



Munich Personal RePEc Archive

Classical Ergodicity and Modern Portfolio Theory

Poitras, Geoffrey and Heaney, John

Simon Fraser University, Simon Fraser University

2015

Online at <https://mpra.ub.uni-muenchen.de/113952/>
MPRA Paper No. 113952, posted 08 Aug 2022 10:21 UTC

3/23/15

Classical Ergodicity and Modern Portfolio Theory

Geoffrey Poitras*
Beedie School of Business
Simon Fraser University
Vancouver B.C.
CANADA V5A 1S6
email: poitras@sfu.ca

and

John Heaney,
Beedie School of Business,
Simon Fraser University

ABSTRACT

What role have theoretical methods initially developed in mathematics and physics played in the progress of financial economics? What is the relationship between financial economics and econophysics? What is the relevance of the ‘classical ergodicity hypothesis’ to modern portfolio theory? This paper addresses these questions by reviewing the etymology and history of the classical ergodicity hypothesis in 19th century statistical mechanics. An explanation of classical ergodicity is provided that establishes a connection to the fundamental empirical problem of using non-experimental data to verify theoretical propositions in modern portfolio theory. The role of the ergodicity assumption in the *ex post / ex ante* quandary confronting modern portfolio theory is also examined.

Keywords: Ergodicity; Ludwig Boltzmann; Econophysics; financial economics, modern portfolio theory.

* The authors are Professor of Finance and Associate Professor of Finance at Simon Fraser University. Helpful comments from Chris Veld, Yulia Veld, Emmanuel Havens, Franck Jovanovic, Marcel Boumans and Christoph Schinckus are gratefully acknowledged.

Classical Ergodicity and Modern Portfolio Theory

"As the physicist builds models of the movement of matter in a frictionless environment, the economist builds models where there are no institutional frictions to the movement of stock prices"

E. Elton and M. Gruber, Modern Portfolio Theory and Investment Analysis (1984, p.273)

1. Introduction

At least since Markowitz (1952) initiated modern portfolio theory (MPT), it has often been maintained that the tradeoff between systematic risk and expected return is the most important theoretical element of financial economics, e.g., Campbell (1996). Extending Mirowski (1984), the static equilibrium methods used to develop propositions in MPT such as the capital asset pricing model can be traced to mathematical concepts developed from the deterministic 'rational mechanics' approach to 19th century physics. In the years since Markowitz (1952), financial economics has also adopted alternative mathematical methods from more recent contributions to physics, especially the diffusion processes employed by Black and Scholes (1973) to determine option prices. The emergence of econophysics during the last decade of the twentieth century, e.g., Roehner (2002); Jovanovic and Schinckus (2013), has provided a variety of theoretical and empirical methods adapted from physics – ranging from statistical mechanics to chaos theory – to analyze financial phenomena. Yet, despite considerable overlap in method, contributions to econophysics have gained limited attention in financial economics. In contrast, econophysicists generally consider financial economics to be primarily concerned with a core theory that is inconsistent with the empirical orientation of physical theory.

Physical theory has evolved considerably from the constrained optimization, static equilibrium approach of rational mechanics which underpins MPT. In detailing historical developments in physics since the 19th century, it is conventional to jump from the determinism of rational mechanics

to quantum mechanics to recent developments in chaos theory, overlooking the relevance of the initial steps toward modeling stochastic behavior of physical phenomena by Ludwig Boltzmann (1844-1906), James Maxwell (1831-1879) and Josiah Gibbs (1839-1903). As such, there is a point of demarcation between the intellectual pre-histories of MPT and econophysics that can, arguably, be traced to the debate over energetics around the end of the 19th century. While the evolution of physics after energetics involved the introduction and subsequent stochastic generalization of ergodic concepts, fueled by the emergence of MPT following Markowitz, financial economics incorporated ergodicity into empirical methods aimed at generalizing and testing the capital asset pricing model and other elements of MPT.¹ Significantly, stochastic generalization of the static equilibrium approach of MPT required the adoption of a restricted class of ergodic processes, i.e., ‘time reversible’ probabilistic models, especially the unimodal likelihood functions associated with certain stationary distributions.

In contrast, from the early ergodic models of Boltzmann to the fractals and chaos theory of Mandelbrot, physics has employed a wider variety of ergodic and non-ergodic stochastic models aimed at capturing key empirical characteristics of various physical problems at hand. Such models typically have a mathematical structure that varies from the constrained optimization techniques underpinning MPT, restricting the straightforward application of many physical models. Yet, the demarcation between the use of ergodic notions in physics and financial economics was blurred substantively by the introduction of diffusion process techniques to solve contingent claims valuation problems. Following contributions by Sprenkle (1961) and Samuelson (1965), Black and Scholes (1973) and Merton (1973) provided an empirically viable method of using diffusion methods to determine, using Ito’s lemma, a partial differential equation that can be solved for an option price.

Use of Ito's lemma to solve stochastic optimization problems is now commonplace in financial economics, e.g., Brennan and Schwartz (1979), including the continuous time generalizations of MPT, e.g., Epstein and Ji (2013). In spite of the considerable progression of certain mathematical techniques employed in physics into financial economics, overcoming the difficulties of applying the wide range of models developed for physical situations to fit the empirical properties of financial data is still a central problem confronting econophysics.

Schinckus (2010, p.3816) accurately recognizes that the positivist philosophical foundation of econophysics depends fundamentally on empirical observation: "The empiricist dimension is probably the first positivist feature of econophysics". Following McCauley (2004) and others, this concern with empiricism often focuses on the identification of macro-level statistical regularities that are characterized by the scaling laws identified by Mandelbrot (1997) and Mandelbrot and Hudson (2004) for financial data. Unfortunately, this empirically driven ideal is often confounded by the 'non-repeatable' experiment that characterizes most observed economic and financial data. There is quandary posed by having only a single observed *ex post* time path to estimate the distributional parameters for the ensemble of *ex ante* time paths needed to make decisions involving future values of financial variables. In contrast to natural sciences such as physics, in the human sciences there is no assurance that *ex post* statistical regularity translates into *ex ante* forecasting accuracy. Resolution of this quandary highlights the usefulness of employing a 'phenomenological' approach to modeling stochastic properties of financial variables relevant to MPT.

To this end, this paper provides an etymology and history of the 'classical ergodicity hypothesis' in 19th century statistical mechanics. Subsequent use of ergodicity in financial economics, in general, and MPT, in particular, is then examined. A modern interpretation of classical ergodicity is provided

that uses Sturm-Liouville theory, a mathematical method central to classical statistical mechanics, to decompose the transition probability density of a one-dimensional diffusion process subject to regular upper and lower reflecting barriers. This ‘classical’ decomposition divides the transition density of an ergodic process into a possibly multi-modal limiting stationary density which is independent of time and initial condition, and a power series of time and boundary dependent transient terms. In contrast, empirical theory aimed at estimating relationships from MPT typically ignores the implications of the initial and boundary conditions that generate transient terms and focuses on properties of a particular class of unimodal limiting stationary densities with finite parameters. To illustrate the implications of the expanded class of ergodic processes available to econophysics, properties of the bi-modal quartic exponential stationary density are considered and used to assess the ability of the classical ergodicity hypothesis to explain certain ‘stylized facts’ associated with the *ex post / ex ante* quandary confronting MPT.

2. A Brief History of Classical Ergodicity

The Encyclopedia of Mathematics (2002) defines ergodic theory as the “metric theory of dynamical systems. The branch of the theory of dynamical systems that studies systems with an invariant measure and related problems.” This modern definition implicitly identifies the birth of ergodic theory with proofs of the mean ergodic theorem by von Neumann (1932) and the pointwise ergodic theorem by Birkhoff (1931). These early proofs have had significant impact in a wide range of modern subjects. For example, the notions of invariant measure and metric transitivity used in the proofs are fundamental to the measure theoretic foundation of modern probability theory (Doob 1953; Mackey 1974). Building on a seminal contribution to probability theory by Kolmogorov (1933), in the years immediately following it was recognized that the ergodic theorems generalize

the strong law of large numbers. Similarly, the equality of ensemble and time averages – the essence of the mean ergodic theorem – is necessary to the concept of a strictly stationary stochastic process. Ergodic theory is the basis for the modern study of random dynamical systems, e.g., Arnold (1988). In mathematics, ergodic theory connects measure theory with the theory of transformation groups. This connection is important in motivating the generalization of harmonic analysis from the real line to locally compact groups.

From the perspective of modern mathematics, statistical physics or systems theory, Birkhoff (1931) and von Neumann (1932) are excellent starting points for a modern history of ergodic theory. Building on the modern ergodic theorems, subsequent developments in these and related fields have been dramatic. These contributions mark the solution to a problem in statistical mechanics and thermodynamics that was recognized sixty years earlier when Ludwig Boltzmann introduced the classical ergodic hypothesis to permit the theoretical phase space average to be interchanged with the measurable time average. For the purpose of contrasting methods from physics and econophysics with those used in MPT, the selection of the less formally correct and rigorous classical ergodic hypothesis of Boltzmann is a more auspicious beginning. Problems of interest in mathematics are generated by a range of subjects, such as physics, chemistry, engineering and biology. The formulation and solution of physical problems in, say, statistical mechanics or particle physics will have mathematical features which are inapplicable or unnecessary for MPT. For example, in statistical mechanics, points in the phase space are often multi-dimensional functions representing the mechanical state of the system, hence the desirability of a group-theoretic interpretation of the ergodic hypothesis. From the perspective of MPT, such complications are largely irrelevant. The history of classical ergodic theory captures the etymology and basic physical interpretation providing

a more revealing pre-history of the relevant MPT mathematics. This arguably more revealing pre-history begins with the formulation of theoretical problems that von Neumann and Birkhoff were later able to solve.

Mirowski (1984; 1989a, esp. ch.5) establishes the importance of 19th century physics in the development of the neoclassical economic system advanced by W. Stanley Jevons (1835-1882) and Leon Walras (1834-1910) during the marginalist revolution of the 1870's. Being derived using principles from neoclassical economic theory, MPT also inherited essential features of mid-19th century physics: deterministic rational mechanics; conservation of energy; and the non-atomistic continuum view of matter that inspired the energetics movement later in the 19th century.² More precisely, from neoclassical economics MPT inherited a variety of static equilibrium techniques and tools such as mean-variance utility functions and constrained optimization. As such, failings of neoclassical economics identified by econophysicists also apply to central propositions of MPT. Included in the failings is an over-emphasis on theoretical results at the expense of identifying models that have greater empirical validity, e.g., Roehner (2002); Schinckus (2010).

It was during the transition from rational to statistical mechanics during the last third of the 19th century that Boltzmann made contributions leading to the transformation of theoretical physics from the microscopic mechanistic models of Rudolf Clausius (1835-1882) and James Maxwell to the macroscopic probabilistic theories of Josiah Gibbs and Albert Einstein (1879-1955).³ Coming largely after the start of the marginalist revolution in economics, this fundamental transformation in theoretical physics had little impact on the progression of financial economics until the appearance of diffusion equations in contributions on continuous time finance that started in the 1960's and culminated in Black and Scholes (1973). The deterministic mechanics of the energistic approach was

well suited to the axiomatic formalization of neoclassical economic theory which culminated in: the von Neumann and Morgenstern expected utility approach to modeling uncertainty; the Bourbaki inspired Arrow-Debreu general equilibrium theory, e.g., Weintraub (2002), and, ultimately, MPT. In turn, empirical estimation and the subsequent extension of static equilibrium MPT results to continuous time were facilitated by the adoption of a narrow class of ergodic processes.

Having descended from the deterministic rational mechanics of mid-19th century physics, defining works of MPT do not capture the probabilistic approach to modeling systems initially introduced by Boltzmann and further clarified by Gibbs.⁴ Mathematical problems raised by Boltzmann were subsequently solved using tools introduced in a string of later contributions by the likes of the Ehrenfests and Cantor in set theory, Gibbs and Einstein in physics, Lebesgue in measure theory, Kolmogorov in probability theory, Weiner and Levy in stochastic processes. Boltzmann was primarily concerned with problems in the kinetic theory of gases, formulating dynamic properties of the stationary Maxwell distribution – the velocity distribution of gas molecules in thermal equilibrium. Starting in 1871, Boltzmann took this analysis one step further to determine the evolution equation for the distribution function. The mathematical implications of this classical analysis still resonate in many subjects of the modern era. The etymology for “ergodic” begins with an 1884 paper by Boltzmann, though the initial insight to use probabilities to describe a gas system can be found as early as 1857 in a paper by Clausius and in the famous 1860 and 1867 papers by Maxwell.⁵

The Maxwell distribution is defined over the velocity of gas molecules and provides the probability for the relative number of molecules with velocities in a certain range. Using a mechanical model that involved molecular collision, Maxwell (1867) was able to demonstrate that, in thermal

equilibrium, this distribution of molecular velocities was a ‘stationary’ distribution that would not change shape due to ongoing molecular collision. Boltzmann aimed to determine whether the Maxwell distribution would emerge in the limit, whatever the initial state of the gas. In order to study the dynamics of the equilibrium distribution over time, Boltzmann introduced the probability distribution of the relative time a gas molecule has a velocity in a certain range while still retaining the notion of probability for velocities of a relative number of gas molecules. Under the classical ergodic hypothesis, the average behavior of the macroscopic gas system, which can objectively be measured over time, can be interchanged with the average value calculated from the ensemble of unobservable and highly complex microscopic molecular motions at a given point in time. In the words of Weiner (1939, p.1): “Both in the older Maxwell theory and in the later theory of Gibbs, it is necessary to make some sort of logical transition between the average behavior of all dynamical systems of a given family or ensemble, and the historical average of a single system.”

3. Use of the Ergodic Hypothesis in Financial Economics

At least since Samuelson (1976), it has been recognized that empirical theory and estimation in economics, in general, and financial economics, in particular, relies heavily on the use of specific unimodal stationary distributions associated with a particular class of ergodic processes. As reflected in the evolution of the concept in economics, the specification and implications of ergodicity have only developed gradually. The early presentation of ergodicity by Samuelson (1976) involves the addition of a discrete Markov error term into the deterministic cobweb model to demonstrate that estimated forecasts of future values, such as prices, “should be less variable than the actual data”. Considerable opaqueness about the definition of ergodicity is reflected in the statement that a “‘stable’ stochastic process ... eventually forgets its past and therefore in the far future can be expected to

approach an ergodic probability distribution” (Samuelson 1976, p.2). The connection between ergodic processes and non-linear dynamics that characterizes present efforts in economics goes unrecognized, e.g., (Samuelson 1976, p.1, 5). While some explicit applications of ergodic processes to theoretical modeling in economics have emerged since Samuelson (1976), e.g., Horst and Wenzelburger (1984); Dixit and Pindyck (1994), financial econometrics has produced the bulk of the contributions.

Initial empirical estimation for the deterministic models of neoclassical economics proceeded with the addition of a stationary, usually Gaussian, error term to produce a discrete time general linear model (GLM) leading to estimation using ordinary least squares or maximum likelihood techniques. In the history of MPT, such early estimations were associated with tests of the capital asset pricing model such as the “market model”, e.g., Elton and Gruber (1984). Iterations and extensions of the GLM to deal with complications arising in empirical estimates dominated early work in econometrics, e.g., Dhrymes (1974) and Theil (1971), leading to application of generalized least squares estimation techniques that encompassed autocorrelated and heteroskedastic error terms. Employing L_2 vector space methods with stationary Gaussian-based error term distributions ensured these early stochastic models implicitly assumed ergodicity. The generalization of this discrete time estimation approach to the class of ARCH and GARCH error term models by Engle and Granger was of such significance that a Nobel prize in economics was awarded for this contribution, e.g., Engle and Granger (1987). By modeling the evolution of the volatility, this approach permitted a limited degree of non-linearity to be modeled providing a substantively better fit of MPT models to observed financial time series, e.g., Beaulieu et al. (2013).

The emergence of ARCH, GARCH and related empirical models was part of a general trend toward

the use of inductive methods in economics, often employing discrete, linear time series methods to model transformed economic variables, e.g., Hendry (1995). At least since Dickey and Fuller (1979), it has been recognized that estimates of univariate time series models for many financial times series reveals evidence of ‘non-stationarity’. A number of approaches have emerged to deal with this apparent empirical quandary.⁶ In particular, transformation techniques for time series models have received considerable attention. Extension of the Box-Jenkins methodology led to the concept of economic time series being $I(0)$ – stationary in the level – and $I(1)$ – non-stationary in the level but stationary after first differencing. Two $I(1)$ economic variables could be cointegrated if differencing the two series produced an $I(0)$ process, e.g., Hendry (1995). Extending early work on distributed lags, long memory processes have also been employed where the time series is only subject to fractional differencing. Significantly, recent contributions on Markov switching processes and exponential smooth transition autoregressive processes have demonstrated the “possibility that nonlinear ergodic processes can be misinterpreted as unit root nonstationary processes” (Kapetanios and Shin 2011, p.620). Bonomo et al. (2011) illustrates the recent application of Markov switching processes in estimating the asset pricing models of MPT.

The conventional view of ergodicity in economics, in general, and financial economics, in particular, is reflected by Hendry (1995, p.100): “Whether economic reality is an ergodic process after suitable transformation is a deep issue” which is difficult to analyze rigorously. As a consequence, in the limited number of instances where ergodicity is examined in economics a variety of different interpretations appear. In contrast, the ergodic hypothesis in classical statistical mechanics is associated with the more physically transparent kinetic gas model than the often technical and targeted concepts of ergodicity encountered in modern economics, in general, and MPT, in particular. For

Boltzmann, the classical ergodic hypothesis permitted the unobserved complex microscopic interactions of individual gas molecules to obey the second law of thermodynamics, a concept that has limited application in economics.⁷ Despite differences in physical interpretation, the problem of modeling ‘macroscopic’ financial variables, such as common stock prices, foreign exchange rates, ‘asset’ prices or interest rates, when it is not possible to derive a theory for describing and predicting empirical observations from known first principles about the (microscopic) rational behavior of individuals and firms. By construction, this involves a phenomenological approach to modeling.⁸

Even though the formal solutions proposed were inadequate by standards of modern mathematics, the thermodynamic model introduced by Boltzmann to explain the dynamic properties of the Maxwell distribution is a pedagogically useful starting point to develop the implications of ergodicity for MPT. To be sure, von Neumann (1932) and Birkhoff (1931) correctly specify ergodicity using Lebesgue integration – an essential analytical tool unavailable to Boltzmann – but the analysis is too complex to be of much value to all but the most mathematically specialized economists. The physical intuition of the kinetic gas model is lost in the generality of the results. Using Boltzmann as a starting point, the large number of mechanical and complex molecular collisions could correspond to the large number of microscopic, atomistic liquidity providers and traders interacting to determine the macroscopic financial market price. In this context, it is variables such as the asset price or the interest rate or the exchange rate, or some combination, that is being measured over time and ergodicity would be associated with the properties of the transition density generating the macroscopic variables. Ergodicity can fail for a number of reasons and there is value in determining the source of the failure. In this vein, there are two fundamental difficulties associated with the classical ergodicity hypothesis in Boltzmann’s statistical mechanics – reversibility and recurrence –

that are largely unrecognized in financial economics.⁹

Halmos (1949, p.1017) is a helpful starting point to sort out the differing notions of ergodicity that are of relevance to the issues at hand: “The ergodic theorem is a statement about a space, a function and a transformation”. In mathematical terms, ergodicity or ‘metric transitivity’ is a property of ‘indecomposable’, measure preserving transformations. Because the transformation acts on points in the space, there is a fundamental connection to the method of measuring relationships such as distance or volume in the space. In von Neumann (1932) and Birkhoff (1931), this is accomplished using the notion of Lebesgue measure: the admissible functions are either integrable (Birkhoff) or square integrable (von Neumann). In contrast to, say, statistical mechanics where spaces and functions account for the complex physical interaction of large numbers of particles, economic theories such as MPT usually specify the space in a mathematically convenient fashion. For example, in the case where there is a single random variable, then the space is “superfluous” (Mackey 1974, p.182) as the random variable is completely described by the distribution. Multiple random variables can be handled by assuming the random variables are discrete with finite state spaces. In effect, conditions for an ‘invariant measure’ are assumed in MPT in order to focus attention on “finding and studying the invariant measures” (Arnold 1998, p.22) where, in the terminology of financial econometrics, the invariant measure usually corresponds to the stationary distribution or likelihood function.

The mean ergodic theorem of von Neumann (1932) provides an essential connection to the ergodicity hypothesis in financial econometrics. It is well known that, in the Hilbert and Banach spaces common to econometric work, the mean ergodic theorem corresponds to the strong law of large numbers. In statistical applications where strictly stationary distributions are assumed, the relevant ergodic transformation, L^* , is the unit shift operator: $L^* \Psi[x(t)] = \Psi[L^* x(t)] = \Psi[x(t+1)]$;

$[(L^*)^k] \Psi[x(t)] = \Psi[x(t+k)]$; and $\{(L^*)^{-k}\} \Psi[x(t)] = \Psi[x(t-k)]$ with k being an integer and $\Psi[x]$ the strictly stationary distribution for x that in the strictly stationary case is replicated at each t .¹⁰ Significantly, this reversible transformation is independent of initial time and state. Because this transformation can be achieved by imposing strict stationarity on $\Psi[x]$, L^* will only work for certain ergodic processes. In effect, the ergodic requirement that the transformation be measure preserving is weaker than the strict stationarity of the stochastic process sufficient to achieve L^* . The practical implications of the reversible ergodic transformation L^* are described by Davidson (1991, p.331): “In an economic world governed entirely by [time reversible] ergodic processes ... economic relationships among variables are timeless, or ahistoric in the sense that the future is merely a statistical reflection of the past”.¹¹

Employing conventional econometrics in empirical studies, MPT requires that the real world distribution for $x(t)$, e.g., the asset return, be sufficiently similar to those for both $x(t+k)$ or $x(t-k)$, i.e., the ergodic transformation L^* is reversible. The reversibility assumption is systemic in MPT appearing in the use of long estimation periods to determine important variables such as the “equity risk premium”. There is a persistent belief that increasing the length or sampling frequency of a financial time series will improve the precision of a statistical estimate, e.g., Dimson et al. (2002). Similarly, focus on the tradeoff between ‘risk and return’ requires the use of **unimodal** stationary densities for transformed financial variables such as the rate of return. The impact of initial and boundary conditions on financial decision making is generally ignored. The inconsistency of reversible processes with key empirical facts, such as the asymmetric tendency for downdrafts in prices to be more severe than upswings, is ignored in favor of adhering to ‘reversible’ theoretical models that can be derived from first principles associated with constrained optimization techniques,

e.g., Constantides (2002).

4. A Phenomenological Interpretation of Classical Ergodicity

In physics, phenomenology lies at the intersection of theory and experiment. Theoretical relationships between empirical observations are modeled without deriving the theory directly from first principles, e.g., Newton's laws of motion. Predictions based on these theoretical relationships are obtained and compared to further experimental data designed to test the predictions. In this fashion, new theories that can be derived from first principles are motivated. Confronted with non-experimental data for important financial variables, such as common stock prices, interest rates and the like, financial economics has developed some theoretical models that aim to fit the 'stylized facts' of those variables. In contrast, the MPT is initially derived directly from the 'first principles' of constrained expected utility maximizing behavior by individuals and firms. Given the difficulties in economics of testing model predictions with 'new' experimental data, physics and econophysics have the potential to provide a rich variety of mathematical techniques that can be adapted to determining mathematical relationships among financial variables that explain the 'stylized facts' of observed non-experimental data.¹²

The evolution of financial economics from the deterministic models of neoclassical economics to more modern stochastic models has been incremental and disjointed. The preference for linear models of static equilibrium relationships has restricted the application of theoretical frameworks that capture more complex non-linear dynamics, e.g., chaos theory; truncated Levy processes. Yet, important financial variables have relatively innocuous sample paths compared to some types of variables encountered in physics. There is an impressive range of mathematical and statistical models that, seemingly, could be applied to almost any physical or financial situation. If the process can be

verbalized, then a model can be specified. This begs the questions: are there transformations – ergodic or otherwise – that capture the basic ‘stylized facts’ of observed financial data? Is the random instability in the observed sample paths identified in, say, stock price time series consistent with the *ex ante* stochastic bifurcation of an ergodic process, e.g., Chiarella et al. (2008)? In the bifurcation case, the associated *ex ante* stationary densities are bimodal and irreversible, a situation where the mean calculated from past values of a single, non-experimental *ex post* realization of the process is not necessarily informative about the mean for future values.

Boltzmann was concerned with demonstrating that the Maxwell distribution emerged in the limit as $t \rightarrow \infty$ for systems with large numbers of particles. The limiting process for t requires that the system run long enough that the initial conditions do not impact the stationary distribution. At the time, two fundamental criticisms were aimed at this general approach: reversibility and recurrence. In the context of financial time series, reversibility relates to the use of past values of the process to forecast future values. Recurrence relates to the properties of the long run average which involves the ability and length of time for an ergodic process to return to its stationary state. For Boltzmann, both these criticisms have roots in the difficulty of reconciling the second law of thermodynamics with the ergodicity hypothesis. Using Sturm-Liouville methods, it can be shown that classical ergodicity requires the transition density of the process to be decomposable into the sum of a stationary density and a mean zero transient term that captures the impact of the initial condition of the system on the individual sample paths; irreversibility relates to properties of the stationary density and non-recurrence to the behavior of the transient term.

Because the particle movements in a kinetic gas model are contained within an enclosed system, e.g., a vertical glass tube, classical Sturm-Liouville (S-L) methods can be applied to obtain solutions

for the transition densities. These classical results for the distributional implications of imposing regular reflecting boundaries on diffusion processes are representative of the modern phenomenological approach to random systems theory which: “studies qualitative changes of the densities of invariant measures of the Markov semigroup generated by random dynamical systems induced by stochastic differential equations” (Crauel et al. 1999, p.27).¹³ Because the initial condition of the system is explicitly recognized, ergodicity in these models takes a different form than that associated with the unit shift transformation of unimodal stationary densities typically adopted in financial economics, in general, and MPT, in particular. The ergodic transition densities are derived as solutions to the forward differential equation associated with one-dimensional diffusions. The transition densities contain a transient term that is dependent on the initial condition of the system and boundaries imposed on the state space. Path dependence, i.e., irreversibility, can be introduced by employing multi-modal stationary densities.

The distributional implications of boundary restrictions, derived by modeling the random variable as a diffusion process subject to reflecting barriers, have been studied for many years, e.g., Feller (1954). The diffusion process framework is useful because it imposes a functional structure that is sufficient for known partial differential equation (PDE) solution procedures to be used to derive the relevant transition probability densities. Wong (1964) demonstrated that with appropriate specification of parameters in the PDE, the transition densities for popular stationary distributions such as the exponential, uniform, and normal distributions can be derived using classical S-L methods. The S-L framework provides sufficient generality to resolve certain empirical difficulties arising from key stylized facts observed in the non-experimental time series from financial economics. More generally, the framework suggests a method of allowing MPT to encompass the

nonlinear dynamics of diffusion processes. In other words, within the more formal mathematical framework of classical statistical mechanics, it is possible to reformulate the classical ergodicity hypothesis to permit a useful stochastic generalization of theories in financial economics such as MPT.

The use of the diffusion model to represent the nonlinear dynamics of stochastic processes is found in a wide range of subjects. Physical restrictions such as the rate of observed genetic mutation in biology or character of heat diffusion in engineering or physics often determine the specific formalization of the diffusion model. Because physical interactions can be complex, mathematical results for diffusion models are pitched at a level of generality sufficient to cover such cases.¹⁴ Such generality is usually not required in financial economics. In this vein, it is possible to exploit mathematical properties of bounded state spaces and one dimensional diffusions to overcome certain analytical problems that can confront continuous time Markov solutions. The key construct in the S-L method is the ergodic transition probability density function U which is associated with the random (financial) variable x at time t ($U = U[x, t | x_0]$) that follows a regular, time homogeneous diffusion process. While it is possible to allow the state space to be an infinite open interval $I_o = (a, b: -\infty \leq a < b \leq \infty)$, a finite closed interval $I_c = [a, b: -\infty < a < b < +\infty]$ or the specific interval $I_s = [0 = a < b < \infty)$ are applicable to financial variables.¹⁵ Assuming that U is twice continuously differentiable in x and once in t and vanishes outside the relevant interval, then U obeys the forward equation (e.g., Gihhman and Skorohod 1979, p.102-4):

$$\frac{\partial^2}{\partial x^2} \{ B[x] U \} - \frac{\partial}{\partial x} \{ A[x] U \} = \frac{\partial U}{\partial t} \quad (1)$$

where: $B[x]$ ($= \frac{1}{2} \sigma^2[x] > 0$) is the one half the infinitesimal variance and $A[x]$ the infinitesimal drift

of the process. $B[x]$ is assumed to be twice and $A[x]$ once continuously differentiable in x . Being time homogeneous, this formulation permits state, but not time, variation in the drift and variance parameters.

If the diffusion process is subject to upper and lower reflecting boundaries that are regular and fixed ($-\infty < a < b < \infty$), the classical ‘‘Sturm-Liouville problem’’ involves solving (1) subject to the separated boundary conditions:¹⁶

$$\frac{\partial}{\partial x} \{ B[x] U[x,t] \} \Big|_{x=a} - A[a] U[a,t] = 0 \quad (3)$$

$$\frac{\partial}{\partial x} \{ B[x] U[x,t] \} \Big|_{x=b} - A[b] U[b,t] = 0 \quad (4)$$

And the initial condition:

$$U[x,0] = f[x_0] \quad \text{where:} \quad \int_a^b f[x_0] = 1 \quad (5)$$

and $f[x_0]$ is the continuous density function associated with x_0 where $a \leq x_0 \leq b$. When the initial starting value, x_0 , is known with certainty, the initial condition becomes the Dirac delta function, $U[x,0] = \delta[x - x_0]$, and the resulting solution for U is referred to as the ‘principal solution’. Within the framework of the S-L method, a stochastic process has **the property of classical ergodicity** when the transition density satisfies:¹⁷

$$\lim_{t \rightarrow \infty} U[x,t | x_0] = \int_a^b f[x_0] U[x,t | x_0] dx_0 = \Psi[x]$$

Important special cases occur for the principal solution ($f[x_0] = \delta[x - x_0]$) and when $f[x_0]$ is from a specific class such as the Pearson distributions. To be ergodic, the time invariant stationary density $\Psi[x]$ is not permitted to ‘decompose’ the sample space with a finite number of indecomposable subdensities, each of which is time invariant. Such irreversible processes are not ergodic, even though

each of the sub-densities could be restricted to obey the ergodic theorem. To achieve ergodicity, a multi-modal stationary density can be used instead of decomposing the sample space using sub-densities with different means. In turn, multi-modal irreversible ergodic processes have the property that the mean calculated from past values of the process are not necessarily informative enough about the modes of the *ex ante* densities to provide accurate predictions.

In order to more accurately capture the *ex ante* properties of financial time series, there are some potentially restrictive features in the classical S-L framework that can be identified. For example, time homogeneity of the process eliminates the need to explicitly consider the location of t_0 .¹⁸ Time homogeneity is a property of U and, as such, is consistent with ‘ahistorical’ MPT. In the sub-class of $\{U\}$ denoted as $\{U^*\}$, a time homogeneous and reversible stationary distribution governs the dynamics of $x(t)$. Significantly, while U is time homogeneous, there are some U consistent with irreversible processes. A relevant issue for econophysicists is to determine which concept – time homogeneity or reversibility – is inconsistent with economic processes that capture: collapsing pricing conventions in asset markets; liquidity traps in money markets; and, collapse induced structural shifts in stock markets. In the S-L framework, the initial state of the system (x_0) is known and the ergodic transition density provides information about how a given point x_0 shifts t units along a trajectory.¹⁹ In contrast, applications in financial econometrics employ the strictly stationary U^* , where the location of x_0 is irrelevant while U incorporates x_0 as an initial condition associated with the solution of a partial differential equation.

5. Density Decomposition Results²⁰

In general, solving the forward equation (1) for U subject to (3), (4) and some admissible form of (5) is difficult, e.g., Feller (1954), Risken (1989). In such circumstances, it is expedient to restrict the

problem specification to permit closed form solutions for the transition density to be obtained. Wong (1964) provides an illustration of this approach. The PDE (1) is reduced to an ODE by only considering the strictly stationary distributions arising from the Pearson system. Restrictions on the associated $\Psi[x]$ are constructed by imposing the fundamental ODE condition for the unimodal Pearson system of distributions:

$$\frac{d\Psi[x]}{dx} = \frac{e_1x + e_0}{d_2x^2 + d_1x + d_0} \Psi[x]$$

The transition probability density U for the ergodic process can then be reconstructed by working back from a specific closed form for the stationary distribution using known results for the solution of specific forms of the forward equation. In this procedure, the d_0, d_1, d_2, e_0 and e_1 in the Pearson ODE are used to specify the relevant parameters in (1). The U for important stationary distributions that fall within the Pearson system, such as the normal, beta, central t , and exponential, can be derived by this method.

The solution procedure employed by Wong (1964) depends crucially on restricting the PDE problem sufficiently to apply classical S-L techniques. Using S-L methods, various studies have generalized the set of solutions for U to cases where the stationary distribution is not a member of the Pearson system or U is otherwise unknown, e.g., Linetsky (2005). In order to employ the separation of variables technique used in solving S-L problems, (1) has to be transformed into the canonical form of the forward equation. To do this, the following function associated with the invariant measure is introduced:

$$r[x] = B[x] \exp \left[- \int_a^x \frac{A[s]}{B[s]} ds \right]$$

Using this function, the forward equation can be rewritten in the form:

$$\frac{1}{r[x]} \frac{\partial}{\partial x} \left\{ p[x] \frac{\partial U}{\partial x} \right\} + q[x] U = \frac{\partial U}{\partial t} \quad (6)$$

$$\text{where: } p[x] = B[x] r[x] \quad q[x] = \frac{\partial^2 B}{\partial x^2} - \frac{\partial A}{\partial x}$$

Equation (6) is the canonical form of equation (1). The S-L problem now involves solving (6) subject to appropriate initial and boundary conditions.

Because the methods for solving the S-L problem are ODE-based, some method of eliminating the time derivative in (1) is required. Exploiting the assumption of time homogeneity, the eigenfunction expansion approach applies separation of variables, permitting (6) to be specified as:

$$U[x,t] = e^{-\lambda t} \varphi[x] \quad (7)$$

Where $\varphi[x]$ is only required to satisfy the easier-to-solve ODE:

$$\frac{1}{r[x]} \frac{d}{dx} \left[p[x] \frac{d\varphi}{dx} \right] + [q[x] + \lambda] \varphi[x] = 0 \quad (1')$$

Transforming the boundary conditions involves substitution of (7) into (3) and (4) and solving to get:

$$\frac{d}{dx} \{ B[x] \varphi[x] \} \Big|_{x=a} - A[a] \varphi[a] = 0 \quad (3')$$

$$\frac{d}{dx} \{ B[x] \varphi[x] \} \Big|_{x=b} - A[b] \varphi[b] = 0 \quad (4')$$

Significant analytical advantages are obtained by making the S-L problem ‘regular’ which involves assuming that $[a,b]$ is a closed interval with $r[x]$, $p[x]$ and $q[x]$ being real valued and $p[x]$ having a continuous derivative on $[a,b]$; and, $r[x] > 0$, $p[x] > 0$ at every point in $[a,b]$. ‘Singular’ S-L problems arise where these conditions are violated due to, say, an infinite state space or a vanishing coefficient

in the interval $[a,b]$. The separated boundary conditions (3) and (4) ensure the problem is self-adjoint (Berg and McGregor 1966, p.91).

The classical S-L problem of solving (6) subject to the initial and boundary conditions admits a solution only for certain critical values of λ , the eigenvalues. Further, since equation (1) is linear in U , the general solution for (7) is given by a linear combination of solutions in the form of eigenfunction expansions. Details of these results can be found in Hille (1969, ch. 8), Birkhoff and Rota (1989, ch. 10) and Karlin and Taylor (1981). When the S-L problem is self-adjoint and regular the solutions for the transition probability density can be summarized in the following (see Appendix for Proof):

Proposition: Ergodic Transition Density Decomposition

The regular, self-adjoint Sturm-Liouville problem has an infinite sequence of real eigenvalues, $0 = \lambda_0 < \lambda_1 < \lambda_2 < \dots$ with:

$$\lim_{n \rightarrow \infty} \lambda_n = \infty$$

To each eigenvalue there corresponds a unique eigenfunction $\varphi_n \equiv \varphi_n[x]$. Normalization of the eigenfunctions produces:

$$\Psi_n[x] = \left[\int_a^b r[x] \varphi_n^2 dx \right]^{-1/2} \varphi_n$$

The $\psi_n[x]$ eigenfunctions form a complete orthonormal system in $L_2[a,b]$. The unique solution in $L_2[a,b]$ to (1), subject to the boundary conditions (3)-(4) and initial condition (5) is, in general form:

$$U[x,t] = \sum_{n=0}^{\infty} c_n \Psi_n[x] e^{-\lambda_n t} \quad (8)$$

$$\text{where: } c_n = \int_a^b r[x] f[x_0] \Psi_n[x] dx$$

Given this, the transition probability density function for x at time t can be reexpressed as the sum of a stationary limiting equilibrium distribution associated with the $\lambda_0 = 0$ eigenvalue, that is linearly independent of the boundaries, and a power series of transient terms, associated with the remaining eigenvalues, that are boundary and initial condition dependent:

$$U[x,t | x_0] = \Psi[x] + T[x,t | x_0] \quad (9)$$

$$\text{where: } \Psi[x] = \frac{r[x]^{-1}}{\int_a^b r[x]^{-1} dx} \quad (10)$$

Using the specifications of λ_n , c_n , and ψ_n , the properties of $T[x,t]$ are defined as:

$$T[x,t | x_0] = \sum_{n=1}^{\infty} c_n e^{-\lambda_n t} \psi_n[x] = \frac{1}{r[x]} \sum_{n=1}^{\infty} e^{-\lambda_n t} \psi_n[x] \psi_n[x_0] \quad (11)$$

$$\text{with: } \int_a^b T[x,t | x_0] dx = 0 \quad \text{and} \quad \lim_{t \rightarrow \infty} T[x,t | x_0] = 0$$

This Proposition provides the general solution to the regular, self-adjoint S-L problem of deriving U when the process is subject to regular reflecting barriers. Taking the limit as $t \rightarrow \infty$ in (9), it follows from (10) and (11) that the transition density of the stochastic process satisfies the classical ergodic property. Considerable effort has been given to determining the convergence behavior of different processes. The distributional impact of the initial conditions and boundary restrictions enter through $T[x,t | x_0]$. From the restrictions on $T[x,t | x_0]$ in (11), the total mass of the transient term is zero so the mean ergodic theorem still applies. The transient only acts to redistribute the mass of the stationary distribution, thereby causing a change in shape which can impact the *ex ante* calculation of the expected value. The specific degree and type of alteration depends on the relevant assumptions made about the parameters and initial functional forms. Significantly, stochastic generalization of static and deterministic MPT almost always ignores the impact of transients by only employing the limiting

stationary distribution component.

The theoretical advantage obtained by imposing regular reflecting barriers on the diffusion state space for the forward equation is that an ergodic decomposition of the transition density is assured. The relevance of bounding the state space and imposing regular reflecting boundaries can be illustrated by considering the well known solution (e.g., Cox and Miller 1965, p.209) for U involving a constant coefficient standard normal variate $Y(t) = (x - x_0 - \mu t) / \sigma$ over the unbounded state space $I_0 = (-\infty \leq x \leq \infty)$. In this case the forward equation (1) reduces to: $\frac{1}{2}\{\partial^2 U / \partial Y^2\} = \partial U / \partial t$. By evaluating these derivatives, it can be verified that the principal solution for U is:

$$U[x,t | x_0] = \frac{1}{\sigma \sqrt{(2\pi t)}} \exp\left[-\frac{(x - x_0 - \mu t)^2}{2\sigma^2 t}\right]$$

and as $t \rightarrow -\infty$ or $t \rightarrow +\infty$ then $U \rightarrow 0$ and the stochastic process is non-ergodic because it does not possess a non-trivial stationary distribution. The mean ergodic theorem fails: if the process runs long enough, then U will evolve to where there is no discernible probability associated with starting from x_0 and reaching the neighborhood of a given point x . The absence of a stationary distribution raises a number of questions, e.g., whether the process has unit roots. Imposing regular reflecting boundaries is a certain method of obtaining a stationary distribution and a discrete spectrum (Hansen and Schienkman 1998, p.13). Alternative methods, such as specifying the process to admit natural boundaries where the parameters of the diffusion are zero within the state space, can give rise to continuous spectrum and raise significant analytical complexities. At least since Feller (1954), the search for useful solutions, including those for singular diffusion problems, has produced a number of specific cases of interest. However, without the analytical certainty of the classical S-L framework, analysis proceeds on a case by case basis.

One possible method of obtaining a stationary distribution without imposing both upper and lower boundaries is to impose only a lower (upper) reflecting barrier and construct the stochastic process such that positive (negative) infinity is non-attracting, e.g., Linetsky (2005); Ait-Sahalia (1999). This can be achieved by using a mean-reverting drift term. In contrast, Cox and Miller [42, p.223-5] use the Brownian motion, constant coefficient forward equation with $x_0 > 0$, $A[x] = \mu < 0$ and $B[x] = \frac{1}{2}\sigma^2$ subject to the lower reflecting barrier at $x = 0$ given in (2) to solve for both the U and the stationary density. The principal solution is solved using the ‘method of images’ to obtain:

$$U[x,t | x_0] = \frac{1}{\sigma \sqrt{2\pi t}} \left\{ \exp\left(-\frac{(x - x_0 - \mu t)^2}{2\sigma^2 t}\right) + \exp\left(-\frac{4x_0\mu t - (x - x_0 - \mu t)^2}{2\sigma^2 t}\right) \right\} \\ + \frac{1}{\sigma \sqrt{2\pi t}} \left\{ \frac{2\mu}{\sigma^2} \exp\left(\frac{2\mu x}{\sigma^2}\right) \left(1 - N\left[\frac{x + x_0 + \mu t}{\sigma\sqrt{t}}\right] \right) \right\}$$

where $N[x]$ is again the cumulative standard normal distribution function. Observing that $A[x] = \mu > 0$ again produces $U \rightarrow 0$ as $t \rightarrow +\infty$, the stationary density for $A[x] = \mu < 0$ has the Maxwell form:

$$\Psi[x] = \frac{2|\mu|}{\sigma^2} \exp\left(-\frac{2|\mu|x}{\sigma^2}\right)$$

Though x_0 does not enter the solution, combined with the location of the boundary at $x = 0$, it does implicitly impose the restriction $x > 0$. From the Proposition, $T[x,t | x_0]$ can be determined as $U[x,t | x_0] - \Psi[x]$.

Following Linetsky (2005), Veerstraeten (2004) and others, the analytical procedure used to determine U involves specifying the parameters of the forward equation and the boundary conditions and then solving for $\Psi[x]$ and $T[x,t | x_0]$. Wong (1964) uses a different approach, initially selecting

a stationary distribution and then solving for U using the restrictions of the Pearson system to specify the forward equation. In this approach, the functional form of the desired stationary distribution determines the appropriate boundary conditions. While application of this approach has been limited to the restricted class of distributions associated with the Pearson system, it is expedient when a known stationary distribution, such as the standard normal distribution, is of interest. More precisely, let:

$$\Psi[x] = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{x^2}{2}\right], \quad I_o = (-\infty < x < \infty)$$

In this case, the boundaries of the state space are non-attracting and not regular. Solving the Pearson equation gives: $d\Psi[x]/dx = -x \Psi[x]$ and a forward equation of the OU form:

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial}{\partial x} xU = \frac{\partial U}{\partial t}$$

Following Wong (1964, p.268) Mehler's formula can be used to express the solution for U as:

$$U[x,t | x_0] = \frac{1}{\sqrt{2\pi(1 - e^{-2t})}} \exp\left[\frac{-(x - x_0 e^{-t})^2}{2(1 - e^{-2t})}\right]$$

Given this, as $t \rightarrow -\infty$ then $U \rightarrow 0$ and as $t \rightarrow +\infty$ then U achieves the stationary standard normal distribution.

6. The Quartic Exponential Distribution

The roots of bifurcation theory can be found in the early solutions to certain deterministic ordinary differential equations. Consider the deterministic dynamics described by the pitchfork bifurcation ODE:

$$\frac{dx}{dt} = -x^3 + \rho_1 x + \rho_0$$

where ρ_0 and ρ_1 are the ‘normal’ and ‘splitting’ control variables, respectively (e.g., Cobb 1978, 1981). While ρ_0 has significant information in a stochastic context, this is not usually the case in the deterministic problem so $\rho_0 = 0$ is assumed. Given this, for $\rho_1 \leq 0$, there is one real equilibrium ($\{dx/dt\} = 0$) solution to this ODE at $x = 0$ where “all initial conditions converge to the same final point exponentially fast with time” (Crauel and Flandoli 1998, p.260). For $\rho_1 > 0$, the solution bifurcates into three equilibrium solutions $x = \{0, \pm\sqrt{\rho_1}\}$, one unstable and two stable. In this case, the state space is split into two physically distinct regions (at $x = 0$) with the degree of splitting controlled by the size of ρ_1 . Even for initial conditions that are ‘close’, the equilibrium achieved will depend on the sign of the initial condition. Stochastic bifurcation theory extends this model to incorporate Markovian randomness. In this theory, “invariant measures are the random analogues of deterministic fixed points” (Arnold 1998, p.469). Significantly, ergodicity now requires that the component densities that bifurcate out of the stationary density at the bifurcation point be invariant measures, e.g., Crauel et al. (1999, sec.3). As such, the ergodic bifurcating process is irreversible in the sense that past sample paths (prior to the bifurcation) cannot reliably be used to generate statistics for the future values of the variable (after the bifurcation).

It is well known that the introduction of randomness to the pitchfork ODE changes the properties of the equilibrium solution, e.g., (Arnold 1998, sec.9.2). It is no longer necessary that the state space for the principal solution be determined by the location of the initial condition relative to the bifurcation point. The possibility for randomness to cause some paths to cross over the bifurcation point depends on the size of volatility of the process, σ , which measures the non-linear signal to white noise ratio. Of the different approaches to introducing randomness (e.g., multiplicative noise), the simplest approach to converting from a deterministic to a stochastic context is to add a Weiner

process ($dW(t)$) to the ODE. Augmenting the diffusion equation to allow for σ to control the relative impact of non-linear drift versus random noise produces the “pitchfork bifurcation with additive noise” (Arnold 1998, p.475) which in symmetric form is:

$$dX(t) = (\rho_1 X(t) - X(t)^3) dt + \sigma dW(t)$$

Applications in financial economics, e.g., Ait-Sahlia (1999), refer to this diffusion process as the double well process. While consistent with the common use of diffusion equations in financial economics, the dynamics of the pitchfork process captured by $T[x,t | x_0]$ have been “forgotten” (Arnold 1998, p.473).

Models in MPT are married to the transition probability densities associated with unimodal stationary distributions, especially the class of Gaussian-related distributions. Yet, it is well known that more flexibility in the shape of the stationary distribution can be achieved using a higher order exponential density, e.g., Fisher (1921), Cobb et al. (1983), Crauel and Flandoli (1999). Increasing the degree of the polynomial in the exponential comes at the expense of introducing additional parameters resulting in a substantial increase in the analytical complexity, typically defying a closed form solution for the transition densities. However, at least since Elliott (1955), it has been recognized that the solution of the associated regular S-L problem will still have a discrete spectrum, even if the specific form of the eigenfunctions and eigenvalues in $T[x,t | x_0]$ are not precisely determined (Horsthemke and Lefever 1984, sec. 6.7). Inferences about transient stochastic behavior can be obtained by examining the solution of the deterministic non-linear dynamics. In this process, attention initially focuses on the properties of the higher order exponential distributions.

To this end, assume that the stationary distribution is a fourth degree or “general quartic” exponential:

$$\Psi[x] = K \exp[-\Phi[x]] = K \exp[-(\beta_4 x^4 + \beta_3 x^3 + \beta_2 x^2 + \beta_1 x)]$$

where: K is a constant determined such that the density integrates to one; and, $\beta_4 > 0$.²¹ Following Fisher 1921, the class of distributions associated with the general quartic exponential admits both unimodal and bimodal densities and nests the standard normal as a limiting case where $\beta_4 = \beta_3 = \beta_1 = 0$ and $\beta_2 = 1/2$ with $K = 1/(\sqrt{2\pi})$. The stationary distribution of the bifurcating double well process is a special case of the symmetric quartic exponential distribution:

$$\Psi[y] = K_S \exp[-\{\beta_2 (x - \mu)^2 + \beta_4 (x - \mu)^4\}] \quad \text{where } \beta_4 \geq 0$$

where μ is the population mean and the symmetry restriction requires $\beta_1 = \beta_3 = 0$. Such multi-modal stationary densities have received scant attention in financial economics, in general, and in MPT, in particular. To see why the condition on β_1 is needed, consider change of origin $X = Y - \{\beta_3 / 4 \beta_4\}$ to remove the cubic term from the general quartic exponential (Matz 1978, p.480):

$$\Psi[y] = K_Q \exp[-\{\kappa (y - \mu_y) + \alpha (y - \mu_y)^2 + \gamma (y - \mu_y)^4\}] \quad \text{where } \gamma \geq 0$$

The substitution of y for x indicates the change of origin which produces the following relations between coefficients for the general and specific cases:

$$\kappa = \frac{8\beta_1\beta_4^2 - 4\beta_2\beta_3\beta_4 + \beta_3^3}{8\beta_4^2} \quad \alpha = \frac{8\beta_2\beta_4 - 3\beta_3^2}{8\beta_4} \quad \gamma = \beta_4$$

The symmetry restriction $\kappa = 0$ can only be satisfied if both β_3 and $\beta_1 = 0$. Given the symmetry restriction, the double well process further requires $-\alpha = \gamma = \sigma = 1$. Solving for the modes of $\Psi[y]$ gives $\pm \sqrt{\{|\alpha| / (2\gamma)\}}$ which reduces to ± 1 for the double well process, as in Ait-Sahlia (1999, Figure 6B, p.1385).

INSERT FIGURE 1 HERE

As illustrated in Figure 1, the selection of a_i in the stationary density $\Psi_i[x] = K_Q \exp\{-(.25 x^4 - .5$

$x^2 - a_i x$ } defines a family of general quartic exponential densities, where a_i is the selected value of κ for that specific density.²² The coefficient restrictions on the parameters α and γ dictate that these values cannot be determined arbitrarily. For example, given that β_4 is set at .25, then for $a_i = 0$, it follows that $\alpha = \beta_2 = 0.5$. ‘Slicing across’ the surface in Figure 1 at $a_i = 0$ reveals a stationary distribution that is equal to the double well density. Continuing to slice across as a_i increases in size, the bimodal density becomes progressively more asymmetrically concentrated in positive x values. Though the location of the modes does not change, the amount of density between the modes and around the negative mode decreases. Similarly, as a_i decreases in size the bimodal density becomes more asymmetrically concentrated in positive x values. While the stationary density is bimodal over $a_i \in \{-1, 1\}$, for $|a_i|$ large enough the density becomes so asymmetric that only a unimodal density appears. For the general quartic, asymmetry arises as the amount of the density surrounding each mode (the sub-density) changes with a_i . In this, the individual stationary sub-densities have a symmetric shape. To introduce asymmetry in the sub-densities, the reflecting boundaries at a and b that bound the state space for the regular S-L problem can be used to introduce positive asymmetry in the lower sub-density and negative asymmetry in the upper sub-density.

Following Chiarella et al. (2008), the stochastic bifurcation process has a number of features which are consistent with the *ex ante* behavior of a securities market driven by a combination of chartists and fundamentalists. Placed in the context of the classical S-L framework, because the stationary distributions are bimodal and depend on forward parameters – such as κ , α , γ and a_i in Figure 1 – that are not known on the decision date, the risk-return tradeoff models employed in MPT are uninformative. What use is the forecast provided by $E[x(T)]$ when it is known that there are other modes for $x(T)$ values that are more likely to occur? A mean estimate that is close to a bifurcation

point would even be unstable. The associated difficulty of calculating an expected return forecast or other statistical estimate such as volatility from past data can be compounded by the presence of transients that originate from boundaries and initial conditions. For example, the presence of a recent structural break, e.g., the collapse in global stock prices from late 2008-early 2009, can be accounted for by appropriate selection of x_0 , the fundamental dependence of investment decisions on the relationship between x_0 and future (not past) performance is not captured by the reversible ergodic processes employed in MPT that ignore transients arising from initial conditions. Mathematical tools such as classical S-L methods are able to demonstrate this fundamental dependence by exploiting properties of *ex ante* bifurcating ergodic processes to generate theoretical *ex post* sample paths that provide a better approximation to the sample paths of observed financial data.

7. Conclusion

The classical ergodicity hypothesis provides a point of demarcation in the pre-histories of MPT and econophysics. To deal with the problem of making statistical inferences from ‘non-experimental’ data, theories in MPT typically employ stationary densities that are: time reversible; unimodal; and, allow no short or long term impact from initial and boundary conditions. The possibility of bimodal processes or *ex ante* impact from initial and boundary conditions is not recognized or, it seems, intended. Significantly, as illustrated in the need to select an a_i in Figure 1 in order to specify the ‘real world’ *ex ante* stationary density, a semantic connection can be established between the subjective uncertainty about encountering a future bifurcation point and, say, the possible collapse of an asset price bubble impacting future market valuations. Examining the quartic exponential stationary distribution associated with a bifurcating ergodic process, it is apparent that this distribution nests the Gaussian distribution as a special case. In this sense, results from classical statistical mechanics can

be employed to produce a stochastic generalization of the unimodal, time reversible processes employed in modern portfolio theory.

Appendix

Preliminaries on solving the Forward Equation:

Due to the widespread application in a wide range of subjects, textbook presentations of the Sturm-Liouville problem possess subtle differences that require some clarification to be applicable to the formulation used in this paper. In particular, to derive the canonical form (6) of the Fokker-Planck equation (1) observe that evaluating the derivatives in (1) gives:

$$B[x] \frac{\partial^2 U}{\partial x^2} + \left[2 \frac{\partial B}{\partial x} - A[x] \right] \frac{\partial U}{\partial x} + \left[\frac{\partial^2 B}{\partial x^2} - \frac{\partial A}{\partial x} \right] U = \frac{\partial U}{\partial t}$$

This can be rewritten as:

$$\frac{1}{r[x]} \frac{\partial}{\partial x} \left[P[x] \frac{\partial U}{\partial x} \right] + Q[x] U = \frac{\partial U}{\partial t}$$

where:

$$P[x] = B[x] r[x] \quad \frac{1}{r[x]} \frac{\partial P}{\partial x} = 2 \frac{\partial B}{\partial x} - A[x] \quad Q[x] = \frac{\partial^2 B}{\partial x^2} - \frac{\partial A}{\partial x}$$

It follows that:

$$\frac{\partial B}{\partial x} = \frac{1}{r[x]} \frac{\partial P}{\partial x} - \frac{1}{r^2} \frac{\partial r}{\partial x} P[x] = 2 \frac{\partial B}{\partial x} - A[x] - \frac{B[x]}{r[x]} \frac{\partial r}{\partial x}$$

This provides the solution for the key function $r[x]$:

$$\frac{1}{r[x]} \frac{\partial r}{\partial x} - \frac{1}{B[x]} \frac{\partial B}{\partial x} = -\frac{A[x]}{B[x]} \quad \rightarrow \quad \ln[r] - \ln[k] = -\int^x \frac{A[s]}{B[s]} ds$$

$$r[x] = B[x] \exp \left[-\int^x \frac{A[s]}{B[s]} ds \right]$$

This $r[x]$ function is used to construct the scale and speed densities commonly found in presentations of solutions to the forward equation, e.g., Karlin and Taylor (1981), Linetsky (2005).

Another specification of the forward equation that is of importance is found in Wong (1964, eq.6-7):

$$\frac{d}{dx} \left[B[x] \rho[x] \frac{d\theta}{dx} \right] + \lambda \rho[x] \theta[x] = 0 \quad \text{with b.c.} \quad B[x] \rho[x] \frac{d\theta}{dx} = 0$$

This formulation occurs after separating variables, say with $U[x] = g[x] h[t]$. Substituting this result into (1) gives:

$$\frac{\partial^2}{\partial x^2}[B g h] - \frac{\partial}{\partial x}[A g h] = g[x] \frac{\partial h}{\partial t}$$

Using the separation of variables substitution $(1/h)\{\partial h / \partial t\} = -\lambda$ and redefining $g[x] = \rho\theta$ gives:

$$\frac{d}{dx} \left[\frac{d}{dx} B g - A g \right] = -\lambda g = \frac{d}{dx} \left[\frac{d}{dx} B[x] \rho[x] \theta[x] - A[x] \rho[x] \theta[x] \right] = -\lambda \rho \theta$$

Evaluating the derivative inside the bracket and using the condition $\{d/dx\} [B\rho] - A\rho = 0$ to specify admissible ρ gives:

$$\frac{d}{dx} \left[\theta \frac{d}{dx} (B\rho) + B\rho \frac{d}{dx} \theta - A\rho\theta \right] = \frac{d}{dx} \left[B\rho \frac{d}{dx} \theta \right] = -\lambda \rho[x] \theta[x]$$

which is equation (6) in Wong (1964). The condition used to define ρ is then used to identify the specification of $B[x]$ and $A[x]$ from the Pearson system. The associated boundary condition follows from observing the $\rho[x]$ will be the ergodic density and making appropriate substitutions into the boundary condition:

$$\frac{\partial}{\partial x} \{B[x] f[t] \rho[x] \theta[x]\} - A[x] f[t] \rho[x] \theta[x] = 0 \quad \rightarrow \quad \frac{d}{dx} [B \rho \theta] - A \rho \theta = 0$$

Evaluating the derivative and taking values at the lower (or upper) boundary gives:

$$\begin{aligned} B[a]\rho[a] \frac{d\theta[a]}{dx} + \theta[a] \frac{dB\rho}{dx} - A[a]\rho[a]\theta[a] &= 0 \\ &= B[a]\rho[a] \frac{d\theta[a]}{dx} + \theta[a] \left[\frac{dB[a]\rho[a]}{dx} - A[a]\rho[a] \right] \end{aligned}$$

Observing the expression in the last bracket is the original condition with the ergodic density serving as U gives the boundary condition stated in Wong (1964, eq.7).

Proof of Proposition:

(a) ψ_n has exactly n zeroes in $[a,b]$

Hille (1969, p.398, Theorem 8.3.3) and Birkhoff and Rota (1989, p.320, Theorem 5) shows that the eigenfunctions of the Sturm-Liouville system (1') with (3'), (4') and (5) have exactly n zeroes in the interval (a,b) . More precisely, since it assumed that $r > 0$, the eigenfunction ψ_n corresponding to the

n th eigenvalue has exactly n zeroes in (a,b) .

(b) For $\psi_n = 0$,
$$\int_a^b \psi_n[x] dx = 0$$

Proof:

For $\psi_n = 0$ the following applies:

$$\psi_n = \frac{1}{\lambda_n} \frac{d}{dx} \left\{ \frac{d}{dx} [B[x] \psi_n] - A[x] \psi_n \right\}$$

$$\therefore \int_a^b \psi_n[x] dx = \frac{1}{\lambda_n} \left\{ \frac{d}{dx} [B[x] \psi_n] \Big|_{x=b} - A[b] \psi_n[b] - \frac{d}{dx} [B[x] \psi_n] \Big|_{x=a} + B[a] \psi_n[a] \right\} = 0$$

Since each $\psi_n[x]$ satisfies the boundary conditions.

(c) For some k , $\lambda_k = 0$.

Proof:

From (8):

$$U[x,t] = \sum_{k=0}^{\infty} c_k e^{-\lambda_k t} \psi_k[x]$$

Since $\int_a^b U[x,t] dx = 1$ then:

$$1 = \sum_{k=0}^{\infty} c_k e^{-\lambda_k t} \int_a^b \psi_k[x] dx$$

But from part (b) this will = 0 (which is a contradiction) unless $\lambda_k = 0$ for some k .

(d) $\lambda_0 = 0$

Proof:

From part (a), $\psi_0[x]$ has no zeroes in (a,b) . Therefore, either $\int_a^b \psi_0[x] dx > 0$ or $\int_a^b \psi_0[x] dx < 0$.

It follows from part (b) that $\lambda_0 = 0$.

(e) $\lambda_n > 0$ for $n \neq 0$. This follows from part (d) and the strict inequality conditions provided in part (a).

(f) Obtaining the solution for $T[x]$ in the Proposition:

From part (d) it follows:
$$\frac{d}{dx} \left\{ [B[x] \psi_0[x,t]]_x - A[x] \psi_0[x,t] \right\} = 0$$

Integrating this equation from a to x and using the boundary condition gives:

$$[B[x] \psi_0[x,t]]_x - A[x] \psi_0[x,t] = 0$$

This equation can be solved for ψ_0 to get:

$$\psi_0 = A [B[x]]^{-1} \exp \left[\int_a^x \frac{A[s]}{B[s]} ds \right] = C [r[x]]^{-1} \quad \text{where: } C = \text{constant}$$

Therefore:

$$\psi_0[x] = \left[\int_a^b r[x] C^2 [r[x]]^{-2} dx \right]^{-1/2} C [r[x]]^{-1} = \frac{[r[x]]^{-1}}{\left[\int_a^b r[x]^{-1} dx \right]^{1/2}}$$

Using this definition and observing that the integral of $f[x]$ over the state space is one it follows:

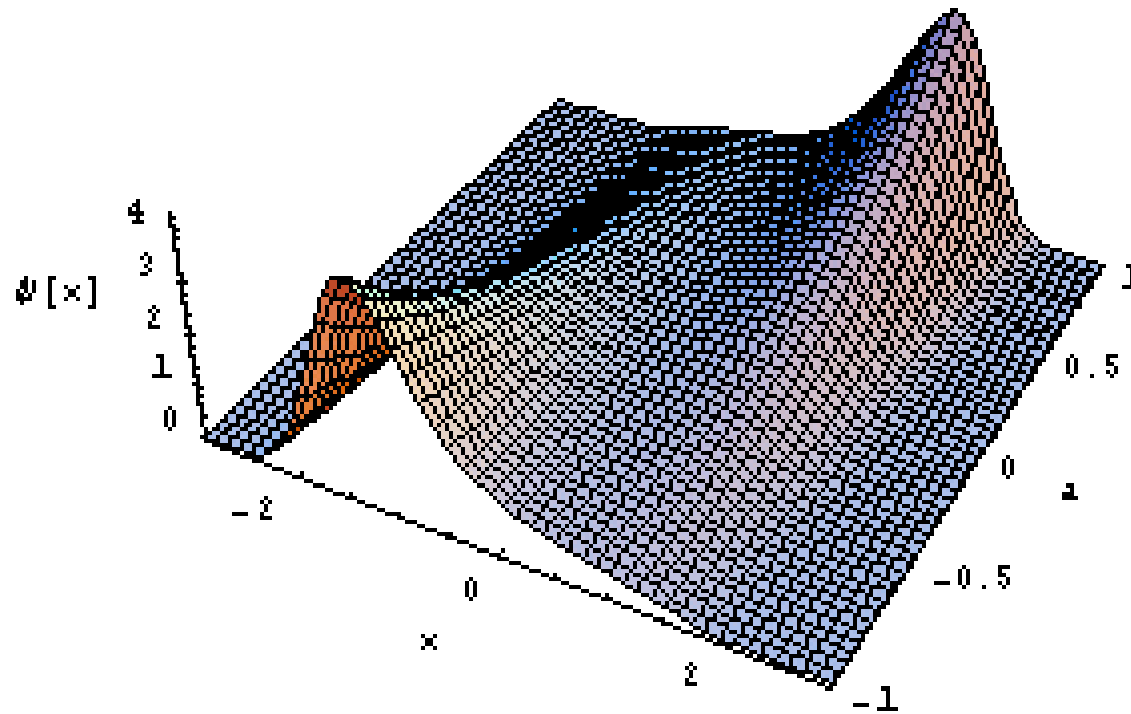
$$c_0 = \int_a^b \left\{ f[x] r[x] \frac{r[x]^{-1}}{\left[\int_a^b r[x] dx \right]^{1/2}} \right\} dx = \frac{1}{\left[\int_a^b r[x] dx \right]^{1/2}}$$

$$\therefore c_0 \psi[x] = \frac{r[x]^{-1}}{\left[\int_a^b [r[x]]^{-1} dx \right]}$$

(g) The Proof of the Proposition now follows from parts (f), (e) and (b).

Figure 1:*

Family of Stationary Densities for $\Psi_i[x] = K_Q \exp\{ -(.25 x^4 - .5 x^2 - a_i x) \}$



* Each of the continuous values for a signifies a different stationary density. For example, at $a = 0$ the density is the double well density which symmetric about zero and with modes at ± 1 .

Bibliography

Aït-Sahalia, Y. (1999), "Transition Densities for Interest Rates and Other Nonlinear Diffusions", Journal of Finance 54: 1361-95.

Arnold, L. (1998), Random Dynamical Systems, New York: Springer-Verlag.

Ball, C. and A. Roma (1998), "Detecting Mean Reversion Within Reflecting Barriers: Applications to the European Exchange Rate Mechanism", Applied Mathematical Finance 5: 1-15.

Beaulieu, M., J. Defour and L. Khalaf (2013), "Identification-Robust Estimation and Testing of the Zro-Beta CAPM", Review of Economic Studies 80: 892–924.

Berg, P. and J. McGregor (1966), Elementary Partial Differential Equations, San Francisco: Holden-Day.

Birkhoff, G. (1931), "Proof of the Ergodic Theorem", Proceedings of the National Academy of Sciences 17: 656-60.

Birkhoff, G. and G. Rota (1989), Ordinary Differential Equations (4th ed.), New York: Wiley.

Black, F. and M. Scholes (1973), "The Pricing of Options and Corporate Liabilities", Journal of Political Economy 81: 637-59.

.Bonomo, M., R. Garcia, N. Meddahi and R. Tédongap (2011) "Generalized Disappointment Aversion, Long-run Volatility Risk, and Asset Prices", Review of Financial Studies 24: 82-122.

Brennan, M. and E. Schwartz (1979), "A Continuous Time Approach to the Pricing of Bonds", Journal of Banking and Finance 3: 133-55.

Brush, S. (1983), Statistical Physics and the Atomic Theory of Matter: From Boyle and Newton to Landau and Onsager, Princeton, NJ: Princeton University Press.

Brush, S. (1976), The Kind of Motion We Call Heat: A History of the Kinetic Theory of Gases in the 19th Century, Amsterdam: North Holland.

Bullard, J. and A. Butler (1993), "Nonlinearity and Chaos in Economic Models: Implications for Policy Decisions", Economic Journal 103: 849-867.

Campbell, J. (1996), "Understanding Risk and Return", Journal of Political Economy 104: 298-345.

Ceragnani, C. (1998), Ludwig Boltzmann: The Man who Trusted Atoms, Oxford, UK: Oxford University Press.

Chiarella, C., X-Z He, D. Wang and M. Zheng (2008), "The stochastic bifurcation behaviour of speculative financial markets", Physica A 387: 3837-3846.

Cobb, L. (1978), "Stochastic Catastrophe Models and Multimodal Distributions", Behavioral Science 23: 360-74.

Cobb, L. (1981), "The Multimodal Exponential Families of Statistical Catastrophe Theory", in C. Taillie, G. Patil and B. Baldessari (eds.), Statistical Distributions in Scientific Work (vol.4), Dordrecht, Holland: Reidel Press.

Cobb, L., P. Koppstein and N. Chen (1983), "Estimation and Moment Recursion Relations for Multimodal Distributions of the Exponential Family", Journal of the American Statistical Association 78: 124-30.

Constantinides, G. (2002), "Rational asset prices", Journal of Finance 57: 1566-82.

Cox, D. and H. Miller (1965), The Theory of Stochastic Processes, London: Chapman and Hall.

Crauel, H. and F. Flandoli (1998), "Additive Noise Destroys a Pitchfork Bifurcation", Journal of Dynamics and Differential Equations 10: 259-74.

Crauel, H., P. Imkeller, and M. Steinkamp (1999), "Bifurcations of One-Dimensional Stochastic Differential Equations", chapter 2 in H. Crauel and M. Gundlach (eds.), Stochastic Dynamics, New York: Springer Verlag.

Davidson, P. (1991), "Is Probability Theory Relevant for Uncertainty? A Post Keynesian Perspective", Journal of Economic Perspectives 5:129-43.

Davidson, P. (1988), "A technical definition of uncertainty and the long-run non-neutrality of money", Cambridge Journal of Economics 12: 329-37.

De Jong, F. (1994), "A univariate analysis of European monetary system exchange rates using a target zone model", Journal of Applied Econometrics 9: 31-45.

Dhrymes (1971), Distributed Lags: Problems of Estimation and Formulation, San Francisco, CA: Holden-Day.

Dhrymes, P. (1974), Econometrics: statistical foundations and applications, New York: Springer-Verlag.

Dickey, D. and W. Fuller (1979), "Distribution of the Estimators for Autoregressive Time Series With a Unit Root", Journal of the American Statistical Association 74: 427-431.

- Dimson, E., P. Marsh and M. Staunton (2002), Triumph of the Optimists, 101 Years of Global Investment Returns, Princeton, NJ: Princeton University Press.
- Dixit, A. and R. Pindyck (1994), Investment under uncertainty, Princeton, NJ: Princeton University Press.
- Doob, J. (1953), Stochastic Processes, New York: Wiley.
- Elliott, J. (1955), “Eigenfunction expansions associated with singular differential operators”, Transactions of the American Mathematical Society 78: 406-25.
- Elton, E. and M. Gruber (1984), Modern Portfolio Theory and Investment Analysis (2nd ed.), New York: Wiley.
- Encyclopedia of Mathematics (2002), “Ergodic Theory”, New York: Springer Verlag / SpringerLink.
- Ehrenfest, P. (1911), Begriffliche Grundlagen der statistischen Auffassung in der Mechanik, Leipzig: B.G. Teubner.
- Engle, R. and C. Granger (1987), “Co-integration and Error Correction: Representation, Estimation and Testing”, Econometrica 55: 251-276.
- Epstein, L and S. Ji (2013), “Ambiguous Volatility and Asset Pricing in Continuous Time”, Review of Financial Studies 26:1740-1786.
- Feller, W. (1954), “Diffusion Processes in One Dimension”, Transactions of the American Mathematical Society 77: 1-31.
- Fisher, R.A. (1921), “On the mathematical foundations of theoretical statistics”, Philosophical Transactions of the Royal Society 222A: 309-68.
- Gallivotti, G. (1995), “Ergodicity, Ensembles, Irreversibility in Boltzmann and Beyond”, Journal of Statistical Physics 78:1571-1589.
- Georgescu-Roegen, N. (1971), The entropy law and the economic process, Cambridge, MA: Harvard University Press.
- Gihman, I. and A. Skorohod (1979), The Theory of Stochastic Processes (vol. II, III), New York: Springer-Verlag.
- Halmos, P. (1949), “Measurable Transformations”, Bulletin of the American Mathematical Society (Nov.): 1015-34.

Hansen, L., J. Scheinkman and N. Touzi (1998), "Spectral Methods for Identifying Scalar Diffusions", Journal of Econometrics 86: 1-32.

Hendry, D. (1995), Dynamic Econometrics, Oxford, UK: Oxford University Press.

Hicks, J. (1939), Value and Capital, Clarendon Press.

Hille, E. (1969), Lectures on Ordinary Differential Equations, London: Addison-WeS-Ley.

Horsthemke, W. and R. Lefever (1984), Noise-Induced Transitions, New York: Springer-Verlag.

Horst, U. and J. Wenzelburger (2008), "On non-ergodic asset prices", Economic Theory 34: 207-34.

Huang, W. (2005), "On the statistical dynamics of economics", Journal of Economic Behavior and Organization 56: 543-65.

Jevons, W.S. (1877), The Principles of Science (2nd ed.), New York: Dover Press reprint (1958).

Jovanovic, F. and C. Schinckus (2013), "Econophysics: A New Challenge for Financial Economics", Journal of the History of Economic Thought 35: 319-52.

Kapetanios, G. and Y. Shin (2011), "Testing the Null Hypothesis of Nonstationary Long Memory against the Alternative Hypothesis of a Nonlinear Ergodic Model", Econometric Reviews 30: 620-45.

Karlin, S. and H. Taylor (1981), A Second Course in Stochastic Processes, New York: Academic Press.

Kolmogorov, A. (1933), Foundations of the Theory of Probability (in German), Springer: Berlin; English Edition, New York: Chelsea (1950).

Linetsky, V. (2005), "On the Transitions Densities for Reflected Diffusions", Advances in Applied Probability 37: 435-60.

Mackey, G. (1974), "Ergodic Theory and its Significance for Statistical Mechanics and Probability Theory", Advances in Mathematics 12: 178-268.

Mandelbrot, B. (1997), Fractals and scaling in finance : discontinuity, concentration, risk, New York: Springer.

Mandelbrot, B. and R. Hudson (2004), The (mis)behavior of markets : a fractal view of risk, ruin, and reward, New York: Basic Books.

Markowitz, H. (1952), "Portfolio Selection", Journal of Finance 7: 77-91.

Matz, A. (1978), "Maximum Likelihood Parameter Estimation for the Quartic Exponential Distribution", Technometrics 20: 475-84.

Maxwell, J.C. (1867), "On the Dynamical Theory of Gases", Philosophical Transactions of the Royal Society 157: 49-88

McCauley, J. (2004), Dynamics of markets : econophysics and finance, Cambridge, UK: Cambridge University Press.

Medio, A. (2005), "Ergodic Theory of Nonlinear Dynamics", in M. Lines (ed.), Nonlinear Dynamical Systems in Economics, New York, Springer: 67-102.

Meitz, M. and P. Saikkonen (2008), "Ergodicity, Mixing and Existence of Moments of a Class of Markov Models with Applications to GARCH and ACD Models", Econometric Theory 24: 1291-1320.

Merton, R. (1973), "The Theory of Rational Option Pricing", Bell Journal of Economics and Management Science, 141-83.

Mirowski, P. (1989a), More Heat than Light, Economics as Social Physics: Physics as nature's economics, Cambridge, UK: Cambridge University Press.

Mirowski, P. (1989b), "The Probabilistic Counter-revolution, or How Stochastic Concepts Came to Neoclassical Economic Theory", Oxford Economic Papers 41: 217-235.

Mirowski, P. (1984), "Physics and the Marginalist Revolution", Cambridge Journal of Economics 8: 361-79.

Nicola, P. (1997), "Real Time Tatonnements and Firm Behaviour", Rivista Internazionale di Scienze Sociali, 105: 403-18.

Reed, M. and B. Simon (1980), Methods of Modern Mathematical Physics, New York: Academic Press.

Risken, H. (1989), The Fokker-Planck Equation, Methods of Solution and Applications, New York: Springer-Verlag.

Roehner (2002), Patterns of Speculation, Cambridge, UK: Cambridge University Press.

Samuelson, P. (1976), "Optimality of Sluggish Predictors under Ergodic Probabilities", International Economic Review 17: 1-7.

Samuelson, P. (1965), "The Rational Theory of Warrant Pricing", Industrial Management Review 6:

13-31.

Samuelson, P. (1947), Foundations of Economic Analysis, Boston, MA: Harvard University Press.

Schinckus, C. (2010), “Is econophysics a new discipline? The neopositivist argument”, Physica A 389: 3814-3821.

Sprenkle, C. (1961), “Warrant Prices as Indicators of Expectations and Preferences”, Yale Economic Essays 1: 178-231.

Theil, H. (1971), Principles of econometrics, New York: Wiley.

Veerstraeten, D. (2004), “The Conditional Probability Density Function for a Reflected Brownian Motion”, Computational Economics 24: 185-207.

von Neumann, J. (1932), “Proof of the Quasi-ergodic Hypothesis”, Proceedings of the National Academy of Sciences 18: 255-63.

Weiner, N. (1939), “The Ergodic Theorem”, Duke Mathematical Journal: 1-18.

Weintraub, E.R. (2002), How Economics Became a Mathematical Science, Durham, NC: Duke University Press.

Wong, E. (1964), “The Construction of a Class of Stationary Markoff Processes”, in Proceedings of Symposium on Applied Mathematics R. Bellman (ed.), Providence, RI: American Mathematical Society.

NOTES

1. For example, Black and Scholes (1973) discuss the consistency of the option pricing formula with the capital asset pricing model. More generally, financial economics employs primarily Gaussian-based finite parameter models that agree with the ‘tradeoff between risk and return’.
2. In rational mechanics, once the initial positions of the particles of interest, e.g., molecules, are known, the mechanical model fully determines the future evolution of the system. This scientific and philosophical approach is often referred to as Laplacian determinism.
3. Boltzmann and Max Planck were vociferous opponents of energetics. The debate over energetics was part of a larger intellectual debate concerning determinism and reversibility. Jevons (1877, p.738-9) reflects the entrenched determinist position of the marginalists: “We may safely accept as a satisfactory scientific hypothesis the doctrine so grandly put forth by Laplace, who asserted that a perfect knowledge of the universe, as it existed at any given moment, would give a perfect knowledge of what was to happen thenceforth and for ever after. Scientific inference is impossible,

unless we may regard the present as the outcome of what is past, and the cause of what is to come. To the view of perfect intelligence nothing is uncertain.” What Boltzmann, Planck and others had observed in statistical physics was that, even though the behavior of one or two molecules can be completely determined, it is not possible to generalize these mechanics to describe the macroscopic motion of molecules in large, complex systems, e.g., Brush (1983, esp. ch.II).

4. As such, Boltzmann was part of the larger: “Second Scientific Revolution, associated with the theories of Darwin, Maxwell, Planck, Einstein, Heisenberg and Schrödinger, (which) substituted a world of process and chance whose ultimate philosophical meaning still remains obscure” (Brush 1983, p.79). This revolution superseded the: “First Scientific Revolution, dominated by the physical astronomy of Copernicus, Kepler, Galileo, and Newton, ... in which all changes are cyclic and all motions are in principle determined by causal laws.” The irreversibility and indeterminism of the Second Scientific Revolution replaces the reversibility and determinism of the First.

5. There are many interesting sources on these points which provide citations for the historical papers that are being discussed. Cercignani (1998, p.146-50) discusses the role of Maxwell and Boltzmann in the development of the ergodic hypothesis. Maxwell (1867) is identified as “perhaps the strongest statement in favour of the ergodic hypothesis”. Brush (1976) has a detailed account of the development of the ergodic hypothesis. Gallavotti (1995) traces the etymology of “ergodic” to the ‘ergode’ in an 1884 paper by Boltzmann. More precisely, an ergode is shorthand for ‘ergomonode’ which is a ‘monode with given energy’ where a ‘monode’ can be either a single stationary distribution taken as an ensemble or a collection of such stationary distributions with some defined parameterization. The specific use is clear from the context. Boltzmann proved that an ergode is an equilibrium ensemble and, as such, provides a mechanical model consistent with the second law of thermodynamics. It is generally recognized that the modern usage of ‘the ergodic hypothesis’ originates with Ehrenfest (1911).

6. Kapetanios and Shin (2011, p.620) capture the essence of this quandary: “Interest in the interface of nonstationarity and nonlinearity has been increasing in the econometric literature. The motivation for this development may be traced to the perceived possibility that nonlinear ergodic processes can be misinterpreted as unit root nonstationary processes. Furthermore, the inability of standard unit root tests to reject the null hypothesis of unit root for a large number of macroeconomic variables, which are supposed to be stationary according to economic theory, is another reason behind the increased interest.”

7. The second law of thermodynamics is the universal law of increasing entropy – a measure of the randomness of molecular motion and the loss of energy to do work. First recognized in the early 19th century, the second law maintains that the entropy of an isolated system, not in equilibrium, will necessarily tend to increase over time. Entropy approaches a maximum value at thermal equilibrium. A number of attempts have been made to apply the entropy of information to problems in economics, with mixed success. In addition to the second law, physics now recognizes the zeroth law of thermodynamics that “any system approaches an equilibrium state” (Reed and Simon 1980, p.54). This implications of the second law for theories in economics was initially explored by Georgescu-Roegen (1971).

8. In this process, the ergodicity hypothesis is required to permit the one observed sample path to be used to estimate the parameters for the *ex ante* distribution of the ensemble paths. In turn, these parameters are used to predict future values of the economic variable.

9. Heterodox critiques are associated with views considered to originate from within economics. Such critiques are seen to be made by ‘economists’, e.g., Post Keynesian economists, institutional economists, radical political economists and so on. Because such critiques take motivation from the theories of mainstream economics, these critiques are distinct from econophysics. Following Schinckus (2010, p.3818): “Econophysicists have then allies within economics with whom they should become acquainted.”

10. Dhyrnes (1974, p.1-29) discusses the algebra of the lag operator.

11. Critiques of mainstream economics that are rooted in the insights of *The General Theory* recognize the distinction between fundamental uncertainty and objective probability. As a consequence, the definition of ergodic theory in heterodox criticisms of mainstream economics lacks formal precision, e.g., the short term dependence of ergodic processes on initial conditions is not usually recognized. Ergodic theory is implicitly seen as another piece of the mathematical formalism inspired by Hilbert and Bourbaki and captured in the Arrow-Debreu general equilibrium model of mainstream economics.

12. In this context though not in all contexts, econophysics provides a ‘macroscopic’ approach. In turn, ergodicity is an assumption that permits the time average from a single observed sample path to (phenomenologically) model the ensemble of sample paths. Given this, econophysics does contain a substantively richer toolkit that encompasses both ergodic and non-ergodic processes. Many works in econophysics implicitly assume ergodicity and develop models based on that assumption.

13. The distinction between invariant and ergodic measures is fundamental. Recognizing a number of distinct definitions of ergodicity are available, following Medio (2005, p.70) the Birkhoff-Khinchin ergodic (BK) theorem for invariant measures can be used to demonstrate that ergodic measures are a class of invariant measures. More precisely, the BK theorem permits the limit of the time average to depend on initial conditions. In effect, the invariant measure is permitted to decompose into invariant ‘sub-measures’. The physical interpretation of this restriction is that sample paths starting from a particular initial condition may only be able to access a part of the sample space, no matter how long the process is allowed to run. For an ergodic process, sample paths starting from any admissible initial condition will be able to ‘fill the sample space’, i.e., if the process is allowed to run long enough, the time average will not depend on the initial condition. Medio (2005, p.73) provides a useful example of an invariant measure that is not ergodic.

14. The phenomenological approach is not without difficulties. For example, the restriction to Markov processes ignores the possibility of invariant measures that are not Markov. In addition, an important analytical construct in bifurcation theory, the Lyapunov exponent, can encounter difficulties with certain invariant Markov measures. Primary concern with the properties of the

stationary distribution is not well suited to analysis of the dynamic paths around a bifurcation point. And so it goes.

15. A diffusion process is ‘regular’ if starting from any point in the state space I , any other point in I can be reached with positive probability (Karlin and Taylor (1981, p.158). This condition is distinct from other definitions of regular that will be introduced: ‘regular boundary conditions’ and ‘regular S-L problem’.

16. The classification of boundary conditions is typically an important issue in the study of solutions to the forward equation. Important types of boundaries include: regular; exit; entrance; and natural. Also important in boundary classification are: the properties of attainable and unattainable; whether the boundary is attracting or non-attracting; and whether the boundary is reflecting or absorbing. In the present context, regular, attainable, reflecting boundaries are usually being considered, with a few specific extensions to other types of boundaries. In general, the specification of boundary conditions is essential in determining whether a given PDE is self-adjoint

17. Heuristically, if the ergodic process runs long enough, then the stationary distribution can be used to estimate the constant mean value. This definition of ergodic is appropriate for the one-dimensional diffusion cases considered in this paper. Other combinations of transformation, space and function will produce different requirements. Various theoretical results are available for the case at hand. For example, the existence of an invariant Markov measure and exponential decay of the autocorrelation function are both assured.

18. For ease of notation it is assumed that $t_0 = 0$. In practice, solving (1) combined with (3)-(5) requires a and b to be specified. While a and b have ready interpretations in physical applications, e.g., the heat flow in an insulated bar, determining these values in economic applications can be more challenging. Some situations, such as the determination of the distribution of an exchange rate subject to control bands (e.g., Ball and Roma 1998), are relatively straight forward. Other situations, such as profit distributions with arbitrage boundaries or output distributions subject to production possibility frontiers, may require the basic S-L framework to be adapted to the specifics of the modeling situation.

19. The mathematics at this point are heuristic. More appropriate would be to observe that U^* is the special case where $U = \Psi[x]$, a strictly stationary distribution. This would require discussion of how to specify the initial and boundary conditions to ensure that this is the solution to the forward equation.

20. A more detailed mathematical treatment can be found in de Jong (1994).

21. In what follows, except where otherwise stated, it is assumed that $\sigma = 1$. Hence, the condition that K be a constant such that the density integrates to one incorporates the $\sigma = 1$ assumption. Allowing $\sigma \neq 1$ will scale either the value of K or the β 's from that stated.

22. A number of simplifications were used to produce the 3D image in Figure 1: x has been centered about μ ; and, $\sigma = K_\rho = 1$. Changing these values will impact the specific size of the parameter values

for a given x but will not change the general appearance of the density plots.