

A Consistent Nonparametric Test of Ergodicity for Time Series with Applications

Ian Domowitz and Mahmoud A. El-Gamal *

July 1997

Abstract

We propose a set of algorithms for testing the ergodicity of empirical time series, without reliance on a specific parametric framework. It is shown that the resulting test asymptotically obtains the correct size for stationary and nonstationary processes, and maximal power against non-ergodic but stationary alternatives. The test will not reject in the presence of nonstationarity that does not lead to ergodic failure. The work is linked to recent research on reformulations of the concept of integrated processes of order zero, and we demonstrate the means to operationalize new concepts of “short memory” for economic time series. Limited Monte Carlo evidence is provided with respect to power against the non-stationary and non-ergodic alternative of unit root processes. The method is used to investigate debates over stability of monetary aggregates relative to GDP, and the mean reversion hypothesis with respect to high frequency data on exchange rates. The test also is applied to other macroeconomic time series, as well as to very high frequency data on asset prices. Both the Monte Carlo and data analysis results suggest that the test has very promising size and power performance.

Respectively: Department of Economics, Northwestern University, Evanston, IL 60208, and Department of Economics, 1180 Observatory Drive, University of Wisconsin, Madison, WI 53706-1393. Email: domowitz@nwu.edu, mahmoud@osiris.ssc.wisc.edu. This draft and future revisions may be obtained at: <http://www.ssc.wisc.edu/~mahmoud/files/contest.ps> We thank Tim Conley, Mark Coppejans, Marty Eichenbaum, Ron Gallant, Clive Granger, Lars Hansen, Yuichi Kitamura, Andy Lo, George Tauchen, Ken West, and the participants at econometrics seminars at Duke, Indiana, Northwestern, and Wisconsin Universities for valuable comments.

A Consistent Nonparametric Test of Ergodicity for Time Series with Applications

Ian Domowitz and Mahmoud A. El-Gamal

1 Introduction

The concept of ergodicity is fundamental in the analysis of economic time series and of dynamic models calibrated by time series data. It is, therefore, surprising that no general testing procedure has been proposed to examine this important hypothesis. The objective of this paper is to fill this gap for the case of Markov processes.

Over the last decade, failure of the ergodic assumption has been largely studied for linear models (see Stock (1995), and Phillips (1995), and the references therein). The emphasis in this literature is on unit roots and the potential cointegration of multiple time series. A number of tests of the unit root hypothesis, in particular, have been proposed and analyzed. Research in the area has been aided enormously by clear definitions and application of the concept of an $I(0)$ random process, signifying stationary ergodic behavior of the process under study.

More recently, however, Granger (1995) demonstrates that the $I(0)$ concept is not well defined for nonlinear models. The importance of the latter class has grown, as computational power and theoretical sophistication in dynamic modeling have both increased. Granger proposes a replacement for $I(0)$, namely “short memory in mean” (SMM), but notes that how a test of this property is to be performed is not clear.

Building on our previous work in Domowitz and El-Gamal (1993), we provide an answer to this question, in the context of a more general testing strategy for the ergodic property in economic time series. In the latter paper, we propose a set of algorithms for testing the ergodicity of a known law of motion, showing that the resulting test asymptotically obtains the correct size and maximal power against stationary nonergodic alternatives. The reasoning behind the test is fairly simple. Stationary ergodicity of a data generating process, which is the most familiar form of ergodicity, is equivalent to the convergence of Cesàro averages of the transition probabilities to a unique invariant measure. A Markov operator on the space of densities is defined, which corresponds to the transition kernel on the state space. Testing the convergence of the Cesàro averages of the transition kernel is equivalent to testing the convergence of Cesàro averages of iterates of all initial densities to the same (unique) stationary density. The test is based on algorithms to draw i.i.d. samples from the Cesàro averages of iterates of any two initial densities through the Markov operator. Another algorithm then is used to randomly choose a pair of initial densities. It then could be shown under the assumption of stationarity of the underlying process that the probability of drawing two densities, whose Cesàro averages converge to the same limit under the alternative of nonergodic behavior, is zero, implying a consistent test.

The contribution of the present paper is to demonstrate that this testing strategy can be made operational when the law of motion is unknown. As a result, we obtain the first test for the ergodic behavior of an observed time series outside the unit root framework, to the best of our knowledge. Beyond our own earlier work, and that of Madsen (1975) on deterministic assessment of ergodicity for finite-dimensional Markov chains with known transitions, we know of no other literature on the problem. The present research is further motivated by several practical considerations.

The SMM concept proposed by Granger (1995) has the potential to be fundamental in extending our knowledge of persistence and cointegration in nonlinear models, if it can be operationalized. SMM is basically a restatement of the proposition that the data generating process has a mixing property. The relationship between testing for ergodicity and a test for mixing is explored conceptually in Domowitz and El-Gamal (1996). The techniques suggested here may be used directly to test the mixing property, replacing convergence of Cesàro averages of iterated densities with the weak convergence of the iterated densities themselves. The potential importance of such a result lies in the modeling question addressed by Granger (1995), namely identification of the proper “balance” of the right and left sides of a dynamic relationship, expressed in terms of persistence properties.

As a practical matter, theoretical modeling issues motivating our analysis go beyond such questions of balance in a posited relationship. The work of Duffie et al. (1994) demonstrates the difficulties of analytically verifying ergodicity in equilibrium models. Ergodic failure of processes arising from dynamic models can be common, however, and often may be impossible to verify analytically. Arthur (1989)’s analysis of production, for example, considers models of market share based on network considerations. Taking the probability of future market share to be dependent on the relative sizes of current market share, he shows that simple positive feedback leads to a failure of the ergodic property, in that the dynamics give many possible outcomes, based on initial conditions, rather than a unique non-cooperative equilibrium solution. Similar considerations enter the work of David (1986). Majumdar et al. (1986) demonstrate that the decision variables from a general class of dynamic programs can have an arbitrarily large number of ergodic subclasses associated with them. This possibility finds additional support in the rational expectations growth models of Eckstein et al. (1991). Other examples exist, most based on nonconvexities in the underlying problem of interest.

Estimation of such models relies on some variant of maximum likelihood or method of moments. The large literature on unit root analysis for linear structures has demonstrated the problems with estimation and inference for a very particular class of nonergodic statistical models. Similar problems with respect to more general models incorporating nonergodic processes, such as the non-normality of asymptotic distributions of estimators and the power properties of standard tests of hypotheses, are pervasive (e.g., Basawa and Scott (1983)). An operational test for the ergodicity of the relevant time series will enable a researcher to identify appropriate inference methods in empirical applications.

In some cases, an attempt to reach policy conclusions is based directly on data, as opposed to precise model formulations. A leading example concerns investigations of wealth distributions (e.g., Durlauf and Johnson (1992), Durlauf and Johnson (1995); Quah (1992)).

A typical exercise is to check whether or not per capita income converges towards a steady state growth path. Quah (1992), in particular, examines such convergence in the context of wealth distributions estimated from the data. Interestingly, the limit distributions found have the characteristics of mixtures, a feature of nonergodic laws of motion (Durrett (1991)). The economic implication is that any convergence across countries potentially depends on country-specific conditions, even in the long run. The finding would suggest refutation of models predicting convergence of wealth distributions internationally. The test proposed here can be used to provide a more rigorous framework for a statistical examination of such problems.

The plan of the paper follows. We introduce the maintained hypothesis behind our analysis and provide two motivating examples in the next section. The first application addresses the question of trend versus difference stationarity of different measures of money velocity; and the second concerns the issue of mean reversion in high frequency exchange rates. Both problems are rephrased in terms of the ergodicity of the process as exhibited by the data series. The basic framework of the test is covered in section 3. Rigorous definitions of the concepts of ergodicity and stationary-ergodicity are given, and an outline of the overall algorithm leading to the test statistic is presented. In section 4, we establish the consistency of the test for nonparametrically estimated laws of motion against the stationary but nonergodic alternative. A variety of nonstationary behavior is permitted under the null hypothesis of ergodicity, and this is formalized for the case of regime switches, in particular. Some practical issues of implementation also are treated in this section, with details of the algorithms underlying the test statistic relegated to an appendix. Section 5 is devoted to considerations of $I(0)$ processes and the short-memory definition of Granger (1995). We present generalizations that permit tests of short-memory related to his original concept. A Monte Carlo analysis of the performance of the test with respect to size and power, against the nonstationary and nonergodic unit root alternative in an AR(1) framework is offered in section 6. We return to the applications of the test to money velocity and exchange rates in section 7. We supplement these examples with other applications to both low and high frequency data. Our intention in the latter case is to further illustrate the type of analysis involved in data applications, and to demonstrate the performance of the test and reasons for rejection of the null (and failure thereof) under a wide variety of circumstances. Some concluding remarks close the paper.

2 Occasional Shocks and Motivating Applications

The maintained hypothesis for this paper is that each time series we consider contains two components: (i) a systematic transition density $p(x_t, \cdot | s_{t+1})$ defining the density of x_{t+1} conditional on its lagged value x_t and a “state” variable s_{t+1} , and (ii) an idiosyncratic “noise” term with density $\nu(\cdot)$ having full support. Since the state is not observable, and the noise term is typically not of economic interest, we restrict our attention to the systematic part of the process $\{x_t\}$. Formally, our goal is to obtain a consistent test of the ergodicity of the systematic part of the marginal process on $\{x_t\}$. The state variable s_t allows us to

encompass regime switching and other forms of nonstationarity which may or may not be accompanied by non-ergodicity. The stochastic process governing $s_t \in \{1, \dots, n\}$ is assumed to have a unique invariant distribution defining the proportion of time s_t spends in each state (e.g. Hamilton (1989), Engle and Hamilton (1990) achieve this through an aperiodic irreducible Markov transition matrix for s_t).

In addition to possible changes in the state s_t , we assume that the observed process $\{x_t\}$ includes occasional idiosyncratic shocks. The transition density of the observed process, may thus be defined as follows:

$$\tilde{p}(x_t, \cdot | s_{t+1}) = \begin{cases} p(x_t, \cdot | s_{t+1}) & \text{if } shock_t = 0 \\ (1 - \alpha)p(x_t, \cdot | s_{t+1}) + \alpha\nu & \text{if } shock_t = 1 \end{cases} \quad (*),$$

where $shock_t = 1$ according to some stochastic process. Unlike the process on s_t which enriches our model, the process on $shock_t$ is motivated by technical considerations. We shall return to the technical assumptions on the processes $\{s_t\}$ and $\{shock_t\}$ in sections 3 and 4. In this section, we wish to motivate this model within the context of two economic examples where policy prescriptions rest on the assumption of stationary ergodicity or lack thereof for particular processes.

2.1 Is There a Stable Relationship between Money and GDP?

It is generally accepted that the velocity of M1 (calculated as nominal GDP/M1) in the U.S. was stable from 1960 to 1980, in the sense of being trend stationary (with an average increase of roughly 3% per year). As seen in Figure 1, the trend seemed to reverse in the early 1980s, and its variance increased noticeably. The “shock” which most argue to have affected the demand for M1 circa 1980 consists of technological advances and new financial options (e.g., money market mutual fund accounts on which customers can write checks) which severed the tie between economic activity and the demand for M1.¹ As a result, many have argued that M1 has become an unreliable tool for monetary policy since 1980 (Friedman (1988), Blinder (1989)). Since the Federal Reserve began in the early 1980s to target M2 instead of M1 as its main measure of the money supply, the velocity of M2 has come under scrutiny. The failure of monetary policy in the early 1990s lent credibility to the hypothesis that the link between M2 and nominal GDP may have been severed (e.g. Friedman and Kuttner (1992)). However, Feldstein (1992) highlighted the fact that the statistical relationship between M2 and nominal GDP had not broken down, but that the Federal Reserve’s ability to target M2 was the source of ineffective monetary policy in the early 1990s. Feldstein and Stock (1993) find further support “that the relation between M2 and nominal GDP is sufficiently strong

¹ ATM networks started in Iowa in 1977, and quickly spread across the U.S. Merrill Lynch launched the first cash management accounts in 1977, and this type of account became very popular in the 1980s. Visa was created in 1977, with 5 million instant customers, leading to a mushrooming credit card industry. Money market funds were also introduced in 1977, with the total balances in such accounts reaching \$122 billion by 1980. All of those financial market innovations initiated in the late 1970s and taking full effect circa 1980 contributed to a reconfiguration of the liquidity profile of various portfolios, rendering M1 a poor predictor of economic activity.

and stable to warrant a further investigation into using M2 to influence nominal GDP in a predictable way” (p.1). They propose a simple Taylor (1985)-like rule of varying M2 in response to movements in nominal GDP which would reduce the volatility of GDP.

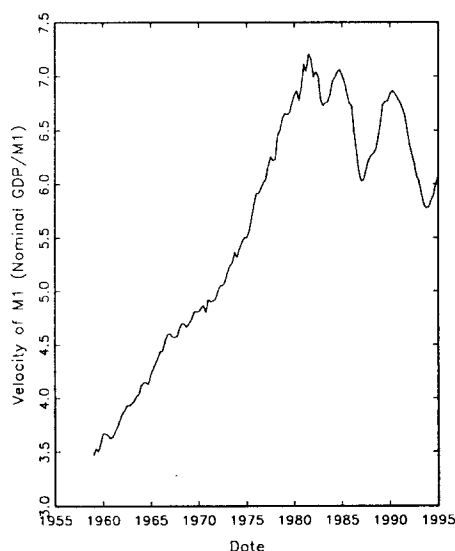


Figure 1: Velocity of M1: 1959:1–1995:2.

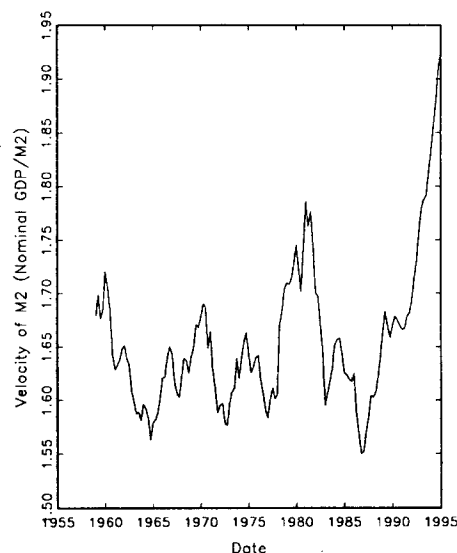


Figure 2: Velocity of M2: 1959:1–1995:2.

It is clear from the literature (e.g. see Edwards (1996)) that the “predictability” of the relationship between a measure of aggregate money supply and nominal GDP is determined in the sense of trend stationarity. In other words, if the detrended series for M_i ($i=1,2$) velocity is stationary and ergodic, we would say that the relationship between M_i and nominal GDP is stable (e.g. velocity of M1 for the 1960-1980 period). We then might infer that targeting the stock of M_i may be a useful tool for monetary policy. The stability of an M_i -to-GDP relationship was studied for M2 in terms of structural breaks in Feldstein and Stock (1993), and in terms of a cointegration relationships between M1, interest rates, and nominal GDP, in Stock and Watson (1993).

As an illustration of our proposed test, we shall abstract from the parametric forms necessary for conducting those types of stability tests, and conduct a non-parametric data-based test on simple transformations of the M_i -velocity series. The results reported later in the paper for quarterly data 1959:1–1995:2 suggest that the velocity of M1 is not trend stationary, but that it is difference stationary over the period, which agrees with the result in Stock and Watson (1993) that M1 and GDP follow unit root processes, and the cointegration relationship includes interest rates. In contrast, we fail to reject the result that the velocity of M2 series is ergodic, which agrees with the results of Feldstein and Stock (1993) and

Feldstein (1992).

2.2 Mean Reversion in High Frequency Exchange Rate Data

Mean reversion in exchange rates has been a topic of practical and academic interest since at least the mid 1980s. Most recently, attention has turned to the role and importance of non-linearity in the relationship between current and past rates in assessing the mean reversion hypothesis (e.g., Hsieh (1989), Engle and Hamilton (1990), O'Connell (1996), Bleaney and Mizen (1996a), as well as references therein). From the economic perspective, motivations for considering potential nonlinearities include agents' forecasting of discrete events, market frictions, and trader behavior (Flood and Garber (1983), O'Connell (1996), and Bleaney and Mizen (1996b), respectively).

Statistical investigations of the associated mean reversion hypotheses are carried out by extending an augmented Dickey-Fuller test to incorporate some form of nonlinearity related to the arguments above. Bleaney and Mizen (1996a), for example, add a third-order polynomial in the level of the lagged rate. Alternatively, a switching regression model, depending on a specification of *large* deviations of the lagged rate from its mean value, is employed. Such models are used in the context of real exchange rates by O'Connell (1996). A drawback of such parametric approaches is the need to tabulate additional critical values of the test statistic for any new specification of the nonlinear relationship. Further, as yet there is no stylized fact to be drawn from the data, with respect to persistence properties.

We use the theoretical and computational results of this paper to extend the empirical investigation of the mean reversion hypothesis in three dimensions. First, we employ high frequency trading data for currency rates, as opposed to the monthly, quarterly, and even annual rates used in other studies. Such data correspond more closely to explanations of nonlinearities based on trading behavior, in particular. Second, we account for any nonlinearity in the Markov process describing the evolution of currency prices through nonparametric estimation of the transition densities that underly our test statistic. This approach not only circumvents the problem of taking a parametric stand on the form of the model, but also obviates the need for extra critical value computations, as the theory will make clear. Finally, we reformulate the problem in terms of general ergodic failure, as opposed to the more specific case of a unit root in the data. We believe that this is more appropriate in nonlinear settings, given the essential linearity of unit root analysis, as discussed in Granger (1995). The occasional shock model used to obtain the power properties of our testing procedure is consistent with the O'Connell (1996), Bleaney and Mizen (1996a), and regime-switching specifications. The model is further used to motivate a decomposition of the series into multiple ergodic subclasses, in line with the suggested explanations for simple nonlinearity in the rate relationships.

We apply our test to the four series of futures currency prices for the British Pound, Deutschmark, Japanese Yen, and Swiss Frank, using 5-minute data for 1984–1993, as shown in Figures 3–6. There is a reasonably strong trend in all four series from 1984 through early 1987, followed by fluctuations around a constant mean. Such behavior suggests multiple

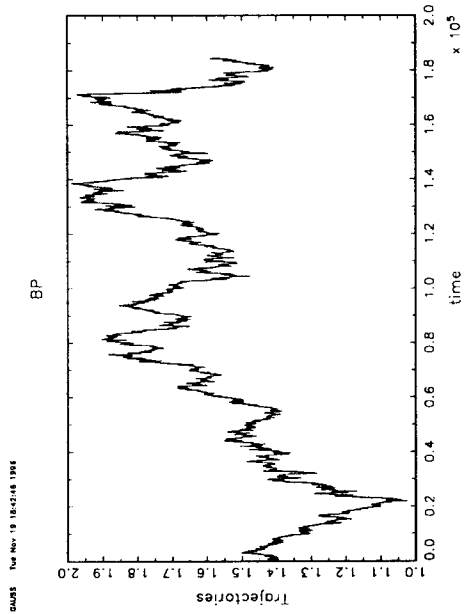


Figure 3: BP 1/3/84-4/30/93, 5 min. data

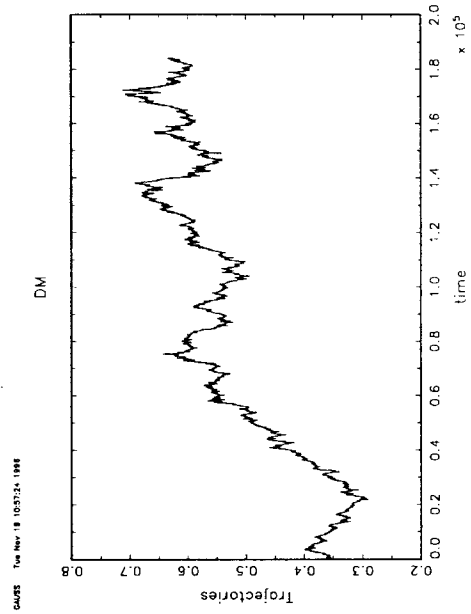


Figure 4: DM 1/3/84-4/30/93, 5 min. data

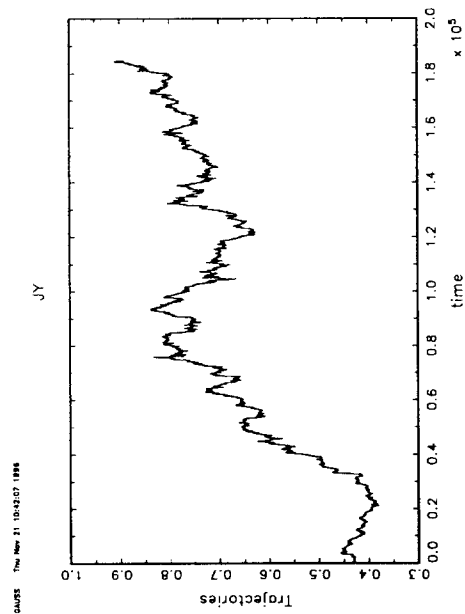


Figure 5: JY 1/3/84-4/30/93, 5 min. data

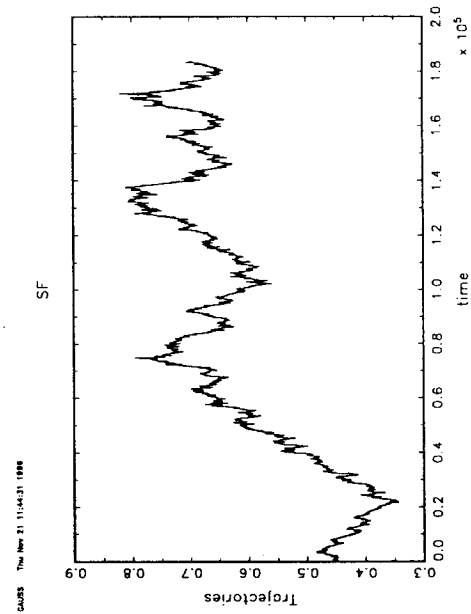


Figure 6: SF 1/3/84-4/30/93, 5 min. data

ergodic subclasses with respect to the underlying process. One such “shock” might be the Louvre Accord amongst G7 countries, reached in February, 1987. The Accord was based on an agreement with respect to stabilizing the behavior of the dollar, relative to other currencies, and was carried out by a series of central bank interventions, especially over the so-called “Louvre period” between 1987 and 1989. The four series exhibit this form of “mean reversion” post 1986. A previous “shock” associated with those series is the September 1985 G5 countries’ Plaza Accord which resulted in the downturn of the rates between 1985 and 1987. This idea of “occasional shocks” leading to switches (not necessarily instantaneous) between ergodic subclasses will be the central assumption for our testing procedure.

Our test results, discussed in Section 6, strongly reject the null hypothesis of ergodicity of those series. This result is consistent with the institutional considerations discussed above (see further discussion in section 6), as well as the models considered by O’Connell (1996) and Bleaney and Mizen (1996a).

3 A Consistent Test of Ergodicity

In this section, we reproduce the framework of Domowitz and El-Gamal (1993) for a known transition function of a univariate time series. We restrict attention to univariate 1st-order Markovian time series $\{x_t\}$. The extension to the multivariate case (and hence to the k^{th} -order Markovian case) is straight-forward, but the implementation of the test becomes much more tedious. We have illustrated in Domowitz and El-Gamal (1993, pp. 592-593) how we can algorithmically convert a multidimensional version of our test into a univariate one. We shall illustrate in Section 4 that a consistent test of the ergodicity of the $\{x_t\}$ process in the presence of possible state s_t transitions can be constructed based only on an estimated Markovian transition on x_t .

We consider a Markov process on \mathbb{R} defined by a transition function $p_T(\xi, A)$ for $\xi \in \mathbb{R}$, and $A \in \mathfrak{B}(\mathbb{R})$, where $\mathfrak{B}(\mathbb{R})$ is the Borel σ -algebra. For the remainder of this section, we shall take p_T as our given estimated law of motion (with density f_T), and we shall discuss the application of our testing strategy to this known law of motion. By following the steps of Domowitz and El-Gamal (1993), we obtain a consistent test of the ergodicity of this given law of motion. In section 4, we shall show that – for a consistent estimator p_T , and under suitable conditions – this will provide a consistent test for the underlying law of motion p as $T \uparrow \infty$.

We assume that for a given ξ , $p_T(\xi, \cdot)$ is a probability measure on $\mathfrak{B}(\mathbb{R})$, and for a given $A \in \mathfrak{B}(\mathbb{R})$, $p_T(\cdot, A)$ is a Borel measurable function. We shall refer to $p_T(\cdot, \cdot)$ as the one-step transition probability. As usual, we define the s -step transition probability recursively by:

$$p_T^{(s)}(\xi, A) = \int_X p_T^{(s-1)}(\xi, d\eta) p_T(\eta, A)$$

We assume that the probability measure $p_T(\xi, \cdot)$ is absolutely continuous, and we denote the corresponding (estimated) transition density by $f_T(\cdot, \cdot)$.

Starting from an initial density $g_0(x)$ on the state space \mathbb{R} , the probability of the process falling in any Borel set A at period s can easily be defined by:

$$Pr_{g_0}\{x_s \in A\} = \int_A g_0(\xi) \cdot p_T^{(s)}(\xi, d\eta) \equiv \int_A g_s(\eta) d\eta$$

This implicitly defines the Markov operator $P_T: D(\mathbb{R}) \rightarrow D(\mathbb{R})$ (via $g_s(\cdot) = P^s g_0(\cdot)$), where $D(\cdot)$ is the space of densities.

If a stationary density g^* exists for P_T , following Loève (1978, p.89), we say the stochastic process defined by $p_T: \mathbb{R} \times \mathfrak{B}(\mathbb{R}) \rightarrow [0, 1]$, or alternatively by $P_T: D(\mathbb{R}) \rightarrow D(\mathbb{R})$, is *stationary-ergodic* if there exists a unique measure π with a corresponding density g^* such that

$$\lim_{s \uparrow \infty} \frac{1}{s} \sum_{i=0}^{s-1} p_T^{(i)}(\xi, A) = \pi(A)$$

for all sets $A \in \mathfrak{B}(X)$, or alternatively

$$\lim_{s \uparrow \infty} \frac{1}{s} \sum_{i=0}^{s-1} P_T^i g(x) = g^*(x)$$

for all $x \in X$, and for all $g \in D(\mathbb{R})$. If there does not exist a stationary density for P_T , we define *ergodicity* by:

$$\lim_{s \uparrow \infty} \left| \frac{1}{s} \sum_{i=0}^{s-1} P_T^i g_1(x) - P_T^i g_2(x) \right| = 0,$$

for all $x \in X$, and for all $g_1, g_2 \in D^2(\mathbb{R})$. The relationships between stationarity, ergodicity, and other concepts are discussed in detail in Section 4. In particular, non-stationary but ergodic processes, which are part of our null hypothesis, are discussed in detail.

We may apply the test of ergodicity of Domowitz and El-Gamal (1993) to the estimated law of motion P_T with the corresponding transition density $f_T(\cdot, \cdot)$. We shall discuss in detail all the algorithms involved in implementing this test in the Appendix. The major steps needed to conduct the test are:

Algorithm 0

1. Choose a compact set in the support of f_T (practical rules will be discussed in Section 4).
2. Randomly draw two initial densities g and g' from the class of polynomial densities of degree k , using *Algorithm A* of the Appendix.
3. Construct two i.i.d. samples of size n from the Cesàro averages $\frac{1}{s} \sum_{i=0}^{s-1} P_T^i g$ and $\frac{1}{s} \sum_{i=0}^{s-1} P_T^i g'$, using *Algorithm B* of the Appendix.
4. Conduct a test of equality of two distributions (e.g. Kolmogorov Smirnov), and obtain the p-value for this test.

5. Repeat steps 2-4 to obtain a number of p-values. Under the null of ergodicity, the p-values so obtained should be uniformly distributed on $[0,1]$.

When considered as a test of the ergodicity of the given (estimated) law of motion P_T , we can consider a single test using steps 1-4 of *Algorithm 0*. The test would reject the null hypothesis of ergodicity if the p-value of this single randomized test is smaller than a pre-specified value. The nature of the alternative hypothesis of non-ergodicity against which we obtain asymptotic power 1 is discussed in Section 5. The following result holds for this test:

Theorem 1 *The test using steps 1-4 of Algorithm 0 obtains the correct asymptotic size under the null hypothesis of ergodicity, and asymptotic power unity (as $s \uparrow \infty$ and then $n \uparrow \infty$) under the alternative hypothesis of non-ergodicity in the presence of stationarity.*

Proof: Domowitz and El-Gamal (1993, Theorem 1, pp. 596-598).²

Remarks

- Step 1 of the algorithm is linked to the necessity of estimating the law of motion. The theoretical rationale, consequences, and the link to structural break methods are discussed in section 4.
- Step 4 leaves open the possibility of using alternative statistics in the procedure. In our applications to follow, we use the Kolmogorov-Smirnov test, based on the distribution of the sup-norm. An alternative might be that of Cramer-von Mises, based on the L^2 norm. Other possibilities are reviewed in Shorack and Wellner (1986).
- Step 5 recommends multiple replications. This is not required for the consistency of the test, as the proof would show. On the other hand, the test is randomized, in the sense of Bierens (1990), for example. Such procedures present the possibility that two researchers, with the same data and method; might reach two different conclusions, based on any single run. The replication recommended in step 5 rules this possibility out through the law of large numbers. In fact, the uniform distribution under the null could conceivably be tested formally, via an additional empirical distribution statistic.
- Step 2 refers to draws from densities represented by polynomials of degree k . As k approaches infinity, the set of such representations is dense in the space of densities. Letting k go to infinity poses no problems for the theory presented here, but moderate k gives very good approximations to most densities.³ We will take $k=10$ in our applications, for example.

² The idea of the proof proceeds as follows: Obtaining the correct size of the test under the null hypothesis follows directly from the equality of the two Cesàro average densities under the null. Obtaining asymptotic power of unity results from the violation of Harris recurrence (Harris (1956, p.115)). This violation makes the probability of drawing two initial densities using *Algorithm A* such that the Cesàro average densities generated from those two converge to the same limit under the alternative hypothesis is zero (Domowitz and El-Gamal (1993, Lemma 2, p. 597)).

³ See, for example, Gallant and Tauchen (1989), who examine the asymptotics, but implement their procedures for very small k .

4 Application of the Test in Time Series Contexts

In this section, we extend our analysis to prove that we can construct an estimator $p_T(\cdot, \cdot)$ of the marginal transition on x , perform the test on p_T following *Algorithm 0*, and obtain a consistent test for the ergodicity of the process $\{x_t\}$. The consistency of the test entails obtaining the correct asymptotic size and power unity against the stationary non-ergodic alternative as $T \uparrow \infty$ first, then as $s \uparrow \infty$, and $n \uparrow \infty$. The following lemma establishes the foundation for applying our test of ergodicity in a time-series context:

Lemma 1 *Let $q(x, \cdot)$ be the transition probability for a Markovian time series and let $p_T(x, \cdot)$ be a consistent estimator which satisfies for all $y \in \mathbb{R}$ and $B \in \mathfrak{B}(\mathbb{R})$:*

$$|p_T(y, B) - q(y, B)| \rightarrow 0$$

in probability as $T \uparrow \infty$, then:

As $T \uparrow \infty$, p_T is ergodic if and only if q is ergodic.

Proof: For the “if” part, let p be ergodic, and by the hypothesized convergence of $p_T(y, B)$ to $q(y, B)$ for all $y \in \mathbb{R}$ and $B \in \mathfrak{B}(\mathbb{R})$, it follows immediately that for all i ,

$$|p_T^{(i)}(y, B) - q^{(i)}(y, B)| \rightarrow 0,$$

in probability as $T \uparrow \infty$, which in turn implies

$$\left| \frac{1}{t} \sum_{i=0}^{t-1} q^{(i)}(y, B) - \frac{1}{t} \sum_{i=0}^{t-1} p_T^{(i)}(y, B) \right| \rightarrow 0,$$

in probability as $T \uparrow \infty$. Now, this implies the convergence of Cesàro iterates of initial measures under p_T to the Cesàro iterates of the same measures under q , and hence p_T inherits the ergodic property of q . The “only if” part follows equally directly by reversing the roles of p_T and q in the above argument. \blacksquare

4.1 The Maintained Hypothesis: Occasional Shocks

Following the standard procedure in econometric analysis, we model our given time series $\{x_t\}$ as consisting of two components: (i) a systematic law of motion, and (ii) an idiosyncratic shock process. Let the law of motion of the observed series be:

$$\tilde{p}(x_t, \cdot | s_{t+1}) = \begin{cases} p(x_t, \cdot | s_{t+1}) & \text{if } shock_t = 0 \\ (1 - \alpha)p(x_t, \cdot | s_{t+1}) + \alpha\nu & \text{if } shock_t = 1 \end{cases} \quad (\star)$$

where ν is a measure with full support, α is some scalar in $(0, 1)$, and $shock_t$ is a stochastic process taking values in $\{0, 1\}$. We assume that the law governing $shock_t$ is independent of $\{x_t, s_t\}$.

We are interested in testing the ergodicity of the systematic component of the marginal process on x . Towards that end, we make the following assumption:

A.0 The stochastic process governing s_t is independent of $\{x_t\}$, and has a unique invariant distribution $\sigma = (\sigma_1, \dots, \sigma_n)$ defining the asymptotic frequency $\sigma_i > 0$ of $s_t = i$.

This assumption is satisfied, for example, by Hamilton (1989)'s setup, where s_t follows an ergodic Markov chain. Now, define

$$p^*(x_t, \cdot) = \sum_{i=1}^n \sigma_i p(x_t, \cdot | s_{t+1} = i),$$

and

$$\tilde{p}^*(x_t, \cdot) = \sum_{i=1}^n \sigma_i \tilde{p}(x_t, \cdot | s_{t+1} = i).$$

The transition \tilde{p}^* defines the average law for the observable marginal process x_t , and the transition p^* defines the systematic component of that marginal process. We are interested in the ergodicity of the systematic law of motion p defined in (\star) . Since the s_t process is not observable, we need our test to depend only on the observable series $\{x_t\}$. Due to the assumed independence of s_t from $\{x_t\}$, the observed process $\{x_t\}$ still follows a Markov process. Moreover, the transition density for $\{x_t\}$ is always defined by one of n possible transitions $\tilde{p}_i(x_t, \cdot) = \tilde{p}(x_t, \cdot | s_t = i)$. Now, the transition $\tilde{p}^*(x_t, \cdot)$ defined above is the expected transition density under the invariant measure σ on $\{s_t\}$, and $p^*(x_t, \cdot)$ is the systematic part of that measure. In what follows, we shall construct the test by first estimating a Markovian transition density p_T on x , and testing the ergodicity of the estimated p_T . We shall make assumptions under which this estimator p_T will be shown to be consistent for \tilde{p}^* . Moreover, we shall make assumptions to guarantee that p_T is also consistent for the systematic component p^* . Using Lemma 1, this will imply that a consistent test of the ergodicity of p_T will be asymptotically consistent for p^* , as $T \uparrow \infty$. In the following lemma, we show that the transition p^* is ergodic if and only if the transition p is ergodic, thus yielding the consistency of the test for p .

Lemma 2 *The transition density $p(x_t, \cdot | s_{t+1})$ is ergodic if and only if the averaged transition $p^*(x_t, \cdot)$ is ergodic.*

Proof: For the first direction, let the process $\{x_t\}$ generated by $x_{t+1} \sim p(x_t, \cdot | s_t)$ be ergodic. This implies that there does not exist an ergodic decomposition of the state space $X = A \cup A^c$ such that if $x_t \in A$, $Pr_p\{x_{t+s} \in A; \forall s\} = 1$. This in turn implies that for all nontrivial measurable sets A, A^c , there exists an $i \in \{1, \dots, n\}$ such that $\int_A \int_{A^c} p(x, y | s = i) dy dx > 0$. Since $\sigma_i > 0, \forall i$, this in turn implies that for all nontrivial measurable sets A, A^c , $\int_A \int_{A^c} p^*(x, y) dy dx > 0$, i.e. that the process defined by the transition p^* is ergodic.

The opposite direction follows by reversing the argument. Assume that the process defined by the average transition p^* was not ergodic. Then, there exist nontrivial measurable sets A, A^c such that $\int_A \int_{A^c} p^*(x, y) dy dx = 0$. Since $\sigma_i > 0, \forall i$, this implies that for all i , $\int_A \int_{A^c} p(x, y | s = i) dy dx = 0$, which implies that the process generated by p is not ergodic. ■

In order to guarantee the consistency of our nonparametric estimate of $\tilde{p}^*(x_t, \cdot)$, we need the shock process to be sufficiently persistent. On the other hand, we need shocks to be

sufficiently infrequent (at least in the limit) so that the $\alpha\nu$ component does not dominate $p^*(x_t, \cdot)$, thus allowing us to obtain a consistent estimator of p^* . Our basic strategy is to employ the following assumptions:

A.1 $\sum_{t=1}^T \text{shock}_t \rightarrow \infty$, almost surely.

A.2 $\frac{1}{T} \sum_{t=1}^T \text{shock}_t \rightarrow 0$, almost surely.

Notice that there is a large number of scenarios under which both of those conditions are satisfied.⁴

We shall prove that we can extend the consistency of our test of ergodicity (as stated in Theorem 1) to the case of an unknown time series process satisfying our maintained hypothesis (\star) and assumptions **A.0**- **A.2**. Towards this end we shall need the following result:

Lemma 3 *Under the maintained hypothesis (\star) and assumption **A.1**, the following is true:*

1. *The marginal process $\{x_t\}$ has a unique stationary sigma-finite measure μ on \mathbb{R} (may be infinite).*
2. *The following weaker version of Doeblin's condition holds:*

For each τ , uniformly in $\xi_\tau \in \mathbb{R}$ (potential values for x_τ), there exists a finite-valued measure λ of sets $A \in \mathfrak{B}(\mathbb{R})$ with $\lambda(\mathbb{R}) > 0$, an integer $\tilde{\tau} \geq \tau$, and a positive ϵ , such that:

$$\tilde{p}^{*(\tilde{\tau})}(\xi_\tau, A) \leq 1 - \epsilon \quad \text{if} \quad \lambda(A) \leq \epsilon$$

3. *Consider a set $X \in \mathfrak{B}(\mathbb{R})$, with $0 < \mu(X) < \infty$, then the "process on X " as defined by Harris (1956, p. 113) is stationary and ergodic.⁵*

Proof: Under assumption **A.1**, condition C of Harris (1956, p.115) is satisfied, since the shock process guarantees that for all starting conditions x_0 , all sets of positive ν -measure (and hence positive Lebesgue measure) will be visited infinitely often with probability 1. Hence, by Harris (1956, Theorem 1, p. 116), the existence and uniqueness (up to constant positive multiples) of a sigma-finite measure μ which is invariant for the transition \tilde{p}^* follows.

⁴ For example, if the shock process arrives at the Fibonacci numbers 1,2,3,5,8,13,21,... Another example would be a Poisson process for shock arrivals, with the mean inter-arrival time diverging to infinity, etc.

⁵ "The process on X " is defined as follows: let $x_0 \in X$, and notice that by part 1 of the Lemma, almost all sequences x_0, x_1, x_2, \dots will have infinitely many elements in X . Construct the process on X as x_0, y_1, y_2, \dots , where each y_i is the first element of $\{x_t\}$ in X after y_{i-1} . In practice, we conduct our test by estimating the transition density $f_T(x, x')$ over a compact interval. Theoretically, we treat that compact interval as though it were fixed ex ante. In practice, we choose the interval by taking the middle 90% of any given sample. For the asymptotic results in this paper to hold, we need to let $T \uparrow$ for the given X , and then let X grow, but the order of the limits is crucial. This is similar in spirit to the practice in the estimation and/or testing for structural breaks, where one typically searches for the break point within the middle 90% of the sample.

The measure $\mu(\mathbb{R})$ may be finite (e.g. if the expected recurrence time is finite), or may be infinite.

We now turn to the proof of part 2. Under **A.1**, $\sum_{t=1}^T shock_t \rightarrow \infty$ almost surely implies that for all τ ,

$$\Pr\{\exists \tau' > \tau | shock_{\tau'} = 1\} = 1.$$

We can rewrite this condition as:

$$\lim_{\tau^* \uparrow \infty} \Pr\{\exists \tau'; \tau^* \geq \tau' > \tau | shock_{\tau'} = 1\} = 1.$$

This convergence in turn implies that for all $\delta > 0$, $\exists \tau^*$ such that:

$$\Pr\{\exists \tau'; \tau^* > \tau' > \tau | shock_{\tau'} = 1\} > \delta.$$

Now, under the maintained hypothesis (\star), construct $\lambda = \alpha\nu + (1 - \alpha)$, and identify ϵ in our Doebelin-type condition with our δ in the previous equation. Then, we have shown that for all $A \in \mathfrak{B}(\mathbb{R})$ with $\lambda(A) < \epsilon$, and for any starting time period τ , there exists a $\tilde{\tau}$ such that $\tau < \tilde{\tau} \leq \tau^*$, and $shock_{\tilde{\tau}} = 1$. For this $\tilde{\tau}$, an upper bound on the probability $\tilde{p}^{*(\tilde{\tau})}(\xi, A)$ is given by $\lambda(A)$ (since, when a shock hits, the maximal probability of $x_{\tau+\tilde{\tau}} \in A$ under the systematic part $p(\cdot, \cdot)$ is $(1 - \alpha)$, and the probability due to the shock is $\alpha\nu(A)$, and (by construction) $\lambda(A) = (1 - \alpha) + \alpha\nu(A)$). Thus, taking $\epsilon < 0.5$, we have shown that there exists a $\tilde{\tau}$ such that $\tilde{p}^{*(\tilde{\tau})}(\xi, A) < \epsilon < 1 - \epsilon$, for all $\xi \in \mathbb{R}$, hence concluding our proof.

For part 3 of the Lemma, notice that for the “process on X ”, the measure μ on X defines a unique invariant probability measure: for $A \in \mathfrak{B}(X)$, $P_\mu(A) = \mu(A)/\mu(X)$.⁶ ■

Now, we propose to estimate the transition density over a compact set X via the kernel density estimator:

$$f_T(x, x') = j_T(x, x')/m_T(x),$$

where

$$j_T(x, x') = \frac{1}{Th_T^2} \sum_{t=1}^{T-1} K\left(\frac{x - x_t}{h_T}\right) K\left(\frac{x' - x_{t+1}}{h_T}\right),$$

and

$$m_T(x) = \frac{1}{Th_T} \sum_{t=1}^T K\left(\frac{x - x_t}{h_T}\right).$$

We assume that the invariant probability measure on X defined by $P_\mu(A) = \mu(A)/\mu(X)$ is absolutely continuous, with density m . By Lemma 2, the process defined by $\tilde{p}^*(\cdot, \cdot)$ on X

⁶ In other words, if we observe the process whenever it is in X , and if the starting condition x_0 is drawn from the unique invariant measure restricted to X , $\mu(\cdot)/\mu(X)$, the average time it spends within any subset of X is defined by the unique sigma-finite invariant measure μ . For example, a random walk $x_t = x_{t-1} + \epsilon_t$ has a unique (infinite) invariant measure μ which coincides with Lebesgue measure on the Real line. Therefore, the process on a compact set X generated by such a random walk, if x_0 is drawn from the uniform measure on X , gives any set $A \in \mathfrak{B}(X)$ probability proportional to its Lebesgue measure. Indeed, this is the fundamental idea in obtaining ergodic theorems for processes which admit an infinite invariant measure, as in Harris and Robbins (1953).

is stationary and ergodic, hence the stationary density $m(\cdot)$ and joint density $j(\cdot, \cdot)$ of the process on X are well defined objects to estimate via $m_T(\cdot)$ and $j_T(\cdot, \cdot)$. We now add the following standard assumptions for consistency of kernel estimation:

A.3 $m(\cdot)$ and $j(\cdot, \cdot)$ are continuous.

A.4 $\lim_{|x| \uparrow \infty} |x|K(x) = 0$.

A.5 $h_T \downarrow 0$, and $Th_T^2 \uparrow \infty$, as $T \uparrow \infty$.

We are now ready for our main result:

Theorem 2 *Under Assumptions A.1-A.5, let*

$$p_T(x, A) = \int_A f_T(x, y) dy,$$

where $f_T(\cdot, \cdot)$ is estimated on a compact set $X \in \mathfrak{B}(\mathbb{R})$ as detailed above. Then, the test in steps 1-4 of Algorithm 0, applied to p_T , obtains the correct asymptotic size under the null of ergodicity of p , and asymptotic power unity against the alternative of non-ergodicity of p , as $T \uparrow \infty$ first, then $X \uparrow \mathbb{R}$, $s = O(T) \uparrow \infty$, and then $n \uparrow \infty$.

Proof: We first demonstrate the consistency of our estimator p_T under the maintained assumptions.⁷ Using the stationary ergodicity under Lemma 2, a direct application of the ergodic theorem to $m_T(x)$ and $j_T(x, y)$ produces the pointwise convergence of the averages: $m_T(x) \rightarrow E[m_T(x)]$, and $j_T(x, y) \rightarrow E[j_T(x, y)]$, in probability.

Under the additional assumptions **A.3-A.5**, it is well known (e.g. Roussas (1969, asymptotic unbiasedness Theorem 2.2, p.75)) that $E[m_T(x)] \rightarrow m(x)$ and $E[j_T(x, y)] \rightarrow j(x, y)$. It follows directly that $f_T(x, y) = j_T(x, y)/m_T(x) \rightarrow \tilde{f}(x, y)$ in probability, which directly produces the required consistency result $p_T(x, A) \rightarrow \tilde{p}^*(x, A)$ in probability, for all $x \in X$, $A \in \mathfrak{B}(X)$.

Now, it is clear that for all $x \in X$ and $A \in \mathfrak{B}(X)$,

$$\text{plim}_{T \uparrow \infty} p_T(x, A) = \Pr\{\text{shock} = 1\} \left[(1 - \alpha)p(x, A) + \alpha\nu(A) \right] + \Pr\{\text{shock} = 0\}p(x, A),$$

⁷ A stronger result of uniform consistency is available under assumptions **A.3-A.5**, and the sufficient condition D_0 ((a) Doeblin's condition holds, and (b) \tilde{p}^* has only a single ergodic set and it contains no cyclically moving subsets). Roussas (1969, Theorem 3.1, p. 77) has shown under those conditions that:

$$\sup_{z, A} |p_T(z, A) - \tilde{p}^*(z, A)| \rightarrow 0,$$

in probability. As we have shown in Lemma 2, the process on X is stationary and ergodic, and a weaker condition similar to Doeblin's condition holds. However, since Doeblin's condition does not hold under assumption A.2, and since uniform consistency is not needed for our purposes, we settle for the weaker pointwise consistency result stated here.

where

$$\Pr\{shock = 1\} = \lim_{T \uparrow \infty} \frac{1}{T} \sum_{t=1}^T shock_t.$$

Since the last term is assumed to converge to zero under **A.2**, this in turn implies that:

$$\text{plim}_{T \uparrow \infty} p_T(x, B) = \text{plim}_{T \uparrow \infty} \hat{p}^*(x, B) = p^*(x, B) ;$$

i.e. $p_T(x, B)$ is a consistent estimator of $p^*(x, B)$.

Under the null hypothesis of ergodicity of p^* , the test applied to p_T obtains the appropriate asymptotic size since p_T inherits the ergodicity of p^* (as well as p) in the limit, as shown in Lemmas 1,2.

Under the alternative hypothesis of non-ergodicity of p (and hence of p^*), we consider the case where p^* has at least one stationary density (which we have called the stationary non-ergodic case) In this first case, since p^* does not satisfy Harris recurrence, it is decomposable (Harris (1956)). Let A and B be two of the ergodic subsets of p^* . As we let $X \uparrow$, eventually A and B will be subsets of X . Then given any initial density g , simulated sequence $\{x_t^g\}$ of length s , and using the estimated transition p_T

$$\lim_{s \uparrow \infty} \frac{1}{s} \sum_{t=0}^{s-1} I_{\{x_{t+1}^g \in B | x_t^g \in A\}} = \lim_{T \uparrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} I_{\{x_{t+1} \in B | x_t \in A\}} \leq \lim_{T \uparrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} shock_t = 0.$$

This proves that under the alternative, the probability of observing transitions between ergodic subclasses of p^* converges to zero as $T \uparrow \infty$. Hence, as in Domowitz and El-Gamal (1993, Theorem 1, Lemma 2, pp. 596-598), the mass under two randomly drawn densities from *Algorithm A* of section 5 must be the same for sets A and B , which requires a linear relationship to hold between the coefficients of the two polynomial densities. This is a probability zero event under *Algorithm A*, and hence, we obtain the asymptotic power of unity against the non-ergodicity of p^* (and, by Lemma 2, of p). ■

Remarks:

- There are four possible properties a time series may possess in terms of stationarity and ergodicity. We have shown under the maintained assumptions (\star) , **A1-A5** that our test obtains the correct asymptotic size under the stationary and non-stationary ergodic cases included in the null hypothesis. Moreover, we have shown that under the same conditions, the test obtains asymptotic power of unity against the stationary non-ergodic case. Given our proof methodology under the maintained assumptions, we are unable to prove that the test will obtain asymptotic power unity against the non-stationary non-ergodic case.
- We have focused the analysis of nonstationarity under the null on behavior most akin to discrete regime shifts, as in Hamilton (1989). The test also will fail to reject in the presence of other forms of nonstationary, but ergodic behavior. These include the asymptotically stationary models of Kámpe de Fériet and Frenkiel (1962), simple

heteroskedasticity, as illustrated by Mokkadem (1987), stable time-varying parameter models, discussed by Rao (1978), and Markov models exhibiting general forms of non-stationarity of the transition kernel that nevertheless admit so-called weak ergodicity (e.g., Isaacson and Madsen (1976)). Formalization for such cases requires an extension of the state space of $\{s_t\}$ from finite to infinite, a complication we avoid to eliminate the need for measure-theoretic niceties that are extraneous to the analysis.

- The reason for indeterminacy of power in our framework under the non-stationary non-ergodic case is due to the necessity of conducting the test on a compact set to ensure stationary ergodicity (using Lemma 2), and consistency of the estimated transition density (first part of the proof of Theorem 2). The theoretical results let $T \uparrow \infty$, and then allow the compact set on which the estimation and testing is implemented to get large. Consider the leading example of a non-stationary non-ergodic process: a random walk $x_t = x_{t-1} + \epsilon_t$. This process admits a unique invariant measure, which is the (infinite) Lebesgue measure. When we consider “the process on (a compact) X ” for such a random walk, this process mimics a random walk with reflecting boundaries (since the periods outside the boundaries of X are ignored by such a process), which is stationary and ergodic. Considering such processes as $T \uparrow \infty$ and as $X \uparrow \mathbb{R}$ is ill-defined since the order of taking the limits leads to different properties.
- In Section 6, we shall conduct a small Monte Carlo analysis of AR(1) processes with unit roots and stationary roots, using designs similar to those in the literature. The Monte Carlo results suggest that our test does obtain reasonable power performance and very good size performance in this environment. In some instances, our test’s size-power combination compares favorably to tests under the unit-root null hypothesis, but a more elaborate Monte Carlo study is needed to confirm those preliminary results. Moreover, the data analysis in Section 7 consistently produce rejections of the null of ergodicity for series which are traditionally considered to have unit-roots. Those results in Sections 6 and 7 suggest that the test may have good power performance in practice against non-stationary non-ergodic alternatives. A theoretical result of consistency against such alternatives may require a different proof methodology for obtaining the correct asymptotic size.
- The use of standard critical values for the two-sample test in step 4 of Algorithm 0 suggests that uncertainty stemming from the estimation of P does not affect the test statistic. This is technically correct, although somewhat unusual in comparison with a variety of other specification tests in the literature. The main sources of uncertainty in the procedure are from a combination of random selection of initial densities and the sampling scheme for Cesàro averages of the density over possible trajectories of the dynamic process. This is quite similar to the analysis of Koop et al. (1996), in the context of nonlinear impulse response functions. In very small samples, estimation uncertainty might be viewed as a more serious problem. The distribution of test p-values then can be bootstrapped via a resampling scheme that permits replication of

the estimation of P . Such a procedure also is commonplace, for example, in impulse response analysis based on vector autoregressions.

4.2 Practical Considerations

In this section, we have extended the asymptotic size and consistency of our test of ergodicity to the case of a time series with kernel estimated transition density. The asymptotics require taking limits as $T \uparrow \infty$ first, then as $s = O(T) \uparrow \infty$, and finally as $n \uparrow \infty$. Since s and n are control parameters in our simulation-based test, this is not particularly restricting, and the crucial asymptotic assumption is the familiar one of $T \uparrow \infty$.

In any given sample of size T , however, the choice of s and n (as well as k) is not pre-specified. Our Monte Carlo analyses in Domowitz and El-Gamal (1993) and Domowitz and El-Gamal (1996) suggest that values of $k = 10$, $s = 50$, and $n = 200$ produce reasonable size and power performance in a particular set of examples. In the Monte Carlo simulations and empirical applications in this paper, we shall always use those values. As we shall see, the size and power performance of the test under those values is quite good in AR(1) and ARMA(1,1) models with stationary roots (for size analysis), and unit roots (for power analysis).

The empirical analysis of macroeconomic and financial time series results in rejection of the null of ergodicity for most series, and failure to reject that null for first differences of the series. In addition to the reasons given in the remarks following Theorem 2, these empirical results suggest that the unit root hypothesis is the most relevant alternative for those series, and our Monte Carlo studies suggest that those values of (s, n, k) are appropriate for obtaining the appropriate size and good power against that alternative. In other applications where a different alternative hypothesis may be more relevant, some Monte Carlo analysis to determine appropriate values of these parameters can be a valuable guide.

5 $I(0)$, Short Memory, and Extended Memory

Deviations from ergodicity in linear models usually involve unstable or explosive behavior, also entailing nonstationarity. There are three related points for nonlinear models. A nonlinear model may exhibit unstable behavior leading to a rejection of the hypothesis of ergodicity, but be stationary. An example is provided by a logistic transformation of a random walk, which exhibits behavior akin to an $I(1)$ variable, but has constant variance. Our test will reject the null in this case. Second, a nonlinear dynamical process may fail to be ergodic without exhibiting explosive behavior, also resulting in rejection of the null. Finally, a non-stationary nonlinear process may possess the ergodic property. An illustration of the last two points is given by the interactive Markov Chain concept developed by Conlisk (1976) and Brumelle and Gerchak (1980).

One response, given in Granger and Swanson (1995), is purely pragmatic: an $I(0)$ process is one that does not fail a powerful test having some general form of $I(0)$ as the null. This is the view implicitly adopted here, for reasons which will be clarified below. In doing so,

we relate Granger (1995)'s substitute for the standard $I(0)$ definition, that of short memory in mean, to the analysis in this paper. In the interest of clarity, we neglect some measure-theoretic niceties, but the discussion can be made entirely rigorous.

The vast bulk of econometric work on failure of ergodicity relies on linearity and the concept of an integrated process, say Y_t , denoted $I(1)$, such that $\Delta Y_t \equiv X_t$ is $I(0)$, i.e., integrated of order zero. The concept rests on a clear definition of what one means by $I(0)$. As part of an ambitious research program on persistence in nonlinear environments, Granger (1995) finds a variety of ambiguities and difficulties in precisely defining $I(0)$, especially for nonlinear data generating processes. His conclusion is that the standard $I(0)$ and $I(1)$ classifications are not sufficient for general analysis. Amongst other things, he notes that the concepts are too linear in form, differencing is a linear operation, and indeed the $I(1)$ concept relies on linear sums of $I(0)$ components.

Granger considers a series Y_t and an information set I_t , which consists simply of the past and present of the series. The unconditional mean of the series, μ , is assumed to exist and be a constant. He defines the short memory in mean concept (SMM), saying that Y_t is SMM if

$$E\left[|E(Y_{t+n}|I_t) - \mu|^2\right] < c_n,$$

where c_n is some sequence that tends to zero as n increases, and n indexes the forecast horizon. As one forecasts into the future, the information in I_t becomes less relevant, i.e., the process loses memory. He notes a possible extension of the concept, SMM of order $\theta > 0$, $SMM(\theta)$, when $c_n = O(n^{-\theta})$, i.e., inclusive of a rate of convergence. Failure of SMM, entailing the notion that the forecast is a function of I_t for all n , is called *extended memory in mean*, denoted EMM. The SMM idea, while conceptually both simple and useful, fails the pragmatic criterion noted above, in that Granger (1995) notes that how such a test is to be performed is not clear.

The SMM formulation is reminiscent of mixing conditions, and Granger indeed notes the equivalence of his SMM concept with that of mixing in mean. More generally, a process is mixing if and only if $\lambda_n \rightarrow 0$ as $n \uparrow \infty$, where, taking T to be the shift operator,

$$\lambda_n = |\eta(B \cap T^{-n}A) - \eta(A)\eta(B)|$$

for any two pairs of events A and B , where η is a probability measure.. The definition implies a form of asymptotic independence, relating to distributions as opposed to means, and a mixing process is certainly mixing in mean. We call a process satisfying such a mixing condition SMM-M, since memory is obviously lost. Clearly, a similar rate of convergence criterion to that proposed by Granger may be imposed here, leading to SMM-M(θ). This is, in fact, common, especially for stronger forms of the mixing assumption (e.g., α -mixing) used in central limit theorem applications for nonstationary processes.

SMM-M does not fail the pragmatic criterion for short memory. An alternative characterization of mixing is given in Domowitz and El-Gamal (1996, Theorem 1): a data-generating process is mixing if and only if $P^n g \Rightarrow f$, $\forall g \in D$, where \Rightarrow signifies weak convergence. A test for mixing is then available by employing the same strategy as the test for ergodicity,

but examining the convergence of $P^n g$ directly, as opposed to using the Cesàro limit. Details of such a test are given in the Appendix. Notice that for such a test, the null hypothesis (of mixing), and the alternative hypothesis (of non-ergodicity) are not exhaustive. This is not uncommon (see Lehman (1986, Chapter 9) for a detailed discussion), and the set of processes that are not mixing, but are ergodic, is very small and not generally of interest.

The link between ergodicity and mixing, and yet another alternative definition of short memory, may be provided by defining a variant of the ergodic coefficient of Dobrushin (1956). Let

$$\xi_n = \frac{1}{n} \sum_{h=0}^{n-1} \eta(B \cap T^{-h}A) - \eta(A)\eta(B).$$

If $\xi_n \rightarrow 0$ as $n \uparrow \infty$ for any two events A and B, the underlying process might be called SMM-E, representing an averaged form of asymptotic independence. The hyphenated E stands for ergodic, since SMM-E is necessary and sufficient for the ergodicity of the process.⁸ The terminology of short memory still is justified, since ergodic processes are certainly characterized by loss of memory, often corresponding to the original SMM concept in the case of stationary Markov processes, in particular. Once again, a rate of convergence could be posited, leading to SMM-E(θ), but this is uncommon practice.

SMM-E satisfies the pragmatic criterion for $I(\theta)$, which is one way to view the contribution of this paper. In this case, the concept of EMM is simply failure of the ergodic property.

We close this section with two related remarks. First, one could further define short memory in terms of geometric ergodicity, SMM-G, say: $\|P^n(x, \cdot) - \pi\| = O(\rho^n)$ for some $\rho < 1$. This concept may fail the pragmatic criterion, however, relating to our second point, which is that while short memory may be testable through the extensions proposed here, short-memory(θ) may not. Verification of rates of convergence typically involve rather precise knowledge of the nature of the underlying process and, to the best of our knowledge, cannot be ascertained directly from data.

6 A Limited Monte Carlo Analysis of the Unit Root vs. Stationary Root AR(1) Case

In this section, we present some Monte Carlo evidence with respect to the performance of the testing strategy previously described under the null of a stationary AR(1), against the alternative of a non-stationary non-ergodic unit-root AR(1). The analysis is limited in two dimensions. First, we make no attempt to cover all possible models containing unit roots that have been studied in the literature. We consider the simple random walk, a model with a moving average component, and a model with an additional stable autoregressive component, all for mean zero series (i.e., without constant terms or deterministic trends). These models

⁸ Proofs of this proposition are commonplace for the case where T is measure-preserving. Simply apply the definition to an invariant set C, i.e. $A = B = C$. Since $P(C) = 0$ or 1, every invariant set is trivial, and the result follows by the ergodic theorem. For results in more complicated environments, see Alpern (1976).

correspond, however, to the leading cases considered by Schwert (1989) and Elliott et al. (1996). Second, the work does not encompass a detailed comparison of various unit root tests relative to the test proposed here. We avoid the complications entailed by the fact that unit root testing schemes generally start from the null of nonergodic behavior, while our test proceeds from the ergodic null.⁹ Standardizations required for such comparisons would involve computing a variety of Monte Carlo experiments for unit root tests themselves, a topic for future research.

All tests are run for $T = 200$ and/or 500 , loosely corresponding to the lengths of post-war quarterly and monthly data series. A total of 1000 tests are computed for each experiment. Rather than report simple rejection frequencies, we summarize the results through a kernel density plot of p-values for the 1000 replications. Under the ergodic null, such densities should be approximately uniform, providing a global, as opposed to purely local, perspective. Under the alternative, the densities should exhibit a strong mode just to the right of $p = 0$, e.g., between 0 and 0.10, say, with a steep decline in the density as $p \rightarrow 1$.

The cases considered are based on the simple autoregressive model,

$$x_t = \rho x_{t-1} + \epsilon_T,$$

under three specifications for the error process. Model I takes the error to be i.i.d. white noise. Model II embodies a moving average assumption, $\epsilon_t = \eta_t - \theta\eta_{t-1}$. Model III posits a stable autoregressive disturbance, namely, $\epsilon_t = \phi\epsilon_{t-1} + \eta_t$. Innovations in the latter two specifications also are taken to be i.i.d. white noise, and all innovations are drawn from a $N(0, 10)$ distribution. Initial conditions x_0 are randomly drawn from $N(0, 100)$. We compute experiments for θ and ϕ each = 0.8 and 0.5, corresponding to cases considered by Schwert (1989) and Elliott et al. (1996). Model I simply takes these coefficients to be equal to zero.

Figures 1 and 2 show the densities of p-values from 1000 tests under Model I: $x_t = \rho x_{t-1} + \epsilon_T$, for $\rho = 0.8, 0.9, 0.95, 0.99, 1.0$, with $T = 200$ in Figure 1 and $T = 500$ in Figure 2. It is clear from these two figures that for all values of ρ up to 0.95, the size of the test is quite reasonable (the density of p-values is roughly uniform over $[0, 1]$). Moreover, the power performance of the test under the alternative $\rho = 1.0$ is quite reasonable, with 33% of the p-values falling under the traditional significance level of 0.05, and 40% falling under 0.1 for $T = 200$. With $T = 500$, 45% of the p-values are below 0.05, and 53% fall below 0.1. A remarkable feature of the size performance under the null hypothesis for $\rho \leq 0.95$ is that the density of p-values is unaltered for $T = 200, 500$. On the other hand, we can clearly see the improvement in power (for $\rho = 1$) as T gets larger. Throughout, we note that for the values of T analyzed here, $\rho = 0.99$ behaves virtually identically to $\rho = 1$. This is a well known feature of near-unit-root processes mimicking unit-root processes for sufficiently small T . We can see from Figures 1 and 2, however, that the size distortion is reasonable up to $\rho = 0.95$, which is considered strongly near-unit-root by most researchers.

Figures 3 and 4 show the densities of p-values from 1000 tests each under Models II and III. Under model II: $x_t = \rho x_{t-1} + \epsilon_t$, $\epsilon_t = \eta_t - \theta\eta_{t-1}$, the test obtains the correct size for the

⁹ Some notable exceptions exist; e.g. Kwiatkowski et al. (1992).

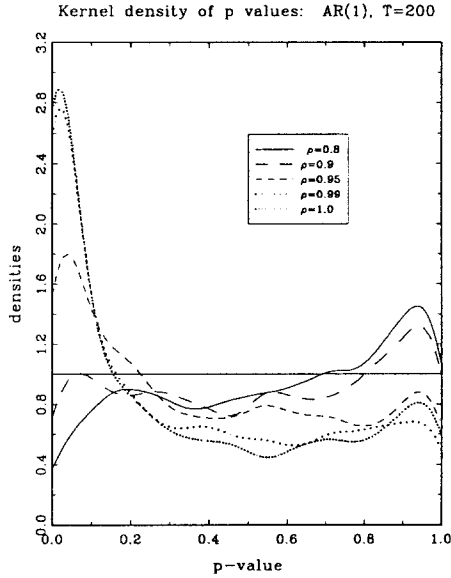


Figure 7: Density of 1000 p-values, T=200.

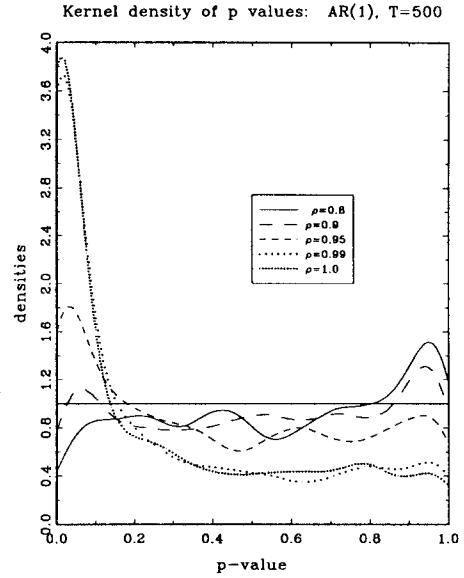


Figure 8: Density of 1000 p-values, T=500.

two cases ($\rho = 0.8, \theta = 0.8$) and ($\rho = 0.8, \theta = 0.5$). However, the performance under the unit root hypothesis mimics the results of Schwert (1989), where the case ($\rho = 1, \theta = 0.8$) is not distinguished by unit root tests (nor by our test) from the stationary root cases, whereas the case ($\rho = 1, \theta = 0.5$) looks more like a pure unit root process, but our power against (like the size of unit root tests) is not as satisfactory as in the pure AR(1) case. We note that this problem for the size of unit root tests and the power of our test against that alternative is a small-sample problem. The attenuation of the explosiveness in the series caused by higher persistence in the residuals causes the unit root process in small samples to mimic stationary processes.

In Figure 4, the error process is $\epsilon_t = \phi\epsilon_{t-1} + \eta_t$. Our test produces reasonable size under $\rho = 0.8$, and moderate power under $\rho = 1$, not dissimilar to the analogous size results in Elliott et al. (1996). For instance, for the case ($\rho = 1.0, \phi = 0.8$) in Figure 4, 47% of the p-values are below 0.05, and 55% of the p-values are below 0.1. In this small sample context, we would conclude from Figures 1-4 that our test has very good power for the cases where unit root tests have good size, and sometimes (as in the case of AR(1) residuals in Figure 4) may also have good power for cases where traditional unit root tests do not have good size performance. Moreover, our test has very good size (e.g. for $\rho = 0.95$ in Figures 1 and 2) for cases where unit root tests have low power against “near-unit-root” stationary alternatives.

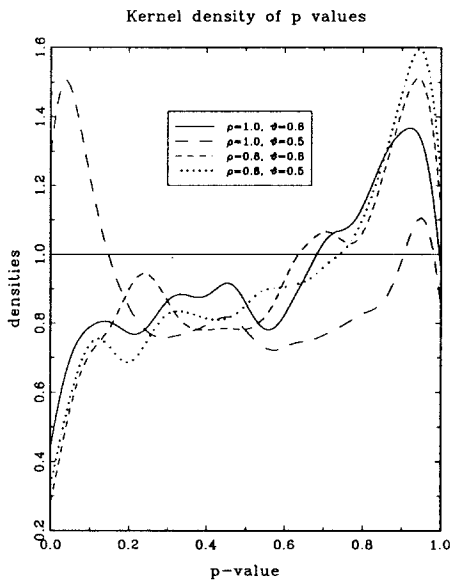


Figure 9: Density of 1000 p-values, T=200, MA(1) residuals.

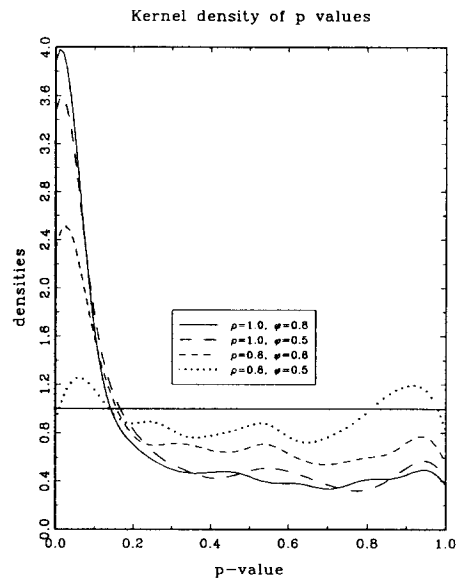


Figure 10: Density of 1000 p-values, T=200, AR(1) residuals.

7 Applications to Low and High Frequency Data

We now return to analyses of the stability of monetary aggregates and the mean reversion hypothesis with respect to high frequency exchange rates, introduced in section 2. Some additional points relevant to the performance of the test are discussed in the context of other applications to both high and low frequency data in a subsection to follow. As with the Monte Carlo analysis, results are presented in graphical form. For each series, a plot of the time series itself is accompanied by graphs of the marginal density of the process and the transition density, which are the essential inputs to the test. The test results themselves are presented via a kernel estimate of the density of p-values over 100 replications of the test.

7.1 Velocities of M1 and M2

We applied the test to the velocity of M1 (M1V) as shown in Figures 7–10. It is clear in Figure 10 that the test rejects a very large proportion of the time at all reasonable significance levels, and hence we strongly reject the null hypothesis of ergodicity of M1V. One observation we mentioned in Section 2.1 was the apparent trend stationarity of M1V between 1960 and 1980, before the trend reversal and increased volatility. Figures 15–18 report the results of applying our test to the detrended M1V for 1959–1995, which again shows a strong rejection of the null hypothesis of ergodicity of the detrended series. This is consistent with the results of Stock and Watson (1993) that the cointegrating vector for M1 and nominal GDP for those series extending beyond the 1980s must include interest rates. The agreement of our results with those of Stock and Watson (1993) is further illustrated in Figures 19–22, where the test fails to reject the null hypothesis of stationary ergodicity for the differenced M1V series, which together with the rejection of stationary ergodicity of the M1V series itself, suggests that M1V is $I(1)$.

The application of our test to M2V is illustrated in Figures 11–14. It is clear that we fail to reject the null hypothesis of stationary ergodicity for M2V. As discussed in Section 2.1, this result is consistent with the conclusions of Feldstein (1992) and Feldstein and Stock (1993) that there exists a stable relationship between M2 and nominal GDP, which renders targeting M2 a useful tool for monetary policy. This conclusion is further strengthened by the Feldstein and Stock (1993) result that the inclusion of interest rates does not eliminate the predictive content of M2.

7.2 Mean reversion of exchange rates

Our results pertaining to the futures prices of foreign exchange are illustrated in figures 23–38. The null of ergodicity is rejected soundly for all cases, as is obvious from the kernel estimate of the density of p-values across 100 realizations of the test statistic. The marginal densities exhibit the same multimodality as observed in the investigation of the money relationships. Wen et al. (1992), for example, identify such multimodality with nonlinear systems under disequilibrium and with transient periods of multistage growth. Such a characterization

may indeed be reasonable for the high frequency currency prices, given the description in Section 2. Multimodality also is consistent with mixtures of distributions, a defining feature of nonergodic behavior Durrett (1991).

Although we believe that our concept of occasional shocks to the process is best illustrated by the institutional considerations discussed in Section 2.3, it also can be linked to the models considered by O'Connell (1996) and Bleaney and Mizen (1996a). In the latter, the relationship between changes in the exchange rate and the lagged level is S-shaped, with a relatively flat interval surrounding an intercept of zero. In this case, a shock to the transition would be such as to place an observation far to the right or left of the flat region, moving the process into another ergodic subclass. In the former, such a shock would be linked to "large" movements of an observation relative to the median of the data.

The motivation for both of the models above is the idea that the process may be ergodic over some realizations of such shocks, but not over others. The issue of integrating structural break analysis of this type into our proposed methodology is beyond the scope of this paper, and is a topic for future research. Nevertheless, something may be said with respect to the results of O'Connell (1996), in particular, based on an intuitive inspection of the available results and the data. He finds that the process is "median-reverting" for small deviations of the exchange rate relative to the sample median, but follows a random walk for large such deviations. A plot of the median of our trading data against the raw series would suggest a similar result, in the sense that the stable behavior observed post-1987 is characterized by small deviations, while the observations in the pre-1987 look like large deviations, in the sense of his model. If we ignore problems associated with estimation of breakpoints, and simply assume that the Louvre Accord signals the break, formal testing would confirm O'Connell's basic results, though with a decidedly different interpretation.

7.3 Other applications

We have thus far illustrated the viability and power of the test via Monte Carlo analysis and applications to low frequency monetary aggregates and high frequency exchange rates. Space constraints preclude the detailed reporting of a variety of other applications undertaken. These include tests on a variety of macroeconomic time series, reminiscent of Nelson and Nelson and Plosser (1982) and Schwert (1987), as well as on an additional eight series of high frequency asset prices, covering markets in stocks, miscellaneous commodities, and short and long term interest rates. Details are available upon request, and we offer only a few comments relevant to the performance properties of the methodology.

The test applied to macroeconomic data series exhibiting well-documented trends (nominal and real GDP and GNP, M1 and M2, and GDP and GNP deflators) uniformly rejects the ergodic null. We do not expect this to be a surprise to anyone. The test rejects strongly, because low values observed early in such series are not revisited later in the sample with any frequency. We simply remark that the moderate power of the test identified by the unit root Monte Carlo evidence does not appear to translate into failure of the test in such empirical applications.

The multimodality of the marginal distribution noted in our previous applications carries over to all such series, consistent with the theory and results of Wen et al. (1992). We further note that multimodality, and associated failure of the ergodic property, may or may not be associated with structural breaks, in the usual sense of that term. This comment is a corollary to the discussion in Section 5, in which it was pointed out that time-varying parameter models may be nonstationary, yet reveal ergodic behavior. Real GDP, for example, exhibits strong multimodality, and the evidence on structural breaks in that series post-war is mixed. An analysis of the first difference of real GDP yields a strictly unimodal density, and a density of p-values for the test that is uniform, indicating failure to reject the ergodic null. Thus, should there be structural breaks in “detrended” GDP, they do not translate into failure of the ergodic property for the differenced series.

Tests pertaining to dependence properties of the high frequency time series involve between 95,000 to 190,000 observations over five to ten year periods. Such data provide a wider variety of behavior in the levels of the series than the macro variables. Stock index and interest rate series exhibit behavior similar to the macro time series, but over much shorter time spans. Commodity prices share a pattern, largely consisting of apparently nonlinear behavior, marked by little growth in levels. The null hypothesis is nevertheless rejected for all but one series in the latter category, for example, as well as for the trending series. For all rejections, the marginal densities exhibit the same multimodality as in the aggregate series, although not all series are trending. These results demonstrate the ability of the test to isolate deviations from ergodic behavior without long time spans of data, unlike unit root tests, as discussed by Campbell and Perron (1991). Further, our findings confirm the ability of the test to reject the null in the absence of any “explosive” behavior, as suggested by the discussion in Section 4. Since the latter is most often associated with failure of the ergodic property by economists, such results are of independent interest.

Finally, a comparison of results across some of the high frequency series suggests that apparent “trends” may not obviate the ergodic property. This possibility is illustrated by results on transaction prices for five-year government notes (FV) in Figures 39–42. Consider this series relative to the SF and DM currency processes, for example. A casual examination of the data might suggest that the SF and DM series are “more stationary”, at least in the sense of trend, than FV. The test strongly rejects the null for the currencies, but not for the interest rate series, however. Although it is tempting to ascribe this result to power properties, there is another reasonable explanation to be gleaned from the data. The SF series, for example, never returns to values below 0.55 after the early part of the sample. Similar behavior is observed for the DM, which fluctuates around the same value later in the period, but never returns to values below 0.4 observed earlier. In contrast, the FV process exhibits occasional returns to earlier price points, with enough frequency to result in failure to reject the null. Such behavior implies that a decomposition into ergodic subclasses is not possible. The process may return from subclass 2, say, to subclass 1, and from subclass 3 to 2, and so on, with positive probability, implying ergodic behavior.

8 Concluding Remarks

We have presented a procedure for a test of the null hypothesis of ergodicity of a process, under the assumption that the process is first-order Markov. It is shown that the test asymptotically obtains the correct size and maximal power against stationary nonergodic alternatives. The test will fail to reject for nonstationary processes that nevertheless have the ergodic property. This implies, for example, that simple heteroskedasticity or time-variation in model parameters will not result in spurious rejections. Examples of such processes are given by Kámpe de Fériet and Frenkiel (1962), Rao (1978), Gray (1988), Mokkadem (1987), and Isaacson and Madsen (1976).

The test may be applied to a known data generating process, in calibration exercises, for example, or to data, for which the specific form of the underlying process is unknown. In that case, the procedure is made operational through the nonparametric estimation of certain transition and marginal densities. Although we rely on an explicitly stochastic framework, Domowitz and El-Gamal (1996) show that the results apply equally well to nonlinear deterministic models. In that case, the theorems of El-Gamal (1991) can be used to generalize the procedure to unknown data generating processes, which are thought to be deterministic based on theoretical considerations.

The restriction to the first order univariate Markov case does not constrain the analysis as much as one might first presume. Higher order Markov processes may be reduced to multivariate first order schemes, as is well known. In that case, and for other considerations mandating multivariate analysis, Domowitz and El-Gamal (1993) provide a dimension-reduction algorithm, and prove that should the resulting univariate process be ergodic (nonergodic), then the multivariate process is ergodic (nonergodic). Those results are completely unaffected by the additional complication of estimating the unknown law of motion dealt with here. Thus, at the expense of adding one more algorithm to those described here, the test is generalized to the multivariate setting.

The test is randomized, in the spirit of Bierens (1990), for example. For this reason, we suggest that the test be run multiple times, basing inference on the density of observed p-values. This should mitigate a common criticism of randomized tests, namely that two researchers could reach very different conclusions based on the same data and procedure. Under the null, the density should be approximately uniform, a condition that could be checked using an additional goodness-of-fit test. Under the alternative, the density is sharply peaked at p-values close to zero, with a steep decline to relatively constant lower values of the statistic. We illustrate this procedure in Monte Carlo experiments against unit root and stable autoregressive processes, as well as in applications to low frequency macroeconomic data and high frequency asset prices.

There are many possible directions for future research, based on the ideas presented here, and we mention only a few. Extensions to nonlinear cointegration problems, one goal of the Granger (1995) research program, may be investigated. The application of the procedure to calibration exercises is yet to be fully explored. Christiano and Harrison (1996) make a start in that direction, but stop short of a formal testing strategy. The relevance of such research

lies in the determination of ergodic behavior in equilibrium, because ergodic behavior implies that macroeconomic policy results only in transitory changes, a prior motivation for the extensive literature on persistence via unit root processes. With respect to the latter class of processes, in particular, a more extensive comparison study of the relative performance of tests for unit roots and our more general procedure is called for, controlling for the differences in null and alternative hypotheses. Finally, we mention an application for which the ergodic properties of a process are known, but where the procedure may still prove useful, namely in Bayesian econometrics. Sampling algorithms used to produce posterior distributions for parameters of interest are stationary and ergodic by construction. It is often difficult to judge, however, whether a simulated distribution is sufficiently close to the unique invariant distribution based on examination of any given sequence in the process. Variants of the procedure introduced here may prove useful in providing a probabilistic assessment as to whether a simulation of any given length produces a sample path distribution close enough to the limit to be empirically useful, i.e., a stopping rule.

Algorithms

This appendix is devoted to a statement of the algorithms used in our test. The implementation of those algorithms in our Monte Carlo analyses and data analysis reported below was conducted in GAUSS version 3.2.29 for SunOS5.x Ultra, and using the GAUSS Maximum Likelihood Version 4.0.22/1 module.¹⁰ All of the computations in this paper were performed on a Sun Ultra2 with two 168Mhz processors.¹¹ For a given series, the code reads in the data and conducts the test by implementing the following steps:

Algorithm 1

1. Calculate cross-validation bandwidth for kernel estimation, using the GAUSS maximum likelihood module, and initializing the maximum likelihood search by a Silverman rule of thumb $h_T = \eta \times T^{-1/5}$, where η was selected to be the difference between the 55th and 45th percentiles (this choice is less sensitive to outliers than Silverman's choice of $\eta = \sigma$, the standard deviation in the series).¹²
2. Obtain kernel density estimates m_T , j_T and f_T as discussed in Section 4, using standard Normal kernels. The estimated transition density $f_T(x, y)$ is estimated on a 100×100 grid, and implies a finite approximation to the Frobenius-Perron operator on density space: $(P_T g)(x) = \int g(y) f_T(y, x) dy$.¹³
3. Perform n_{tests} tests (in the empirical applications in Section 7, $n_{tests} = 100$). Each test is constructed as follows (where, as discussed earlier, in all applications in this paper, $s = 50$, $n = 200$, $k = 10$):
 - (a) Use *Algorithm A* (below) to generate random coefficients (c_1, \dots, c_k) , and (c'_1, \dots, c'_k) , draw n data points, each i.i.d., from $g(x) = \sum_{i=1}^k c_i x^i$ and $g'(x) = \sum_{i=1}^k c'_i x^i$.
 - (b) Use *Algorithm B* (below) to randomly draw n i.i.d. samples each from $(1/s) \sum_{i=1}^s P_T^{i-1} g$ and $(1/s) \sum_{i=1}^s P_T^{i-1} g'$.
 - (c) Use *Algorithm C* to perform a Kolmogorov-Smirnov test of the equality of the distributions of the two samples, and obtain p-value for the test.
4. Plot the time series, the estimated marginal density and transition density, and a kernel smoothed density of the p-values so obtained. Under the null-

¹⁰ GAUSS is a trademark of Aptech Systems, Inc.

¹¹ However, the code is easily portable to other platforms under which GAUSS code can be run. The current code is available upon request from the authors, who aim to make a fully portable version available on the internet in the near future.

¹² For an excellent survey of different methods of bandwidth selection for kernel density estimation, see Jones et al. (1996). For a discussion of the applicability of essentially the same bandwidth selection methods in time series contexts, see Bosq (1996, pp.88-91).

¹³ For a justification of using finite approximations of the Frobenius-Perron operator on the density space, see Li (1976). Alternative analysis justifying finite approximations to continuous processes is contained in Guihenneuc-Jouhaux and Robert (1996).

hypothesis of ergodicity, this density should be uniform over $[0,1]$, and under the alternative, it should put a significant weight on low values of p .

We now proceed to a documentation of the three component algorithms A , B and C . The steps in those three algorithms are very similar to those implemented in Domowitz and El-Gamal (1993) and Domowitz and El-Gamal (1996), and we include them for completeness.

Algorithm A

1. Choose a compact interval $[a, b]$, the support of initial densities.
2. Generate a random vector p_0, \dots, p_k on the $k + 1$ -dimensional simplex as follows:
 - (a) Generate U_1, \dots, U_k , i.i.d. $U[0,1]$.
 - (b) Generate the order statistics $U_{(1)}, \dots, U_{(k)}$, and $U_{(0)} = 1$, and $U_{(k+1)} = 1$.
 - (c) Set $p_i = U_{(i+1)} - U_{(i)}$.
3. Compute c_0, \dots, c_k :
 - (a) For $i = 1, \dots, k$, generate V_i i.i.d. $U[0,1]$.
 - (b) Set $c_i = (i + 1)p_i$ if $V_i > 0.5$, otherwise, set $c_i = -(i + 1)/i p_i$.
 - (c) Set $c_0 = p_0 - \sum_{i:c_i < 0} c_i$.
4. Generate i.i.d. from the density $g(x) = \sum_{i=0}^k c_i x^i$ using the algorithm of Ahrens and Dieter (1974).
5. Transform the i.i.d. draws (on $[0,1]$) by multiplying them by b and adding a , to produce i.i.d. draws on the compact set $[a, b]$.

Algorithm A provides us with samples of *i.i.d.* draws from a polynomial density of order k whose coefficients are drawn randomly. The product of each application *Algorithm A* is a sample of n i.i.d. draws from the density $g(x) = \sum_{i=0}^k c_i x^i$. To obtain a sample of n i.i.d. draws from the Cesàro averaged density: $\sum_{j=0}^{s-1} P_T^j g$, we use the following algorithm:

Algorithm B

1. For each of the n observations x_i^g , draw a value $t_i \in \{0 \dots s\}$ with probability $1/(s + 1)$ for each value.
2. For the i^{th} observation, iterate for $j = 1, \dots, t_i$:
 - (a) Given x_i^g , normalize $f_T(\tilde{x}_i^g, \cdot)$ to sum to one; where \tilde{x}_i^g is the closest point on the 100×100 grid to x_i^g .
 - (b) Replace x_i^g with a random draw from the multinomial random variables with probabilities defined by the normalized $f_T(\tilde{x}_i^g, \cdot)$.
 - (c) Repeat steps (a) and (b) t_i times.

Now that we can apply *Algorithm A* and *Algorithm B* twice each to produce two i.i.d. samples from the Cesàro averages of two randomly drawn initial densities, we apply the two-sample Kolmogorov-Smirnov test of equality of two distributions to those two samples. *Algorithm C* was implemented by coding the algorithms *kstwo* and *probks* from Press et al. (1988, pp. 493-4) in GAUSS (using the built-in sorting function in GAUSS). We note that as in Domowitz and El-Gamal (1996), a test of mixing would be constructed by setting all the t_i 's in step 1 of *Algorithm B* at $t_i = s$. Note, however that such a test would attain the correct asymptotic size under the null hypothesis of mixing, but it is only guaranteed to be consistent (i.e. have asymptotic power of unity) against nonergodic (and not necessarily all non-mixing) alternatives.

References

- Ahrens, J. and U. Dieter. 1974. Computer methods for sampling from gamma, beta, poisson, and binomial distributions. *Computing* 12:223–246.
- Alpern, S. 1976. New proofs that weak mixing is generic. *Inventiones Mathematicae* 32:263–279.
- Arthur, W. 1989. Competing technologies and lock-ins by historical events. *The Economic Journal* 99:116–131.
- Basawa, I. and D. Scott. 1983. *Asymptotic optimal inference for non-ergodic models*. New York: Springer-Verlag.
- Bierens, H. 1990. A consistent conditional moment test of functional form. *Econometrica* 58:1443–1458.
- Bleaney, M. and P. Mizen. 1996a. Are real exchange rates stationary? evidence from sterling bilateral rates 1973–93. *Economic Notes* 25:465–482.
- Bleaney, M. and P. Mizen. 1996b. Non-linearities in exchange rate dynamics. *Economic Record* 72:36–45.
- Blinder, A. 1989. Comment. *Brookings Papers on Economic Activity* 1:111–114.
- Bosq, D. 1996. *Nonparametric statistics for stochastic processes: Estimation and prediction*. New York: Springer.
- Brumelle, S. and Y. Gerchak. 1980. A stochastic model allowing interaction among individuals and its behavior for large populations. *Journal of Mathematical Sociology* 7:73–90.
- Campbell, J. and P. Perron. 1991. Pitfalls and opportunities: What macroeconomists should know about unit roots. In Blanchard, O. and S. Fischer, eds., *NBER Macroeconomics Annual 1991*. MIT Press, Cambridge, MA.
- Christiano, L. and S. Harrison. 1996. Chaos, sunspots, and automatic stabilizers. *Journal of Monetary Economics* forthcoming.
- Conlisk, J. 1976. Interactive markov chains. *Journal of Mathematical Sociology* 4:157–185.
- David, P. 1986. Narrow windows, blind giants, and angry orphans: The dynamics of systems rivalries and dilemmas of technology policy. Center for Economic Policy Research working paper no. 10. Stanford University.
- Dobrushin, R. 1956. Central limit theorems for non-stationary Markov chains I and II. *Theory of Probability and Its Applications* 1:56–80,329–383.

- Domowitz, I. and M. El-Gamal. 1993. A consistent test of stationary ergodicity. *Econometric Theory* 9:589–601.
- Domowitz, I. and M. El-Gamal. 1996. Is your model chaotic? a consistent simulation-based test. In Lakshmikantham, V., ed., *Proceedings of the First World Congress of Nonlinear Analysts*. Walter de Gruyter.
- Duffie, D., J. Geanakoplis, A. Mas-Colell, and A. McLennan. 1994. Stationary markov equilibria. *Econometrica* 62:745–782.
- Durlauf, S. and P. Johnson. 1992. Local versus global convergence across national economies. Financial Markets Group Discussion Paper no. 131. London School of Economics.
- Durlauf, S. and P. Johnson. 1995. Multiple regimes and cross-country growth behavior. *Journal of Applied Econometrics* 10:365–384.
- Durrett, T. 1991. *Probability: Theory and examples*. Pacific Grove, CA: Wadsworth and Brooks/Cole Advanced Books.
- Eckstein, Z., C. Foulides, and T. Kollintzas. 1991. On the many kinds of growth. Center for Economic Policy Research Discussion Paper no. 592. Stanford University.
- Edwards, F. 1996. *The new finance: Regulation and financial stability*. Washington D.C.: AEI Press.
- El-Gamal, M. 1991. Non-parametric estimation of deterministically chaotic systems. *Economic Theory* 1:147–167.
- Elliott, G., T. Rothenberg, and J. Stock. 1996. Efficient tests for an autoregressive unit root. *Econometrica* 64:813–864.
- Engle, C. and J. Hamilton. 1990. Long swings in the dollar: Are they in the data and do markets know it? *American Economic Review* 80:689–713.
- Feldstein, M. 1992. The recent failure of U.S. Monetary Policy. NBER working paper #4236.
- Feldstein, M. and J. Stock. 1993. The use of a monetary aggregate to target nominal gdp. NBER working paper #4304.
- Flood, R. and P. Garber. 1983. A model of stochastic process switching. *Econometrica* 51:537–551.
- Friedman, B. 1988. Lessons on Monetary Policy from the 1980s. *Journal of Economic Perspectives* 2:55–72.
- Friedman, B. and K. Kuttner. 1992. Money, income, prices and interest rates. *American Economic Review* 82:472–492.

- Gallant, A. and G. Tauchen. 1989. Semiparametric estimation of conditionally constrained heterogeneous processes. *Econometrica* 57:1091–1120.
- Granger, C. 1995. Modelling nonlinear relationships between extended-memory models. *Econometrica* 63:265–280.
- Granger, C. and N. Swanson. 1995. Further developments in the study of cointegrated variables. Economics Discussion Paper 95-11. U.C. San Diego.
- Gray, R. 1988. *Probability, random processes, and ergodic properties*. New York: Springer Verlag.
- Guihenneuc-Jouhaux, C. and C. Robert. 1996. Discretizations of continuous state space markov chains for mcmc convergence assessment. CREST working paper 9630, Institut Nationale de la Statistique et des Etudes Economique.
- Hamilton, J. 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57:537–84.
- Harris, T. 1956. The existence of stationary measures for certain markov processes. In Neyman, J., ed., *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume II*, pages 113–124. University of California Press, Berkeley.
- Harris, T. and H. Robbins. 1953. Ergodic theory of markov chains admitting an infinite invariant measure. *Proceedings of the National Academy of Science* 39:860–864.
- Hsieh, D. 1989. Testing for nonlinear dependence in daily foreign exchange rates. *Journal of Business* 62:339–368.
- Isaacson, D. and R. Madsen. 1976. *Markov chains: Theory and applications*. New York: John Wiley and Sons.
- Jones, M., J. Marron, and S. Sheather. 1996. A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association* 91:401–407.
- Kámpe de Fériet, J. and F. Frenkiel. 1962. Correlation and spectra for nonstationary random functions. *Mathematical Computation* 16:1–21.
- Koop, G., M. Pesaran, and S. Potter. 1996. Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics* 74:119–147.
- Kwiatkowski, D., P. Phillips, P. Schmidt, and Y. Shin. 1992. Testing for the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics* 54:158–178.
- Lehman, E. 1986. *Testing statistical hypotheses*. New York: John Wiley and Sons.
- Li, T. 1976. Finite approximation for the Frobenius-Perron operator: A solution to Ulam's conjecture. *Journal of Approximation Theory* 17:177–186.

- Loève, M. 1978. *Probability theory II, 4th edition*. New York: Springer-Verlag.
- Madsen, R. 1975. Decidability of $\alpha(p^k) > 0$ for some k . *Journal of Applied Probability* 12:333–340.
- Majumdar, M., T. Mitra, and Y. Nyarko. 1986. Dynamic optimization under uncertainty: Non-convex feasible sets. Manuscript. Cornell University.
- Mokkadem, A. 1987. Sur un modèle autoregressif non lineaire, ergodicité, et ergodicité géométrique. *Journal of Time Series Analysis* 8:195–204.
- Nelson, C. and C. Plosser. 1982. Trends and random walks in macroeconomic time series. *Journal of Monetary Economics* 10:139–162.
- O'Connell, P. 1996. Market frictions and relative traded-goods prices. Technical report, Harvard University. manuscript.
- Phillips, P. 1995. Fully modified least squares and vector autoregression. *Econometrica* 63:1023–1078.
- Press, W., B. Flannery, S. Teukolsky, and W. Vetterling. 1988. *Numerical recipes in C*. Cambridge: Cambridge University Press.
- Quah, D. 1992. Empirical cross-section dynamics in economic growth. Macroeconomics Discussion Paper no. 75. Federal Reserve Bank of Minneapolis.
- Rao, M. 1978. Covariance analysis of nonstationary time series. In Krishnaiah, P., ed., *Developments in Statistics, Volume 1*. Academic Press, New York.
- Roussas, G. 1969. Nonparametric estimation in markov processes. *Annals of the Institute of Statistical Mathematics* 21:73–87.
- Schwert, G. 1987. Effects of model specification on tests for unit roots in macroeconomic data. *Journal of Monetary Economics* 20:73–103.
- Schwert, G. 1989. Tests for unit roots: A monte carlo investigation. *Journal of Business and Economics Statistics* 7:147–159.
- Shorack, G. and J. Wellner. 1986. *Empirical processes with applications to statistics*. New York: John Wiley and Sons.
- Stock, J. 1995. Unit roots, structural breaks, and trends. In Engle, R. and D. McFadden, eds., *Handbook of Econometrics, Volume 4*. North Holland, Amsterdam.
- Stock, J. and M. Watson. 1993. A simple estimation of cointegrating vectors in higher order integrated systems. *Econometrica* 61(4):783–820.

Taylor, J. 1985. What would Nominal GNP do to the business cycle? *Carnegie-Rochester Conference Series on Public Policy* 22:61-84.

Wen, K., P. Chen, and Z. Zhang. 1992. Evolving multi-humped distributions and switching business modes – empirical observations of disequilibrium behavior. Working Paper #92-09-03. IC2 Institute, University of Texas at Austin.

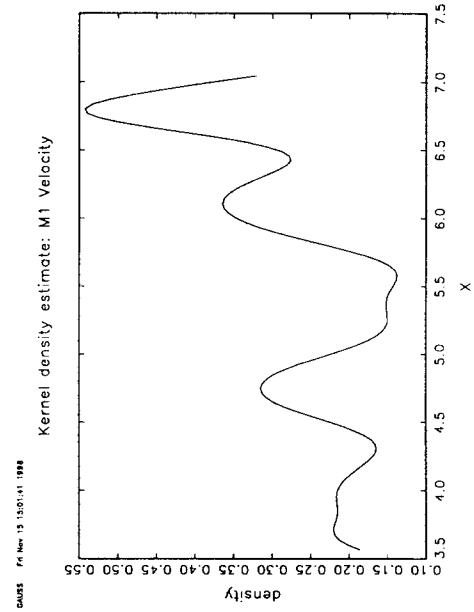
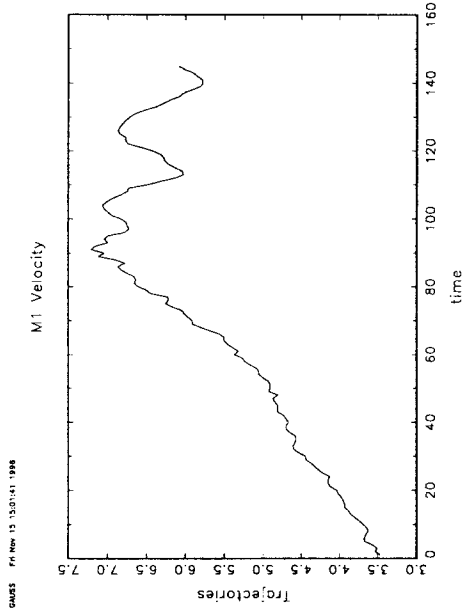


Figure 11: Trajectory of M1 velocity, 1959:1–1995:2 Figure 12: Marginal density of M1 velocity, 1959:1–1995:2

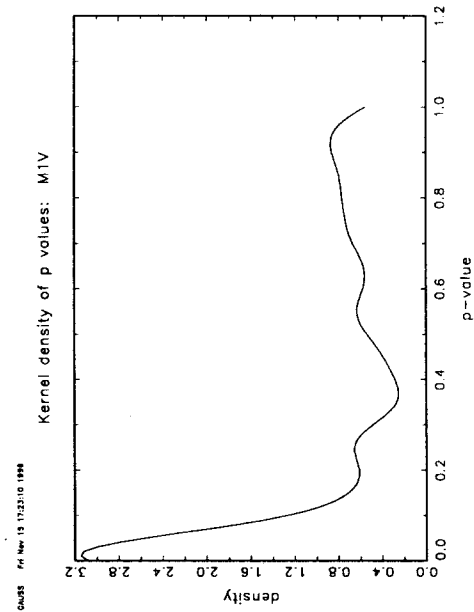
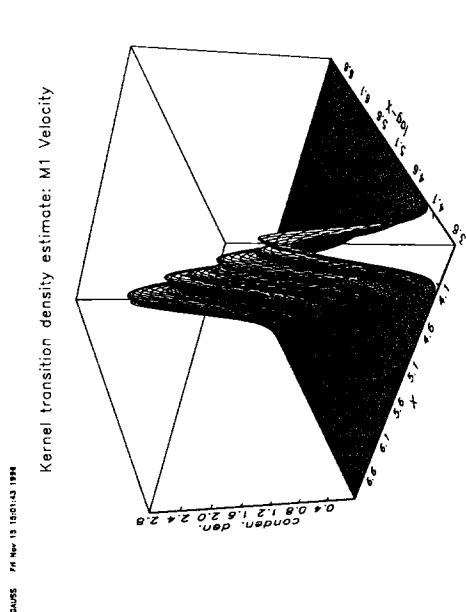


Figure 13: Transition density of M1 velocity, 1959:2–1995:2 Figure 14: Density of p-values of test for M1 velocity, 1959:1–1995:2

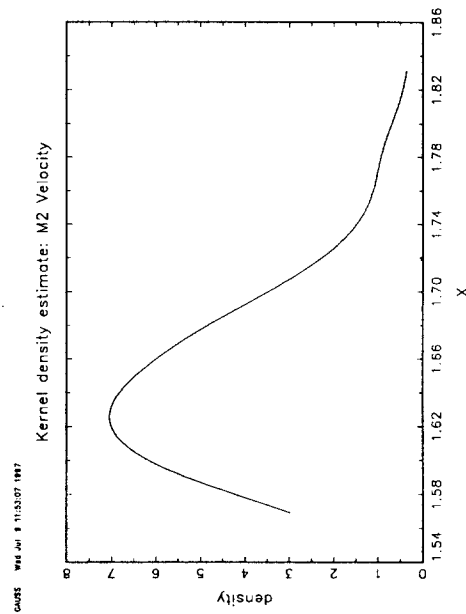
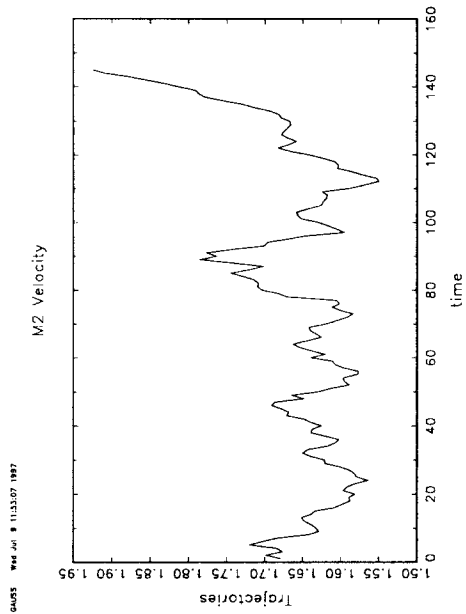


Figure 15: Trajectory of M2 velocity, 1959:1–1995:2

Figure 16: Marginal density of M2 velocity, 1959:1–1995:2

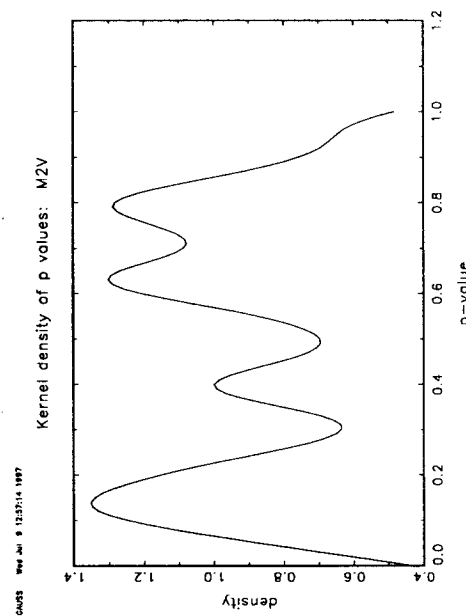
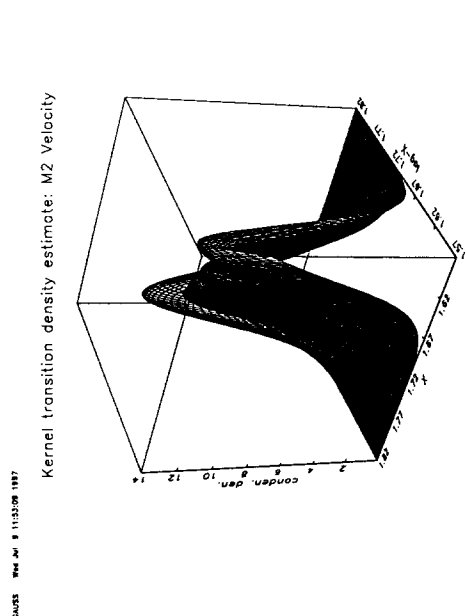


Figure 17: Transition density of M2 velocity, 1959:2–1995:2

Figure 18: Density of p-values of test for M2 velocity, 1959:1–1995:2

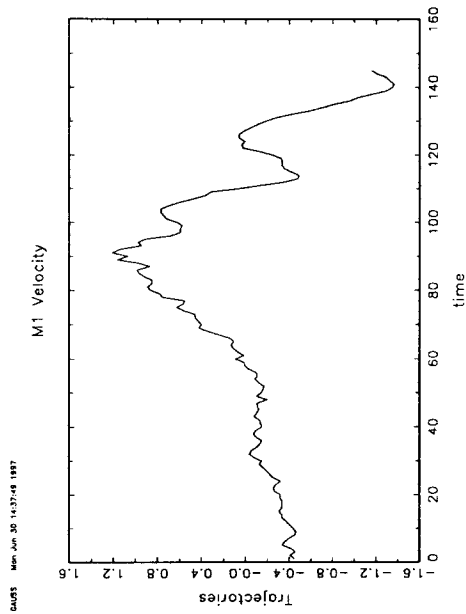


Figure 19: Trajectory of detrended of M1 Velocity, 1959:2–1995:2

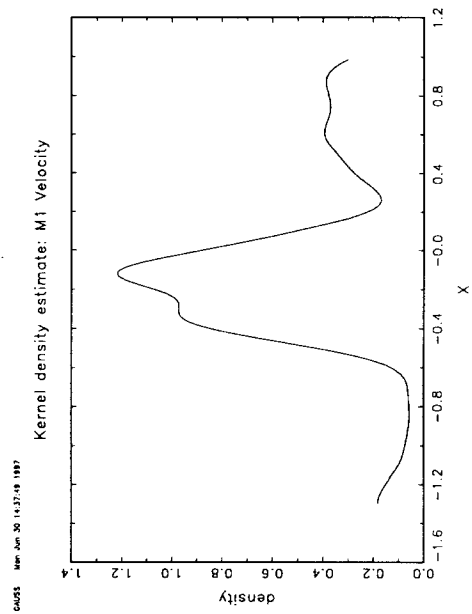


Figure 20: Marginal density of detrended M1 velocity, 1959:2–1995:2

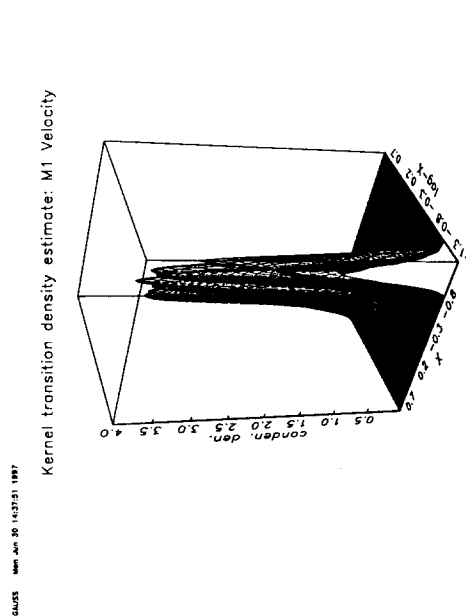


Figure 21: Transition density of detrended of M1 Velocity, 1959:2–1995:2

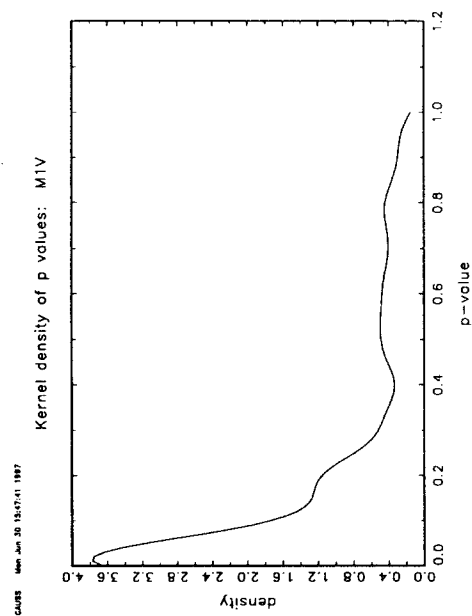


Figure 22: Density of p-values of test for detrended M1 velocity, 1959:2–1995:2

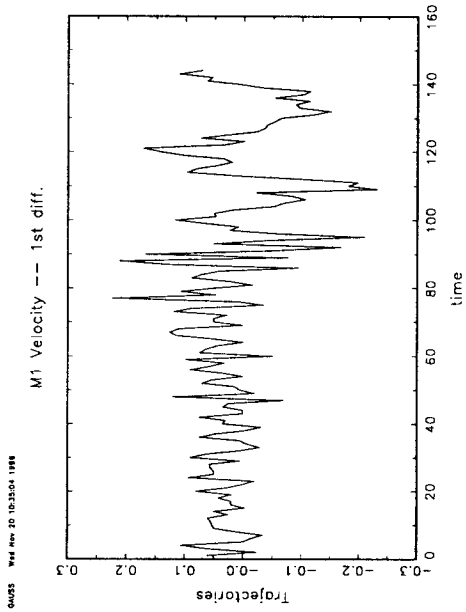


Figure 23: Trajectory of 1st differences of M1 Velocity, 1959:2–1995:2

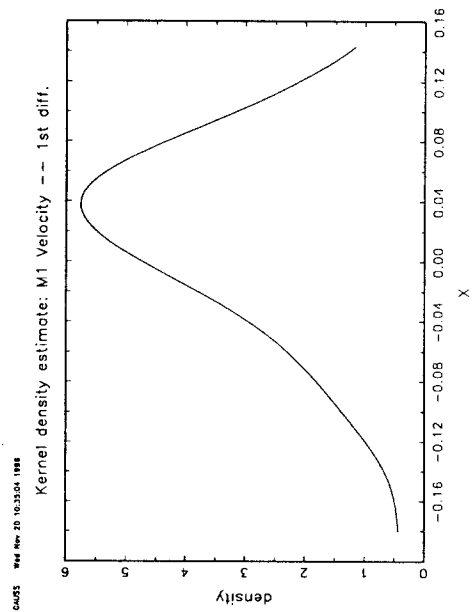


Figure 24: Marginal density of 1st differenced M1 velocity, 1959:2–1995:2

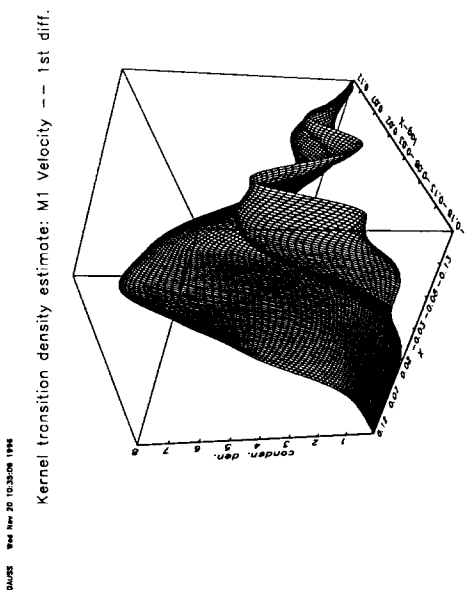


Figure 25: Transition density of 1st differences of M1 Velocity, 1959:2–1995:2

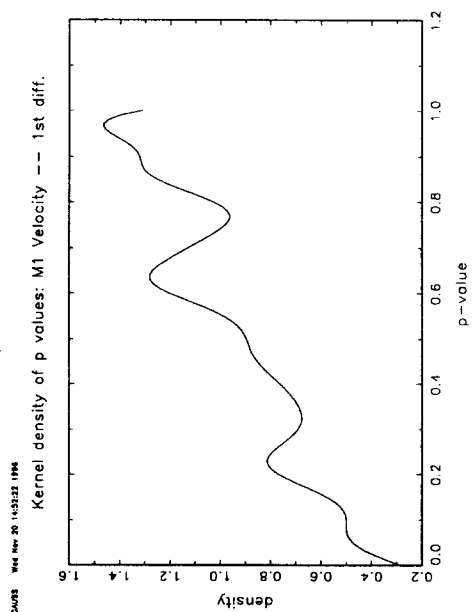


Figure 26: Density of p-values of test for 1st differenced M1 velocity, 1959:2–1995:2

British Pound: 1/3/84 – 4/30/93, 5 Minute Data

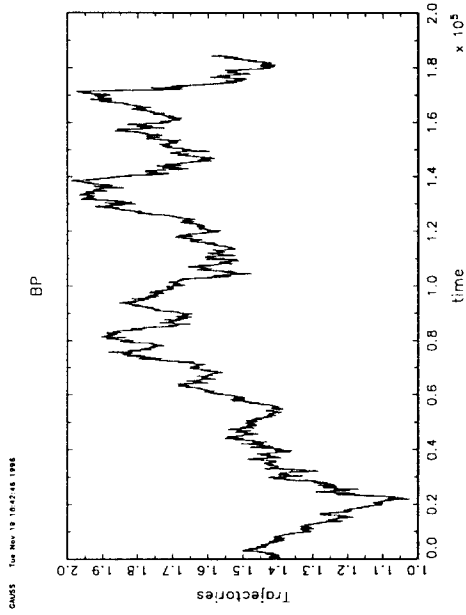


Figure 27: Trajectory of BP

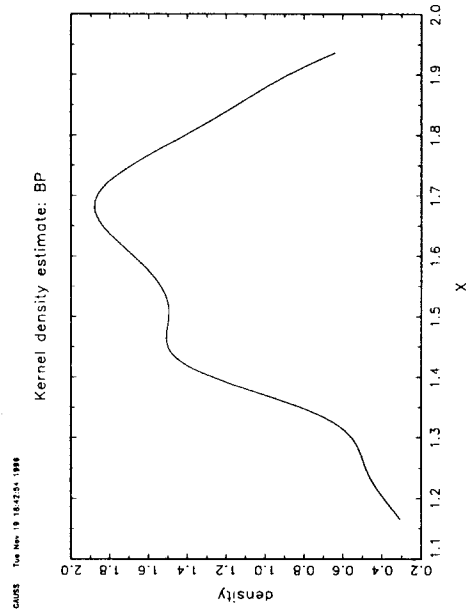


Figure 28: Marginal density of BP

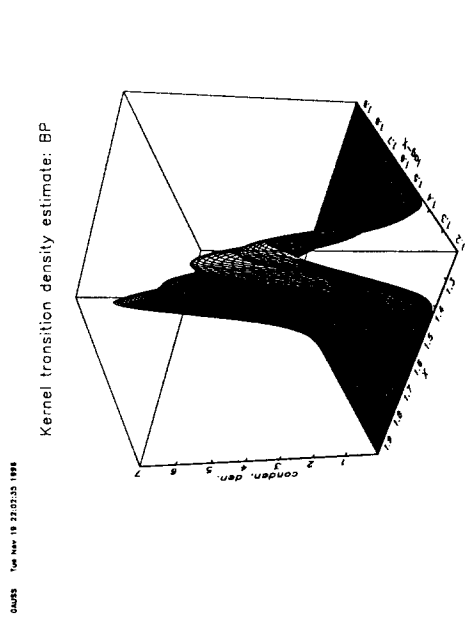


Figure 29: Transition density of BP

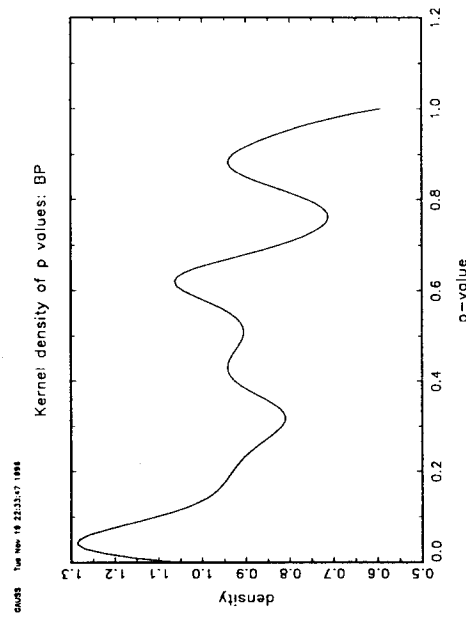


Figure 30: Density of p-values of test for BP

Deutschemark: 1/3/84 – 4/30/93, 5 Minute Data

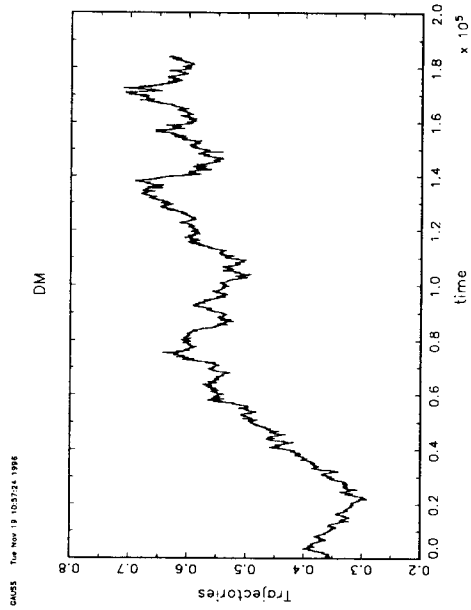


Figure 31: Trajectory of DM

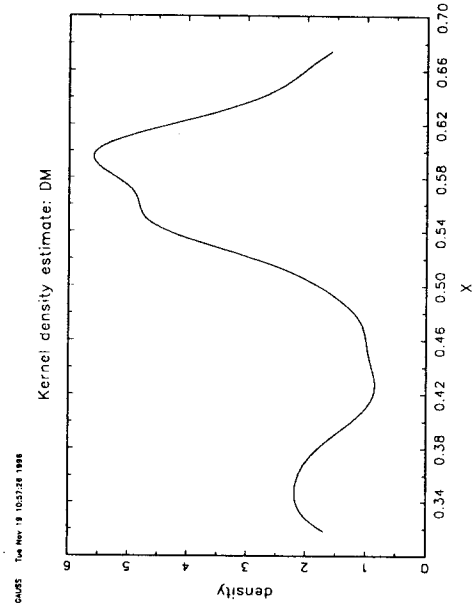


Figure 32: Marginal density of DM

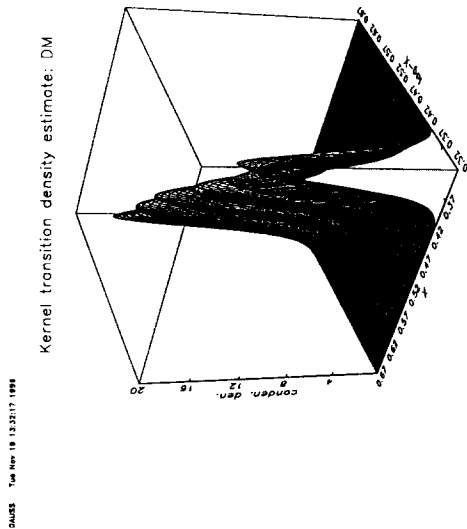


Figure 33: Transition density of DM

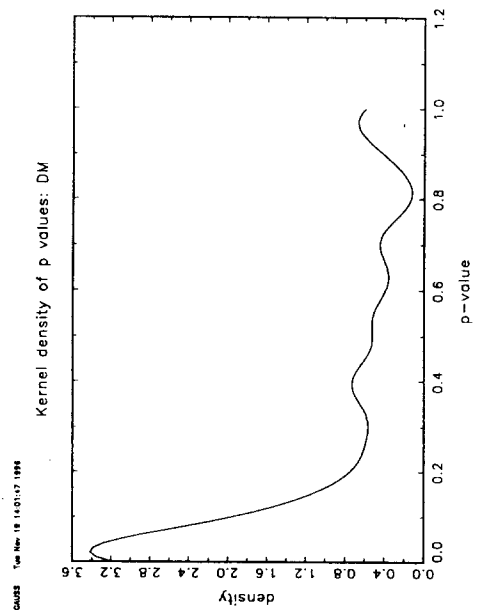


Figure 34: Density of p-values of test for DM

Japanese Yen: 1/3/84 – 4/30/93, 5 Minute Data

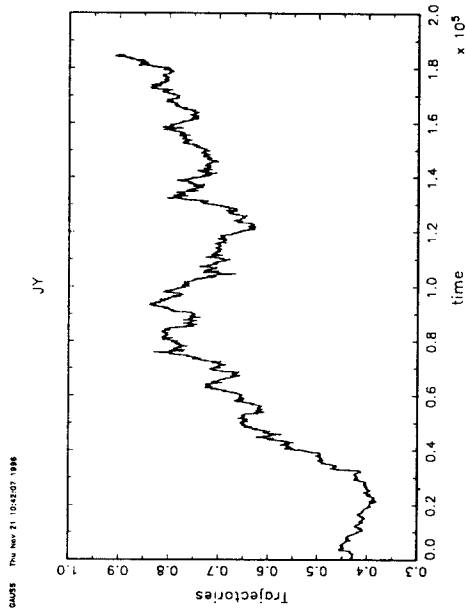


Figure 35: Trajectory of JY

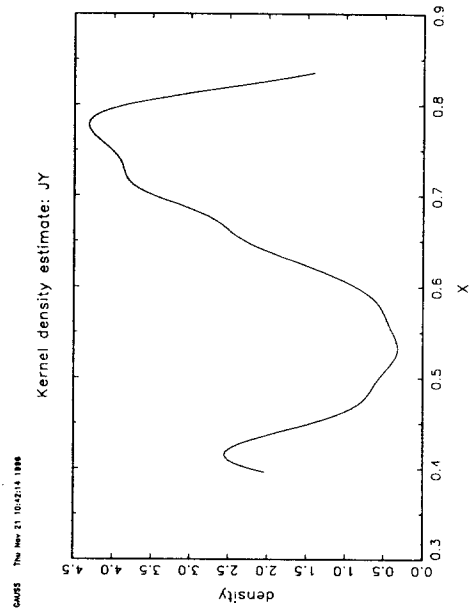


Figure 36: Marginal density of JY

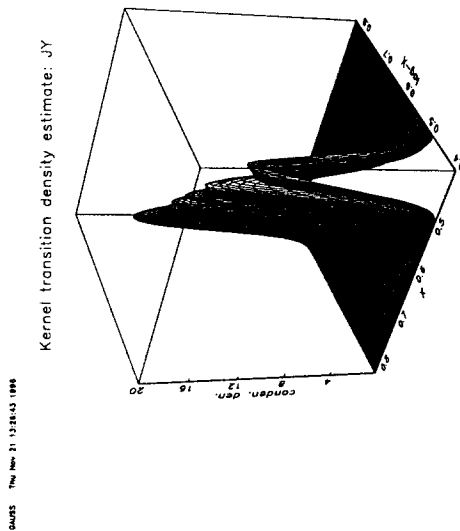


Figure 37: Transition density of JY

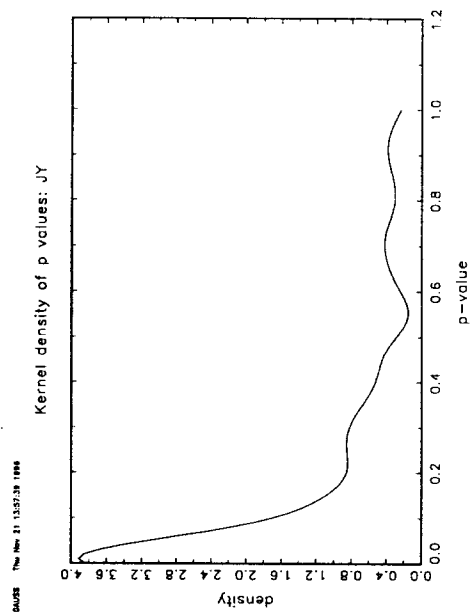


Figure 38: Density of p-values of test for JY

Swiss Franc: 1/3/84 – 4/30/93, 5 Minute Data

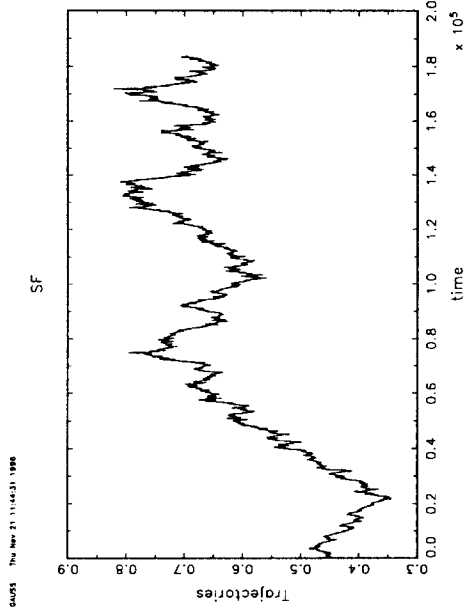


Figure 39: Trajectory of SF

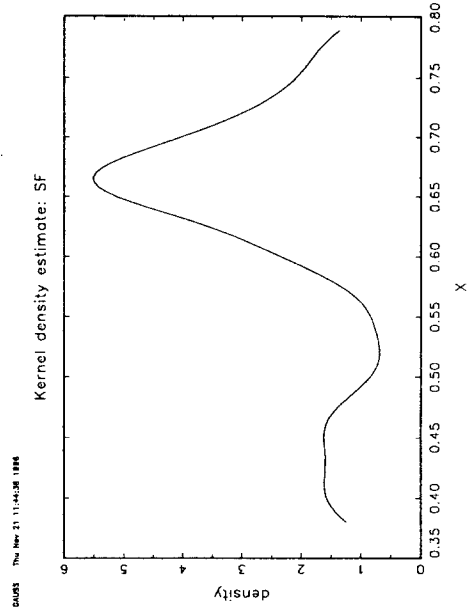


Figure 40: Marginal density of SF

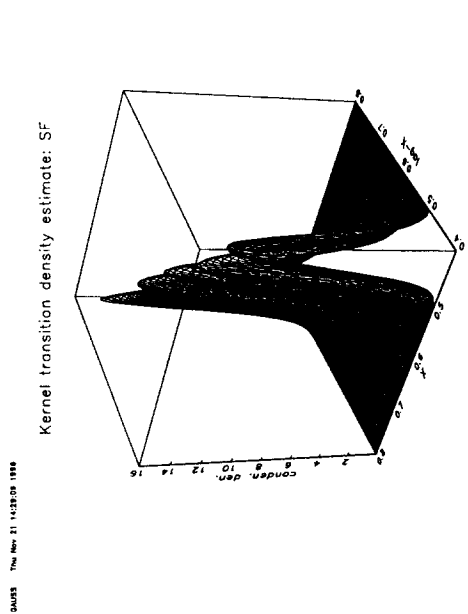


Figure 41: Transition density of SF

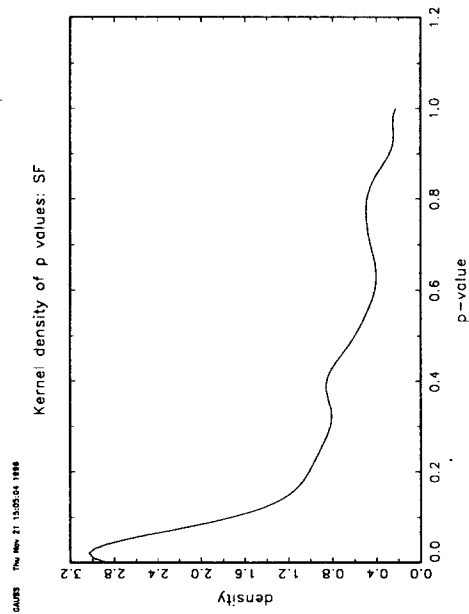


Figure 42: Density of p-values of test for SF

5-Year Treasury Notes: 6/1/88 – 4/30/93, 5 Minute Data

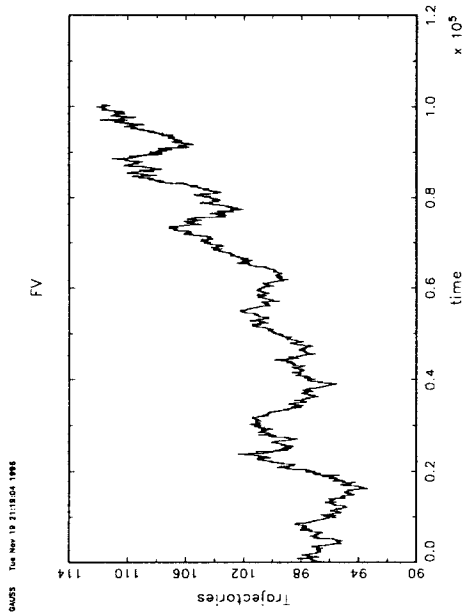


Figure 43: Trajectory of FV

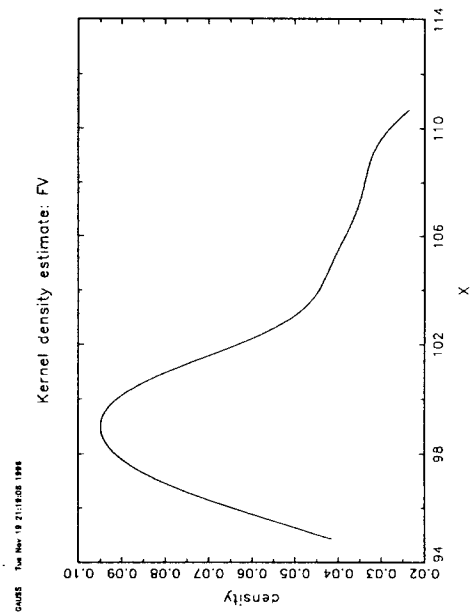


Figure 44: Marginal density of FV

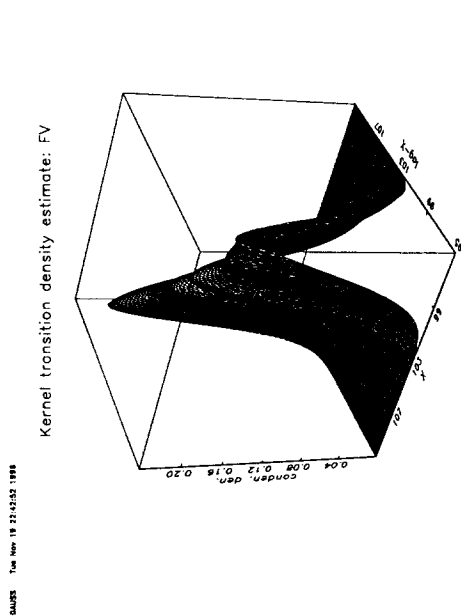


Figure 45: Transition density of FV

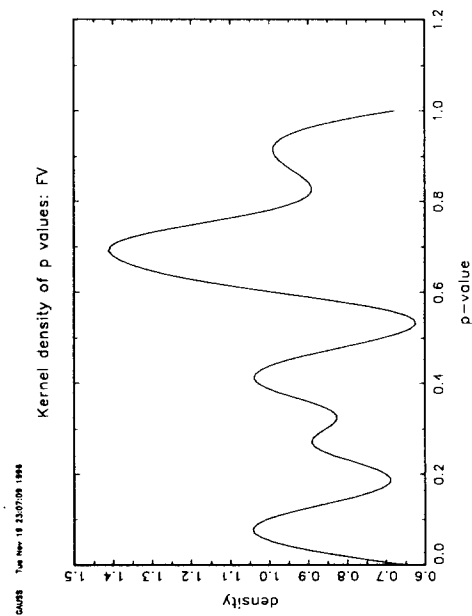


Figure 46: Density of p-values of test for FV