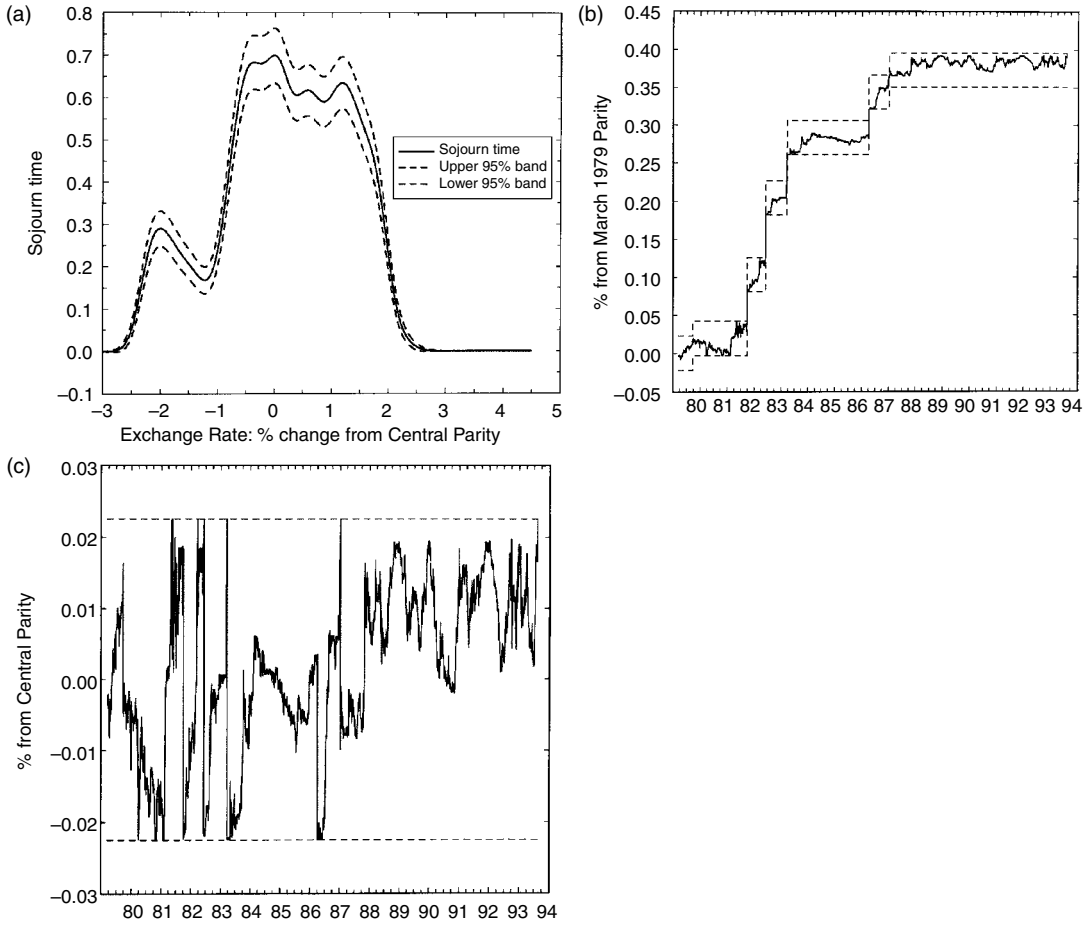

Figure 6. (a) Spatial density of FF/DM exchange rate; (b) FF/DM exchange rate data; (c) FF/DM data about central parity

spatial density estimate and reveals a primary concentration of variation around the 16% inflation rate, a mode around 9% inflation and a further mode indicating a peak in deflation around −2.5%.

Figure 7(c) gives hazard rate estimates for US inflation. Inflation hazards clearly peak at low levels (around 3%), intermediate levels (around 6%) and low two-digit levels (around 10–12%). The 3% and 10–12% peaks are both statistically significant and there is a fall-off in the hazard rate around 8%, between the peaks. The rising hazard around 16% is insignificant and can be ignored, being dependent on only a few observations. Overall, these estimates indicate that, conditional on there being inflation, historical experience over the last 60 years in the USA indicates that the predominant inflation risks are at low levels and low two-digit levels.

Figure 7(d) gives hazard rate estimates for US deflation. These are based on an estimate of the left-sided hazard function (21). The unbroken line in Figure 7(d) measures the conditional risk of inflation at a particular rate, given that inflation is no greater than that particular rate. Thus, when that inflation rate is negative, the curve measures the conditional risk of deflation. The estimates

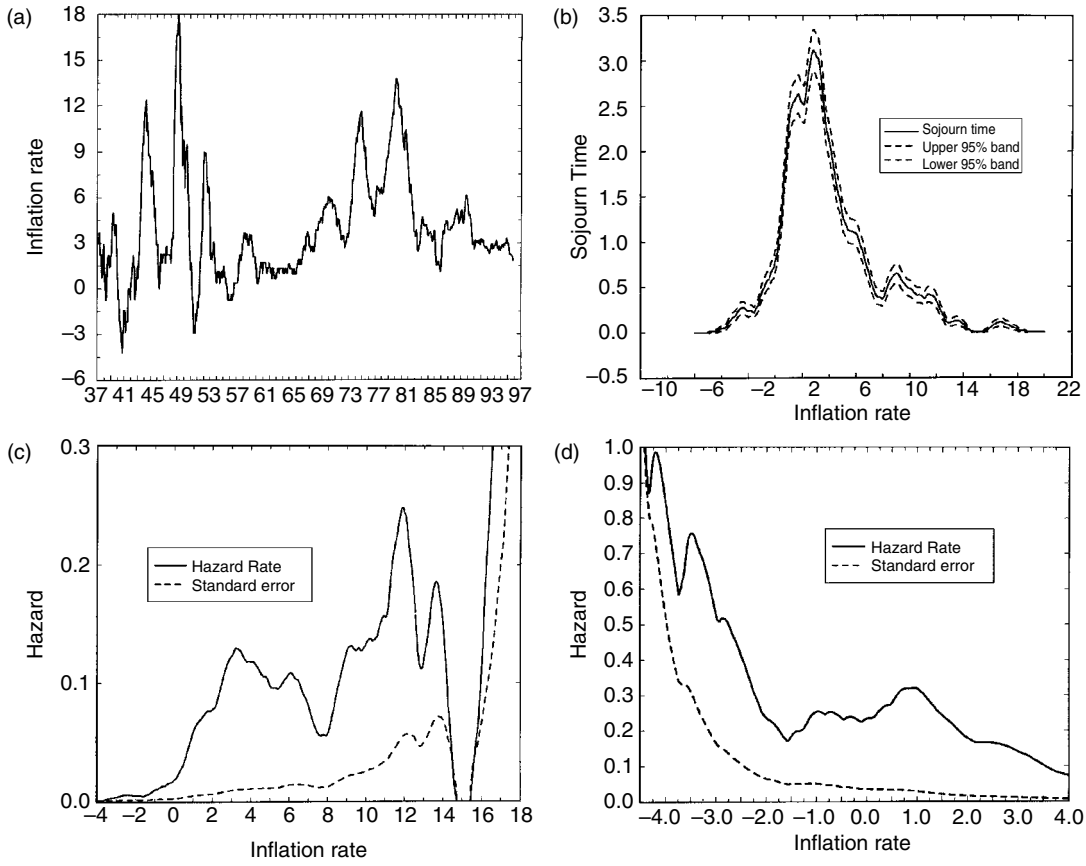

Figure 7. (a) US inflation 1934–97; (b) Spatial density of US inflation 1934–97; (c) US inflation hazard 1934–37; (d) US deflation hazard 1934–97

shown in Figure 7(d) indicate that there is a significant risk of deflation around the -1% level. The risk then falls off and subsequently rises again to another peak around -3.5%. This peak and the subsequent higher rates of deflationary risk are statistically insignificant and are based on only a single episode of high deflationary rates experienced in the interwar period. In sum, we may take these estimates as indicative of a non-negligible risk of low levels of deflation (around -1%) based on historical US experience.

4.3. Opinion Poll Data: Nixon and Clinton Approval Ratings

We end this empirical section with a brief application to political opinion poll data. Like many economic time series, there is substantial evidence that presidential opinion poll data are well modelled by unit root processes (e.g. Blood and Phillips, 1995, 1997). As such, this type of data is amenable to the type of descriptive statistical analysis considered in this paper. In some respects, our approach is particularly useful with data of this type because the data is usually not equally spaced in time, making conventional discrete time modeling difficult. Figures 8(b) and 9(b) show

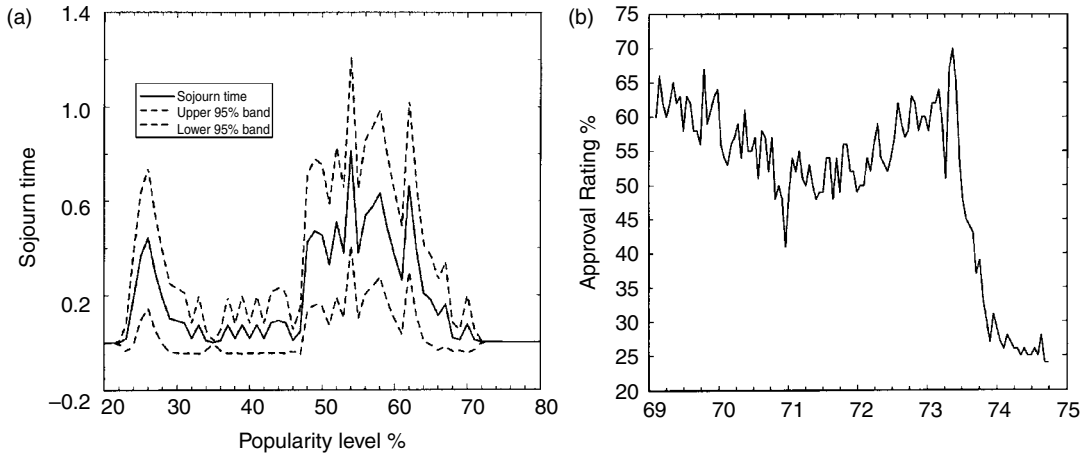

Figure 8. (a) Spatial density of Nixon approval ratings; (b) President Nixon approval ratings

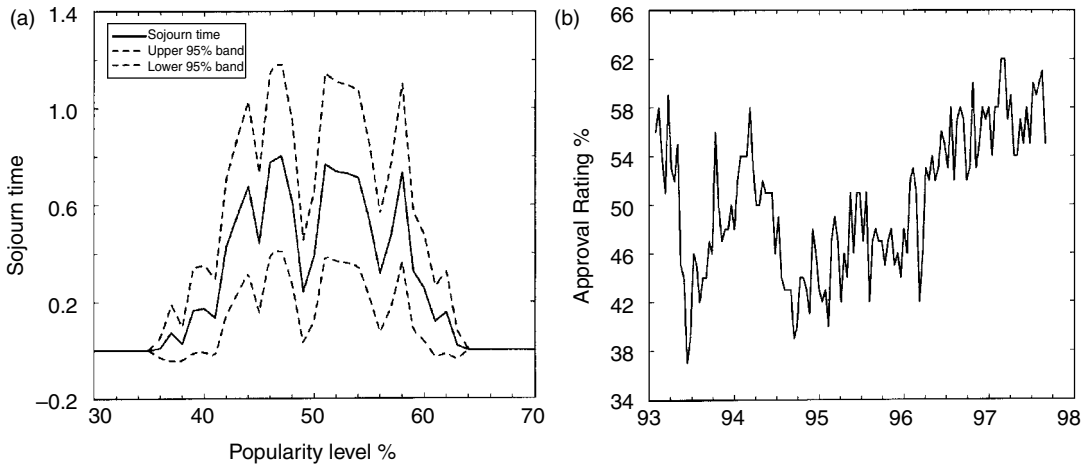

Figure 9. (a) Spatial density of Clinton approval ratings; (b) President Clinton approval ratings

opinion poll data for the Nixon and Clinton presidencies. From these data, the spatial densities were computed and are shown in Figures 8(a) and 9(a).

As is apparent from the time plot of the data in Figure 8(b), the latter part of the Nixon presidency was characterized by a sharp falling off in the ratings as the Watergate crisis culminated. For obvious reasons, Nixon approval ratings do not follow a stationary process. Nevertheless, we may conduct a spatial density analysis along the lines we have discussed. The spatial density estimate shown in Figure 8(a) manifests the phenomenon of the effects of the fulminating Watergate crisis on the Nixon presidency in a remarkably clear bimodality in the Nixon approval ratings, with a significant lower mode in approval around 25%.

Figure 9(b) gives a time plot of the approval ratings for President Clinton, up to November 1997, prior to the breaking of the Lewinski scandal. The spatial density of Clinton approval ratings are

shown in Figure 9(a). Apparently, it is hard to reject that Clinton approval ratings are uniformly distributed over the region [40%, 60%], indicating a clear difference in the spatial distribution of approval ratings between the two presidencies. Obviously, it will be of interest to use these methods to assess the effects of the Lewinski scandal on the spatial distribution of Clinton approval ratings.

5. CONCLUDING COMMENTS ON THE RELATION TO NON-PARAMETRIC METHODS

Econometric work, like other applications of statistics, involves data reduction. Even descriptive techniques, like the spatial densities and hazard rates that are applied here, necessitate a loss of information. Their usefulness comes from the need to discover regular features of the data and convenient means of expression for them, tasks that seems to be much more difficult for non-stationary data than they do for stationary data. However, while we no longer have a framework of time-invariant characteristics to rely upon when the data are non-stationary, we can find convenient quantitative representations of their sample characteristics without being dependent on the use of a specific model. Thus, whereas we no longer have fixed population moments or a time-invariant probability density to rely upon, we do have a well-defined concept of spatial location that has meaning beyond the immediate sample data. Our analysis shows that it is possible to construct quantitative measures of spatial density and apply these measures in an informative way to a variety of different data sets. Once this has been done, it is possible to use these measures in further constructive ways to estimate interesting functionals of spatial densities like hazard functions.

The techniques discussed and illustrated in this paper have a certain role to play in the ongoing evolution of econometric methods. In recent years, much of econometrics has been concerned with an attempt to achieve generality wherever possible, without sacrificing specificity where it connects most closely to underlying economic ideas. One way of attaining generality that has become increasingly popular in both microeconometric and time series studies is the use of non-parametric and semiparametric techniques. These techniques seek to avoid precise formulations or specific functional representations wherever generality is considered desirable and, thereby, it is hoped that the techniques will sit more comfortably with abstract propositions of economic theory. Nevertheless, existing validation of the use of these techniques has rested on the presence of invariant functional quantities, like a probability density or a spectrum, that can be estimated. What the methods of this paper show is that the notion of a general non-parametric approach to data analysis continues to retain validity even when the data are manifestly non-stationary and there are no underlying time-invariant quantities to estimate. What changes is not the approach to data analysis, but the interpretation of the empirical quantities that emerge from a non-parametric analysis. For non-stationary series, these quantities simply reflect variational decompositions across space rather than probability decompositions.

6. NOTATION

$\rightarrow_{a.s}$	almost sure convergence
$\rightarrow_{p}$	convergence in probability
$=_{d}$	distributional equivalence
$B(r)$	Brownian motion
$BM(\sigma^2)$	Brownian motion with variance σ^2

$\mathbf{1}(A)$	indicator of A
$MN(0, G)$	mixed normal distribution with mixing variate G
$\Rightarrow, \rightarrow_d$	weak convergence
$[\cdot]$	integer part of
$r \wedge s$	$min(r, s)$
$\equiv$	equivalence
$o_p(1)$	tends to zero in probability
$o_{a.s.}(1)$	tends to zero almost surely

ACKNOWLEDGEMENTS

Some of the methods and empirical results given here were first reported in one of the author's *Journal of Applied Econometrics* Annual Lectures at the University of Wisconsin in April 1998. Thanks go to the Department of Economics at the University of Wisconsin for hosting these lectures and to the JAE for extending the invitation to present them. The author thanks Zhian-Hua Zhu for supplying the opinion poll data, and Alex Maynard for help in obtaining the exchange rate data. Thanks also go to the NSF for research support under Grant Nos, SES 94-22922 and SBR-9730295. The computations were performed by the author in GAUSS.

REFERENCES

Blood DJ, Phillips PCB. 1995. Recession headlines, consumer sentiment, the state of the economy and presidential popularity: A time series analysis 1989–1993. *International Journal of Public Opinion Research* **7**: 2–22.

Blood DJ, Phillips PCB. 1997. Economic headline news on the agenda: new approaches to understanding causes and effects. In *Communication and Democracy: Exploring the Intellectual Frontiers in Agenda-Setting Theory*, McCombs M, Shaw DL, Weaver D (eds). Lawrence Erlbaum: London.

Bosq D. 1998. *Nonparametric Statistics for Stochastic Processes* (2nd edn). Springer: New York.

Csörgö M, Horváth L. 1993. *Weighted Approximations in Probability and Statistics*. Wiley: New York.

Dixit A. 1993. *The Art of Smooth Pasting*. Harwood Academic Publishers: Chur.

Espasa A, Sargan JD. 1977. The spectral estimation of simultaneous equation systems with lagged endogeneous variables. *International Economic Review* **18**: 583–605.

Feller W. 1957. *An Introduction to Probability Theory and its Applications* (Vol. I, 2nd edn). Wiley: New York.

Hannan EJ. 1963. Regression for time series. In *Time Series Analysis*, Rosenblatt M (ed.). Wiley: New York; 17–37.

Härdle W, Linton O. 1994. Applied nonparametric methods. In *The Handbook of Econometrics* (Vol. IV), McFadden DF, Eagle RF III (eds). North-Holland: Amsterdam.

Krugman P. 1991. Target zones and exchange rate dynamics. *Quarterly Journal of Economics* **106**: 669–682.

Krugman P. 1992. Exchange rates in a currency band: a sketch of the new approach. In *Exchange Rate Targets and Currency Bands*, Krugman P, Miller M (eds). Cambridge University Press: Cambridge.

Pesaran MH, Ruge-Murcia FJ. 1999. Analysis of exchange-rate target zones using a limited dependent rational expectations model with jumps. *Journal of Business and Economic Statistics* **17**: 50–66.

Phillips PCB. 1986. Understanding spurious regressions in econometrics. *Journal of Econometrics* **33**: 311–340.

Phillips PCB. 1987. Time series regression with a unit root. *Econometrica* **55**: 277–301.

Phillips PCB. 1988. Multiple regression with integrated processes. In *Statistical Inference from Stochastic Processes, Contemporary Mathematics*, Prabha NV (ed.). **80**: 79–106.

Phillips PCB. 1998a. Econometric analysis of Fisher's equation. Cowles Foundation Discussion Paper, No. 1180. *Yale University*.

Phillips PCB. 1998b. New tools for understanding spurious regressions. *Econometrica* **66**: 1299–1326.
Phillips, Solo. 1992. Asymptotics for Linear processes. *Annals of Statistics* **20**: 971–1001.
Phillips PCB, Park JY. 1998. Nonstationary density estimation and kernel autoregression. Cowles Foundation Discussion Paper No. 1181, Yale University.
Revuz D, Yor M. 1994. *Continuous Martingales and Brownian Motion* (2nd edn). Springer-Verlag: New York.
Sargan JD. 1953. An approximate treatment of the properties of the correlogram and peridogram. *Journal of the Royal Statistical Society, Series B* **15**: 140–152.
Sargan JD. 1958. The estimation of economic relationships using instrumental variables. *Econometrica* **26**: 393–415.
Sargan JD. 1959. The estimation of relationships with autocorrelated residuals by the use of the instrumental variables. *Journal of the Royal Statistical Society, Series B* **21**: 91–105.
Silverman BW. 1986. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall: London.
Sims CA. 1980. Macroeconomics and reality. *Econometrica* **48**: 1–48.
Svensson LEO. 1992. An interpretation of recent research on exchange rate target zones. *Journal of Economic Perspectives* **6**: 119–144.
Tyurin K, Phillips PCB. 1999. The occupation density of fractional Brownian motion and some of its applications. Yale University, mimeo.
Geman D, Horowitz J. 1980. Occupation densities. *Annals of Probability* **8**: 1–67.

