DOI: 10.14754/CEU.2016.01

Boundary Crossing Counting Processes

Theory and Applications in Statistics and Finance

Peter Farkas

A dissertation submitted in partial fulfilment of the
requirements for the
Dissertation for the Degree of Doctor of Philosophy
at
Central European University
Budapest, Hungary

Supervisor: Laszlo Matyas
Associate supervisor: Peter Kondor

DOI: 10.14754/CEU.2016.01

Abstract

This thesis presents some new results in the field of statistics and finance. As for the former, we discuss how to make nonparametric inference without relying on asymptotic approximation. As for the latter, we solve the optimal portfolio choice problem without describing security prices with a parametric model.

These results are accomplished by representing and analyzing the data using a new, state-dependent, perspective. More precisely, we relate to the sampling frequency in a new way. Let us consider Equation (1) below.

$$\Delta X_t = X_t - X_{t-c}, \tag{1}$$

where X_t is a stochastic process with memory and c is the sampling frequency. In many econometric studies, the data generating process is represented and analyzed as if it were sampled using some constant sampling frequency. Typically, c is chosen to be one, that is $\Delta X_t = X_t - X_{t-1}$: for data published with monthly frequency, c is equal to one months, for daily observations c is equal to one day.

This thesis takes an inverse approach. We represent and analyze the data generating process as if it were sampled by a specific random frequency. More precisely, we exogenously fix $\Delta X_t = X_t - X_{t-c}$ to be either some predefined positive, U, or negative, L, number and allow the sampling frequency to vary.

$$\Delta X_t = X_t - X_{t-T^A} = \begin{cases} U \\ L \end{cases} \tag{2}$$

where T^A, as explained later, represents a boundary crossing moment. Thus, in our representation, the sampling frequency, c, is random

DOI: 10.14754/CEU.2016.01

and the data is represented using boundary crossing events. This representation requires us to introduce new stochastic processes which characterize these boundary crossing events.

This new perspective opens up new opportunities in the field of statistics and in finance. As for the former, using this representation, nonparametric inference can be made without relying on asymptotic approximation. As for the latter, we can solve the optimal portfolio choice problem without describing the security prices with a parametric model.

The thesis consists of three chapters. Each chapters is a self-standing article intended for publication in peer-reviewed journals. Thus, they are kept as separate entities. Consequently, sometimes the content of the thesis is repetitive although an effort was made to reduce redundancy as much as possible.

The first chapter aims to provide a brief theoretical foundation hence its results are applied throughout the thesis. It also discusses univariate unit root testing from this new perspective. The next chapter extends some of the results of the first chapter to panel data settings. The last chapter applies the theoretical results of the first chapter to solve the optimal portfolio choice problem. The abstract of each chapters is as follows.

Chapter 1. Counting Process Generated by Boundary Crossing Events: Theory and Applications in Nonparametric Statistics

This chapter introduces and analyzes a new class of stochastic process, named the Boundary Crossing Counting (BCC) process. It shows how to obtain the upper and lower crossing distribution which counts how many times a stochastic process crosses some exogenously defined boundaries. Also, it derives the upper minus lower crossing

distribution using a binomial grid. The methods of estimation are calibrated by comparing analytical and estimated BCC distributions. The next part of the chapter shows how to use boundary crossing events to test for unit roots. Our Monte Carlo studies show that the proposed test is more powerful than the Augmented Dickey-Fuller test or the Phillips-Perron test in time series settings when the error term has t-distribution and the time-dimension is relatively short. It is also more powerful than the Variance ratio test. We conclude with a financial application in which we show that based on Shiller's data, the excess total return based on S&P500 exhibits mean reverting behavior.

Chapter 2. Testing for Unit Roots in Panel Data with Boundary Crossing Counts, which is joint work with Laszlo Matyas.

This chapter introduces a new, distribution free, non-asymptotic, approach for unit root testing based on boundary crossing counts. Using this approach, we develop two versions of a panel unit root test. The first can be applied in the case of cross-sectionally independent panel data, while the second is designed for cross-sectionally dependent panels. As for the results, the first version of the newly proposed test dominates the IPS test and the Maddala-Wu test in case of relatively short, cross-sectionally independent panel data. The second version is more powerful than existing second generation panel data tests, such as Bai and Ng's PANIC unit root test or Pesaran's CADF test in case the data is generated by a multi-factor model and the time dimension is relatively short. Next, we show that the unit root hypothesis cannot be rejected on real exchange rate data hence we do not find supportive evidence for the PPP hypothesis. Finally, we discuss various methodological issues related to this newly proposed test.

DOI: 10.14754/CEU.2016.01

Chapter 3. Portfolio Choice Without Distributional Assumptions: State-dependent Rebalancing in the Nonparametric Domain

We solve the portfolio choice problem without distributional assumptions by extending the use of state-dependent rebalancing to nonparametric settings. We propose a specific, state-dependent rebalancing and show how it is related to the Kelly criterion. Under this rebalancing, the full distribution of the portfolio's terminal value can be approximated by a well-behaving and discrete probability distribution based on boundary crossings. When applied to parametric specifications under transaction costs, the method replicates the baseline results of the geometric Brownian motion. As for nonparametric applications, first, we show that the log-optimal allocation in the US was a leveraged purchase; next we find that leveraged returns were significantly different in various epochs. We continue by explaining how this newly-proposed method can be used for density forecast and conclude with some additional technical details.

DOI: 10.14754/CEU.2016.01

Acknowledgements

First of all, I am very grateful to my supervisor, Laszlo Matyas for his friendly and professional support over the years which helped me to complete the thesis.

I am also very thankful to my associate supervisor, Peter Kondor for showing me how to structure scientific problems and Peter Medvegyev for early discussions on this topic.

At various points during the development of this thesis, I benefited from suggestions and advices from several faculty members and fellow doctoral students at the Central European University for which I am very thankful.

The reviews of Timo Teräsvirta and Robert Lieli were extremely helpful for which I am also very thankful.

I am indebted to my fellow doctoral students, especially Anna Adamecz, Dzsamila Vonnak and Peter Zsohar for helping me in the transition from a business to an academic life, especially during the early stages of the program. I am also thankful to the excellent faculty as well as visiting professors at CEU for their outstanding courses.

I am also deeply grateful to George Soros for founding and financing CEU.

All remaining errors are my own.

DOI: 10.14754/CEU.2016.01

Contents

DOI: 10.14754/CEU.2016.01

DOI: 10.14754/CEU.2016.01

List of Figures

DOI: 10.14754/CEU.2016.01

List of Tables

DOI: 10.14754/CEU.2016.01

Chapter 1

Counting Process Generated by Boundary Crossing Events

Abstract

This chapter introduces and analyzes a new class of stochastic process, named the Boundary Crossing Counting Process. It shows how to obtain the upper and lower crossing distribution which counts how many times a stochastic process crosses some exogenously defined boundaries. Also, it derives the upper minus lower crossing distribution using a binomial grid. The methods of estimation are calibrated by comparing analytical and estimated BCC distributions. The next part of the chapter shows how to use boundary crossing events to test for unit roots. Our Monte Carlo studies show that the proposed test is more powerful than the Augmented Dickey-Fuller test or the Phillips-Perron test in time series settings when the error term has t-distribution and the time-dimension is relatively short. It is also more powerful than the Variance ratio test. We conclude with a financial application in which we show that based on Shiller's data, the excess total return based on S&P500 exhibits mean reverting behavior.

DOI: 10.14754/CEU.2016.01

1.1 Introduction

Technological innovation and the IT revolution have brought us into a new era of data abundance. This previously unseen richness of data creates an opportunity for nonparametric methods, especially in fields where there is a genuine need for flexible stochastic modeling.

In this chapter, we discuss a new class of stochastic process which may prove to be a promising tool for nonparametric data analysis. These stochastic processes are called Boundary Crossing Counting Processes or BCC-processes. The study of these BCC-processes is motivated by their useful applications. These stochastic processes appear to have diverse applications in the field of statistics, finance and management sciences, but here, we focus only on their statistical applications.

As far as we know, repeated boundary crossing behaviour has not been described before using boundary crossing counting distributions. Also, the mapping between the first exit time distribution and the BCC distributions is new and non-trivial. As such, they are still novelties in probability theory. Naturally, the statistical tests and the financial applications are also new.

These stochastic processes are characterized by an underlying stochastic process, X_t, with memory, enclosed by a lower boundary, L_t and an upper boundary, U_t. Boundary crossing events are defined as the first event when some stochastic process crosses either of the boundaries. We derive a new stochastic process, called restarted process, indicated by X_t^*, from the original one by restarting it at some initial value X_t^0 upon each boundary crossing event. Note that the initial value, X_t^0 may be a fixed value or a random variable, as long as it is enclosed by the lower and the upper boundary. Boundary Crossing Counting Processes are those discrete processes which count the number of boundary crossing events for these restarted stochastic

processes. More specifically,

1. upper crossing counting, Y_t^U, counts the number of upper crossing events: $Y_t^U = Y_t^U + 1$ if $X_t^* = U_t$ and $X_{t+\epsilon}^* = X^0$.

2. Also, lower crossing counting Y_t^L counts the number of lower crossing events: $Y_t^L = Y_t^L + 1$ if $X_t^* = L_t$ and $X_{t+\epsilon}^* = X^0$,

where ϵ is an arbitrarily small positive number. Let us now introduce two additional boundary crossing counting distributions.

1. $Y_t^A = Y_t^U + Y_t^L$ counts all events, both the upper and lower crossing events.

2. Finally, $Y_t^D = Y_t^U - Y_t^L$ describes the difference between the number of upper and lower crossing events.

Finally, Y_t is used to refer to all these stochastic processes. The counting process is the function of the restarted process, $Y_t(X_t^*(X_t, L_t, U_t, X_t^0))$, but this dependence is suppressed for ease of notation. Figure 1.1 illustrates an upper and lower crossing counting process.

The main innovation of our work is two-fold: on the one hand, the second section extends the existing theoretical results of probability by introducing and analyzing these new Boundary Crossing Counting stochastic processes. To our knowledge, BCC-processes have not been dealt with or characterized in this manner. On the other hand, we show how to use boundary crossing events for unit root testing. We find that our test is more powerful than the Augmented Dickey-Fuller test or the Phillips-Perron test in time series settings when the error term has t-distribution and the time-dimension is relatively short. It is also more powerful than the Variance

DOI: 10.14754/CEU.2016.01

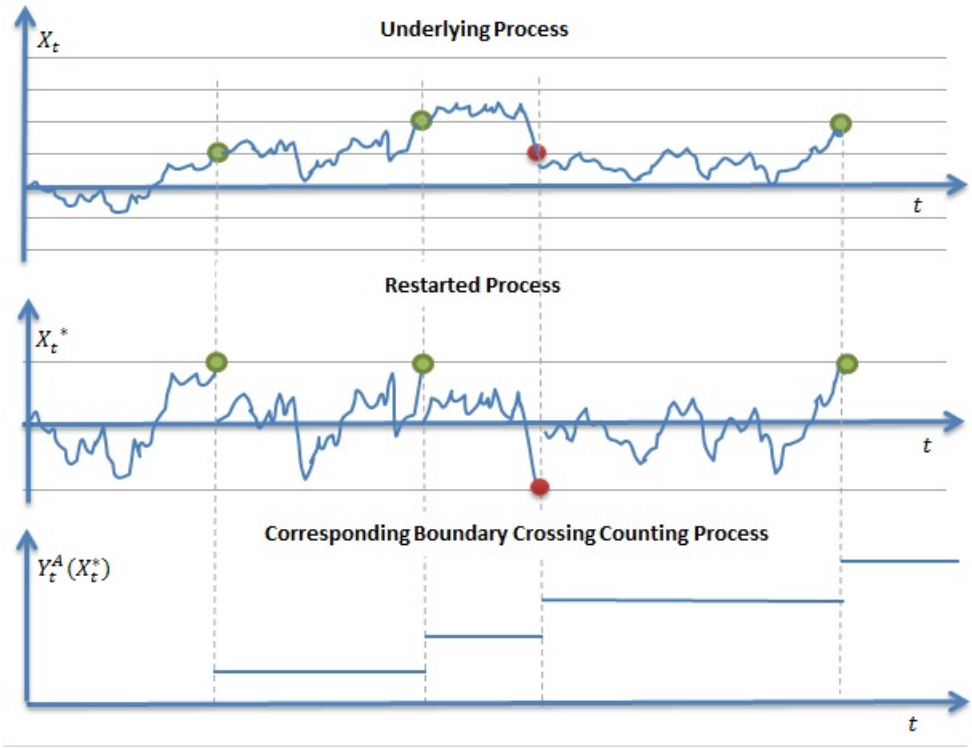


Figure 1.1: *An upper and lower crossing counting process. The original stochastic process, X_t is restarted upon each boundary crossing events. The counting process, $Y_t^A(X_t^*)$, counts the number of restarts.*

ratio test. We also find that based on Shiller's data, a total return index based on the S&P500 exhibits mean reverting behavior.

The BCC test has several desirable properties. Besides the usual favorable properties of nonparametric methods, our method is a non-asymptotic one. Therefore it does not suffer from asymptotic size-distortion. Naturally, the BCC test suffers from some minor drawbacks. The BCC distribution is a discrete one, hence similarly to Fisher's exact test, selecting the critical values for the usual significance levels is somewhat problematic and we have to make use of the closest available discrete value. Also, boundary selection at this stage is exogenous.

DOI: 10.14754/CEU.2016.01

The theory applied here builds heavily on the mathematical results related to boundary crossing events. There is no clear consensus on how to name the various concepts, therefore, let us briefly review the terminology. "First passage time" or "hitting time" is typically used in situation, where there is only one boundary. "Expected first passage time" describes the expected amount of time needed to reach that boundary. "First passage time distribution" aims to characterize the full distribution. The case of two boundaries is usually referred to as "first exit time" or "double-barrier hitting time" although the notation "first exit time" is also used to describe first passage time[1]. "Exit times" should not be confused with "first range time", as range is generally used to describe the difference between the maximum and the minimum value. In this thesis, we follow the terminology of Borodin and Salminen (2002) who use the term "first exit time" to describe the case of double boundaries.

The first wave of literature on this topic is by classical authors, like Bachelier (1900) or Kolmogoroff (1931) and was motivated by the gamblers ruin problem in finance and by repeated (independent) sampling problem in statistics.

The second wave of literature aimed to formalize these early results, along with some corresponding results in physics which is summarized by Feller (1971). Another general treatment is given by Karlin and Taylor (1981), who characterizes boundary crossing probabilities and expected first passage-times using certain functionals and the concept of scale function and speed function under fairly general assumptions. A less general approach focusing on the geometric Brownian motion is given by Karlin and Taylor (1998).

[1]This terminology is used for example in Wilmott (1998, p. 144).

A third wave of literature consists of articles partially motivated by pricing certain financial options (barrier-options). Lin (1998) for example, proposes to use the Gerber-Shiu technique, Gerber and Shiu (1994), along with the Laplace Transforms for calculating the first exit time (double-barrier hitting time) distributions. Linetsky (2004) proposes to use a spectral expansion approach for calculating hitting time distributions. The results for common continuous stochastic processes have been summarized in a handbook format by Borodin and Salminen (2002). Here, they also discuss the theory used in deriving these results, although actual proofs are generally not included. Valov (2009) proposes an even more general approach which connects the theory of boundary crossing events with the theory of integral equations.

In general, the theory of exit times and hitting times are much more developed for continuous processes than for discrete ones, which is why applying simulations when dealing with discrete data is fairly common. For example, Valenti et al. (2007) compare empirical and simulated hitting time distributions. More precisely, they find substantial deviation between the hitting time distributions derived from the Brownian motion, from the GARCH model and from the Heston model and the hitting time distribution obtained from actual data on security prices. Yet, they have not devised an exact test to quantify these deviations.

We use the BCC processes for statistical testing. Our test is basically a nonparametric specification test, or, in other words, a nonparametric method for model validation. One of the first articles in this relatively new field was written by Ait-Sahalia (1996), who compared the marginal density estimator with the nonparametric sample-based alternative. His method has certain drawbacks in finite samples, as shown by Pritsker (1998), notably, it requires

a relatively large sample. Consequently, researchers have developed several alternative tests.

A common method for these nonparametric tests is to devise and compare some measure for both the parametric and the nonparametric specifications. Naturally, rejection occurs when the discrepancy between the two measures is substantial. Papers essentially differ in how to measure the discrepancy. Transition density based comparisons have been used, for example, by Gao and King (2004) or Hong and Li (2005). Alternatively, a comparison based on the likelihood function has been proposed by Fan et al. (2001) who developed a generalized likelihood ratio test for this purpose. Anderson (1993) suggested a comparison based on the spectral densities. Recently, Song (2011) recommended a comparison based on the infinitesimal operator. These methods have been reviewed, for example, by Fan et al. (2005) and Zhao et al. (2008).

Our method differs from these nonparametric tests because we do not rely on asymptotic approximations. In this regard, our method is similar in spirit to the literature on exact statistics, such as the work of Fisher (1932)'s on binomial test and the method proposed by Dufour and Farhat (2001).

The first chapter is organized as follows. The second section introduces and characterizes BCC-processes and connects them to the existing literature on stochastic processes. The third section discusses estimation-related issues. In the fourth section, we devise a univariate unit root test which is based on boundary crossings. We finish the chapter with a financial application in which we analyze mean-reversion in total excess returns based on S&P500. The last section offers concluding thoughts.

DOI: 10.14754/CEU.2016.01

1.2 Theory of BCC Processes

This section is a non-technical discussion on the theory of BCC-processes. In particular, we do not prove the existence of the concepts we introduce; rather we provide references for readers interested in such existential proofs. The main innovation here is a recursive algorithm, which allows us to calculate the upper and lower crossing counting distribution, Y_t^A, from the first exit time distribution and the upper minus lower crossing distribution using Y_t^A and the upper crossing probabilities. The procedure allows us to obtain BCC distributions even in cases where only one realization of the underlying stochastic process is observed.

1.2.1 Concepts and Assumptions

The theory of first exit times is more developed for continuous-time processes; therefore first, we begin the discussion with the continuous case. The concepts used in the subsection are known in the literature on probability and stochastic processes.

Concepts

Definition 1 *Let X_t be some stochastic process with memory.*

A commonly used example for X_t is the Wiener process. The stochastic process is enclosed by some lower, L_t and upper U_t boundaries.

Definition 2 *Let L_t and U_t be two measurable functions, $L_t < U_t$.*

For example, $L_t = -1$ and $U_t = 1$.

Potentially, there are three kinds of boundaries[2]. The simplest case is the one of constant boundaries, where $L_t = L$ and $U_t = U$. Also, in certain applications, it may be useful to define boundaries as a function of physical time. Finally, boundaries may be stochastic as well.

Depending on the actual applications, boundaries may be chosen exogenously or endogenously, as a solution to the appropriate optimal (stochastic) control problem. For example, in statistical applications one may ask which boundary functions would maximize the statistical power of the Boundary Crossing Counting test. In this dissertation, we are not going to solve such problems under general settings because doing so would require a lengthy technical discussion and at this stage we do not see any significant additional benefits of going down this path.[3] Instead, we work with exogenously chosen constant boundaries denoted by L and U. Hence, our solutions are suboptimal and they lack the elegance of the optimal solutions. Nevertheless, they may prove to be useful in certain cases. Besides, we also carry out robustness exercises in order to evaluate the importance of the boundaries' role.

The restarted process is derived from the original stochastic processes and the double boundaries.

Definition 3 *Let X_t^0 be a fixed or a stochastic restarting value, $L < X_t^0 < U$.*

[2]Boundary classification may be found in Karlin and Taylor (1981, p. 234), where they differentiate between "regular", "absorbing", "natural" and "entrance" types. The type of boundary applied in our dissertation does not have a one to one correspondence to any of these cases: They could be called "restarting boundaries". If one must classify, restarting boundaries are attainable and regular boundaries, where the process is restarted upon boundary crossing events.

[3]Interested readers may find the description of these methods, for example, in the book of Kirk (2004).

Each time the stochastic process crosses the boundaries, its value is "almost immediately" reset to this restarting value, $X^*_{T^A+\epsilon} = X^0_{T^A+\epsilon}$*, where* T^A *is a boundary crossing moment and* ϵ *is an arbitrarily small positive real number. If no boundary crossing occurs, then the change of the restarted process equals to the change of the original process,* $X_{t2} - X_{t1} = X^*_{t2} - X^*_{t1}$*, where no boundary crossing moments exists in the* $[t1, t2]$ *interval.*

For example, if X_t is the Wiener process, $L_t = -1$ and $U_t = 1$ and the restarting value is zero, then X^*_T is a restarted Wiener process which is a Wiener process almost everywhere except in boundary crossing moments.

The notation "almost immediately" assumes that the process is reset to its initial value exogenously. Therefore, we know that at the first exit time T^A, the stochastic process is at $X^*_{T^A} \notin (L, U)$ and we know that at time $X^*_{T^A+\epsilon}$, the process is at $X^0_{T^A+\epsilon}$, yet we do not deal with what happens in the time-period $(T^A, T^A + \epsilon)$.

How should the BCC distribution be calculated? We can distinguish between two cases. First, if we have a parametric process with known parameters, then we can simulate a large number of sample paths, count the number of boundary crossing events at each path, and finally approximate the BCC distribution by a histogram built from the number of events observed in each sample path. The other, more interesting, case is when we only have one realization of the underlying stochastic process. In this case, we can still obtain the BCC distribution using the concept of first exit time.

Definition 4 *Let first exit time,* T^A *be defined as the first time-moment in which some stochastic process first crosses either of the boundaries.*

$$T^A = \begin{cases} inf(t : X^*_t \notin (L, U) \text{ if } t \text{ is finite} \\ \infty \text{ otherwise} \end{cases} \tag{1.1}$$

Note that the literature typically uses the T^{ab} notation, while we applied the T^A convention for the sake of consistency. It is important that we take the first occasion on which the boundary is crossed. Take[4] a standard Wiener process W_t, $L_t = -1$ and $U_t = 1$ as an example. If $W_t = 1$ at some time t, then, by the Blumenthal 0-1 law, W_t crosses the upper boundary infinitely many times in an arbitrarily small neighborhood of t. From all of these crossing events, we choose the first one which makes the boundary crossing events well defined. For the same reason, we also assume that the restarting value is strictly larger than the lower boundary and strictly smaller than the upper boundary.

Definition 5 *The cumulative first exit time distribution describes the probability that the first exit time is smaller or equal than* t*, that is* $FET(t) = P(T^A \leq t)$*. As for the first exit time distribution,* $\int_0^t fet(t)dt = FET(t)$*.*

The distribution function's dependence on the boundaries and on the restarted process is suppressed for ease of notation. As an example, the first exit time distribution for Brownian motion has the shape of the Inverse-Gaussian distribution as described by Feller (1971, p. 52) or by Lin (1998). Also, this distribution is shown in Figure 1.3. We postpone the somewhat technical discussion on how to calculate the first exit time distribution until the end of this section.

Finally, in order to characterize the upper minus lower crossing counting distribution, we also need to introduce the following concept.

Definition 6 *Let upper boundary crossing probability,* p*, be defined as the probability that the stochastic process reaches the upper boundary before*

[4]The review of Robert Lieli has proposed this insightful example.

hitting the lower one.

$$p = P(X^*_{T^A} = U) \tag{1.2}$$

To conclude, we rely on the concept of first exit time and on the concept of upper crossing probability. Both concepts are known in the literature on probability and on stochastic processes. Our contribution is to use these concepts to characterize repeated boundary crossing behavior.

Assumptions

Assumption 1 *Let T^A be positive and finite.*

In the standard literature on stochastic processes, this assumption is frequently a theorem derived from more elementary assumptions. The finiteness of the first exit time is a well-known property for martingales, as explained, for example, by Medvegyev (2007). The typical proof for non-martingales is to convert the process to a martingale, as shown, for example by Karlin and Taylor (1998). The non-zero property of the first exit time is only problematic if the limits of the boundaries are equal to the initial value of the process, a case which is not dealt with here. A similar problem for hitting times is discussed by Valov (2009). In any case, this is a non-elementary assumption, it imposes restrictions on the underlying stochastic process and the boundary structure, and not all stochastic processes and corresponding boundaries will satisfy this assumption.

We also need the following simplifying assumption.

Assumption 2 *Let us assume that the boundary crossing counting distributions are unconditional. Also, let the first exit time distribution as well as the upper-boundary crossing probability also be unconditional.*

The term "unconditional" is adopted from the user guide of the Matlabs Econometric toolbox.

The following examples will elaborate further on Assumption 2. Let us first take the Brownian motion with drift as an example. The first exit time distribution in general, as explained by Borodin and Salminen (2002, p. 640), is as follows:

$$fet^c(t) \approx cc_t(\frac{U+L-2X_t^*}{2}, \frac{U-L}{2}) \tag{1.3}$$

where $fet^c(.)$ is the conditional first exit time distribution, cc_t is defined in formula (1.22). Note that $fet^c(.)$ depends on X_t. Assumption 2 restricts X_t to be the restarting value. Hence, the unconditional first exit time distribution for the Brownian motion with drift is as follows.

$$fet(t) \approx cc_t(\frac{U+L-2X_{T^A+\epsilon}^0}{2}, \frac{U-L}{2}) \tag{1.4}$$

where $fet(.)$ is the unconditional first exit time distribution, T^A is a boundary crossing moment and $T^A + \epsilon < t$.

Moreover, let us consider the Brownian motion without drift as another example. As explained by Feller (1971), the upper crossing probability is as follows:

$$P^c(X_{T^A+1}^* = U) = \frac{X_t^* - L}{U - L} \tag{1.5}$$

where $P^c(.)$ is the conditional upper crossing probability, T^A and $T^A + 1$ are boundary crossing moments and $T^A + \epsilon < t < T^A + 1$. Essentially, in this case, the upper crossing probability only depends on the relative distance of X_t from the boundaries. Assumption 2 restricts X_t^* to be the restarting value.

$$P(X_{T^A+1}^* = U) = \frac{X_{T^A+\epsilon}^0 - L}{U - L} \tag{1.6}$$

For example, if the restarting value, $X^0_{T^A+\epsilon}$, is zero, the stochastic process is currently at 50, that is $X_t = 50$, $U = 100$, $L - 100$, then the conditional upper crossing probability, conditioned on the current value of the stochastic process, is 0.75. The unconditional probability, which is based on the restarted value, is 0.5.

1.2.2 Upper and Lower Crossing Counting

In this subsection, we discuss how to calculate the upper and lower crossing counting distribution using the first exit time distribution. This is a new contribution to the literature.

The probability that no boundary crossing event occurs until time T is simply the probability that the first boundary crossing event occurs at a later time:

$$p^A_0(T) = 1 - \int_T^\infty fet(t)\,\mathrm{d}t. \tag{1.7}$$

It can be shown by induction that the probability of exactly k boundary crossing events occurs until time T can be calculated as:

$$p^A_k(T) = \int_0^T fet(t) \times p^A_{k-1}(T-t)\,\mathrm{d}t. \tag{1.8}$$

Now, if the first exit time is known, then $p^A_0(T)$ can be calculated directly from the first exit time distributions, $p^A_1(T)$ can be calculated from $p^A_0(T)$ and $fet(.)$, and so on. Therefore, by applying (1.8) recursively, we can characterize the BCC distribution completely, using the first exit time distributions. Carrying out this recursion analytically in continuous time is challenging even for simple stochastic processes. But the procedure is relatively straightforward in discrete time to be discussed next.

Restarting the process in discrete time is problematic as it may happen that the process crosses two boundaries between two observations. The following simplifying assumption excludes this possibility.

Assumption 3 *The probability that boundaries were crossed twice within one time interval is negligibly small. Assuming that boundary crossing has occurred between t and $t+1$, $p(L < X_t < U$ and $X_{t+1} > 2U$ or $X_{t+1} < 2L) = \epsilon'$, where ϵ' is an arbitrarily small positive real number.*

This assumption can be justified in a number of different ways. If the underlying process is continuous, then it implies that the sampling is sufficiently frequent so that the probability of double boundary crossing is negligibly small. Thus, in case of continuous processes, this is really an issue related to sampling frequencies. Also, the underlying process may be discontinuous as long as the size of the jumps is restricted to be less than the size of the boundaries.

In any case, if we have T observations, then the maximum number of boundary crossing events is T. In this case we can characterize the boundary crossing counting distribution with the help of the following P^A matrix:

$$P^A = \begin{pmatrix} p_0^A(1) & p_1^A(1) & \cdots & p_n^A(1) \\ p_0^A(2) & p_1^A(2) & \cdots & p_n^A(2) \\ \vdots & \vdots & \ddots & \vdots \\ p_0^A(T) & p_1^A(T) & \cdots & p_n^A(T) \end{pmatrix} \tag{1.9}$$

where the subscript indicates the number of boundary crossing events, for example $p_i^A(t)$ indicates that until period t, exactly i boundary crossing events have occurred. The rows of P^A describe the BCC distribution at a given moment of time, while the columns of P^A are also meaningful: they

describe the probability that exactly $1, 2, ..., t, ..., T$ period is needed for some i boundary crossing events to occur.

The first column can be calculated from the first exit time distribution using Equation (1.7). Any other column, j, may be calculated recursively using:

$$P^A(:, j) = F_2(j) \times F_1 \tag{1.10}$$

In this expression, F_1 is a static matrix composed of the first exit time distributions while $F_2(j)$ can be expressed recursively, using probabilities obtained in the previous steps. More precisely,

$$F_1 = [FET(1), (FET(2) - FET(1)), ..., (FET(T) - FET(T-1))]', \tag{1.11}$$

while

$$F_2(j) = \begin{pmatrix} p^A_{j-1}(0) & p^A_{j-1}(-1) & \cdots & p^A_{j-1}(-T) \\ p^A_{j-1}(1) & p^A_{j-1}(0) & \cdots & p^A_{j-1}(-T+1) \\ \vdots & \vdots & \ddots & \vdots \\ p^A_{j-1}(T-1) & p^A_{j-1}(T-2) & \cdots & p^A_{j-1}(0) \end{pmatrix} \tag{1.12}$$

A general term of $F_2(j)$ is given by $p^A_{j-1}(t-(t-r+c))$ where t is the number of observations, r is the row number and c is the column number and the following conventions are respected.

- if $j - 1 > t - (t - r + c)$ then $p^A_{j-1}(t - (t - r + c)) = 0$ since the number of boundary crossing events between two observations is at most one.
- if $j - 1 = t - (t - r + c) = 0$ then $p^A_{j-1}(t - (t - r + c)) = 1$

The advantage of this matrix-formulation over a brute-force combinatorial calculation is the reduction in the calculation-complexity: while the brute-force combinatorial method would require T^3 steps, the matrix-formulation described above reduces the required number of steps to T^2. Thus, the algorithm becomes slow, yet feasible for not very large sample sizes.

1.2.3 Upper Minus Lower Crossing Counting

Next, we show that the upper minus lower crossing counting distribution, defined on page 4 of this chapter, denoted by Y_t^D, can be obtained from the upper boundary crossing probabilities and the upper and lower crossing distribution. This is also a new contribution to the literature.

The idea is to use a tree-based approach frequently applied in option pricing. More specifically, Y_t^D can be represented in a "random-time binomial tree"[5]. The term "random time" is appropriate because the time needed to move from one state to the next is random.

In tree-based models, stochastic processes are modeled with discrete states. The time to move from one state to the next is typically non-stochastic. Compared to classical binomial trees where the stochastic variable may either go up or down, here we allow for three options: the stochastic process may either go up, go down, or remain in that particular state.

Intuitively, Y_t^D depends on two factors. On the one hand, we need to know how many boundary crossing events occur. On the other hand, we also need to know the probability of moving up. A node of the tree $B_t(i, j)$ can be described by the number of boundary crossing events, i is the number of upper crossings, j is the number of lower crossings. Note that the grid itself also changes dynamically as time changes. Figure 1.2 is essentially a snapshot taken at a given point of time.

Next, the grid is characterized with a V_t matrix. Note that some of

[5]Such random-time binomial tree could also be represented by a classical trinomial tree. From an IT point of view, trinomial tree would be a less efficient representation in a sense that the number of redundant representation leading to the same outcome would be higher.

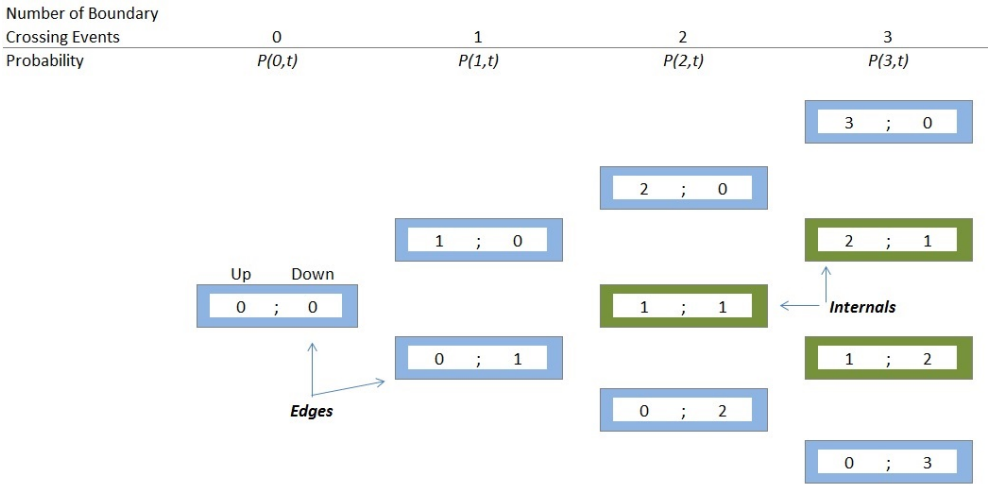


Figure 1.2: *Random-time binomial tree. In contrast to a classical binomial tree, here, the stochastic process may not need to terminate at the last column but it can terminate anywhere in the grid. The boundary crossing probabilities characterize the horizontal dimension while the upper crossing probability characterizes the vertical one.*

the states in this matrix have zero probability which does not influence the results. Characterizing the grid can be done in two steps. The vertical location, V_t, can simply be described by the number of boundary crossings.

$$V_t = \begin{pmatrix} P_t^A(0) & P_t^A(1) & \cdots & P_t^A(t) \\ P_t^A(0) & P_t^A(1) & \cdots & P_t^A(t) \\ \vdots & \vdots & \ddots & \vdots \\ P_t^A(0) & P_t^A(1) & \cdots & P_t^A(t) \end{pmatrix} \tag{1.13}$$

where $P_t^A(k)$ is the probability to observe exactly k boundary crossing events until time t. Conditioned on the vertical location, the horizonal location, $H_t = [h(0), h(1)...h(j)...h(t)]$, can simply be described using upper boundary crossing probabilities. For the simplest case of constant[6] upper

[6]Mean-reversion as well as autocorrelation for example would imply non-constant

crossing probability, the constant-probability binomial distribution can be used. For j number of boundary crossings,

$$h(j) = \begin{pmatrix} 0 \\ \vdots \\ \binom{j}{j} p^j \\ \binom{j}{j-1} p^{j-1} \times (1-p) \\ \vdots \\ \binom{j}{0} (1-p)^j \\ \vdots \\ 0 \end{pmatrix} \tag{1.14}$$

, where p is the upper crossing probability. Since the vertical and the horizontal location is independent, the grid can be characterized as:

$$B_t = V_t \odot H_t, \tag{1.15}$$

where $\odot$ indicates element by element multiplication or the Hadamard product. Obtaining the distribution of Y_t^D from B_t simply involves collecting terms where $i - j$ are equal.

$$p(Y_t^D = k) = \sum_{l=k}^{n} B_t(l, l-k) \tag{1.16}$$

Overall, Y_t^D can be calculated based on Y_t^A and p, that is based on the upper and lower crossing counting distribution and on the upper crossing probability.

1.2.4 Further Details on Analytical Solutions

This subsection briefly summarizes the relevant literature on the analytical methods for calculating the first exit time distribution and the upper crossing

upper-crossing probability. In this case, the binomial distribution may be replaced by an appropriately chosen recursive algorithm.

probability.

The analytical solution for the first exit time distribution can largely be facilitated by assuming constant boundaries. The cases of non-constant or stochastic boundaries are somewhat more challenging and to our knowledge, these cases have not been solved in the literature. One possible solution may be to adopt the techniques which have been used to solve for the first passage time distribution.[7]

A potential procedure for deriving the first exit time distribution analytically is described as follows. First, by subtracting the expected value from the original stochastic process, we obtain a martingale. If the initial value of the martingale is known, then we can express the expected value of the martingale at the first exit time. This is so because according to the Optimal Sampling Theorem, or Doobs lemma, the expected value of a martingale conditioned on the information available at time zero is equal to its initial value at any time-period. By rearranging this expected value, we can obtain the Laplace transforms of the first exit times. The probability distribution functions can then be derived by inverting these Laplace transforms.

The formula for first exit time distribution functions can be found in the handbook of Borodin and Salminen (2002) for several processes, such as for the Brownian motion with no drift (p. 212), for Brownian motion with drift, (p.309), for Bessel process of order 0.5 (p.309) and finally for geometric Brownian motion (p.627). In page 109, the authors also discuss briefly a general method for deriving these distributions. More detailed proof for the geometric Brownian motion is discussed in Lin (1998).

Let us finish this subsection by briefly discussing how to calculate the

[7]These methods are discussed in Redner (2001) and in Valov (2009).

upper crossing probability. For nonparametric cases, the probability can be estimated similarly to how the success rate is estimated in the case of binomial distributions. On the other hand, for many frequently used[8], parametric processes, the boundary crossing probability can be expressed using scale functions, as explained for example in Karlin and Taylor (1981).

$$p = \frac{S(X_0) - S(L)}{S(U) - S(L)} \tag{1.17}$$

where

$$S(x) = \exp(-\int^x \frac{2\mu(y)}{\sigma^2(y)}dy) = \exp(\times - S(X)) \tag{1.18}$$

is the scale function, $\mu(.)$ and $\sigma^2(.)$ are the infinitesimal moments and finally $\int^x \frac{2\mu(y)}{\sigma^2(y)}dy = S(X)$, is the scale function. The lower limit of the integrals does not play a significant role and thus, it is omitted which is a typical convention in the corresponding literature. This equations essentially shows that once the process has been appropriately scaled, the probability of upper (or lower) boundary crossing depends only on the initial points relative distance from the lower and upper boundaries.

1.3 Estimating BCC Distributions

This section focuses on how to actually estimate BCC distributions. As the BCC distribution has not been introduced in the literature beforehand, we cannot compare our results against standard benchmarks. Instead, we calibrate our method by developing and comparing three different estimation procedures.

Throughout this section, we use T to indicate the lengths of a sample path and N to indicate the number of sample paths. Moreover, we introduce

[8] This formulation is true for Ito processes, for details see Karlin and Taylor (1981)

the count operator, #(.), which essentially counts the number of cases meeting certain criteria.

We primarily focus on Y_T^A as it is also fundamental to estimating Y_T^D. There are two main methods, the first one requires many realizations of the stochastic process, the second method can also be applied in cases where only one realization of the stochastic process is available.

1.3.1 Direct Estimation

The following fairly straightforward method uses the fact that the number of boundary crossing events can simply be counted in the data. The probability that we observe exactly k boundary crossing events is as follows:

$$P(Y_T^A = k) = \frac{\#(Y_T^A = k)}{N} \tag{1.19}$$

If the data generating process suggests that this is a smooth distribution, then it may be appropriate to apply some smoothing algorithm, for example fitting kernel density using inverse normal distribution. The advantage of the method of direct estimation is its simplicity. In particular, it does not require any specific assumption on the data generating process.

On the other hand, this method often produces only a single point estimate, as in actual time series data, we can typically observe only one realization of the stochastic process.

1.3.2 Estimation Using First Exit Time Distribution

Obtaining the BCC distribution in this case can be done in two steps. The first step is to derive or estimate the first exit time distribution. The next step is to apply the recursion described in the previous section. For analytically given stochastic processes, the method to derive the first exit

time distribution has already been discussed in the previous chapter. For nonparametric processes, the first exit time distribution can be obtained by the following counting estimation.

$$FET(t) - FET(t-1) = fet^{\#}(t) = \frac{\#(t-1 < T^A \leq t)}{\sum_{i=1}^{N} Y_i^A(T)}, \tag{1.20}$$

where $(.)^{\#}$ is the counting estimator, T^A represents a boundary crossing moment, $Y_i^A(T)$ is the number of boundary crossing events for cross-section i and N is the number of cross sections. If we are willing to assume that $fet(t)$ is a smooth function, then it may be appropriate to apply some kind of smoothing algorithm on $fet^{\#}(t)$, for example fitting a kernel density using the inverse normal distribution.

Once the first exit time distribution has been estimated, we can carry out recursions described in Equation (1.10). With the help of this second method, we can reconstruct the number of boundary crossing events even in case where there is only one sample path. Besides, it is relatively more accurate in case of rare events which is of course highly important for statistical specification testing. Intuitively, this method uses more information than the method of direct estimation as in this case, we not only consider the number of boundary crossing events, but also the timing of the crossings.

1.3.3 Calibrating Estimation Methods

The following subsection highlights a few practical problems related to the estimation of the BCC distribution. As explained in the review of Fan et al. (2005) and in the corresponding commentaries, one of the most important problems of nonparametric statistical testing is the issue of sampling biases which is analyzed first. Next, we analyze the effect of small sample size which

results in unregistered boundary crossing events and small sample bias. We conclude with some further issues.

Sampling issues

Sampling bias occurs when the null hypothesis is a continuous stochastic process that is to be tested on discretely sampled data. Let us illustrate this issue by comparing an analytical and a nonparametric estimation. In order to obtain an analytical solution, let us assume that X_t is generated by sampling from a continuous Wiener process, using some constant sampling frequency.

$$X_t = X_{t-1} + \mu, \tag{1.21}$$

where $\mu \sim N(0,1)$. The first exit time distribution in this case can be expressed using the following theta function as shown in Borodin and Salminen (2002, p. 640):

$$cc_t(v,z) \approx \sum_{k=-k^*}^{k=k^*} (-1)^k \frac{z+v+2k\times z}{\sqrt{2\times\pi}t^{\frac{3}{2}}} \exp\frac{-(z-v+2k\times z)^2}{2t}, v<z, \tag{1.22}$$

where v and z can be calculated from the boundaries and from the initial value while k describes the precision with which the calculation is carried out. Note that the approximation improves as $k^* \to \infty$. In our calculations, we have used $k^* = 1000$ although $k^* = 50$ already provides reasonable approximation. The first exit time for upper or lower crossing can be calculated as follows:

$$fet^A(t) \approx cc_t(\frac{U+L-2X_t^0}{2}, \frac{U-L}{2}) \tag{1.23}$$

Note that $U + L - 2X_0 < U + L - 2L = U - L$, therefore the condition in formula (1.22) is fulfilled.

As for the nonparametric approach, let us simulate 1000 sample paths, each having the length of 2520 observations. Estimations were done using equations (1.20), and the counting estimator was smoothed using kernel density estimate.

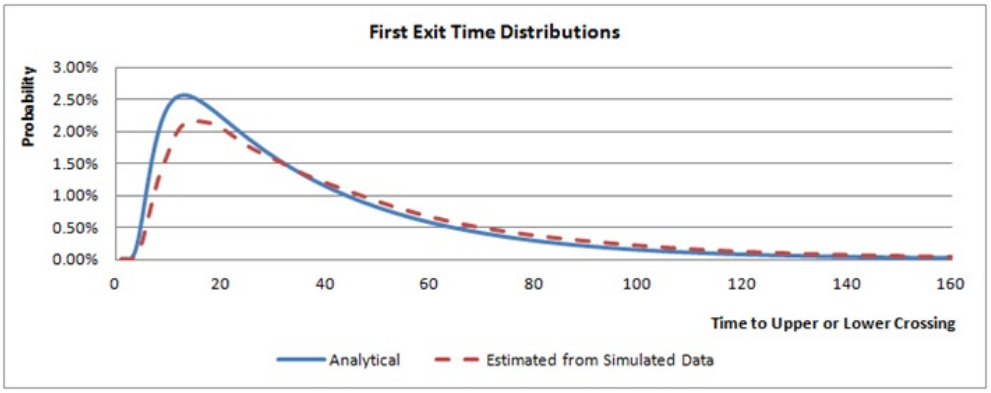


Figure 1.3: *Analytical and nonparametric first exit time distribution of the standard Brownian motion for boundaries L = -6 and U =6. The two distributions differ because of the sampling bias.*

Figure 1.3 reveals that the nonparametric and the analytical first exit time distributions are not equal, the difference appears to be significant.

The nonparametric first exit time distribution tends to underestimate the probability of shorter exit times and overestimate the probability of longer exit times. The estimation does not improve as the sample size increases therefore this is not a small-sample bias. What causes this difference? It is due to sampling: the analytical solution treats time as a continuous variable, while the other solution works with discretely sampled observations. In the latter case, the value of the random variable between the two observations is unknown. It may very well be possible that the random variable crosses the boundaries between the two observations, yet at the moment of observation, the random variable is no longer outside the boundaries. Such unobserved

crossings are not registered as boundary crossings in the case of discrete sampling.

The upper and lower crossing counting distributions have the shape of a normal distribution. It is a discrete distribution[9]. Since the BCC distribution is based on the first exit time distribution, it inherits its sampling bias.

The solution to the problem of sampling bias is to take into account both the minimum and the maximum values. This, for example, can easily be done in the case of financial data when not only closing prices, but when the minimum and the maximum prices are also recorded. In case of simulated data, minimum and maximum prices can be obtained by increasing the sampling frequency and taking into account certain extreme values.[10]. That is to say, instead of generating 2 520 observations using mean zero and unit standard deviations, we simulate 252 000 observations using mean zero and standard deviation of 1/100 from which we not only select the closing values, but the minimum and the maximum values as well.

As shown in Figure 1.4, the first simulated BCC distribution underestimates the number of boundary crossing events. This is in-line with the explanation given above, namely that the simulated distribution only counts those boundary crossing events where the random variable remains outside of the boundaries at the moment of sampling. On the other hand, the bias in the second BCC distribution is largely reduced. Therefore, sampling

[9]Still, we often use solid lines in the diagrams, so that readers can differentiate between the various distributions.

[10]We acknowledge that this method underestimates the maximum value and overestimates the minimum value. There are more sophisticated methods for generating extreme prices as explained, for example, in Mcleish (2002). As here we use extreme values for illustrative purpose only, we would not substantially benefit from using more sophisticated models.

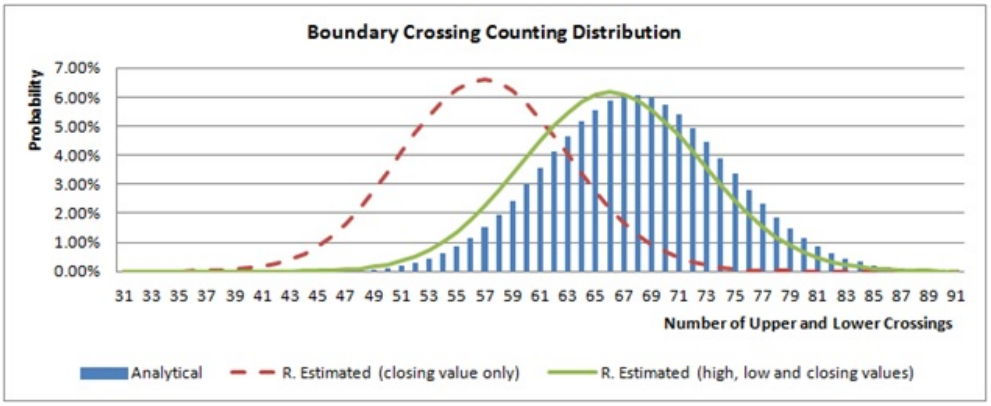


Figure 1.4: *Analytical and recursively estimated nonparametric upper and lower crossing counting distribution for boundaries L = -6 and U =6. The sampling bias can be reduced by taking into account minimum and maximum values.*

bias can be substantially reduced by taking into account the minimum and the maximum values as well.

To conclude, the BCC distribution is sensitive to sampling. Therefore, potentially, the difference in the number of boundary crossing events may be due to difference in the data generating process or to the difference in the sampling frequency.

Small sample bias

Small sample bias is due to the fact that the counting estimator does not take into account the evolution of the stochastic process after the last boundary crossing event. In other words, we do not observe where and when the stochastic process crosses the boundaries after it has been restarted for the last time. Naturally, as sample size increases, the role of this last boundary crossing observation diminishes.

The comparison of the analytical and the nonparametric distribution is useful in quantifying this bias. For this experiment, we have simulated 5000 sample paths with sample lengths of 252 observations, which is approximately one year in the financial price-series. We have simulated minimum and maximum values as well. As each sample path results in one observation, overall we have 5000 observations. The direct estimation for the BCC distribution is essentially a normalized, non-smoothed histogram based on these 5000 observations, which is compared to the analytical BCC distribution.

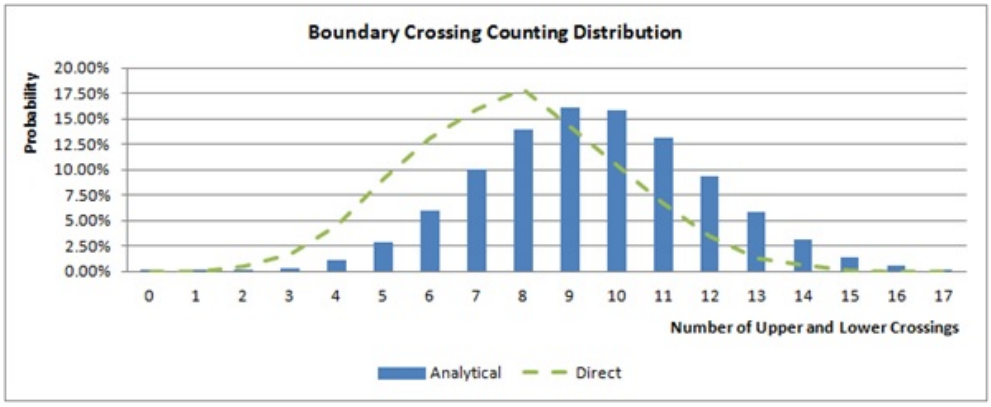


Figure 1.5: *Analytical and direct upper and lower crossing counting distribution for 252 observation for the standard Brownian motion in case of constant boundaries set to L = -5 and U = 5. The two distributions differ due to small sample bias.*

Naturally, as sample size increases, the role of this last boundary crossing observation diminishes. In data having lengths of 1000 observations, for boundaries placed at five standard deviations, the difference between the Direct and the Analytical distribution in case of the BM(0,1) process is almost completely eliminated.

Selecting from a discrete distribution

The next issue is related with selecting critical values from a discrete distribution. Often, test-statistics have continuous distributions. Therefore, selecting the desired critical values resulting in the usual probability for type-1 errors is relatively straightforward.

On the other hand, the BCC distribution, analogously for example to the binomial distribution which is being used in the Fisher's exact test, is discrete. Thus, we are unable to find the exact critical values matching the desired type-1 error probabilities.

As a convention, we chose the critical values in a way that the resulting probability for type-1 error lies as close to the desired value as possible. Hence, although the BCC test is a non-asymptotic test and does not suffer from asymptotic size distortion, yet the actual size and the nominal size is not necessarily equal.

Boundary selection

Throughout the dissertation, we assume that the data-generating process is a continuous process which is being sampled as described in Equation (1.24).

$$\Delta X_t = X_t - X_{t-c}, \tag{1.24}$$

where c is the sampling frequency. In most econometric studies, the following (often implicit) assumptions are applied when representing and analyzing the DGP:

- First of all, c is constant

- It is exogenously given or chosen.
- Typically $c = 1$, that is the change in the state variable is measured as $X_t - X_{t-1}$.

How do we chose c? Most of the time, it is set to be one in the unit in which the data is published. But sometimes, based on theoretical reasoning, c is exogenously chosen to be larger than one as on page 186. in the book of Shiller (2005). There, the dataset has monthly frequency but the sampling frequency is exogenously chosen to be 120 months, that is $c = 120$.

It is also important to mention that the number of data points available under such sampling frequency is very limited which is also mentioned by the author: *"The relation between price-earnings ratios and subsequent returns appears to be moderately strong, though there are questions about its statistical significance since there are fewer than twelve non-overlapping ten-year intervals in the 115 years worth of data."* The choice of the sampling frequency drives the number of data points available for inference just as the choice on boundaries limit the number of boundary crossing events in our approach.

Finally, it worth mentioning that the frequency using which the data is published limits the choice of c. Monthly data cannot be used to analyze how daily changes behave. Similarly, in our case, the sampling frequency of the data limits the choice on boundaries.

The main difference between our study and typical econometric studies is that in our dissertation, we represent and analyze the data-generating process using a random sampling frequency. This random frequency is driven by exogenously chosen boundaries. The choice on the boundaries is limited by the frequency of the data. Hence in these aspects, our approach does not improve the standard method. Table 1.1 below summarizes the discussion on

the role of boundary selection.

	Sampling mechanism	
	Time-dependent	**State-dependent, boundary based**
Change in the state variable	Flexible	Limited, either U or L
Sampling frequency	Constant, typically $c = 1$	Flexible, described by the first exit time distribution
	Exogenous, driven by how the data is published or by economic theory or by the interest of the researcher	

Table 1.1: *Time-dependent and state-dependent, boundary based sampling.*

To conclude, the choice of boundaries in our paper, similar to the decision on the sampling frequency in many econometric studies, is exogenous. Therefore, it should be based on the research question. However, as explained in the sampling subsection above, the choice on the boundaries is very much limited by the sampling frequency of the data.

1.4 Univariate Unit Root Testing

1.4.1 Baseline Model

We begin our analysis by demonstrating how to apply a boundary crossing based method for univariate unit-root testing, which is an important question in statistics as well as in economics. Consequently, authors, starting perhaps from Dickey and Fuller (1981) and Phillips and Perron (1988), have devised a large number of tests for identifying unit-roots. As reviewing these tests would require a separate article and this has already been done, for example

by Phillips and Xiao (1998), we do not attempt to provide a comprehensive review here.

Instead we treat the most simple (and hence typically unrealistic) case of a single time series with i.i.d. errors in detail. Our goal here is to demonstrate how a new statistical tool operates in an environment that is familiar to most researchers.

The baseline model is as follows.

$$x_t = \mu + \alpha x_{t-1} + u_t, \tag{1.25}$$

where $\alpha = 1$ or $\alpha = 0.9$, which is a typical choice in simulation studies as in Maddala and Wu (1999). For simplicity, we assume that $\mu = 0$ and u_t is independent but not necessarily identically or normally distributed. Finally, let us assume that x_t starts from minus infinity. The null hypothesis is as follows.

$$H_0 : \alpha = 1 \tag{1.26}$$

The alternative hypothesis is as follows.

$$H_1 : \alpha < 1 \tag{1.27}$$

The alternative hypothesis typically assumes that $\alpha < 1$ while here, we could test for explosive null-hypothesis as well. Nevertheless, we chose the non-explosive case so as to be able to compare our results to common unit-root procedures.

1.4.2 Testing Procedure

We aim to test for unit roots using boundary crossing events. Boundaries are chosen to be symmetric, $L = -U$, and the restarting value is zero for all t,

that is $X_t^0 = 0$. Finally, let the restarted process be defined over $\widetilde{X}_t = X_t - X_0$ for ease of notation.

Under these conditions, $Y_{k-1}^D > 0$ implies that $\widetilde{X}_t$ moves in a positive range in between the two boundary crossing events. Also, if $Y_{k-1}^D < 0$, then $\widetilde{X}_t$ moves in a negative range. Finally, if $Y_{k-1}^D = 0$, then $\widetilde{X}_t$ fluctuates around zero.

Furthermore, let Z_k describe the k^{th} boundary crossing event in a way that $Z_k = 1$ in case of an upper crossing and $Z_k = -1$ is case of lower crossing. We aim to exploit the relationship between Y_{k-1}^D and Z_k.

Under the null hypothesis, $\delta x_t = \mu + u_t$. Consequently, the following upper-crossing probabilities are equal:

$$H_0 : \underbrace{p(Z_k = -1 | Y_{k-1}^D < 0)}_{p_{11}} = \underbrace{p(Z_k = -1 | 0 < Y_{k-1}^D)}_{p_{12}} \tag{1.28}$$

Likewise, the lower crossing probabilities below are also equal:

$$H_0 : \underbrace{p(Z_k = 1 | Y_{k-1}^D < 0)}_{p_{21}} = \underbrace{p(Z_k = 1 | 0 < Y_{k-1}^D)}_{p_{22}} \tag{1.29}$$

By combining Equation (1.31) and Equation (1.32), we obtain the following equality under the null:

$$H_0 : p_{11} + p_{22} = p_{12} + p_{21} \tag{1.30}$$

Under the stationary alternative hypothesis, the stochastic process has the tendency to return to the mean. Since $(\alpha - 1) < 0$, a lower crossing event is more likely in case $Y_{k-1}^D > 0$, than in case $Y_{k-1}^D < 0$.

$$H_1 : \underbrace{p(Z_k = -1 | Y_{k-1}^D < 0)}_{p_{11}} < \underbrace{p(Z_k = -1 | 0 < Y_{k-1}^D)}_{p_{12}} \tag{1.31}$$

Also, an upper crossing event is more likely in case $Y^D_{k-1} < 0$, than in case $Y^D_{k-1} > 0$.

$$H_1 : \underbrace{p(Z_k = 1|Y^D_{k-1} < 0)}_{p_{21}} > \underbrace{p(Z_k = 1|0 < Y^D_{k-1})}_{p_{22}} \tag{1.32}$$

Consequently, the alternative hypothesis can be described as

$$H_1 : p_{11} + p_{22} < p_{12} + p_{21} \tag{1.33}$$

In the data, we can observe five kinds of events, which are summarized in Table 1.2.

		Cumulative Upper minus Lower Crossing		
		$Y^D_{k-1} < 0$	$Y^D_{k-1} = 0$	$0 < Y^D_{k-1}$
Next BC Event	$Z_k = -1$	$E_{11} + 0.25$ (Divergence)	E_{00} (Non Informative)	$E_{12} + 0.25$ (Convergence)
	$Z_k = 1$	$E_{21} + 0.25$ (Convergence)		$E_{22} + 0.25$ (Divergence)

Table 1.2: *Contingency table based on boundary crossing events.* E_{jk} *indicates the number of events observed in the data. Note that 0.25 is added to each cells for technical reasons to avoid any division with zero.*

We essentially differentiate between three cases. First, in case of events E_{11} or E_{22}, the stochastic process drifts further away from the origin, in other words, it diverges. Also, in the case of events E_{12} and E_{21}, it converges back towards the origin. Finally, in the case of events E_{00}, the stochastic process is close to the origin, thus these boundary crossing events are considered to be noninformative in this regard and hence, they are not taken into account.

From what has been noted above, the right hand side of Equation (1.30) expresses the convergence probabilities, $p_c = p_{12}+p_{21}$, which can be estimated

as follows:

$$p_c^{\#} = \frac{E_{12} + E_{21} + \frac{1}{2}}{E_{11} + E_{12} + E_{21} + E_{22} + 1} = \frac{B_c}{B_T}, \tag{1.34}$$

where $(.)^{\#}$ is the counting estimator, B_c is the number of convergence events and $B_T > 0$ is the total number of informative boundary crossing events.

To conclude, under the null hypothesis, the convergence probability is 0.5. Under the stationary alternative, the convergence probability is greater than 0.5. Note that we could also analyze the explosive alternative hypothesis, which would imply that the convergence probability is less than 0.5 as well as the joint stationary or explosive alternatives, which would imply that $p_c^{\#} \neq 0.5$, but these cases are not discussed due to space constraints.

We continue by discussing how to test the $p_c^{\#} = 0.5$ hypothesis. Essentially, the test statistics can be obtained by a quasi-binomial distribution as each boundary crossing event can be interpreted as a Bernoulli trial which takes the value of one upon convergence event and the value of zero upon divergence event.

The number of Bernoulli trials are $B_T = E_{11} + E_{12} + E_{21} + E_{22} + 1$, the number of successful trials are $B_c = E_{12} + E_{21} + 0.5$ and the success probability is 0.5. The difference between our case and the pure binomial distribution is that here, the number of trials is stochastic. This can, however, be easily accounted for as the test distribution can be conditioned on the realized number of trials: the resulting conditional distribution is a binomial one,

$$(B_c | B_T) \overset{H_0}{\sim} Bin(B_c, 0.5 | B_T), \tag{1.35}$$

where $Bin(.)$ denotes the binomial distribution. Since the stationary alternative states that $p_c^{\#} > 0.5$ the test is one sided.

1.4.3 Monte Carlo Analysis

In this subsection, we carry out two Monte Carlo studies. The first analyzes the role of boundary selection for the BCC-test. The second compares the BCC-test to other, commonly used unit root tests. The data-generating process is as described by Equation (1.25). In both studies, we consider two different models. The first is the case of unit root, the second is the stationary case with $\alpha_i = 0.9$. Also, μ is assumed to be zero for simplicity.

The design of the first Monte Carlo study is as follows. We simulate 2000 sample paths, each consisting of either 100, 500 or 1000 observations. We consider six different set of boundaries, measured in the standard deviations of Δx_t, ranging from $+/-1$ to $+/-6$. After counting the number of boundary crossing events, we obtain the p–values from the corresponding binomial distribution. We accept the null hypothesis if the p–value is larger than 5%. Rejection frequencies are calculated as the ratio of the number of cases when the null hypothesis is rejected and the total number of simulated sample paths. Hence, the first part of the table shows the actual size of the test, while the second part shows the power.

Table 1.3 reveals that as sample size increases the ideal boundaries widen. For small sample sizes, wider boundaries are not practical since such setup does not generate enough boundary crossing events for inference. Hence, neither the null, nor the alternative hypothesis can be rejected. For larger sample sizes, having enough boundary crossing event is less of an issue and hence, in this case, we can focus on having more informative events.

Characterizing the optimal boundaries analytically is beyond the scope of this paper, but clearly, the goal, as far as possible, is to minimize size distortion while maximiz the power of the test. Here, we settle for the

			Probability of Rejecting the Null Hypothesis					
DGP			Lower boundaries at the standard deviation of Δx_{1t}					
α	N	T	[-1,1]	[-2,2]	[-3,3]	[-4,4]	[-5,5]	[-6,6]
1	1	100	0.0780	0.0490	0.0280	0.0060	0.0000	0.0000
1	1	500	0.1040	0.0600	0.0550	0.0500	0.0330	0.0310
1	1	1000	0.0920	0.0480	0.0410	0.0410	0.0430	0.0460
0.9	1	100	0.2600	0.1730	0.0280	0.0000	0.0000	0.0000
0.9	1	500	0.6530	0.6270	0.7170	0.7980	0.5620	0.1420
0.9	1	1000	0.6750	0.6570	0.7450	0.8390	0.8860	0.6520

Table 1.3: *Boundary selection for univariate unit root test. The table shows the rejection frequencies. The nominal significance level is 5%. Rejection occurs when the p-value is less than 0.05.*

following heuristic rule.

$$U_i = \begin{cases} \sigma \text{ if } T < 100 \\ (1 + \frac{(T-100)}{225}) \times \hat{\sigma} \text{ otherwise,} \end{cases} \tag{1.36}$$

where T is the number of observations and $\hat{\sigma}$ is the sample standard deviation. The test provides reasonable differentiating power while, at the same time, the empirical size is close to the nominal one. In the remaining part of this chapter, we use this heuristic rule for setting up the boundaries.

We continue with the comparative Monte Carlo study. The BCC test is compared to several parametric unit-root tests. In particular, we compare the BCC test to the Augmented Dickey-Fuller test by Dickey and Fuller (1981), further referred to as ADF test, to the Phillips-Perron test by Phillips and Perron (1988), further referred to as PP test and finally to the variance ratio test by Lo and MacKinlay (1988) further referred to as VR test. We implement these tests using the corresponding build-in Matlab functions. As

for the BCC-test, we select boundaries according to the rule (1.36).

The design of the experiment is as follows. We simulate 2000 sample paths, each consisting of either 50, 100 or 200 observations. As for the error term, u_t is assumed to follow a t-distribution whose parameters match the log-returns of the S&P500. The table show the rejection frequencies. The first part of the table shows the actual size of the test, while the second part of the table shows the power.

DGP			Unit Root Tests			
α_i	N	T	ADF	PP	VR	BCC
1	1	50	0.0535	0.0535	0.051	0.0445
1	1	100	0.0455	0.0455	0.0505	0.0705
1	1	200	0.0605	0.0605	0.048	0.061
0.9	1	50	0.1085	0.1085	0.0565	0.1415
0.9	1	100	0.321	0.321	0.0500	0.345
0.9	1	200	0.8865	0.8865	0.0915	0.4975

Table 1.4: *Monte Carlo results for time series unit root tests. The table shows the rejection frequencies for the BCC test as well as other, commonly used, time series unit root tests. The nominal significance level is 5%. Rejection occurs when the p-value is less than 0.05.*

Table 1.4 reveals that the BCC test performs better than the standard ADF and PP test in case the sample size is small, 50 in our case, while the ADF and the PP test performs better for larger sample sizes. Although the differentiating power in a small sample is modest, by combining multiple cross-sections in the case of panel-data, even this small difference may result in sizable gain of statistical power for the panel data case. This possibility will be explored in the next chapter. Also, the BCC test performs better than the

Variance Ratio tests. Note that the Variance Ratio test has been primarily designed to identify heteroscedasticity and it is not primarily designed to differentiate between unit root and near unit root processes.

Furthermore, the statistical power of the BCC test in larger samples may be improved further by perfecting the counting mechanism. The idea is as follows. Right now, we discard those boundary crossing events for which the Y_t^D is zero. Yet, one of the characteristics of stationary processes is that they are more likely to cross the long-term mean than a unit root process. Hence, the number of such events, E_{00} in Table 1.2, also contains information. Incorporating this information into the testing procedure may improve performance further, especially for larger samples.

To conclude, the BCC test is relatively powerful in case the sample size is small. For larger samples, tests based on the ADF regression dominate the current version of the BCC test.

1.5 Financial Application

In this section, we consider a financial application which analyzes the possibility of mean reversion in financial markets. The starting point is based on the first equation of Balvers et al. (2000).

$$\log(R_{t+1}) - \log(R_t) = \mu + \beta(P_t - P_t^*) + u_t \tag{1.37}$$

where R_t is the total return index, P_t^* is the long term mean or the equilibrium value of the market which is unobserved and P_t is the price level. Moreover, we assume that u_t is independent but not necessarily identically or normally distributed. The null hypothesis is as follows.

$$H_0 : \beta = 0 \tag{1.38}$$

The alternative hypothesis is as follows:

$$H_1 : \beta < 0 \tag{1.39}$$

The fundamental problem is that P_t^*, that is the long-term mean or the equilibrium value to which the stochastic process supposedly returns, is unobserved. Hence, parametric estimation may be problematic.

The above-described nonparametric method is useful in overcoming this problem. In our approach, we assume that we can infer $(P_t - P_t^*)$, that is the difference between the equilibrium value and the current market value, based on some fundamental measure, F_t. As for this fundamental measure, we use the price earnings ratio as suggested by Shiller (2005, p. 186).

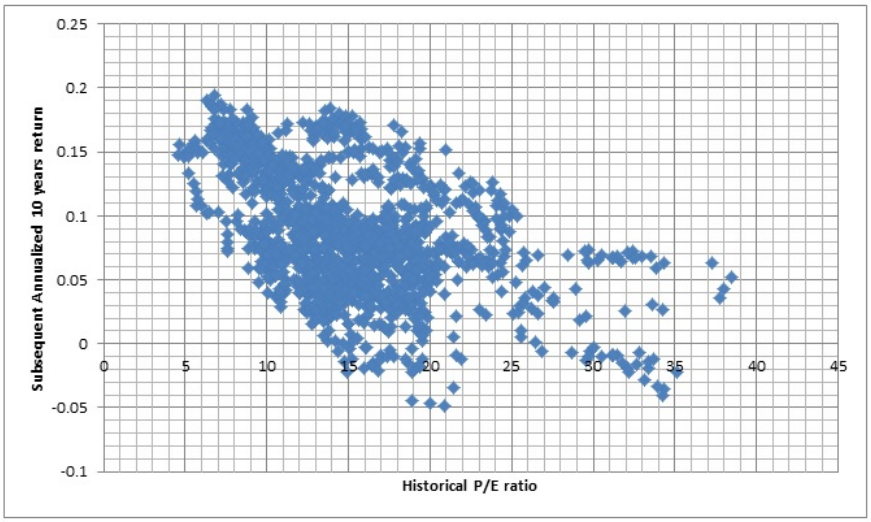


Figure 1.6: *P/E ratio and annualized returns. This figure is based on Shiller (2005, p. 186). The data is downloaded from Shiller's website. Note that observations are overlapping.*

Shiller states that there is probably a weak relationship between the price earnings ratio and the long-run stock returns. However, quantifying this relationship is problematic because the observations are overlapping.

Our approach for measuring this relationship differs from Shiller's as we do not work with overlapping observations for the returns. Instead, we measure returns using boundary crossing events. Consequently, we do not work with constant sampling frequency but rather with random sampling frequency.

Also, in our approach, we do not need to specify the exact relationship between the equilibrium value and this fundamental measure, it is sufficient to assume the following:

Assumption 4 ***Market is overvalued,*** *that is* $P_t - P_t^* > 0$ *if the price earnings ratio,* F_t*, is sufficiently above its long-term average, that is* $F_t > \bar{F} + C$*, where* $\bar{F}$ *is the median P/E ratio and* C *is an exogenously chosen constant.*

Likewise,

Assumption 5 ***Market is undervalued****, that is* $P_t - P_t^* < 0$ *if the price earnings ratio is sufficiently below its long-term average, that is* $F_t < \bar{F} - C$*.*

If the price-earnings ratio is above its long-term average, then it is a sign that the market is overvalued, that is P_t is above its equilibrium value, P^*. Likewise, if the price-earnings ratio is below its long-term average then the market is undervalued, hence P_t is below its equilibrium value, P^*.

Under the null hypothesis, the upper crossing probability is not influenced by the market fundamentals. Under the alternative hypothesis, an upper crossing event is more likely when the market is undervalued than when it is overvalued. Likewise, a lower crossing event is more likely when the market is overvalued than when it is undervalued. We test the null hypothesis using the convergence probability described in equations (1.34).

As for the data, we use monthly S&P500 data as obtained from Shiller's website[11]. We measure the returns as excess log returns. As for the risk-free rate, we use the 10-year rate as provided by Shiller. As for the boundaries, $U = 0.2070 = 5 \times \sigma^{telr}$, where $telr$ indicates total excess log returns and $L = -0.2070$. As for F_t, that is for the fundamental value, we use Shiller's cyclically adjusted price earnings ratio, or CAPE ratio, as calculated on the spreadsheet provided by Shiller. This is the 10 year moving average of the real price-earnings ratios.

As for $\bar{F}$, we use the median value of the CAPE ratios which is 16.0. We chose $C = 1.5$, that is we assume that the market is undervalued if the $F_t < 14.5$. Likewise, we assume that the market is overvalued if $F_t > 17.5$. The baseline results are shown in the following contingency table.

		Market valuation		
		undervaluation $(F_t < \bar{F} - C)$	neutral $(\bar{F} - C \leq F_t \leq \bar{F} + C)$	overvaluation $\bar{F} + C < F_t$
Next BC	$Z_k = -1$	8	7	10
Event	$Z_k = 1$	28	8	14

Table 1.5: *Stock market valuation and boundary crossing events. In total, we observe 75 boundary crossing events over the periods of 135 years. Thus, the average holding period is 1.80 years. The convergence probability is 0.63 which is significantly above 0.5 at 5% significance level, the p-value is 0.0259.*

The data rejects the null hypothesis. The convergence probability as defined in Equation (1.34) is 0.63 which is significantly above 0.5. The p-value is 0.0259. In total, we observe 75 counting events over 135 years. Hence, the average holding period is 1.80 years which is much lower than the holding

[11] `http://www.econ.yale.edu/~shiller/data.htm`

period of 10 years suggested by Shiller. We could increase the average holding period by applying wider boundaries. However, such an increase would reduce the number of boundary crossing events which would make inference more difficult.

The results can be interpreted as follows. In the past 135 years, under the current boundary setting, there were 25 occasions on which the excess return of the S&P500 over some random investment horizon were -20.7%. One could have avoided 17 occasions, that is approximately 68% of the cases by exiting the market when the CAPE ratio rises above 14.5. Of course, the price to pay for such market timing strategy is to avoid 22 occasions in which the market increased by 20.7%.

Let us conclude the application by analyzing how the choice on the parameters influences our results. We consider two factors: boundary selections and the choice of C for determining over and undervaluation. As for the choice on boundaries, we also consider upper boundaries placed at 4 and 6 standard deviation distance, that is for example in case of log returns to $U = 0.1657 = 4 \times \sigma^{telr}$ and $U = 0.2485 = 6 \times \sigma^{telr}$, $L = -U$. Finally, as for the choice on C, we also consider $C = 1$ and $C = 2$. The results are shown in Table 1.6.

p–values of the BCC-test			
Boundaries in	C		
standard deviation	1.0	1.5	2.0
-4,4	0.0272	0.0140	0.0147
-5,5	0.0178	0.0259	0.0240
-6,6	0.0871	0.0586	0.0266

Table 1.6: *Robustness exercise for the BCC-test on mean-reversion.*

Table 1.6 further confirms what Shiller proposes in his book: the relationship between the seasonally adjusted price-earnings ratio and the excess returns is probably significant. None of the settings accept the null hypothesis at 10% significance level. This finding may be interpreted as supportive evidence against strong efficient market hypothesis and in favour for fundamental analysis and market timing strategies.

1.6 Summary of Chapter 1.

In this chapter, we introduce and characterize a new stochastic process which counts how many times the original stochastic process crosses boundaries. We discuss some common pitfalls when estimating BCC-distributions.

Next, we develop a testing procedure based on boundary crossing counts. We apply this procedure for univariate unit root testing. We find that the BCC test is more powerful than the Augmented Dickey-Fuller test and the Phillips-Perron test when the sample size is small. Also, it is more powerful than the Variance ratio test.

We finish this chapter with an application in which we analyze the possibility of mean reversion in the excess returns for the S&P500. We identify the unobserved mean using Shiller's CAPE ratio. Our test supports mean reversion which can be interpreted as evidence against strong efficient market hypothesis.

We have identified several opportunities for further development. First of all, the counting procedure for the unit root test can be improved. Also, the test procedure can be extended to panel data. The latter will be covered in the next chapter.

Finally, statistical tests could be developed based on the number of

upper and lower crossing events. These tests would essentially compare the number of boundary crossing events under the null and the alternative hypothesis. By using this version of the BCC test, one can evaluate the probability that an observed realization of a potentially unknown stochastic process is generated by some other, parametrically or non-parametrically defined stochastic process. A financial analyst may, for example, calculate the probability that the DGP for the price of some security is well represented by the Brownian motion or by some GARCH(p,q) process.

Initial Monte Carlo studies not detailed in this dissertation however show that such tests are not particularly powerful. More specifically, these additional Monte Carlo studies show that unit root test based on the number of upper and lower crossing events is less powerful than the usual parametric alternatives. Also, such a specification test is less powerful than the likelihood based tests when identifying GARCH models. Nevertheless, in some cases,[12] as explained in Andrews (2001), the LR test is problematic because of the nonnegativity constraints on the GARCH coefficients. In such cases, when the LR test statistic does not have the usual chi-squared asymptotic null distribution, one could still use the BCC-test. These options may be investigated further in additional papers.

[12]The review of Robert Lieli has proposed this insightful example.

DOI: 10.14754/CEU.2016.01

Chapter 2

Testing for Unit Roots in Panel Data with Boundary Crossing Counts

This is a joint work with Laszlo Matyas.

Abstract

In this chapter, we introduce a new, distribution free, non-asymptotic, approach for unit root testing based on boundary crossing counts. Using this approach, we develop two versions of a panel unit root test. The first one can be applied in the case of cross-sectionally independent panel data, while the second is designed for cross-sectionally dependent panels. As for the results, the first version of the newly proposed test dominates the IPS test and the Maddala-Wu test in case of relatively short, cross-sectionally independent panel data. The second version is more powerful than existing second generation panel data tests, such as Bai and Ng's PANIC unit root test or Pesaran's CADF test in case the data is generated by a multi-factor model and the time dimension is relatively short. Next, we show that the

unit root hypothesis cannot be rejected on real exchange rate data hence we do not find supportive evidence for the PPP hypothesis. Finally, we discuss various methodological issues related to our newly proposed test.

2.1 Introduction

The problem of unit roots in economic time series has been studied in econometrics for decades, both from a methodological, as well as a strictly economic point of view. Many statistical and econometric techniques, for example the Box and Jenkins analysis, assume that we are able to differentiate between stationary and unit root processes. Also, verifying economic theory empirically — such as, for example, the purchasing power hypothesis or various other convergence hypotheses, often involves testing for unit roots. Hence, developing and refining statistical tests in order to verify or falsify the unit root hypothesis has long been an important item on the agenda of econometrics.

In this paper, we develop a new second generation panel data unit root test. Depending on the underlying data generating process (DGP), the literature considers two basic model structures. The first is suited for data which do not exhibit a deterministic trend, such as, for example, real exchange rates, inflation rates or interest rates, etc. The second can be used to formalize DGP with deterministic trends, such as the GDP, etc. Let us take, as a starting point, the first structure without deterministic trend, which is also discussed in the review of Breitung and Pesaran (2008, p. 295):

$$X_{it} = \mu_i + \alpha_i X_{it-1} + u_{it} \tag{2.1}$$

where X_{it} are the data series to be analysed, μ_i are the individual-specific fixed effect and u_{it} are the composite error terms, with the number of cross

sections being $i = 1, \ldots, N$ and the number of time dimensions $t = 1, \ldots, T_i$, that is we cater for unbalanced panels as well.

In line with the literature, as in Levin and Lin (1992), Maddala and Wu (1999) and Im et al. (2003), our null hypothesis assumes that all α_i are 1:

$$H_0 : \alpha_1 = \cdots = \alpha_N = 1 \text{ for all } i = 1, \ldots, N \tag{2.2}$$

We consider the heterogeneous alternative where some, but not necessarily all, cross sections are stationary:

$$H_1 : \exists\ N_0\text{: } \alpha_{N_0} < 1,\ \ 0 < N_0 \leq N \tag{2.3}$$

As for the individual effects, they may or may not be nil under the null hypothesis. Under the former, the individual effects are typically defined as $(1 - \alpha_i) \times \mu_i$, which is the discrete counterpart of the Ornstein-Uhlenbeck process, as detailed in Szimayer and Maller (2004). This formulation (described for time series as the "second case" in Hamilton (1994, p. 490)) facilitates the analysis in two different ways. First, under this specification, the Dickey-Fuller asymptotics can be used for the OLS estimates, while with individual effect present in the null (described for time series as the "third case" in Hamilton (1994, p. 497)) it cannot. Second, fixed effects in the dynamic model result in biased OLS estimates, moreover, consistency in N cannot generally be established. Since in this paper, we do not rely on asymptotic distributions, we do not restrict the individual effects under the null hypothesis. Instead, we discuss both, the specific as well as the more general formulation.

As for the composite error terms, Hurlin et al. (2007) highlight two main approaches. First, authors may use various factor structures, such as in Choi (2006) or in Pesaran (2007). Others, for example Chang (2002), propose

to work with the residuals covariance matrix and to rely on instrumental variables. Here, we apply a covariance matrix-based technique as well, but instead of relying on the residual covariance matrix, we work with the covariance matrix of variables describing boundary crossing events.

The panel unit root testing literature, in general, is structured as follows (see Banerjee (1999), Baltagi and Kao (2001), Hurlin et al. (2007) and Breitung and Pesaran (2008)). First, there are first generation tests (see Levin and Lin (1992)), where the error terms of the model are assumed to be independent across i, and second generation tests (see Pesaran (2007)), where the errors terms are allowed to be contemporaneously correlated.[1] The independence assumption can be quite problematic, as cross-sectional dependence may arise in many applications, for example in the case of output growth equations, as in Pesaran (2004), or due to spatial dependence as in Baltagi et al. (2007).

Second, the alternative hypothesis may be homogeneous or heterogeneous. The former, used for example by Levin et al. (2002), assumes that $\alpha_1 = ... = \alpha_i = ...\alpha_n$. This homogeneous alternative is somewhat restrictive, for example in the case of convergence hypothesis for different countries in a macro model, this would imply that all countries or regions converge at the same rate if indeed they converge at all. Consequently, the less restrictive heterogeneous alternative which allow for cross-sectional differences has been introduced, for example, by Im et al. (2003).

Finally, tests differ in how they aggregate across different cross sections. There is in fact quite a variety of different aggregational techniques to combine individual cross sections. Maddala and Wu (1999), for example,

[1] A more elaborate classification of first and second generation tests is provided in Hurlin et al. (2007, p. 3, Table 1.).

suggest making use of the early results of meta analysis described in Tippett et al. (1931) and Fisher (1932), or more recently by Wolf (1986) who combines individual significance levels. Alternatively, Im et al. (2003) propose to merge individual t-statistics. What we propose in this paper is to aggregate by counting the number of boundary crossing events.

To summarize: The test we propose in this paper can be classified as a second generation unit root test with heterogeneous alternatives which aggregates across individuals using the number of boundary crossing events (to be introduced below).

The intuition behind the test is very simple. Let us assume that a series X_{it} is enclosed by an upper and a lower boundary. If the process is stationary, a boundary crossing event is less likely than if it is unit root, as the demeaned process, unlike the unit root one, has the tendency to return to zero. By counting the boundary crossings therefore we can distinguish between the two processes.

Formally, let us introduce a new class of discrete stochastic process called boundary crossing counting process or BCC process, Y_{it}, which counts the number of boundary crossing events. Let U_i be some upper boundary and L_i be some lower boundary (the decision on the boundaries will be discussed in the next section). Also, upon each boundary crossing, the underlying stochastic process is restarted at some restarting value, X_{it}^0. (the choice of this restarting value will also be detailed in the next section). Let the restarted process be denoted by X_{it}^*. Note that the process may be restarted several times.[2] Let us differentiate between the following counting processes.

[2]Boundary classification may be found in Karlin and Taylor (1981, p. 234), where they differentiate between "regular", "absorbing", "natural" and "entrance" types. The type of boundary applied in our paper does not have a one to one correspondence to any of these cases: They could be called "restarting boundaries". If one must classify, restarting

1. $Y_{it}^{U}(X_{it}^{*})$ counts the number of upper crossing events, that is how many times X_{it}^{*} needs to be restarted after an upper-crossing event.

2. $Y_{it}^{L}(X_{it}^{*})$ counts the number of lower crossing events, that is how many times X_{it}^{*} needs to be restarted after a lower-crossing event.

3. $Y_{it}^{A}(X_{it}^{*}) = Y_{it}^{U}(X_{it}^{*}) + Y_{it}^{L}(X_{it}^{*})$ counts all crossing events.

4. $Y_{it}^{D}(X_{it}^{*}) = Y_{it}^{U}(X_{it}^{*}) - Y_{it}^{L}(X_{it}^{*})$ is the difference between the number of upper and lower crossing events.

5. Finally, sometimes there is a need to refer to all of these processes at once, in this case we use the notation $Y_{it}(X_{it}^{*})$.

Thus, the counting process is a function of the restarted process, $Y_{it}(X_{it}^{*})$. Also, the restarted process is a function of the underlying data, the two boundaries and finally the restarting value, $Y_{it}(X_{it}^{*}(X_{it}, U_i, L_i, X_{it}^{0}))$. These dependencies are suppressed in the rest of the paper for ease of notation.

Our approach has several desirable properties. Besides the usual favorable properties of nonparametric tests, our method is non-asymptotic (although we briefly discuss the large sample properties as well). Moreover, the technique can also be used in the case of unbalanced panels, or panels with missing values, or when the data generating process is sampled with an uneven frequency. Also, the test is relatively powerful when the error term of the underlying data generating process follows a t-distribution rather than a normal one.

Naturally, the BCC test suffers from certain drawbacks. Similar to Fisher's exact test, the distribution of the test statistics is a discrete one.

boundaries are attainable and regular boundaries, where the process is restarted upon boundary crossing events.

Consequently, selecting the usual 1%, 5% and 10% as critical value is somewhat problematic, and we have to make use of the closest available discrete value.

Our paper is structured as follows. Section 2 sets out the model and discusses some estimation issues. Section 3 compares the finite sample properties of this newly introduced test to other frequently used unit root tests using Monte Carlo simulations. Section 4 is dedicated to an application. Section 5 discusses additional technical details while the last section provides a conclusion.

2.2 Testing for Unit Roots Using Boundary Crossing Events

In this section, we show how to use the number of boundary crossing events for testing for unit roots in panel data. We proceed with the derivation in two steps. First, we discuss how to construct the test statistics in an ideal case when errors are independent. We continue by extending the derivation for cases when the error terms are not independent.

We assume that the DGP is characterized by Equation 2.1. Moreover, let us assume that X_{it} starts from minus infinity. Also, for the time being, we assume that the individual effects[3] are zero. Moreover, boundaries are chosen to be symmetric, $L_i = -U_i$, and the restarting value is zero for all i and t, that is $X_{it}^0 = 0$. Finally, let the restarted process be defined over $\widetilde{X}_{it} = X_{it} - X_{i0}$ for ease of notation.

[3]This assumption is discussed in detail at the end of this chapter.

2.2.1 Test Statistics in case of Independent Errors

This case is a straight-forward extension of the time series version of the test. If errors are independent then boundary crossing events can be pooled over the cross sections. Formally, under the null hypothesis, Equation (1.32) can be modified as follows.

$$\underbrace{p(Z_{ik} = -1|Y^D_{ik-1} < 0)}_{p_{11}} = \underbrace{p(Z_{ik} = -1|0 < Y^D_{ik-1})}_{p_{12}} \tag{2.4}$$

Likewise, the lower crossing probabilities below are also equal:

$$\underbrace{p(Z_{ik} = 1|Y^D_{ik-1} < 0)}_{p_{21}} = \underbrace{p(Z_{ik} = 1|0 < Y^D_{ik-1})}_{p_{22}} \tag{2.5}$$

By combining Equation (2.5) and Equation (2.4), we obtain equality (1.30) under the null. Under the stationary alternative, since $(\alpha_i - 1) < 0$, a lower crossing event is more likely in case $Y^D_{ik-1} > 0$, than in case $Y^D_{ik-1} < 0$. Also, an upper crossing event is more likely in case $Y^D_{ik-1} < 0$, than in case $Y^D_{ik-1} > 0$. Consequently, the alternative hypothesis can be described as in Equation (1.33).

Similarly to the time series case of Chapter 1, we observe five kinds of events as described in Table 2.1.

These events are now based on pooled observations. The convergence probability, which is defined in Equation (1.34) and the test statistics which is defined in Equation (1.35) can be obtained as in the time series case of Chapter 1.

As for the boundary settings, we can modify Equation 1.36 as follows.

$$U_i = \begin{cases} \sigma_i \text{ if } T < 100 \\ (1 + min(1, \frac{N}{100}) \times \frac{(T_i - 100)}{225}) \times \hat{\sigma}_i \text{ otherwise,} \end{cases} \tag{2.6}$$

		Cumulative Upper minus Lower Crossing		
		$Y^D_{ik-1} < 0$	$Y^D_{ik-1} = 0$	$0 < Y^D_{ik-1}$
Next BC Event	$Z_{ik} = -1$	$E_{11} + 0.25$ (Divergence)	E_{00} (Non Informative)	$E_{12} + 0.25$ (Convergence)
	$Z_{ik} = 1$	$E_{21} + 0.25$ (Convergence)		$E_{22} + 0.25$ (Divergence)

Table 2.1: *Contingency table based on boundary crossing events.* E_{jk} *indicates the number of events observed in the data. Note that 0.25 is added to each cells for technical reasons in order to avoid any division with zero.*

where N is the number of cross sections, T_i is the number of observations in cross-section i and finally $\hat{\sigma}_i$ is the sample standard deviation for cross section i. The test provides reasonable differentiating power while, at the same time, it's empirical size is close to the nominal one. In the remaining part of the dissertation, we use this heuristic rule for setting up the boundaries unless explicitly stated otherwise.

2.2.2 Test Statistics in case of Dependent Errors

The following section analyzes the case of dependent errors. If errors are dependent, then the boundary crossing events can no longer be described by independent Bernoulli trials and hence, the binomial distribution cannot be used anymore. Yet, under cross-sectional dependence, the null hypothesis described in Equation 1.30 is still valid under the null, only the variance of the test distribution is affected. Let us begin by dealing with cross-sectional dependence. We return to the question of autocorrelation later.

Additional Monte Carlo studies show that applying the binomial distribution which ignores cross-sectional dependence results in size distortion.

Depending on the strengths of the dependence and the number of cross sections, the actual size may be between 5% and 40% in case the nominal size is 5%. Consequently, the test procedure needs to be modified for panel data with cross-sectional dependence.

Potentially, there are three methods to adjust for cross-dependence. The first one is to rely on the law of dependent large numbers. The second one, which is somewhat theoretical in nature, aims to restore the independence of the Bernoulli trials by modifying the counting procedure. Finally, the last one is built on the fact that under the null hypothesis, the variance of the sum of the variables describing the individual trials can be estimated from the individual boundary crossing events. We continue our analysis with this last solution while the first two approaches are discussed in the fifth section.

Next, we show how to capture the cross-dependence with the covariance matrix of the individual trials. First, for some boundary crossing k, let us define C_{ik} in the following way:

$$C_{ik} = -1 \quad \text{if} \qquad \begin{cases} Z_{ik} = 1 \text{ and } Y^D_{ik-1} > 0 \text{ or} \\ Z_{ik} = -1 \text{ and } Y^D_{ik-1} < 0. \end{cases} \tag{2.7}$$

Also, in case the boundary crossing event points toward convergence:

$$C_{ik} = 1 \quad \text{if} \qquad \begin{cases} Z_{ik} = 1 \text{ and } Y^D_{ik-1} < 0 \text{ or} \\ Z_{ik} = -1 \text{ and } Y^D_{ik-1} > 0. \end{cases} \tag{2.8}$$

Using these notations, the null hypothesis can be restated as

$$H_0 : \sum_{i=1}^{N} \sum_{k=1}^{B_i} C_{ik} = S_c = 0 \tag{2.9}$$

where B_i is the total number of quasi-Bernoulli[4] trials for cross-section i, that is for those boundary crossing events where $Y^D_{ik-1} \neq 0$. From now on, we refer

[4]In case of Bernoulli trials, the outcome is either +1 or zero. Here, the outcome is either +1 or -1.

to C_{ik} as convergence dummies and S_c as convergence sum. For the ease of notation, we suppress the indexes of the summations. Under the stationary alternative, the convergence sum is greater than zero:

$$H_1 : \sum\sum C_{ik} > 0 \tag{2.10}$$

Next, let us define C as follows:

$$C = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{T1} & c_{T2} & \cdots & c_{Tn} \end{pmatrix} \tag{2.11}$$

where the elements of C may be one, zero or minus one:

$$c_{it} = \begin{cases} 0 \text{ if } Y_{it}^A = Y_{it-1}^A \text{ or } Y_{it-1}^D = 0 \\ C_{ik(t)} \text{ otherwise.} \end{cases} \tag{2.12}$$

where the subscript in $C_{ik(t)}$ indicates that the k^{th} boundary crossing event occurs in time t. Also, let $\Sigma = C'C./(1_{n\times 1}B)$, where $B = [B_1, B_2, ...B_N]$ describes the number of quasi-Bernoulli trials for each cross sections and ./ indicates element by element division. Note that if no boundary crossing events are observed for a particular cross section, then it needs to be removed from the sample. The variance of the test statistics can be expressed using Σ:

$$var(S_c) = E({S_c}^2) = E(B \times \Sigma \times 1_{n\times 1})) \tag{2.13}$$

where the first equality is due to the fact that $E(S_c) = 0$ under the null hypothesis, while the second equality is due to the fact that adding zeros to a sum does not modify its value.

We construct the empirical distribution by simulation. The idea is to make use of the fact that under the null, if we simulate $\sum(B_i)$ random

numbers having zero mean and Σ as covariance matrix, then the sum of these simulated random numbers will have the same mean and the same variance as S_c. Hence, under the null, we can approximate the confidence interval for S_c using the sum of these simulated random numbers. In theory, we could draw the elements for the summation from an arbitrarily distribution, yet in practice, the convenient choice for this simulated distribution is the normal distribution.

Potentially, depending on the given application, there may be many different methods to carry out the above-described simulation. Here, we show a solution which is based on the Cholesky decomposition. The algorithm to obtain the test distribution and its critical value is the following:

1. Count the number of boundary crossing events. Multiple boundary crossings between two observations shall be recorded as two consecutive boundary crossing events.[5]

2. Estimate Σ from the sample.

3. Simulate 1000 correlated random numbers using normal distribution with mean zero and covariance matrix Σ and a sample size of B_T. In practice, especially when N is large and T is small, it may happen that Σ is not positive definite. Since we are dealing with matrixes containing only 1,0,-1, we are more and more likely to observe perfect dependence by chance as the number of cross sections increases even if the underlying DGP does not involves dependent cross sections. In practice, this can be corrected by finding the nearest symmetric positive semi-definite matrix as described by Higham (1988).[6]

[5]This is in fact a sampling issue, the sampling frequency is not fine enough to record what happens between two observations.

[6]The procedure was implemented based on John D'Errico's Matlab code. We would

4. The test statistics can be obtained by summing up each simulated sample, the critical value being the 95% percentile of these sums.

5. The null hypothesis is rejected if the convergence sum observed in the sample is larger than this critical value.

To sum up: the above-described algorithm simulates random numbers having the same mean and the same variance as the sample convergence sum. Hence, although they may differ in higher moments, they are likely to be distributed similarly, and so the simulated sums can be used to approximate the true test distribution.

Our Monte Carlo studies show that this method indeed reduces size distortion. The actual size of this simulation-based BCC test on a factor model was less than 10% for a nominal size of 5%. Naturally, the power against the stationary alternative is somewhat also reduced by these measures, yet the test is still quite capable to differentiate between the stationary and the unit root processes.

Regarding the BCC-test, there are a couple of additional issues[7] of interest. First of all, the inference is based on constructing the empirical distribution of the test statistic by simulation. Hence, the covariance matrix is not necessarily generated under the null hypothesis because it is computed from the panel. Note that our test is very similar to many of the tests analyzed above as they all suggest capturing dependence based on the data.

Moreover, the dependence may also be captured using principle component analysis. In fact, the ideal method to capture dependence varies from application to application. If the method to capture the dependencies is inappropriate, then the BCC-test also suffers. Here, we only have room to

like to take this opportunity to acknowledge his contribution.

[7]These insightful comments are based on the review of Timo Teräsvirta.

describe one potential solution for illustrative purposes.

In addition, we use the normal distribution to construct the empirical distribution. Although we present a technical device to correct for possible singularity, if the matrix is near-singular then the normality assumption may not be appropriate.

2.3 Comparative Monte Carlo Analysis

This section compares the performance of the BCC test with other, commonly used unit root tests. It is important to mention that, strictly speaking, direct power comparisons between the different tests are not valid, since they have different null and/or alternative hypotheses. These differences are discussed in more detail in the fifth section. Yet, we still present the actual size and the power of the different tests in one table, but these results need to be interpreted with caution.

The design of the Monte Carlo experiments aims to simulate data for which nonparametric, in a distribution-free sense, methods may be reasonably applied. In particular, throughout the experiments, we assume that error terms have t-distribution with three degrees of freedom, which is approximately equal to the estimated degree of freedom of the daily log returns of the S&P500.

The DGP may be unit root or stationary with $\alpha_i = 0.9$, which is a typical choice also taken by Maddala and Wu (1999). For the stationary case, we assume that the initial observation is close to the long-term mean.

The Monte Carlo trials are repeated 2000 times, and for each repetition we carry out the tests at 5% significance level, the probability of rejecting the null hypothesis is obtained by dividing the number of cases when the null

is rejected by the total number of Monte Carlo trials.

We begin by analyzing the first generation panel unit root tests and conclude with the second generation tests. The simulations were implemented in Matlab,the code was inspired by Hurlin's Matlab codes[8] for which we are very grateful.

2.3.1 First Generation Panel Unit Root Tests

We begin by comparing the BCC test with three different first generation tests under cross-sectional independence. First, we consider several versions of Im et al. (2003)'s IPS tests. More specifically, we calculate the w-bar test which is based on the t-values, the t-bar test which is based on the moments of the DF distribution and finally the z-bar test which is based on the assumption of no autocorrelation of the residuals. Since the error term in the DGP is not autocorrelated, the results of w-bar, t-bar and z-bar tests should not be substantially different.

In addition, we consider two versions of Maddala and Wu (1999)'s test, further referred to as MW test, and Choi (2001)'s test, further referred to as CH test. The two tests differ in how they combine the individual p-values. In Maddala and Wu (1999), the p-values are calculated based on the critical values of Fisher's statistics while in Choi (2001), they are based on the individual ADF statistics. Since these tests rely on meta-analysis-based techniques to combine p-values, from now on, we will refer to them as meta-analysis based tests. For both tests, we consider two versions. In the first version, the autocorrelation is estimated from the simulated data, while in the second version, the lag parameter is set to be zero.

Finally, as for the BCC test, since the Bernoulli trials are independent,

[8]The libraries were downloaded from the website of Orlean's University.

we rely on the binomial distribution-based version. The test statistics are obtained by pooling the number of boundary crossing events over the cross sections and the p-values are obtained from the right hand side of the corresponding binomial distribution.

As for the DGP, besides the baseline assumptions detailed in the beginning of this section, we assume that the cross sections are independent

DGP				First Generation Unit Root Tests							
f_i	α_i	N	T	IPS (w-bar)	IPS (t-bar)	IPS (z-bar)	MW (lag = 0)	MW (DF-lag)	CH (lag = 0	CH (DF-lag)	BCC
0	1	12	25	0.1030	0.1505	0.0605	0.0985	0.0325	0.1190	0.0460	0.0630
0	1	20	25	0.1015	0.1720	0.0590	0.1155	0.0430	0.1350	0.0515	0.0575
0	1	12	50	0.0550	0.0545	0.0575	0.0380	0.0405	0.0480	0.0505	0.0855
0	1	20	50	0.0595	0.0570	0.0580	0.0400	0.0420	0.0505	0.0505	0.0750
0	0.9	12	25	0.3455	0.4205	0.2830	0.2995	0.1635	0.3335	0.1960	0.5010
0	0.9	20	25	0.4300	0.5615	0.4175	0.3945	0.2320	0.4315	0.2665	0.6855
0	0.9	12	50	0.8440	0.8395	0.8230	0.6710	0.6230	0.7105	0.6735	0.9150
0	0.9	20	50	0.9745	0.9720	0.9660	0.8755	0.8365	0.8995	0.8615	0.9880

Table 2.2: *Monte Carlo results for first generation panel data unit root tests. The table shows the probability of rejecting the null hypothesis for the BCC test and other, commonly used, first generation panel data unit root tests in case of balanced panels and cross-sectional independence. The nominal significance level is 5%. Rejection occurs when the p-value is less than 0.05.*

The table reveals that the BCC test has the highest power. As the sample size grows, the difference in the power between the existing tests and the BCC test diminishes. These results are in line with the findings of the time series analysis detailed in the previous chapter where we have shown that the BCC test is more powerful than the ADF test in case the sample size is small. Moreover, pooling the number of boundary crossing events over all

cross-sections seems to be an effective aggregational technique. Combining relatively powerful individual tests in an effective way results in a panel data test which has high differentiating power.

Moreover, in this particular Monte Carlo setup, the t-bar version of the IPS-tests has the highest power in small samples, even in the case T is small, it also suffers from minor size-distortion. Also, the IPS tests typically show a somewhat higher differentiating power than the tests based on meta-analysis. As for the meta-analysis based tests, the test of Choi (2001) has somewhat higher differentiating power than the test of Maddala and Wu (1999). Finally, by providing additional information on the lag structure, the power of the meta-analysis based tests, especially in small sample, can be improved.

Additional Monte Carlo experiments show that missing data causes size-distortion in parametric first generation tests. In case of BCC test, some size-distortion is also present but to a much lesser degree. Under time series settings, the ADF test does not exhibit significant size distortion, hence the problem is probably caused by aggregational techniques.

Overall, the binomial BCC test can be used, when the cross-sections are independent and it is advised to be used when the panel has missing observations.

2.3.2 Second Generation Panel Unit Root Tests

We continue by studying second generation unit root tests. Since Maddala and Wu (1999) already conducted a set of experiments in which the simulated data is spatially dependent, here, we focus on factor models. Table 1. in Hurlin et al. (2007)'s review differentiates between two main approaches to deal with cross-sectional dependencies. First, authors may use various factor models, such as in Choi (2006) or in Bai and Ng (2002). Others, for example Chang

(2002), propose to work with the residuals' covariance matrix.

Factor models assume that the dependence is captured by one or more factors. In the case of a single common factor, Pesaran (2007) proposes to deal with cross-sectional dependence by further augmenting the Augmented Dickey-Fuller regressions by both the cross section average of the lagged levels and of the lagged first differences. These cross-sectionally augmented Augmented Dickey-Fuller equations, CADFs, are estimated by OLS, and the individual t–ratios of the OLS estimates are combined to obtain the test statistics. The advantage of this approach is its simplicity, while it may not be able to fully capture those cross-dependence structures which consist of several factors.

Multiple common factors are typically quantified using principle components. Bai and Ng (2004), for example, propose to first separate the common factors and the idiosyncratic terms and then to test them separately. The advantage of their method is that the properties of the common components may also be of economic interest, not just the those of the original data. Moon and Perron (2004), on the other hand, promote testing for unit roots on the de-factored series, which allows for a rather general specification of the common components. Since both multi-factor approaches described above rely on principle component analysis, the results may depend on the scale on which the variables are measured. Also, the differentiating power of these tests in finite samples when N is large and T is small may in some cases be limited.

Alternatively, cross-dependence may be captured via the covariance matrix. Fundamentally, the difficulty arises from the fact that the limiting distribution of the OLS or GLS estimators is dependent on certain nuisance parameters and hence, the usual Wald type of test cannot directly be

applied. There are, however, some methods to overcome this problem. First, bootstrap-based estimators starting perhaps from Maddala and Wu (1999) may be used. Also, Chang (2002) proposes using a special, non-linear instrumental variable estimator and make use of the fact that the proposed individual IV estimates for the t–ratio statistics are asymptotically independent even for dependent cross-sectional units. Finally, Demetrescu et al. (2006) explain how to combine individual p–values in the cases where there is constant correlation among the p–values of the individual estimates.

In the next Monte Carlo exercise, we are interested in how tests perform under general conditions. Hence, we simulate data using the following multiple common factor model:

$$u_{it} = f_i^1 \times \Theta_t^1 + f_i^2 \times \Theta_t^2 + f_i^3 \times \Theta_t^3 + \epsilon_{it}, \tag{2.14}$$

where Θ_t^1, Θ_t^2, and Θ_t^3 are i.i.d. random unobserved common components and f_i^1, f_i^2, and f_i^3 are the factors. The random components are assumed to be drawn from a t-distribution with three degrees of freedom. As for the value of f_i^1, f_i^2, and f_i^3, we assume that they are randomly chosen from the uniform distribution centered around some predefined constants, which are detailed in Table 2.3. Hence, the loadings are different for each cross section.

	f_i^1		f_i^2		f_i^3	
	neg. dep.	pos. dep.	neg. dep.	pos. dep	neg. dep.	pos. dep.
w.f.d.	-0.2	0.4	-0.1	0.3	-0.3	0.2
s.f.d.	-0.6	0.8	-0.5	0.7	-0.75	0.7

Table 2.3: *Assumption for factor loadings on the individual cross sections in case cross-dependence arise out of multiple common factors.*

The first row, $w.f.d$, abbreviates weaker factor dependence while the second row, $s.f.d.$ abbreviates stronger factor dependence. For example, in

the first row, for the first factor loading, half of the cross-sections are assumed to have f_i^1 around -0.2 while for the remaining cross sections, f_i^1 is assumed to be around 0.4. The first row models a case when more than half of the variation is driven by the idiosyncratic term. For the second row, most of the variation is driven by the unobserved common factors.

We compare the BCC test to the PANIC unit root test of Bai and Ng (2004), further referred to as BNG test, and the Cross-sectionally Augmented Dickey-Fuller test, or CADF test, of Pesaran (2007), further referred to as PS test. As for the former unit root test, once the factor structure has been removed, the p-values of the different cross sections are aggregated either by Choi (2001)'s method, shown in the first column, or by Maddala and Wu (1999)'s method, shown in the second column. As for the latter unit root test, the first column, titled $CIPS$, shows the cross-sectionally augmented version of the IPS test, which is based on t-bar statistics while the second column, titled $CIPS^*$, shows the suitably truncated version of the cross-sectionally augmented DF-statistics. As for the BCC test, we rely on the simulation-based version. The test statistics are obtained by pooling the convergence dummies from the cross sections. The p-values are obtained from the distribution of the simulated convergence sums.

Table 2.4 shows the rejection frequencies. The first part of the table shows the actual size of the tests and the second part of the table shows the power. The nominal size is set at 5.0% for all tests.

DGP				Second Generation Panel Unit Root Tests				
f_i	α_i	N	T	BNG (Choi)	BNG (MW)	PS (CIPS)	PS (CIPS*)	BCC (sim)
w.f.d	1	12	25	0.066	0.057	0.045	0.041	0.096
w.f.d	1	20	25	0.046	0.039	0.054	0.050	0.077
w.f.d	1	12	50	0.043	0.036	0.042	0.040	0.088
w.f.d	1	20	50	0.032	0.025	0.046	0.045	0.105
w.f.d	0.9	12	25	0.251	0.217	0.094	0.094	0.520
w.f.d	0.9	20	25	0.416	0.381	0.129	0.129	0.692
w.f.d	0.9	12	50	0.827	0.784	0.331	0.331	0.906
w.f.d	0.9	20	50	0.974	0.967	0.570	0.570	0.971
s.f.d	1	12	25	0.072	0.061	0.084	0.084	0.097
s.f.d	1	20	25	0.047	0.040	0.100	0.097	0.105
s.f.d	1	12	50	0.060	0.046	0.081	0.080	0.109
s.f.d	1	20	50	0.044	0.037	0.093	0.093	0.134
s.f.d	0.9	12	25	0.276	0.241	0.149	0.149	0.482
s.f.d	0.9	20	25	0.388	0.353	0.225	0.225	0.557
s.f.d	0.9	12	50	0.801	0.768	0.405	0.405	0.820
s.f.d	0.9	20	50	0.963	0.949	0.571	0.571	0.884

Table 2.4: *Monte Carlo results for second generation panel unit root tests. The table shows the probability of rejecting the null hypothesis for the BCC test and other, commonly used, second generation panel data unit root tests for balanced panels in case dependence arises out of multiple common factors. The nominal significance level is 5%. Rejection occurs when the p-value is less than 0.05.*

BCC test continues to be the most powerful when the sample size is

small. For larger sample sizes, the BNG test dominates the BCC test in a sense that it has comparable power while the size of the BNG test is closer to the nominal size than the size of the BCC test.

As for the BNG test, Choi's aggregational technique is slightly more powerful than the alternative method. The test of Bai and Ng (2004) is more powerful than the test of Pesaran (2007) which is in line with the expectation as the former is a test specifically designed for multiple factor models.

As for the BCC-test, Table 2.4 suggest that its empirical size increases with the amount of information. This is related to how the dependence is captured. Hence for larger panels, we may need to develop additional techniques for capturing dependence. The use of principle component analysis for example could improve the properties further.

To conclude, the BCC test can be applied in cases when the cross-sectional dependence arises out of a common unobserved component. It dominates existing tests when the sample size is small. However, its performance weakens as the sample size increases. Hence, it may be reasonable to combine the BCC-technique with factor-analysis based procedures. This possibility is briefly outlined in the last section when the direction for further research is discussed.

2.4 Empirical Application

This section applies the new test to evaluate whether the PPP hypothesis holds. This is a common application for the above-reviewed panel unit tests as in Chang (2002) or in Pesaran (2007). We use the data of Pesaran (2007) as downloaded from the data archive of the Journal of Applied Econometrics for comparability. This panel covers the period of 1974 -1998 for 17 OECD

countries. The test is applied to log real exchange rates which are computed as $x_{it} = s_{it} + p_{ust} - p_{it}$, where s_{it} is the log of the nominal exchange rate of the currency of country i in terms of US dollars and p_{ust} and p_{it} are logarithms of consumer price indices for the United States and country i respectively.

We define the counting process over $\widetilde{x}_{it} = x_{it} - x_{i0}$. We chose the boundaries as described in Equation (2.6). We construct the empirical distribution by simulation as described in the section entitled "test statistics in case of dependent errors." The results are shown in Table 2.5.

		Cumulative Upper minus Lower Crossing		
		$Y^D_{ik-1} < 0$	$Y^D_{ik-1} = 0$	$0 < Y^D_{ik-1}$
Next BC	$Z_{ik} = -1$	161	93	102
Event	$Z_{ik} = 1$	206		91

Table 2.5: *Testing the PPP hypothesis using boundary crossing events. The convergence probability is* 0.5500 *which is not significant at the usual significance levels. The p-value is 0.1722.*

Based on the data, the unit root hypothesis cannot be rejected. Therefore our test does not support the PPP hypothesis. This result is in-line with some of the findings in the literature while it contradicts others. For example, Pesaran (2007) found that the CIPS test does not reject the unit root hypothesis for the same dataset. On the other hand, using a similar dataset, the test of Chang (2002) strongly rejects the unit root hypothesis.

Our findings are robust to the parameter settings. In particular, as shown in Table 2.6, applying narrower or wider boundaries results in a similar conclusion.

Note that some caution may be needed when interpreting this result. In particular, the presence of autocorrelation, which is discussed in the next

L_i	U_i	convergence probability	p-value
-1	1	0.5419	0.1796
-1.18	1.18	0.5500	0.1722
-2	2	0.5749	0.1116

Table 2.6: *Sensitivity analysis for the BCC test on the PPP hypothesis. The null hypothesis of unit root cannot be rejected even if applying narrower or wider boundaries.*

section, may influence this finding. Thus, further development of the test may be needed to fully confirm this result.

2.5 Discussion

In this section, we discuss several additional issues. First, we analyze the role of individual effects. We continue by briefly discussing methodological issues. Next, we analyze how to deal with autocorrelation and finally, we conclude with a brief discussion on the large sample properties.

2.5.1 The Role of Individual Effects

Let us begin by analyzing what happens if the individual effects are not zero under the null hypothesis of unit root. Table 2.7 summarizes the necessary additional notations.

Thus, δ captures the effect of μ_i. Let us substitute $E_{12} = (0.5 - \delta) \times B_T \times (1 - p_{22})$ and $E_{21} = (0.5 + \delta) \times B_T \times p_{21}$ in Equation(1.34).

$$p_c^{\#} = \frac{(0.5 - \delta) \times B_T \times (1 - p_{22}) + (0.5 + \delta) \times B_T \times p_{21}}{B_T} \tag{2.15}$$

	Cumulative Upper minus Lower Crossing	
	$Y^D_{ik-1} < 0$	$0 < Y^D_{ik-1}$
$Z_{ik-1} = -1$	$E_{11} = (\frac{1}{2} - \delta) \times B_T \times (1 - p_{21})$	$E_{12} = (\frac{1}{2} + \delta) \times B_T \times (1 - p_{22})$
$Z_{ik-1} = 1$	$E_{21} = (\frac{1}{2} - \delta) \times B_T \times p_{21}$	$E_{21} = (\frac{1}{2} + \delta) \times B_T \times p_{22}$
Total	$(\frac{1}{2} - \delta) \times B_T$	$(\frac{1}{2} + \delta) \times B_T$

Table 2.7: *The role of individual effect for the panel BCC test.*

Simplifying yields

$$p_c^{\#} = \frac{1}{2} + \frac{1}{2}(p_{22} - p_{21}) + \delta(1 - (p_{21} + p_{22})). \tag{2.16}$$

Under the unit root hypothesis, assuming away from autocorrelation for the time being, $p_{21} = p_{22}$. As for the individual effect of Equation (2.1), there are three cases.

1. If $\mu_i = 0$, then $\delta = 0$ and $p_c^{\#} = 0.5$.

2. If $\mu_i > 0$, then $\delta > 0$ and $(p_{21} + p_{22}) > 1$. Hence, $p_c^{\#} < 0.5$.

3. Finally, if $\mu_i < 0$, then $\delta < 0$ and $(p_{21} + p_{22}) < 1$. Hence, $p_c^{\#} < 0.5$.

To conclude, if the null hypothesis assume that $\mu_i = 0$ and in the true process, $\mu_i <> 0$, then the actual size of the BCC test will be less than the nominal size. Note that this is similar to other unit root tests which are based on the Dickey-Fuller asymptotics since in this case, they use the DF-statistics when in reality, they should be using the standard OLS t-statistics. In both cases, the actual inference[9] is unlikely to be negatively affected by the potential misspecification of the individual effect in the null hypothesis because economics theory, most of the time, postulates that a

[9]This issue is discussed further in the next subsection.

stochastic process is either stationary or unit root. If the true process is unit root with a drift then it is easier to identify the lack of stationarity.

Let us continue by analyzing the stationary case. Here, the critical assumption is that X_{it} starts from minus infinity which, for simplicity, is quantified as $X_{i0} = \mu_i/(1-\alpha_i)$. What happens if the process does not start from minus infinity? Substituting $X_{i0} = X_{i0} + \sum_{j=1}^{t-1} \Delta X_{ij}$ to Equation 2.1 results

$$X_{it} = \mu_i + \alpha_i(X_{i0} + \sum_{j=1}^{t-1} \Delta X_{ij}) + u_{it} \tag{2.17}$$

Assuming that $X_{i0} = (\mu_i + \gamma_i)/(1-\alpha_i)$, where γ_i captures the difference between the process initial value and its long-term mean, Equation (2.17) can be reformulated as follows.

$$X_{it} = \mu_i + \alpha_i(\frac{\mu_i + \gamma_i}{1-\alpha_i} + \sum_{j=1}^{t-1} \Delta X_{ij}) + u_{it} \tag{2.18}$$

Substituting $\mu_i = \mu_i \times (1-\alpha_i)/(1-\alpha_i)$ and $\widetilde{X}_{it} = X_{it} - X_{i0}$ results

$$\widetilde{X}_{it} = -\gamma_i + \alpha_i \sum_{j=1}^{t-1} \Delta X_{ij} + u_{it}. \tag{2.19}$$

Thus, if the process is assumed to start at minus infinity, that is $\gamma_i = 0$, then it is free of individual effects in the sample period. Consequently, $\delta = 0$ in Equation (2.16) and the convergence probability captures only the difference in p_{11} and p_{12}. On the other hand, if the process starts far away from the long-term mean, that is $\gamma_i <> 0$, then the individual effect is not zero, thus $\delta <> 0$ and this effects the convergence probability.

Note that depending on the initial value, X_{i0}, the stochastic process may exhibit fairly different behavior in finite samples, which is illustrated in Figure 2.1. In the first subdiagram of Figure 2.1, the stochastic process does not appear to exhibit mean-reverting behavior. Since the BCC test is

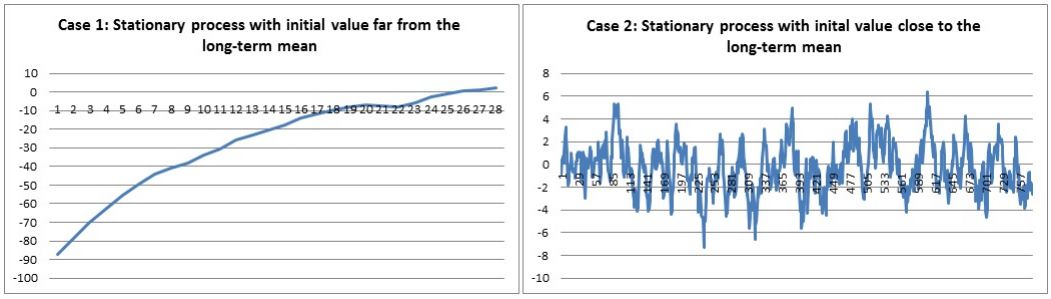


Figure 2.1: *Stationary process with an off-equilibrium and near-equilibrium initial value. The DGP for both cases is as in Equation* (2.1) *with* $\alpha_i = 0.9$. *In the first case, the stationary process is started from an off-equilibrium position while in the second case, the first observation is close to its long-term mean.*

designed to measure the tendency to return to the mean, it would not reject the null hypothesis in this case. Note that economic theory often postulates the case shown in the second subdiagram of Figure 2.1: It not only predicts that a stochastic process is stationary, but the theory often also implicitly infers that it is close to its long-term mean. Hence, the BCC test incorporates this implicit assumption as well.

To conclude, the BCC test essentially quantifies the tendency to return to the mean. If the null hypothesis is accepted, then there is no such tendency, which implies that the DGP is either unit root process, or a stationary process for which the initial value is far away from the process's long term mean. In other words, the stochastic process has been initiated from an out of equilibrium position and it has not been measured for a sufficiently long time period for the process to return to its long-term mean. The BCC test is less suitable for differentiating between these two cases. If the null hypothesis is rejected, then there is a tendency to return to the mean and hence, the DGP is likely to be stationary near to its long-term mean.

2.5.2 Methodological Issues

In this chapter, we mirror closely the typical assumptions made by the panel unit root tests, such as in Im et al. (2003), Chang (2002) or in Breitung and Pesaran (2008), including their strengths as well as their weaknesses. Potentially, there are several meaningful research questions, for example:

1. Is the DGP unit root without deterministic trend or stationary process without deterministic trend?

2. Is the DGP unit root with drift or trend-stationary?

Although the second question is fully relevant, in this paper, we only deal with the first question due to space constraints. In the literature, it is common to begin the discussion with the first case as seen in Im et al. (2003), Chang (2002) or in Breitung and Pesaran (2008). Additional cases may be discussed in a separate paper.

When dealing with the first question, we make two implicit assumptions. First, we assume that the deterministic trend is excluded based on economic theory as well as based on the nature of the problem being modeled. Moreover, we assume that economic theory predicts stationarity. These implicit assumptions are claimed to be true for many commonly tested hypothesis such as the PPP hypothesis.

Using Carl Popper's terminology, in order to test any theory, one needs to try to falsify it. Thus, in order to test for stationarity, we need to assume a non-stationary null hypothesis. There are many options for choosing this null hypothesis. We considered the one without individual effect because among all the non-stationary models described by Equation (2.1), it has the highest convergence probability. If $\mu_i \neq 0$ then as explained in Equation (2.16), the convergence probability is less than half. Consequently, if we reject the null

hypothesis under the model with no individual effect then we would reject it in cases of positive or negative individual effect as well.

On the other hand, accepting the null hypothesis implies that there is at least one realistic non-stationary model which is supported by the data. Hence, economic theory suggesting stationarity cannot be confirmed. (It cannot necessarily be falsified either due to potential issues related to statistical power but that is a separate issue.) Of course, there may be infinite many non-stationary models which could be rejected. However, using Popper's terminology, in order to falsify the economic theory, it is sufficient to show one realistic counter-example.

2.5.3 Boundary Crossing Counts and Autocorrelation

Let us discuss the case when there is autocorrelation in the error term in Equation (2.1). Typically, such as in Bai and Ng (2002), Choi (2006), Im et al. (2003), Chang (2002), Levin et al. (2002), Maddala and Wu (1999) and Pesaran (2007), panel unit root tests are based on individual ADF tests and the number of lags are estimated from the data. This can safely be done as (Said and Dickey (1984)) the variable of interest in the ADF test and in the DF test have the same limiting distribution even in the case when the number of lags, m, is unknown, if $m^3/T \to \infty$ as $T \to \infty$. This result holds under more general conditions as well, as discussed in Chang and Park (2002). Alternatively, as in Moon and Perron (2004), instead of using the ADF test, one may incorporate the lag structure into the factor model as well.

In case of BCC test, based on Equation (2.16), autocorrelation effects the convergence probability in finite samples. In particular, in the case of positive autocorrelation, the convergence probability is less than 0.5 while in

the case of negative autocorrelation, the convergence probability is greater than 0.5 under the null hypothesis in finite samples. Thus, autocorrelation may induce size distortion. Similarly, in the case of stationary process, autocorrelation may induce loss of power.

We continue by outlining two potential solutions to the problem of autocorrelation in finite samples. Note that we do not provide detailed resolution due to space constrains. The first option is to filter out the autocorrelation of the error term in Equation (2.1) and continue by carrying out the BCC test on the autocorrelation-adjusted data.

In order to implement this option, the first step is to assume that the autocorrelation structure can be captured by the following model.

$$\Delta X_{it} = \mu_i + (\alpha_i - 1)X_{it-1} + \rho_{i1}\Delta X_{it-1} + \rho_{i2}\Delta X_{i,t-2} + ...\rho_{im_i}\Delta X_{i,t-m_i} + \epsilon_{it} \quad (2.20)$$

where ϵ_{it} is free of autocorrelation and m_i is the number of lags. The convergence probability can be restored to 0.5 by first estimating the autocorrelation structure and second by defining the counting process over the autocorrelation-adjusted differences as shown in Equation (2.21).

$$\widetilde{X'}_{it} = \begin{cases} 0 \text{ if } t \leq m_i \\ \sum_{j=m_i+1}^{t} \Delta X_{ij} - (\rho_{i1}^E \Delta X_{it-1} + ... + \rho_{im_i}^E \Delta x_{i,t-m_i}) \text{ for } t > m_i. \end{cases} \quad (2.21)$$

where $\rho_{i1}^E, ...\rho_{im_i}^E$ are the estimated autocorrelation coefficients. This option may be problematic because, besides the usual issues such as selection of lags as discussed in Harris (1992), it requires a parametric estimation which is somewhat alien to the original nonparametric philosophy of the test.

Alternatively, we can account for potential autocorrelation via the contingency table. The idea is similar to what has been discussed in section

two, but instead of analyzing the relationship between the full history of the process, Y^D_{ik-1}, and the next boundary crossing event,Z_{ik}, as in Table 2.1, here, we analyze the relationship between the "immediate history", represented by the last, Y^U_{ik-1}, or last few, boundary crossing events, and the next boundary crossing event.

More precisely, the condition $Y^U_{ik-1} = 1$ implies that the DGP's immediate history was characterized by positive shocks. If there is no autocorrelation, this information should not affect the next boundary crossing event.

$$p(Z_{ik} = 1|Y^U_{ik-1} = 0) = p(Z_{ik} = 1|Y^U_{ik-1} = 1); \tag{2.22}$$

Likewise, the lower crossing probabilities below should also not depend on the previous boundary crossing events:

$$p(Z_{ik} = -1|Y^U_{ik-1} = 0) = p(Z_{ik} = -1|Y^U_{ik-1} = 1); \tag{2.23}$$

By combining these two equations, we obtain the following equality under the null:

$$\begin{aligned} &p(Z_{ik} = 1|Y^U_{ik-1} = 0) + p(Z_{ik} = -1|Y^U_{ik-1} = 1) = \\ &= p(Z_{ik} = -1|Y^U_{ik-1} = 0) + p(Z_{ik} = 1|Y^U_{ik-1} = 1) \end{aligned} \tag{2.24}$$

For ease of notation, let us introduce an additional variable which describes the effect of the process's immediate history.

$$A_{ik} = -1 \text{ if } \begin{cases} Z_{kt} = 1 \text{ and } Y^U_{ik-1} = 0 \text{ or} \\ Z_{kt} = -1 \text{ and } Y^U_{ik-1} = 1. \end{cases} \tag{2.25}$$

Also, the events of the right hand side of Equation (2.24) are denoted as follows:

$$A_{ik} = 1 \text{ if } \begin{cases} Z_{kt} = 1 \text{ and } Y^U_{ik-1} = 1 \text{ or} \\ Z_{kt} = -1 \text{ and } Y^U_{ik-1} = 0. \end{cases} \tag{2.26}$$

Using these notations, the null hypothesis, which is somewhat analogue with the null hypothesis of no autocorrelation in the parametric case, stating that the stochastic process's immediate history does not affect its next realization, can be described as follows:

$$H_0 : p(A_{ik} = 1) = p(A_{ik} = -1) = \frac{1}{2} \tag{2.27}$$

From now on, we refer to $p(A_{ik} = 1)$, as autocorrelation probability. Under the alternative hypothesis of autocorrelation, these probabilities are no longer equal, the stochastic process's immediate history has an effect on the next event.

$$H_1 : p(A_{ik} = 1) \neq p(A_{ik} = 0) \tag{2.28}$$

More precisely, in case of positive correlation, $p(A_{ik} = 1) > p(A_{ik} = 0)$, while in case of negative autocorrelation, $p(A_{ik} = 1) < p(A_{ik} = 0)$. All this is summarized in Table 2.8.

		Last Boundary Crossing Event	
		$Y^U_{ik-1} = 0$	$Y^U_{ik-1} = 1$
Next BC Event	$Z_{ik} = -1$	$A_{ik} = 1$	$A_{ik} = -1$
	$Z_{ik} = 1$	$A_{ik} = -1$	$A_{ik} = 1$

Table 2.8: *Contingency table describing the effect of autocorrelation.*

In case of the parametric approach, one can include additional lags into the autocorrelation structure. In this nonparametric, state-based approach, it is also possible to include additional states. Assuming we find significant autocorrelation probability in the above-described first step, we can check for additional autocorrelation by adding an additional states. For example, we may ask if

$$p(Z_{ik} = 1|Y^U_{ik-1} = 0, Y^U_{ik-2} = 0) \stackrel{?}{=} p(Z_{ik} = 1|Y^U_{ik-1} = 0, Y^U_{ik-2} = 1)? \tag{2.29}$$

The null hypothesis in this case would state that once we have controlled for the immediate history of the process by controlling for Y^U_{ik-1}, the additional history represented by $Y^U_{ik-2} = 0$ does not affect the next boundary crossing probabilities significantly, while the alternative hypothesis would state that the additional history is of importance.

Finally, let us discuss how to combine the effect of the process's immediate and full history. More precisely, let us assume that we have found significant autocorrelation probability in the first state but did not find a significant relationship in the second state. Table 2.9 summarizes the potential states for this case.

<table>
<tr><td colspan="2"></td><td colspan="3">Cumulative Upper minus Lower Crossing</td><td colspan="2"></td></tr>
<tr><td colspan="2"></td><td>$Y^D_{ik-1} < 0$</td><td>$Y^D_{ik-1} = 0$</td><td>$0 < Y^D_{ik-1}$</td><td colspan="2"></td></tr>
<tr><td rowspan="4">Next
Boundary
Crossing
Event</td><td>$Z_{ik} = -1$</td><td>$C^1_{ik} = -1$</td><td rowspan="4">Non
Informative</td><td>$C^1_{ik} = 1$</td><td rowspan="2">$Y^U_{ik-1} = 0$</td><td rowspan="4">Previous
Boundary
Crossing
Event</td></tr>
<tr><td>$Z_{ik} = 1$</td><td>$C^1_{ik} = 1$</td><td>$C^1_{ik} = -1$</td></tr>
<tr><td>$Z_{ik} = -1$</td><td>$C^2_{ik} = -1$</td><td>$C^2_{ik} = 1$</td><td rowspan="2">$Y^U_{ik-1} = 1$</td></tr>
<tr><td>$Z_{ik} = 1$</td><td>$C^2_{ik} = 1$</td><td>$C^2_{ik} = 1$</td></tr>
</table>

Table 2.9: *Contingency table describing the logic of the BCC unit root test in case of autocorrelation.*

The remaining steps are almost identical to what has been described above for the case of no autocorrelation. The null hypothesis states that:

$$H_0 : p(C^1_{ik} = 1) + p(C^2_{ik} = 1) = p(C^1_{ik} = -1) + p(C^2_{ik} = -1), \tag{2.30}$$

while alternative hypothesis states that

$$H_1 : p(C^1_{ik} = 1) + p(C^2_{ik} = 1) > p(C^1_{ik} = -1) + p(C^2_{ik} = -1). \tag{2.31}$$

In case of cross-sectionally independent error terms, the distribution of the test statistics can be calculated by using the fact that the sum of two

binomial distributions is also binomial. In case of cross-sectional dependence, the simulation-based methods can be used. Further elaborating on the method outlined above is well beyond the scope of our paper. At this stage, we can conclude that potentially, we can adjust for autocorrelation in a nonparametric manner as well.

2.5.4 Large Sample Properties

Here, we briefly discuss the large sample properties of the BCC test under the assumptions stated in the second section. The structure of the problem in our case differs, to a certain extent, from a typical asymptotic analysis. Here, the properties of test statistics depend basically on the number of restarts, that is on the number of Bernoulli trials, which only indirectly depends on the sample size of the original data.

Fundamentally, we analyze the data in two steps. The first step is to characterize the original data using boundary crossing events. This step converts the original data (having some unknown distribution) to random variables having Bernoulli distribution. The second step is to estimate the convergence probabilities based on the resampled data.

The properties of this second-step estimator depend on the properties of the boundary crossing events, namely, on its dependence structure and on its sample size. These properties, in turn, depend both on the original data as well as on the counting procedure, specifically on how the boundaries are selected and how the counting is carried out. The structure of the problem is summarized in Table 2.10.

Let us begin by the first case, when the error terms in Equation (2.1) are independent. In this case, the necessary condition for consistency requires that the resampled data's sample size converges in probability to infinity

	Original DGP		Counting		BC events		Estimator of $p(C_{ik}=1)$
	Sample size	Dep. of u_{it}	Restart	Boundaries	Sample size	Dep. of C_{ik}	
1	$T \to \infty$ or $N \to \infty$	Ind.	Full	Constant	$B_T \to \infty$	Ind.	Consistent
2	$T \to \infty$ and N is finite	Strong crosss-dep.	Full	Constant	$B_T \to \infty$	Dep.	Consistent, as dep. LLN applies
3	$T \to \infty$ and N is finite	Strong cross-dep.	Rand.	Constant	$B_T \to \infty$	Ind.	Consistent
4	$N \to \infty$ and T is finite	Strong cross-dep.	Full	Constant	$B_T \to \infty$	Dep.	Not consistent, dep. LLN does not apply
5	$N \to \infty$ and T is finite	Strong cross-dep.	Rand.	Constant	B_T is finite	Ind.	Not consistent, sample size is finite
6	$N \to \infty$ and T is finite	Weak cross-dep.	Full	Constant	$B_T \to \infty$	Dep.	Consistent, as dep. LLN applies
7	T, N finite	Ind.	Full	$b \to \infty$	$B_T \to \infty$	Ind.	Probably inconsistent

Table 2.10: *What drives the large sample properties of the BCC test?*

when the underlying data converges to infinity. This is ensured if the probability of observing a boundary crossing event is positive and either $N \to +\infty$ and T is finite, or $T \to +\infty$ and N is finite, finally when $N \to +\infty$ and $T \to +\infty$ regardless of how N/T behaves.

There are basically two methods to deal with the problem of cross-dependence. The first method (cases 2, 4 and 6 in Table 2.10) is to carry out the counting the usual way, that is to restart the counting process immediately after the boundary crossing event has been observed. As a result, the resampled data will consist of dependent Bernoulli trials. The consistency of the convergence probability estimator in this case depends on whether the

law of dependent large number applies, or not.

While, in a way, independence is unique, dependence comes in many different forms. Consequently, there are many different dependent LLNs, as found in Andrews (1988), Hansen (1991) or De Jong (1995). The fundamental idea behind these theorems is essentially very similar: if the dependence between the observations decrease sufficiently quickly as the distance between them increases then the LLN applies.

Such a decrease in dependency occurs in the second case due to the increasing time-distance and in case 6 when the error terms are by definition weak-dependent. This latter case may be used to model spatial dependence when the dependence between the cross-sections decreases as physical distance increases. On the other hand, the law of dependent LLN would probably not apply in case 4 when there is strong cross-dependence. Such strong cross-dependence may arise as a result of common unobserved factors.

So far, we have used general dependent LLN. Alternatively, it may be possible to make use of the specific law of large numbers for dependent Bernoulli trials. To our knowledge, such law does not exists under general specification, but special models have been analyzed. For example, James et al. (2008) examines a special case when the success probability of the trials is conditioned on the total number of successes achieved up to that point. Unfortunately, this model cannot be directly adapted to the case of BCC test as the counting procedure induces effects which are not captured by this model. Yet, a more general model along these lines may provide further insights on the large sample properties in the future.

The second method to deal with cross-dependence, shown as case 3 and 5 of Table 2.10, is to alter the counting procedure in a way that the resulting

counting events consist of independent Bernoulli trials. The idea is as follows:

- Start the counting procedure at the first cross section.
- Assuming that we observe a boundary crossing event in time T_1^*, we do not restart the procedure for this cross-section but continue the counting on the next cross section. More precisely, we continue the counting on $\widetilde{X}_{2t} = \sum_{j=T_1^*+1}^{t} \Delta X_{2j}$.
- Likewise, each time we observe a boundary crossing event, the counting continues in the next cross section. Naturally, if the number of cross sections are finite and the counting on the last cross section is finished, then the counting continues in the first cross section again.
- Finally, if we do not observe boundary crossing for some cross-section, the counting also continues in the next cross section. Simply, such cross sections are ignored.

This procedure re-establishes the independence of the Bernoulli trials as the resampled variables are calculated based on observations which come from different time-periods and hence which are independent. Therefore, the only remaining necessary condition for consistency is to make sure that the counting does not stop so that the sample size of the resampled data goes to infinity as the sample size of the original data goes to infinity. This condition holds for example in case 3 but does not hold in case 5, when the time dimension is not sufficiently large.

Finally, let us examine case 7, which is interesting from a theoretical point of view. In this case, the sample size in the original data is finite, but $b \to 0$ where $U_i = b \times \sigma_i$ and $L_i = -U_i$. First of all, in order to be able to carry out the counting, we would need to assume that the underlying

DGP is continuous from which we can obtain an infinitely fine sample. The dilemma is as follows. If $b \to 0$, then the conjecture[10] is that the sample size of the resampled data goes to infinity. Thus, we would be able to estimate the convergence probability with arbitrary precision. Such precision would contradict statistical intuition, as it is unlikely that the limitation posed by the lengths of the observation period could be overcome by resampling.

2.6 Conclusion

In this paper, we introduce and analyze a new unit root test which is based on counting boundary crossing events. We detail two versions of the test. The first is designed for cross-sectionally independent panel data. The second is able to take into account cross-sectional dependence as well. Our Monte Carlo studies find that in small samples, the proposed tests are more powerful than existing unit root tests in cases where the error term has t-distribution.

During the analysis, we identified two opportunities for further improvements. First of all, our method can very well be combined with principle component analysis. Also, the method for choosing boundaries may be improved further. These ideas may be pursued further in additional papers.

[10]In order to verify this hypothesis, the first step would be to examine the limit of the first exit time distribution as the boundary goes to zero. The limiting function is likely to be similar to the Dirac delta function. The second step would be to show that the sample size of the resampled data converge in probability to infinity for any finite T, which would follow from the fact that boundary crossing counting distribution essentially involves repeated convolutions as discussed in the first chapter.

DOI: 10.14754/CEU.2016.01

Chapter 3

Portfolio Choice Without Distributional Assumptions Using Boundary Crossing Counts

Abstract

We solve the portfolio choice problem without distributional assumptions by extending the use of state-dependent rebalancing to nonparametric settings. We propose a specific, state-dependent rebalancing and show how it is related to the Kelly criterion. Under this rebalancing, the full distribution of the portfolio's terminal value can be approximated by a well-behaving, discrete, probability distribution based boundary crossings. When applied to parametric specifications under transaction costs, the method replicates the baseline results of the geometric Brownian motion. As for nonparametric applications, first, we show that the log-optimal allocation in the USA was a leveraged purchase, next we find that leveraged returns were significantly

different in various epochs. We conclude by explaining how this newly proposed method can be used for density forecast.

3.1 Introduction

The optimal portfolio choice is an important problem for theoreticians, for financial practitioners as well as for any non-professional with an investment decision to make. The history of security markets as well as the recent financial crisis highlight the importance of relying on robust assumptions and techniques when dealing with financial markets.

In this chapter, we discuss how to solve the portfolio choice problem without assuming a parametric process for the returns of the assets. We exogenously assume a specific, state-dependent, constant, double boundary based rebalancing mechanism similar in spirit to the one used in the literature on optimal inattention and transaction costs.[1] Then, we show that finding the optimal strategy within this family needs only the characterization of the portfolio's boundary crossing probability and frequencies. The terminal value's distribution can be characterized by a well-behaving, discrete, probability distribution called boundary crossing counting distribution. Finally, we find the optimal trading strategy within this family using this distribution.

Our focus is to find the optimal portfolio weights without assuming a parametric process for the returns. Hence, we do not analyze rebalancing behavior in detail. Rebalancing is exogenous in our model: we assume a constant, double boundary based rebalancing, which may or may not be

[1]Optimal inattention originates from Baumol (1952) and Tobin (1956), their method was later was taken up in the literature on transaction costs starting by Constantinides (1986) and Dumas and Luciano (1991).

optimal. Also, we exogenously assume that the entire portfolio is rebalanced at the same time, when the change in the portfolio's value since the last rebalancing reaches some pre-defined, critical level. We provide a heuristic argument for selecting the optimal portfolio weights from a pre-defined set of options given this rebalancing rule if the returns are nonparametric, defined by data, and investor's utility is either constant relative risk aversion, CRRA, or logarithmic, potentially constrained by the VAR of the terminal value.

The proposed approach has several advantages. As we discuss later, boundary crossing counting distributions can be estimated directly from data, therefore we can derive the optimal portfolio weights directly, without parametric assumptions on the returns. Consequently, we can incorporate many features typical of financial data, such as fat tails. Moreover, besides finding the optimal portfolio, we can also characterize its evaluation in time through density forecasts as we characterize the portfolio's full distribution, not only its expected value. Consequently, we can also calculate value at risk, VAR, limits directly from data, without simulations. Finally, we can easily deal with important practical issues, such as transaction costs, the issue of ruin conditions for leveraged positions, or the cost of margin financing. In particular, under certain not very restrictive assumptions, the method allows us to analyze leveraged positions and short-selling in discrete time under zero ruin probability. This is especially important for today's economy since economic conditions have led to a record-high level in margin loans in the USA. It appears that many investors choose to hold leveraged positions.

Our approach has certain limitations. First of all, we implicitly assume that prices are continuous, therefore at this stage we do not allow for discontinuities. Also, by assuming that the portfolio can be rebalanced at

the boundaries, we essentially abstract away from execution risk.[2]

The specific, state-dependent rebalancing we propose shows that the Kelly criterion and the Merton's model essentially differ only in how they model rebalancing. Moreover, our method has both analytical and nonparametric applications. As for the former, we solve for the optimal portfolio under geometric Brownian motion analytically, without simulations, so as to find that results are similar to the baseline model of continuous rebalancing. Therefore, we gain more flexibility without losing the insights provided by the simpler approach. As for the latter, we show that the profitability of leveraged positions differs in the various epochs of the stock market in the USA. There are expansion and consolidation periods. In the former, leveraged positions are profitable and the log-optimal investment is a leveraged purchase. In the latter, the log-optimal allocation is close to the buy and hold strategy. The difference in the leveraged positions' profitability between these periods is closed to being statistically significant.

In the literature on optimal portfolio choice, authors often begin by specifying and estimating a parametric model for the returns. We choose to avoid this step as it has been proven notoriously difficult to come up with an accurate parametric model.[3] Parametric specifications usually describe the asset's returns almost everywhere, while we focus only on those points which drive the terminal value's distribution and these points do not require

[2] Incorporating discontinuities as well as execution risk is possible, yet would induce additional complexities and we do not see the benefit of going down this path at this stage.

[3] Classical papers such as Markowitz (1952), Merton (1969), Merton (1971), Samuelson (1969), Malkiel and Fama (1970), often assume that prices follow a geometric Brownian motion and abstract away from financial frictions. The GBM hypothesis is often rejected by the data as discussed by Lo and MacKinlay (1999), by Cont (2001) or by Campbell and Thompson (2008).

a parametric specification.

The closest approach to ours is the one described in the paper of Dumas and Luciano (1991) which, under a different specification including only two assets and a parametric specification for the return process, shows that if the portfolio is rebalanced each time the ratio of the risk-free and the risky asset falls below or exceeds the optimally chosen boundaries, then it is optimal to choose fixed boundaries. They also provide a heuristic argument for determining the level of these fixed boundaries. A similar state-dependent mechanism was applied by Balduzzi and Lynch (2000), who show that if we take into account predictability and heteroscedasticity, then the optimal boundaries are no longer constant. Also, under another, fairly different, set of assumptions including dynamic consumption and rebalancing between transaction accounts and investments, Abel et al. (2013) discuss the merits of time-dependent versus state-dependent rebalancing. As pointed out by Dumas and Luciano, and elaborated further by Buss and Dumas (2013), many of these models suffer from a logical quasi-inconsistency as they assume an exogenous process (geometric Brownian motion which is an infinite variation process) which is inconsistent with the optimization problem itself: it is not clear how a universe populated by participants trading based on state-dependent rules would result in such an infinite variation process. As we solve the problem without the GBM assumption, this quasi-inconsistency does not directly apply to our case.

Nonparametric studies rely on time-dependent adjustment, such as Brandt (1999), Brandt (2003) or the universal portfolio approach invented by Cover (1991), which has been extended, for example, by Blum and Kalai (1999) to be able to incorporate transaction costs, and further explored by Gyorfi and Vajda (2008) and by Horváth and Urbán (2011). State dependent

adjustment, up to our knowledge, has only been used in parametric studies. Here, we extend the use of state-dependent rebalancing to nonparametric specifications.

The chapter is structured as follows. The second section begins by describing the portfolio choice problems under infrequent adjustment. Next, we explain how to use boundary crossing counting processes (BCC processes) to characterize these problems. We begin the application by solving the well-known problem of simple portfolio choice under geometric Brownian motion analytically, without simulations. Here, we also compare ours to the standard solution. Next, we continue with the nonparametric specifications where we assume that the data generating process is well represented by the historical data of the USA. Thereafter, we discuss some additional technical details. Finally, the last section provides a balanced conclusion.

3.2 Rebalancing and BCC Distributions

In this section, we discuss how to solve the optimal portfolio choice under infrequent adjustment assuming a specific, state dependent, boundary crossing based rebalancing mechanism. After writing up the problem generally, we reformulate and solve it using boundary crossing counting processes.

3.2.1 Portfolio Choice Under Infrequent Adjustment

Under general settings, the optimal portfolio choice problem entails searching for the expected utility maximizing portfolio weights.

$$\begin{aligned} &\max_{W} \quad E(U(V_T(W,P))) \\ &\text{where} \quad w_{it} = \frac{q_{it} \times p_{it}}{V_t}, \qquad \sum_{i=1}^{N} w_{it} = 1 \text{ for every t.} \end{aligned} \tag{3.1}$$

where N is the number of securities, T is the investment horizon, W is a $T \times N$ matrix of the portfolio weights, $w_{it} \in W$ is a weight of a security i in time t, $E(.)$ indicates the expected value, $U(.)$ is a non-path dependent utility function, V_T is the portfolio's terminal value, P is a $T \times N$ matrix of the security prices, $p_{it} \in P$ is a price of a security i at time t and finally q_{it} indicates quantity, that is the number of shares held in the portfolio security i in time t. The weights, unless $w_{it} = 1$ or 0 for any t, are typically not constant in time. Hence, controlling for portfolio weights involves changing the quantities.

This act is typically referred to as rebalancing which may be continuous or infrequent. The former assumes that quantities are changed continuously in such a way which keeps the weights constant. Most of the papers in the field including both the mean-variance paradigm of Markowitz (1952) and the continuous time approach based on Bellman equation starting from Merton (1969), Merton (1971) and Samuelson (1969) take this path. Continuous rebalancing may be problematic in the presence of transaction costs as pricing processes are typically assumed to have infinite variations, hence even a small proportional transaction cost would result in infinite costs over any finite interval. Besides, securities are traded in discrete quantities in the exchanges, prices are discrete and evaluation of the historical performance is based on discrete observations.

Models of infrequent rebalancing can make up for some of these weaknesses. These mechanisms are often characterized as time-dependent or state-dependent. Both approaches can be described in the language of boundary crossing. Since rebalancing is infrequent, quantities are only adjusted occasionally, each time some rebalancing signal[4] X^*_{it} crosses either an upper or a lower boundary.

In the case of time-dependent rebalancing, the rebalancing signal is chosen to be the economic time[5], $X^*_{it} = t_{economic}$, and the upper boundary is an exogenously selected constant, for example one day in case of analyzing daily data. Also, rebalancing is typically simultaneous, $X^*_{it} = X^*_{jt}$. The lower boundary does not play a role as economic time is a non-decreasing function of time, hence it can be an arbitrarily negative number.

In the state-dependent case, the rebalancing signal differs from application to application. In case of optimal portfolio choice, Dumas and Luciano (1991) for example propose to rebalance based on the ratio of the risk-free asset and risky asset. A similar problem arises in case of modeling delta hedge in option pricing. Authors, for example Martellini and Priaulet (2002), typically propose to rebalance based on the option's delta.

In portfolio choice, time-dependent rebalancing is problematic for various reasons. First of all, portfolios including leveraged positions or short-selling are inadmissible over any discrete time interval if the underlying distribution describing the price change is unbounded as the possibility that prices drop arbitrarily close to zero in case of leveraged purchase or rise arbitrarily high in case of short-selling implies a positive probability that an investor cannot repay the loan. As bankruptcy would imply infinite disutility

[4]Note that the rebalancing signal is essentially a restarted process.

[5]The time passed between Friday and Monday is typically assumed to be equal to the time passed between Monday and Tuesday.

under many commonly used utility functions, such as CRRA utility or log-utility, investors will not choose such positions under these assumptions. If we want to model market participants who do choose to take leveraged positions or to sell short, then either we have to restrict the underlying distribution to be bounded, introduce some kind of insurance against ruin events or abandon time-dependent rebalancing altogether.

Moreover, the optimal behavior of an investor facing transaction costs is mostly state-dependent, therefore time-dependent rebalancing may overestimate transaction costs. Finally, under time-dependent adjustment, transaction costs introduce path-dependence, therefore, even if one finds the optimal allocation, it is challenging to characterize the portfolio's full distribution.

The state-dependent rebalancing mechanism we propose overcomes many of these issues. First, by assuming away from potential liquidity constrains and price-discontinuities, we can establish no-ruin conditions for any finite weights. As for transaction costs, the optimal behavior under parametric specification is mostly state-dependent. Assuming that the insights of the parametric case carries over to nonparametric specifications, we can probably get closer to the optimal case using state-dependent rebalancing strategies. Finally, as we describe the portfolio's value with a single discrete stochastic variable, we are able to characterize the optimal portfolio's full distribution. Consequently, we can carry out statistical inference or create nonparametric density forecasts.

3.2.2 Rebalancing Based on the Portfolio's Value

We begin by describing a rebalancing scheme based on the change in the portfolio's value, similar in spirit but different in focus to the one used in

the literature on optimal inattention. We assume that the entire portfolio is rebalanced at once, when the cumulative change in the portfolio's value since the last rebalancing reaches some critical level.

$$X_t^* = \frac{V'_{t^*+t'}}{V_{t^* \in T^A}} = \begin{cases} G_U & \text{for upper crossings} \\ G_L & \text{for lower crossings} \end{cases} \tag{3.2}$$

Also, we assume that boundaries are constant[6], $G_U > 1$ and $0 < G_L < 1$. In the formula above, as well as in the rest of the chapter, we need to differentiate between four different types of "time":

1. First, t indicates time in general, the economic time as it is normally used.

2. Next, $t^* \in T^A$ indicates the boundary crossing moments.

3. Moreover, $t^{'}$ indicates the time passed since the last boundary crossing, $t^{'} = t - max(t^*)$.

4. Finally, $T \in T^A$ indicates the investment horizon.

The chronology of rebalancing is as follows. At some time t, the rebalancing signal, X_t^*, reaches the critical level. The portfolio is rebalanced at $t+\epsilon$, where ϵ is sufficiently small so that $P_{t+\epsilon} \approx P_t$. Overall, the rebalancing occurs almost instantaneously. Thus, some ϵ time after the rebalancing event, the portfolio is perfectly balanced and the value of the rebalancing signal is one.

$$X^*_{t^*+\epsilon} = \frac{V_{t^*+\epsilon}}{V_{t^*}} = 1; \tag{3.3}$$

Thus, the portfolio's value can be approximated by a step function. Or to put it another way, this step function describes the portfolio's value

[6]Consequently, all our results can be considered as a lower bound for the case of full optimization where boundaries are non-constant.

for an observer who only observes the portfolio's value at rebalancing and approximates the portfolio's value for another observer who observes the portfolio's value in economic time. Since boundaries are constant, the change

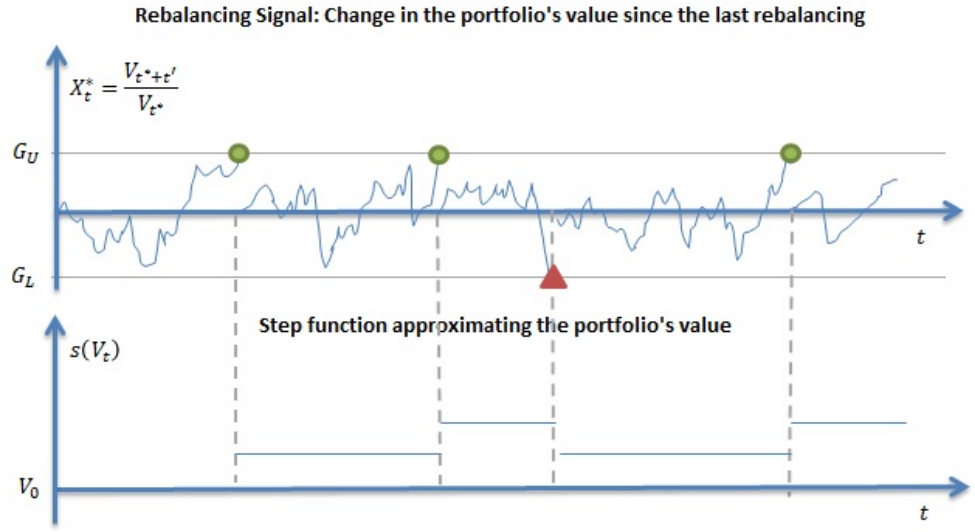


Figure 3.1: *Approximating the portfolio's value by a step function using state-dependent rebalancing. The portfolio is assumed to be rebalanced each time the change since the last rebalancing reaches some upper or lower boundary.*

in the portfolio's value upon boundary crossing can only take two discrete values. Consequently, the portfolio's terminal value at some investment horizon T can be described as follows:

$$V_T = V_0 \times G_U^{Y_T^U} \times G_L^{Y_T^L} \tag{3.4}$$

where V_0 is the portfolio's initial value and the stochastic elements of the problem are captured by boundary crossing counting distributions:

1. Y_t^U counts the number of upper crossing events: $Y_{t^*}^U = Y_{t^*}^U + 1$ if $X_{t^*}^* = G_U$ and $X_{t^*+\epsilon}^* = 1$.

2. Y_t^L counts the number of lower crossing events: $Y_{t^*}^L = Y_{t^*}^L + 1$ if $X_{t^*}^* = G_L$ and $X_{t^*+\epsilon}^* = 1$.

Let us introduce two additional boundary crossing counting distributions for further use.

1. $Y_t^A = Y_t^U + Y_t^L$ counts the number of upper and lower crossing events, that is it counts how many times the portfolio needs to be rebalanced altogether.

2. $Y_t^D = Y_t^U - Y_t^L$ described the difference between the upper and the lower crossing events.

Finally, Y_t is used to refer to these stochastic processes at once. In this formulation, no-ruin conditions only require finite weights for the risky assets as $V_T \geq V_0 \times G_L^k > 0$ for any k, where k is a natural number. Without loss of generality, we can normalize[7] the initial portfolio's value to one.

In order to finish the formulation, let us impose the $G = G_U = 1/G_L$ restriction[8] on the boundaries. Also, in the rest of this chapter, we use the G notation instead of the G_U and G_L when we want to emphasize that $G_U = 1/G_L$. Imposing this restriction on Equation (3.4) results in the following expression for the portfolio's value:

$$V_T = G^{Y_T^D} \tag{3.5}$$

Here, the randomness of the problem is captured using one stochastic

[7] It would also make sense to take into account the cost of entry by setting the initial values to $V_0 = \sum |w_i| \times (1 - tc)$. Here, we aim to analyze annualized returns over long horizons, therefore we abstract away from these initial costs.

[8] Our rebalancing mechanism essentially allows us to use tree-based pricing similar to the one used for pricing options. This restricting ensures that trees are recombining.

variable. Hence, the optimization problem can be described as follows:

$$\max_{W,G,T} \quad \sum_{i=-m}^{m} P(Y_T^D(W,G,T)=i) \times U(G^i) \tag{3.6}$$

where m describes the precision at which the infinite sum is evaluated and $P(Y_T^D(W,G,T) = i)$ is the probability that $Y_T^D(W,G,T) = i$. We have assumed that the investment horizon is also a decision variable[9] for the sake of completeness. Similarly to the parametric case, the optimization involves deciding which distribution we want to sample from. Here, however, the distribution is discrete, endogenous and it is the result of the choice on the control variables, T, W and G.

So far, we have assumed a general utility function, There is one utility function, the logarithmic, where the optimization is even simpler: it does not require to know the upper minus lower crossing counting distribution. Instead, it is sufficient to know its expected value. This is an advantage from a statistical point of view. Due to this advantage as well as other favorable properties[10], the rest of the chapter will mainly use the logarithmic utility as the base case. In finance, returns are mostly measured in annualized

[9]This possibility could for example be used to model retirement decisions.

[10] Logarithmic utility is considered to be an important benchmark case having many theoretically appealing properties as detailed by the large number of papers starting from Kelly (1956), and Breiman (1961), reviewed recently for example by Christensen (2005) or in MacLean et al. (2011). In particular it has been shown by Breiman (1961) and by Long (1990) that there exists a portfolio (growth-optimal portfolio or numeraire portfolio or log optimal portfolio) for which the price of any other portfolio denominated in the price of the growth-optimal portfolio becomes supermartingale. Furthermore, growth-optimal strategy maximizes the probability that the portfolio is more valuable than any other portfolio, therefore has a certain selective advantage as detailed by Latane (1959). Among all admissible portfolios, the growth-optimal portfolio minimizes the expected time needed to reach, for the first time, any predetermined constant as shown by Merton and Samuelson (1992). If claims are discounted using the growth-optimal portfolio, then expectation needs

growth rate for convenience. Therefore, the baseline optimization problem is as follows:

$$\max_{W_t} \quad \Big(E(\frac{\log(V_T)}{T}) = E(r_T) = E(\frac{Y_T^D}{T}) \times \log(G)\Big) \tag{3.7}$$

where r_T is the annualized growth rate. Furthermore, the expected value of the portfolio's return can be expressed as follows.

$$E(\frac{Y_T^D}{T}) \times \log(G)) = \frac{n}{T} E(\frac{Y_T^A}{n}(\frac{Y^U}{Y^A} - \frac{Y^L}{Y^A})) \times \log(G) \tag{3.8}$$

where n is the sample size, T is the holding period. For example, when analyzing a 10 year investment using daily data, $n = 2520$ and $T = 10$. By letting $q = E(\frac{Y^A}{N})$ be the frequency of boundary crossing and letting $p = E(\frac{Y^U}{Y^A})$ be the probability of upper crossing, Equation(3.8) can be rewritten as follows.

$$E(r_T) = \frac{n}{T} \times q \times \underbrace{(2p-1)}_{Kelly-ratio} \times \log(G) \tag{3.9}$$

This formulation connects the gambling literature's well-known formula of Kelly (1956), also discussed in more detail in Poundstone (2010), with the portfolio choice literature.

The intuition of the formulation above can be understood by considering the following analogy. The investor walks in to a casino where there are a large number of different tables each offering the option of doubling or losing the initial bet. These tables differ in three aspects:

- The speed at which the game is played which corresponds to q,

to be taken with respect to historical probability measures as explained in Long (1990) or Bajeux Besnainou and Portait (1997) therefore these portfolios may provide a unifying framework for asset prices as shown by Platen (2006).

- the probability of winning, represented by p,
- and finally the amount of stake required, which is $\log(G)$ in our formulation.

At which table will the log-optimal investor sit? Regardless of the table chosen, the optimal bet for $p > 0.5$ is $(2p - 1)$, the amount to win or lose at each round is $\log(G)$. The games, in average, are repeated $n \times q$ times. Since the investor has only a limited amount of time to play, which corresponds to the investment horizon, T, he will choose the table at which the expected amount of gain combining these three factors is at its maximum.

Accordingly, the Kelly criterion and Merton's classical solution for the optimal portfolio choice under logarithmic utility essentially differs in the rebalancing behavior. The former assumes infrequent, state-dependent rebalancing while the latter operates with continuous rebalancing. Hence, our approach allows us to represent the portfolio choice problem as a repeated game, and this allows us to solve the problem without having to build a parametric model for the asset returns.

Overall, we have shown that if we can obtain the boundary crossing counting distribution, which is discussed in detail in the previous chapter, then we can also solve the optimal portfolio choice without assuming a parametric process for the returns.

3.3 Applications

In this section, we consider the case of optimal portfolio choice with two asset classes. We begin by analytically solving the simple portfolio choice when there is only one single risky asset whose price follows geometric Brownian motion (GBM), and a single composite (risk-free) asset which does not pay

interest. Besides being a frequently used pricing model, the GBM is a good starting point as it has analytical solutions for both the newly introduced boundary crossing based rebalancing as well as for the standard continuous case, which is very helpful in validating our method. We continue with the nonparametric case where we analyze the profitability of leveraged positions in the various epochs of the USA.

3.3.1 Analytical Solutions for Simple Portfolio Choice Under Transaction Cost and Geometric Brownian Motion

Obtaining an analytical solution requires a simplifying assumption: In order to be able to work with constant boundaries, we abstract away from the interest paid on the risk-free asset. Moreover, we define the counting process over $\log(X_t^{GBM})$, that is $X_t^*(\log(X_t^{GBM}), U, L, 0))$. For constant boundaries over Brownian motion, the first exit time distribution as well as the upper crossing probability can be calculated analytically. In order to be able to calculate the expected return, we need to express how much the portfolio's value changes upon boundary crossing, meaning that we need to express G_U and G_L as a function of the boundaries.

The logic of the derivation is as follows. We begin by exogenously defining the lower boundary, L. Next, we express G_L from L and G_U from G_L since $G_U = 1/G_L$. Finally, we express U from G_U. Note that as a result of transaction costs, the boundaries are slightly asymmetric, that is, $U \neq -L$.

The lower boundaries are chosen exogenously, $L = 7 \times \sigma$. In this case, as shown in the appendix, the change in the portfolio's value upon lower

crossing can be described as follows:

$$G_L = \frac{1 + w \times \exp(L) \times (1 + pct) - w}{1 + w \times ptc}, \tag{3.10}$$

where ptc is the proportional transaction cost. Since $G_U = 1/G_L$, the change in the portfolio's value upon upper crossing can be described as follows:

$$G_U = \frac{1 + w \times ptc}{1 + w \times \exp(L) \times (1 + ptc) - w}. \tag{3.11}$$

Finally, the change in the portfolio's value upon upper crossing determines the value of the upper boundary

$$U = log(\frac{(1 + w \times ptc)^2 - (1 - w)(1 + w \times exp(L) \times (1 + ptc) - w)}{(1 + w \times exp(L) \times (1 + ptc) - w) \times w \times (1 + ptc)}), \tag{3.12}$$

The first exit time distribution for the Brownian motion with variance normalized to one can be obtained by substituting L and U into the formula of Borodin and Salminen (2002):

$$fet(t', L, U) = e^{-ms^2 t'/2}(e^{msL} ss(U, U-L)dt + e^{msU} ss(-L, U-L)dt') \tag{3.13}$$

where $ss(.)$ is the theta function and $ms = \frac{\mu}{\sigma}$. Using Equation (3.13) and the recursion described in Equation (1.10), we can obtain the Y_T^A.

Substituting the scale function of the Brownian motion results in the following formula for upper boundary crossing probability.

$$P(X_{t^*}^{BM} = U) = \frac{1 - \exp(-L \times \frac{2\mu}{\sigma^2})}{\exp(-U \times \frac{2\mu}{\sigma^2}) - \exp(-L \times \frac{2\mu}{\sigma^2})} \tag{3.14}$$

Using Equation(3.13) and Equation (3.14) as detailed in Equation (1.15) and in Equation (1.16), we can characterize the upper minus lower crossing distribution. Substituting this distribution and G from Equation (3.11) to Equation (3.7) characterizes the solution.

The analytical solution for logarithmic utility and continuous rebalancing – detailed briefly in the appendix – is well-known, it is derived under more general utility by Merton (1969) put into perspective for example, by Peters (2011). Figure 3.2 compares the expected growth rate for BCC-based and continuous rebalancing policies using the diffusion parameters estimated based on the closing prices of the Dow Jones Industrial Average between 1928 and 2012 by the standard maximum likelihood method, detailed, for example, in Gourieroux and Jasiak (2001), resulting in $\mu^{ML} = 0.0437$ and $\sigma^{ML} = 0.1849$.

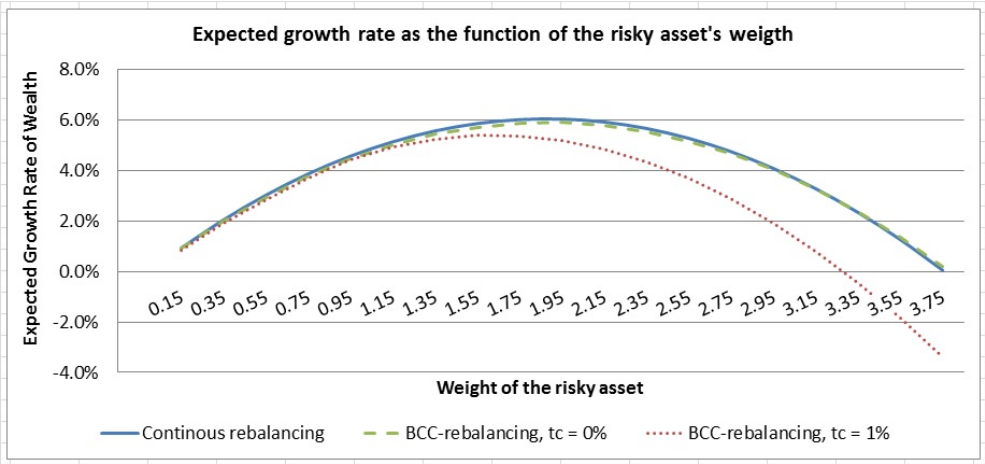


Figure 3.2: *Continuous and boundary crossing based rebalancing under geometric Brownian motion. The two approaches result in an almost identical solution in case there is no transaction cost. The BC-based approach is able to take into account proportional transaction costs as well.*

The diagram reveals little difference between the continuous and the BCC-based rebalancing under the setup without transaction costs. Therefore, replacing the continuous rebalancing of the Mertons model with the boundary crossing based one ceteris-paribus does not lead to significantly different result. Hence, the new approach keeps the intuition of the simpler model;

results are not driven by the BCC-based rebalancing using exogenous boundaries placed at seven standard deviation distance. This means that we can extend the analysis to nonparametric specifications without losing the insights provided by the simpler parametric case.

Although logarithmic form offers certain advantage in the BCC-framework, yet the method can be adopted without much difficulty to other utility functions as well. For example, assuming utility has CRRA form, the optimal investment can be approximated as follows:

$$E(\frac{V_{3.96}^{1-\theta}}{1-\theta}) = \sum_{i=-150}^{150} P(Y_T^D = i) \times \frac{G^{i\times(1-\theta)}}{1-\theta} \tag{3.15}$$

where θ is the risk aversion coefficient. Evaluating this equation for portfolio weights in the [0.15, 0.25...3.75] range and selecting the weight where the utility is at its maximum provides us with the following approximation for optimal weight.

Rebalancing	tc	Risk Aversion Coefficient				
		3	3.5	4	4.5	5
Continuous	0%	0.63	0.54	0.47	0.42	0.38
Infrequent,	0%	0.65	0.55	0.45	0.45	0.35
BC-based	1%	0.65	0.55	0.45	0.45	0.35

Table 3.1: *Optimal portfolio weights under CRRA utility and geometric Brownian motion for continuous and boundary crossing based rebalancing. The two approaches result in an almost identical solution.*

Once again, besides the approximation error which could be reduced by applying a finer grid, the results under BC-rebalancing are similar to the baseline continuous case. Transaction costs do not influence the results significantly in this case since rebalancing is not very expensive for

non-leveraged portfolios.

For both utility functions described above, our solution is comparable with that of Dumas and Luciano. Both solutions are analytical yet they provide a closed form solution while here, we provide an algorithmic approximation. Both models assume that rebalancing is infrequent, state-based, yet we do not assume that investors are optimally inattentive.

3.3.2 Simple Portfolio Choice Under Nonparametric Data Generating Process

We are now in the position to discuss the optimal portfolio choice if the data generating process is unknown but well represented by historical data. For simplicity, we discuss the simple portfolio choice, where the DGP has two components: for risky assets, we use the $S\&P500$ gross total return index as obtained from Shiller's website for the period 1934-1988 and 1988 onwards, we use Chicago Board of Trade's TRI. As for the composite asset, we use the secondary market three months treasury bill rate as obtained from FRED's database under ticker TB3MS as the reference rate. Naturally, we implicitly assume that the dependence between these two assets is also well represented by the historical DGP.

We obtained the cost of financing by an educated guess based on Fortune (2000), Fortune (2001) papers; we assume that deposits are 10 basis point lower than the reference rate but at least zero, borrowing costs 200 basis points above the reference rate. Regarding further data-related issues, when combining data sampled with different frequencies, we adjust to the highest frequency using Brownian Bridge for risky assets and step-function for the composite assets.

Furthermore, we use logarithmic utility in this subsection. The point

estimation is obtained using the counting estimator. Transaction costs are assumed to be 1%. The change in the portfolio's value upon lower crossings are defined as follows:

$$G_L = \frac{1 + w \times exp(7 \times \sigma) \times (1 + 0.01) - w}{1 + w \times 0.01} \tag{3.16}$$

where $\sigma = 0.1849$ is the estimated standard deviation for this period. We set $G_U = 1/G_L$. We select the optimal portfolio weight from a predefined range of [0.15, 0.25, ... ,5].

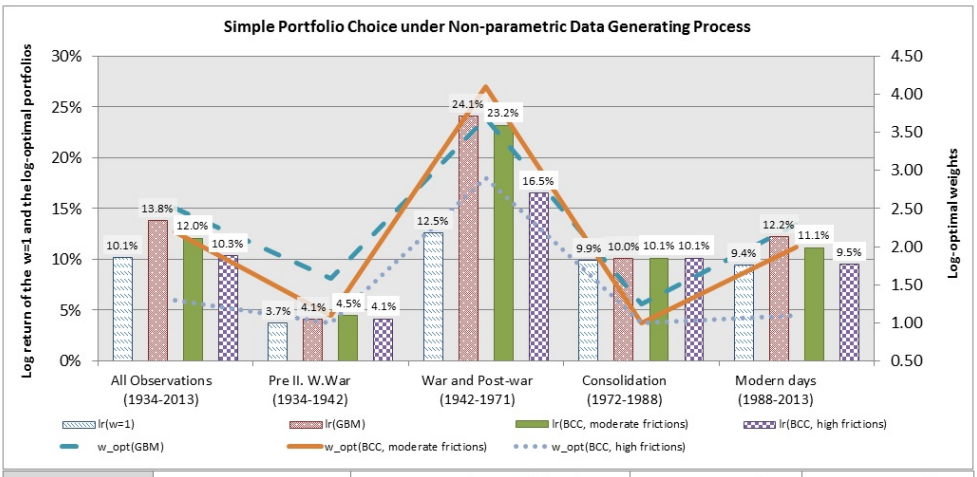
Diagram 3.3 shows the log-optimal allocation in major epochs of the USA's stock markets:

- The first epoch starts[11] in 1934 and finishes with the attack on Pearl Harbour in 1941.
- The next period finishes with Nixon's Shock in 1971, which is often considered as the date of abolishing the gold standard.
- The third epoch starts 1971 and finishes in 1987. This is the epoch of consolidation characterized by the oil-price shock as well as monetary stabilization policies.
- Finally, the last period is the modern days, which we categorize from 1987 to 2013.

Note that the years are rounded to the nearest integer when shown in Diagram 3.3. For example, the second epoch ends on December 7, 1941, but it is rounded and displayed as 1942.

Diagram 3.3 reveals three interesting observations. First, the GBM under continuous rebalancing gives very similar optimal weight and returns

[11]We could not including the 1929 crisis because the time series for the short term rate started in 1934.



	DGP			GBM, Continous rebalancing				BCC, ptc =1%, mf = 2%		BCC, ptc =2%, mf = 4%	
	From	To	lr(1)	r^e	σ^{ML}	w*	lr(w*)	w*	lr(w*)	w*	lr(w*)
All Observations	1934	2013	10.11%	8.16%	3.26%	2.50	13.79%	2.20	12.03%	1.30	10.33%
Pre- II. W.War	1934	1942	3.70%	4.94%	3.12%	1.59	4.06%	1.10	4.45%	1.00	4.09%
War and Post-war	1942	1971	12.54%	11.75%	3.19%	3.68	24.07%	4.10	23.16%	2.90	16.47%
Consolidation	1972	1988	9.86%	4.22%	3.38%	1.25	10.04%	1.00	10.08%	1.00	10.08%
Modern days	1988	2013	9.40%	7.61%	3.32%	2.30	12.16%	2.00	11.15%	1.10	9.48%

Figure 3.3: *Log-optimal allocation (right axis) and the corresponding annualized returns (left axis) in the various epochs of the USA's stock market. The lr(1) is the log return in case the weight for the risky asset is one, r^e is the excess return and σ^{ML} is the estimated standard deviation. Also, w* is the optimal weight and lr(w*) is the log-optimal return. Finally, ptc is the proportional transaction cost and mf is the cost of margin financing. The data between 1934 and 1987 is from Shiller which has monthly frequency. The data from 1987 onwards is from CBOT which has daily frequency. The years are rounded to the nearest integer. For example, the second period starts on December 7, 1941, but it is rounded and displayed as 1942.*

as the nonparametric BCC with 1% transaction cost. It appears that despite its simplicity, the GBM model operates reasonably well. Also, relatively large transaction costs are needed to deviate the optimal weight considerably from those of the baseline GBM model. Second, the profitability of leveraged

positions varies greatly in the different epochs. There are periods, the pre-war period and the consolidation period, when leveraged positions are not very profitable and the log-optimal allocation is close to the $w = 1$ strategy. There are also expansion periods, when leveraged positions are profitable and the log-optimal investment involves extensive use of leverage. Third, the leveraged returns in some of these epochs appear to be higher than one would expect. For instance, between 1942 and 1971, over 20% annual return[12] prevailed for almost 30 years. Therefore, leveraged equity investments appear to be "too profitable."

How do these findings fit into the literature? Some authors argue that the log-optimal portfolio weight for the risky asset should be close to one. For example Peters (2011) argues that *"it is unlikely that the simple strategy of borrowing money and investing it in the S&P500 would outperform the market. It is equally unlikely that investing only part of ones money in the S&P500 would outperform the market."* Yet, we show that in certain epochs, leveraged returns outperform the market even after taking into account transaction costs as well as the cost of margin financing.

Our findings are similar in spirit to those of the equity premium literature which originates from Mehra and Prescott (1985). There are however some technical differences in measurement. The equity premium literature typically compares the real return on the market index to the risk-free returns. Here, we compare the log-optimal and the market returns. Based on the results of the equity premium literature, our findings are somewhat less surprising. It is actually reasonable that if equity premium

[12]George Soros, considered to be among one of the most successful hedge-fund managers, returned roughly 20% return to investors annually, according to internet sources: http://www.investopedia.com/financial-edge/0912/buffet-vs.-soros-investment-strategies.aspx.

exists then leveraged equity investments are more profitable than one would expect but only to a certain degree.

Authors suggest many explanations on why equity premium prevails, as summarized in Mehra and Prescott (2003). The closest in spirit to what one sees in Diagram 3.3 are those articles which emphasize the difference between ex-ante and ex-post measurements. The equity premium as well as the log-optimal weight and the corresponding returns are ex-post measurements while only ex-ante difference would be truly puzzling. Indeed, Fama and French (2002) found that the average returns were higher than expected in the second half of the twentieth century.

Why do ex-ante and ex-post figures differ? Stock prices are under several types of risk, namely diffusion risk, jump risk and the risk of parameter uncertainty. It has been shown by Santa-Clara and Yan (2010) using option-price implied risk and Gabaix (2012) using calibrated macro model that unrealized events, jumps or disasters, which could have happened but did not happen may cause such a difference.

The next section investigates whether the profitability of diversified investments is constant in time or not. In other words, we test for structural breaks. The existence of such breaks would explain why an unexpectedly high ex-post leveraged return can prevail for an extended period of time. If structural breaks create different regimes and if investors do not know in advance which regime will emerge in the future, then the ex-post measurement does not reflect the regime uncertainty. Consequently, ex-post figures such as the one analyzed in the equity premium literature or the one described in Diagram 3.3 may be higher than one would expect.

In particular, we analyze whether the profitability between the second period is different to the other periods. Under the null hypothesis,

investments in the period between 1942 and 1971 did not have significantly higher returns. Under the alternative hypothesis, investments in this period had significantly higher returns. Since the annualized growth rates are not invariant to the portfolio weights, we carry out the tests under two different set of weights. The mechanism[13] for calculating the test statistics is as follows.

1. Based on the observations ranging from 1934 to 1942 and from 1971 to 2014, we estimate the full distribution of the annualized log returns for a holding period of 29.58 years using the method outlined in the first chapter under the subsection "Recursive estimation using first exit time distribution." Specifically, we first estimate the first exit time distribution and the upper crossing probability from the data. We next obtain the upper and lower crossing distribution. Finally, we obtain the annualized log returns from the upper minus lower crossing distribution.

2. Next, we calculate the probability that the annualized returns observed in the second period come from this distribution.

3. Finally, we reject the null hypothesis if the p-values calculated in the previous step are lower 0.05.

The results in Figure 3.4 reveal that returns on leveraged investments in the second period, between 1942 and 1971, are likely to be significantly higher than in the complementary periods. Hence, investors are likely to face regime-switching risk. This may explain why ex-post leverage premium prevails for an extended period of time. Indeed, it has been shown for example by Maenhout (2004) that ex-ante demand for a robust investor who also takes

[13]This mechanism was greatly improved based on the review of Timo Teräsvirta.

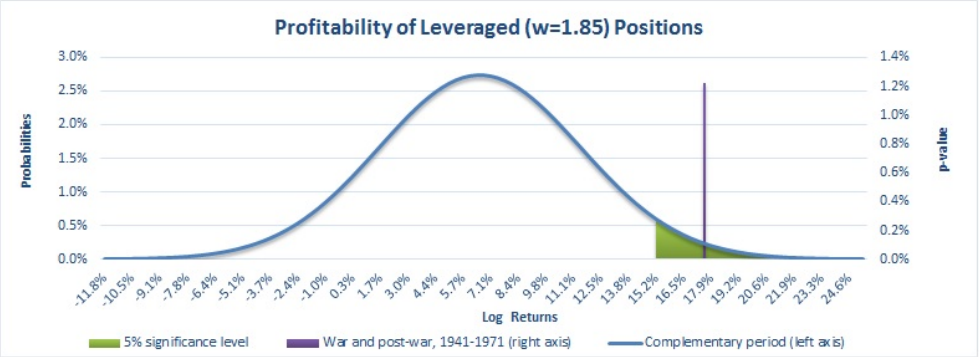


Figure 3.4: *Testing the null hypothesis of no structural breaks in the USA stock market. The distribution is calculated based on the data between 1934 and 1942 and between 1971 and 2013. The point estimation is calculated based on the data between 1942 and 1971. The null hypothesis assumes that the observed log-returns between 1942 and 1971 come from this distribution. The null hypothesis is rejected at 5.0% significance level.*

into account the risk of parameter uncertainty is significantly lower than the ex-post demand under no parameter uncertainty.

Table 3.2 below shows how the choice of the parameters influences these results.

From Table 3.2, it is clear that only leveraged returns have been significantly higher which essentially implies that the excess return over the risk-free rate were significantly higher. Naturally, further investigation would be needed to identify the source of the differences in the ex-post returns. Here, we only have space to mention a few possible explanations. The lower returns of the pre-war period may be explained by the stock market crash of 1929. Basically, the public may have lost interest[14] in investments. The high

[14]Historical records characterize this period as "boring." It is interesting to notice that Jesse Lauriston Livermore, the famous speculator committed suicide in the pre-war period, in 1940.

w	L	U	p-value
1.85	-4	4	0.8%
1.85	-5	5	1.2%
1.85	-6	6	1.7%
0.55	-4	4	21.7%
0.55	-5	5	12.8%
0.55	-6	6	13.8%

Table 3.2: *Sensitivity analysis for the test on structural breaks in financial data. Regardless of how boundaries are chosen, the null hypothesis of no structural breaks is rejected for leveraged positions. On the other hand, the null hypothesis cannot be rejected for a conservative portfolio where the weight of the risky asset is 0.55. Hence, it is the excess return which was significantly higher between 1942 and 1971.*

returns of the war and the post war period may be explained by favorable real economic conditions such as high GDP growth, low unemployment and low inflation.

It may also be the case that the relatively high log-optimal returns between 1942 and 1971 was a one time event. It may have been the result of credit constraints and it is possible that due to the lack of financial infrastructure, companies could not obtain an optimal financial structure. Hence, investors could gain from leveraging. As a result of financial innovation, the introduction of high-yield bonds, etc., this imperfection has been eliminated. This possibility is somewhat supported by for example Saretto and Tookes (2013), who show that companies with traded CDS contracts on their debt are able to maintain higher leverage ratios.

3.3.3 Extension: Nonparametric Density Forecast

Since the BCC technique allows us to describe the complete distribution of the portfolio's value, not just its expected value, and we are also capable of producing density forecasts. Moreover, our method does not suffer from some of the commonly encountered density forecast pitfalls. In particular, we do not need to assume Gaussian innovations. Also, we do not need to capture the dependence between the returns with simplified models. The following diagram shows how an unleveraged, conservative portfolio's density evolves in time.

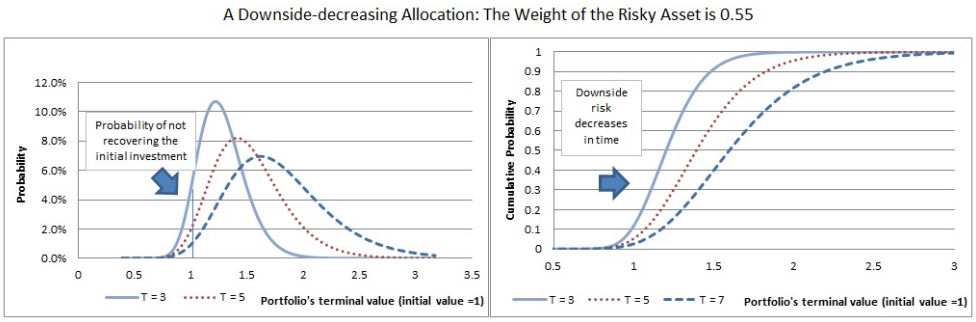


Figure 3.5: *Density forecast for non-leveraged allocation if the data-generating process is as the S&P total return index between 1934 and 2013.*

This diagram complements the debate on whether the optimal fraction of wealth invested into stocks is horizon-dependent or not. For the conservative portfolio shown in the diagram above, the probability of not recovering the investment at least in nominal term decreases in time. For more leveraged allocation, it may not be the case. Overall, there are downside decreasing and downside increasing allocations. Hence, if an investor's preference includes downside risk, which is highly possible otherwise guaranteed investment funds would not be so popular, then

investment rules are horizon-dependent.

This observation is in line with common wisdom, described by Malkiel and Fama (1970), which states that the broker's typical recommendation is a horizon-dependent one: "The longer period over which you can hold on to your investment, the greater should be the share of common stocks in your portfolio". Early academic papers such as Samuelson (1969) and Merton (1971) have often derived horizon-independent rules which go against this common wisdom. Consequently, authors, for example Brennan et al. (1997), Liu and Loewenstein (2002), Cocco et al. (2005), Brandt (2009), finally Wachter and Yogo (2010), proposed many adjustments to these early models, such as time-varying investment opportunities, time-varying parameters, transaction costs, borrowing constrains and human capital, predictability of dividend's growth or heterogeneity in wealth in order to explain horizon-dependence.

Here, we complement these findings by noticing the difference in the downside risk's evolution: in order to explain horizon-dependence, it is sufficient to assume that older investors are more concerned with preserving their wealth, at least in nominal terms, and put more emphasise on downside risk, while younger investors are more focused on the potential upside. Besides noticing this fact, we also provide a method to solve such problems.

Figure 3.5 also explains why one may argue that long-run investments are less risky. Risk is often measured as the portfolio's standard deviation which indeed increases over time. However, if we choose to measure risk with some other measures, for example by the probability that the investment does not lose value, at least in nominal terms, then the risk may indeed be decreasing over time. Overall, the answer depends on how we measure risk.

3.4 Discussion

Here, we discuss some additional details. First, we briefly discuss volatility clustering before outlining the logic of optimization under nonparametric DGP.

3.4.1 Volatility Clustering

At this stage, we solve the problem without explicitly considering volatility clustering. Our method however could be used to solve the portfolio choice problem under volatility clustering which may be covered in a separate paper. The main idea is as follows.

Similarly to conditional/unconditional volatility, one can differentiate between conditional upper crossing probabilities and unconditional upper crossing probabilities. Conditional upper crossing probabilities (which could be used to obtain conditional portfolio weights) are not constant in time. This is due to conditional heteroscedasticity, in high volatility periods, upper crossing probabilities are lower as described in Equation (3.17) and in Equation (3.18). Hence, conditional optimal portfolio weights are not constant either.

More specifically, optimal conditional portfolio weights in high-volatility periods are likely to be lower than the average. Likewise, the optimal portfolio weights are likely to be higher than average in low volatility periods. I will try to confirm this hypothesis in a separate paper.

3.4.2 Factors Affecting the Optimization

In our formulation, the randomness of the problem is captured by the upper minus lower crossing distribution. The optimization takes into account three

factors: the investment horizon, the exposure to risky asset(s) and boundary selection. The remainder of this section illustrates the role of these factors for simple portfolio choice, when the risk-free asset does not pay interest, the investment horizon is approximately 4 years (1000 observations) and finally the risky asset's data generating process (further referred as DGP) is the geometric Brownian motion (GBM) with annualized moments of $\mu = 0.0642$ and $\sigma = 0.1849$, which also implies, as we shall see later, that $p > 0.5$.

Investment horizon

The effect of ceteris paribus changing the investment horizon is trivial. The number of boundary crossing events increases, while the upper boundary crossing probability is unaffected. Consequently, the distribution of the terminal value shifts right in time, and widens.

Exposure to risky asset

The effect of increasing exposure to risky asset is similar yet ceteris paribus increasing exposure not only increases the number of boundary crossing events but also decreases the upper boundary crossing probability. Intuitively, increasing the weight of the risky asset increases the first moment linearly and the second moment quadratically, therefore larger weight increases the randomness of the stochastic process. This can also be shown formally for Ito processes. As explained in Karlin and Taylor (1981), the boundary crossing probability is typically expressed using scale functions:

$$p = P(X_{t^*} = U) = \frac{S(X_0) - S(L)}{S(U) - S(L)} \tag{3.17}$$

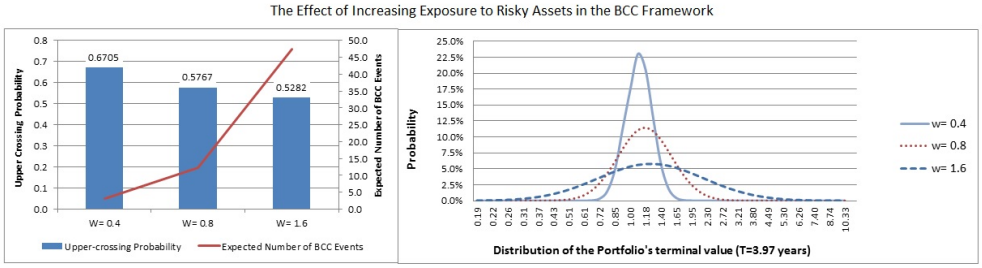


Figure 3.6: *The relationship between the exposure to risky assets and the portfolio's terminal value in the BCC framework*

where $t^* \in T^A$ indicates a boundary crossing moment, $S(.)$ is the scale function, X_0 is the initial value, U is the upper boundary and finally L is the lower boundary. As for the scale function,

$$S(X) = \exp(-\int^X \frac{w \times 2\mu(y)}{w^2 \times \sigma^2(y)} dy) = \exp(w^{-1} \times -SX) \tag{3.18}$$

where, $mu(.)$ and $\sigma^2(.)$ are the infinitesimal moments and $\int^x \frac{2\mu(y)}{\sigma^2(y)} dy = S(X)$. In other words, $S(X)$ indicates the scale function taken at some variable, X. The lower limit of the integrals does not play a significant role thus is omitted in accord with the literature.

Equation (3.18) essentially shows that once the process has been appropriately scaled, the probability of upper (or lower) boundary crossing depends only on the initial values relative distance from the lower and upper boundaries. The derivative of the boundary crossing probability with respect to the risky asset's weight is equal to:

$$\frac{\partial p}{\partial w} = \frac{e^{\frac{S(L)+S(U)-S(X)}{w}}}{\left(e^{\frac{S(L)}{w}} - e^{\frac{S(U)}{w}}\right)^2 w^2} \times (e^{\frac{S(X)}{w}}(S(L)+S(U)) + e^{\frac{S(U)}{w}}(S(L)-S(X)) + e^{\frac{S(L)}{w}}(-S(U)+S(X))) \tag{3.19}$$

where $S(L)$ is the scale function taken at a lower limit and SU is the scale function taken at the upper limit. The derivative is always negative as the

denominator and the first term of the numerator is trivially positive while the second term of the numerator is negative due to Jensen's inequality as $-S(L) + S(U) + S(L) - S(X) - S(U) + S(X) = 0$ and the exponential function is convex. Consequently, the upper boundary crossing probability is a decreasing function of the risky asset's weight, w.

This finding is very intuitive. For a pure random process with zero drift, the probability of upper crossing is 0.5, while for a pure deterministic, monotone increasing process, the probability of upper crossing is one. Therefore, by increasing the weight of the risky asset w, we increase the randomness of the process and hence we decrease the probability of upper boundary crossing. Essentially, we balance between how many times we collect the gain or the loss and the probability of collecting a gain instead of a loss. Overall, when choosing optimal portfolio weight for the risky asset under exogenously given rebalancing rules, the tradeoff is to have a higher number of rebalancing and a lower probability of favorable boundary crossing, or lower number of rebalancing and a higher probability of favorable boundary crossings.

Boundary selection

A change in the boundary structure alters three factors simultaneously:

- First, widening the boundaries trivially decreases the number of boundary crossing events.
- Next, it increases the magnitude of the gain or losses at boundary crossing.
- Finally, it increases the upper boundary crossing probabilities.

The overall effect combines these factors. Essentially, the choice on the set of boundaries determines the bins in the histogram describing the portfolio's terminal value. If boundaries are narrowed, then the histogram will consist of more bins. Yet, under typical circumstances, such as 1% proportional transaction costs or usual market volatility, it will not describe a fundamentally different distribution.

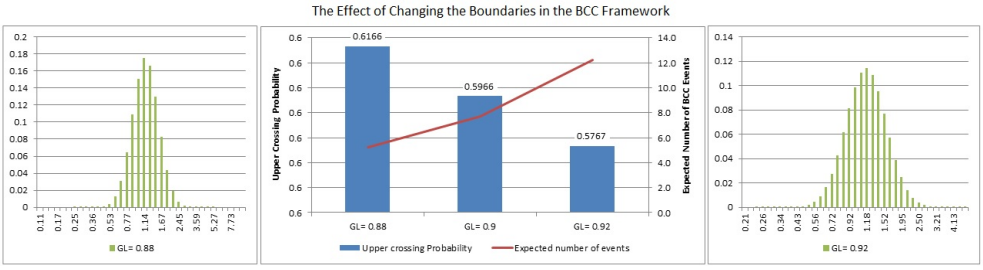


Figure 3.7: *The relationship between the rebalancing boundaries and the portfolio's terminal value in the BCC framework*

In the parametric domain, researchers are mostly interested in optimizing between the effect of transaction costs and the frequency of rebalancing. For nonparametric applications however, both financial and statistical considerations need to be taken into account.

Optimization based on the data itself presents some challenges in comparison to the parametric case mainly because the properties of the DGP is not defined axiomatically. Rather, these properties are captured by stylized facts as in Cont (2001) or is Teräsvirta and Zhao (2011). One of the stylized fact relevant to this topic is as follows:

Rama Cont: "Aggregational Gaussianity *As one increases the time scale Δt over which returns are calculated, their distribution looks more and more like a normal distribution. In particular, the shape of the distribution is not the same at different time scales. "*

Formally, Cont discusses how the distribution of $\Delta X_t = X_t - X_{t-c}$ depends on c. He claims that as c increases the distribution of ΔX_t looks more and more like a normal distribution. Although this statement is not explained in detail by Cont, it appears to be plausible since financial data is essentially tick by tick data aggregated over some c period. As c increases, so does the average number of observations being condensed into one data point. The number of these observations is likely to influence the properties of the data.

Naturally, similar things occur when widening the boundaries. In our dissertation, the time scale over which ΔX_t is calculated is random:

$$T \approx Y^A \times c \tag{3.20}$$

where T is the time dimension, Y^A is the number of boundary crossing events. Consequently, if boundaries widen then $E(Y^A)$ decreases, $E(c)$ increases. As $E(c)$ increases, the distribution looks more and more normal by Aggregational Gaussianity. Of course, we implicitly assume that Aggregational Gaussianity also applies to our sampling scheme.

Besides, the sampling frequency of the data limits the choice on the boundaries. For tighter boundaries, one also needs finer sampling. For analyzing small changes, one may need intra-day data, while for wider boundaries, daily data may suffice.

From what has been said above, choosing optimal boundaries is more involved than choosing optimal weight for the risky asset as the choice needs to take into account many different aspects. Reconsolidating all these aspects is well beyond the scope of our work and we do not see the benefit of favouring one aspect over the other. Therefore, boundaries are chosen in a way in which all of these aspects are acknowledged to some extent.

To conclude, optimization in the BCC framework essentially involves

calculating the expected utility derived from the portfolio's terminal value, where the probabilities can be obtained from BCC distributions. A change in the risky asset's weight alters the distribution of the portfolio's terminal value. The utility function, depending on the level of risk aversion[15] penalizes negative outcomes, which is reflected in the expected value. The optimal value is found by searching for those values where the expected utility is maximal.

3.5 Summary of Chapter 3.

In this chapter, we have introduced a new, nonparametric, technique to solve the optimal portfolio problem. This is therefore a methodological study which aims to provide an approach through which problems related to portfolio choice can be modeled under less restrictive assumptions. The novelty of our work is twofold, as we introduce a new method and we obtain some new results using this method.

As explained earlier, this is a methodological study. Therefore, the content of this chapter is definitely not intended to be received as an investment advice. Moreover, it is also not meant to be an advice to portfolio managers on how to rebalance. Rather, it advises quants on how to model rebalancing.

It is important to note that rebalancing is almost always an exogenous technical device in financial models. For example, the Black-Scholes model should not be interpreted as advice for option traders on how to hedge options. Nor does it try to describe how options are hedged in practice. It actually delivers advice to quants on how to calculate the value of an option.

[15]Note that not all utility functions have well-defined solution, risk-neutral utilities for example may result in corner solutions.

This dissertation is similar in this regard.

On the other hand, papers on endogenous rebalancing such as Constantinides (1986) or Dumas and Luciano (1991) actually find that under parametric settings, constant boundaries-based rebalancing is optimal. Assuming that their insights carry over to nonparametric cases, it is not necessarily a bad idea to rebalance in a way described in this dissertation. Yet from a methodological point of view, rebalancing assumptions does not necessarily need to be realistic or fully optimal.

The new method we introduced finds that by assuming a specific, state dependent rebalancing, we can solve the portfolio choice problem without parametric assumptions on the return process. Hence, we can incorporate many important features of financial data, such as fat-tails or volatility-clustering. Furthermore, we can deal with important practical issues, such as proportional transaction costs or the cost of margin financing. Moreover, our method describes the portfolio's value using a discrete stochastic variable, which allows us to calculate not only the expected value, but the full distribution as well. This is useful in nonparametric density forecasts and statistical inference. Finally, the method can also be used to solve the baseline model of geometric Brownian motion analytically, hence the model generalizes the standard solution without losing the insights provided by the simpler cases.

As for the results of this method, historical data of the USA's stock market suggests that not particularly extensive leveraged purchase of a diversified stock index is log-optimal therefore leveraged purchase does not imply risk-seeking behavior: risk averse investors may also rely on this technique if their risk-aversion is not too high. There are epochs, for example between 1942 and 1971 when leveraged positions are probably more profitable

than in consolidation periods, such as the one between 1971 and 1987.

Interpretation of these findings can be done from both a theoretical and a policy level. As for the former, researchers starting perhaps with Bogen and Krooss (1960) often argue that the level of leveraged positions play an important role in financial stability. Our contribution is to show that there is a tendency in USA's stock market to favour those, who hold leveraged positions, and these leveraged positions may also be held by risk-averse, log-optimal investors. As for the latter, some authors, for example Shiller (2005), or Hardouvelis and Theodossiou (2002) argue that the FED should return to more active margin policy, such that between 1934 and 1974. Based on our findings, log-optimal investment strategies may involve leveraged purchases. Thus, these policies may result in reduction of leveraged positions held by log-optimal investors.

The method offers many additional opportunities. For example, volatility clustering at this stage is not fully taken into account. Our method however could be used to solve the portfolio choice problem under volatility clustering. The idea is that we now work with constant upper crossing probabilities. However, this approach could be refined by predicting upper crossing probabilities using conditional variances. This may be covered in a separate paper.

3.6 Appendix on the Counting Procedures

The following section discussed the relationship between the boundaries and the change in the portfolio's value upon boundary crossing. We consider two separate moments of time. The first moment is right after rebalancing, the second moment lies between rebalancing moments. Variables describing the

second moment are indicated with $(.)'$. Also, q^j indicates the quantity of stock j right after rebalancing. The change in the portfolio's value between these two periods can be described as follows.

$$V' = V + \Pi - TC + DBS \tag{3.21}$$

where Π is the profit or loss, TC is the transaction cost while DBS, which abbreviates deposits, borrowing, short-selling, describes the cost of financing. Let us substitute each element one by one. Assuming that the total number of asset is m, the profit of the loss between two periods can be described as follows.

$$\Pi = \sum_{j=1}^{m} q^j \times \Delta p^j = \sum_{j=1}^{m} \frac{w^j V}{p^j} \times \Delta p^j = V \times \sum_{j=1}^{m} w^j \times (\exp(lnr^j) - 1) \tag{3.22}$$

where lnr^j indicates cumulative log change of asset i since the last boundary crossing. Since there is no change in the quantity between two boundary crossing events, the profit is the weighted average price change. The cost of financing is the cumulative interest received or paid, which can be obtained directly from the data.

$$DBS = V \times dbs; \tag{3.23}$$

Finally, the transaction cost is equal to:

$$TC = \sum_{j=1}^{m} abs(q'^j - q^j) \times p'^j \times ptc = \sum_{j=1}^{m} (q'^j - q^j) \times p'^j \times ptc^j \tag{3.24}$$

where the proportional transaction cost, $ptc^j = ptc$ if $q'^j > q^j$ and $ptc^j = -ptc$ otherwise. Substituting out quantities yields:

$$TC = \sum_{j=1}^{m} (\frac{w'^j \times V'}{p'^j} - \frac{w^j \times V}{p^j}) \times p'^j \times ptc^j \tag{3.25}$$

Simplifying yields

$$TC = V' \times \sum_{j=1}^{m} w'^{j} \times ptc^{j} - V \times \sum_{j=1}^{m} w^{j} \times ptc^{j} \times \exp(lnr^{j}) \tag{3.26}$$

Substituting Π, DBS and TC results

$$V'(1+\sum_{j=1}^{m} w'^{j} \times ptc^{j}) = V \times (1+\sum_{j=1}^{m} w^{j} \times (\exp(lnr^{j})-1)+dbs+\sum_{j=1}^{m} w^{j} \times ptc^{j} \times \exp(lnr^{j})) \tag{3.27}$$

From which the rebalancing signal, that is the change in the portfolio's value is equal to:

$$X^{*}_{t^{*}} = \frac{V'}{V} = \frac{1 - \sum_{j=1}^{m} w^{j} + dbs + \sum_{j=1}^{m} w^{j} \times \exp(lnr^{j}) \times (1 + ptc^{j})}{1 + \sum_{j=1}^{m} w'^{j} \times ptc^{j}} \tag{3.28}$$

The following technical assumption on the timing of the payments concludes the derivation. In particular, we assume that interest on margin financing as well as for short-selling is collected at the beginning of the period while interests paid on deposits are accounted at the end of the period. For lower-crossing events, the log change is calculated using the closing price and the next period's minimum price while for upper-crossing events the closing price and next period's maximum price. It is assumed that the rebalancing signal is initiated at one, $X_0 = 1$. Then, the following steps needs to be repeated for each number of observations.

1. Increment the value of *dbs* by the coming period's borrowing and shortselling costs.

2. Calculate the critical price level at which lower crossing event would occur.

3. Update the log change using the minimum price.

4. Evaluate if a lower-crossing event has occurred by comparing the cumulative log change with the critical price. If the cumulative log change is lower then the critical price, then carry out the boundary crossing steps and return to the previous step.

5. Calculate the critical price level at which upper crossing event would occur.

6. Update the log change using the maximum price.

7. Evaluate if an upper-crossing event has occurred by comparing the cumulative log change with the critical price. If the cumulative log change is higher then the critical price, then carry out the boundary crossing steps and return to the previous step.

8. Update the value of the cumulative log change using the closing price

9. Finally, ad the interest received on the deposit, if any.

Upon each boundary crossing event, the following steps are required.

1. Increment the appropriate counting variable, $Y_t = Y_t + 1$

2. Record how much time was needed for the crossing event to occur This information can be used to estimate the first exit time distribution.

3. Reset the value of the cumulative variables, such as DBS and the cumulative log change to zero.

Overall, based on Equation (3.28), constant change in the domain of portfolio's value upon rebalancing requires non-constant boundaries in the log-change domain: the interest collected or paid increase as t' increases, therefore the log change needs to be adjusted accordingly.

First exit time distribution and upper boundary crossing probabilities are not well defined for non-constant boundaries. In this case, we need to assume that no interest paid/received on the composite asset. Yet, we can account for proportional transaction costs. The derivation for the simplified case is as follows. The change in the portfolio's value is as follows.

$$V' = V + \Pi - TC \tag{3.29}$$

where V is the value of the portfolio upon rebalancing, Π is the profit or loss and finally TC is the transaction cost. The profit is as follows.

$$\Pi = q \times \Delta p = \frac{V \times w}{P} \times \Delta p = V \times w \times (exp(lnr) - 1) \tag{3.30}$$

The transaction cost, assuming ptc is negative for purchases and positive for sales is equal to:

$$TC = abs(q' - q) \times p' \times ptc = (\frac{V' \times w}{p'} - \frac{V \times w}{p}) \times p' \times ptc \tag{3.31}$$

Substituting yields:

$$V' = V + V \times w \times (exp(lnr) - 1) - (V' \times w \times ptc - V \times w \times exp(lnr) \times ptc) \tag{3.32}$$

Rearranging yields:

$$V' \times (1 + w \times ptc) = V \times (1 + w \times exp(lnr) \times (1 + ptc) - w) \tag{3.33}$$

From which, the change in wealth is equal to

$$\frac{V'}{V} = X_t = \frac{1 + w \times exp(lnr) \times (1 + ptc) - w}{1 + w \times ptc} \tag{3.34}$$

The change in wealth upon lower crossing, G_L, is equal to:

$$G_L = \frac{1 + w \times exp(L) \times (1 + ptc) - w}{1 + w \times ptc} \tag{3.35}$$

If boundaries are defined in the portfolio's value domain, then change in the log-return domain can be calculated as follows.

$$\log(\frac{G_L \times (1 + w \times ptc) - 1 + w}{w \times (1 + ptc)}) = L \tag{3.36}$$

The change in wealth upon upper crossing is equal to:

$$G_U = \frac{1}{G_L} = \frac{1 + w \times ptc}{1 + w \times exp(L) \times (1 + ptc) - w} \tag{3.37}$$

From which the value of the upper bound in the log-change domain is equal to.

$$\frac{1 + w \times exp(U) \times (1 + ptc) - w}{1 + w \times ptc} = \frac{1 + w \times ptc}{1 + w \times exp(L) \times (1 + ptc) - w} \tag{3.38}$$

Finally, the value of the upper bound can be expressed as follows.

$$U = log(\frac{(1 + w \times ptc)^2 - (1 - w)(1 + w \times exp(L) \times (1 + ptc) - w)}{(1 + w \times exp(L) \times (1 + ptc) - w) \times w \times (1 + ptc)}) \tag{3.39}$$

Constant boundaries therefore require us to abstract away from the interest paid on the composite asset. If we do want to take financing into account, then we either have to solve the problem for non-constant boundaries which is somewhat more involved, or rely on simulations instead.

3.7 Supplementary Appendix on Optimal Simple Portfolio for Geometric Brownian Motion under Continuous Rebalancing

The portfolio consisting of a risky asset and a risk-free composite asset follows:

$$dP_t = (\mu_{rr} + w\mu_e)P_t dt + w\sigma P_t dW_t \tag{3.40}$$

where μ_{rr} is the return on the risk-free assets assumed to be zero for the purpose of this exercise, μ_{rm} is the market return, $\mu_e = \mu_{rm} - \mu_{rr}$ is the excess return, w is the weight of the risky asset and finally W_t is the Wiener process. This formulation implicitly assumes that investors can keep a constant fraction of their wealth in the risky asset which would require continuously adjusting the number of shares, unless $w = 1$. The question of interest is log-optimal value of w. Using Ito formula:

$$dln(P_t) = (\mu_{rr} + w\mu_e - \frac{1}{2}w^2\sigma^2)dt + w\sigma dW_t \tag{3.41}$$

The expected value of the Wiener process is zero, the expected value of the exponential growth rate can be expressed as:

$$E(g) = E(\frac{dln(P_t)}{dt} = (\mu_{rr} + w\mu_e - \frac{1}{2}w^2\sigma^2) \tag{3.42}$$

Solving for the optimal value and substituting for $\mu_{rr} = 0$ yields:

$$w^{opt} = \frac{\mu}{\sigma^2} \tag{3.43}$$

DOI: 10.14754/CEU.2016.01

Bibliography

Abel, Andrew B., Janice C. Eberly, and Stavros Panageas, 2013, Optimal inattention to the stock market with information costs and transactions costs, *Econometrica* 81, 1455–1481.

Ait-Sahalia, Yacine, 1996, Testing continuous-time models of the spot interest rate, *Review of Financial studies* 9, 385–426.

Anderson, Theodore W, 1993, Goodness of fit tests for spectral distributions, *The Annals of Statistics* 830–847.

Andrews, Donald WK, 1988, Laws of large numbers for dependent non-identically distributed random variables, *Econometric theory* 4, 458–467.

Andrews, Donald WK, 2001, Testing when a parameter is on the boundary of the maintained hypothesis, *Econometrica* 683–734.

Bachelier, Louis, 1900, *Théorie de la spéculation* (Gauthier-Villars).

Bai, Jushan, and Serena Ng, 2002, Determining the number of factors in approximate factor models, *Econometrica* 70, 191–221.

Bai, Jushan, and Serena Ng, 2004, A panic attack on unit roots and cointegration, *Econometrica* 72, 1127–1177.

Bajeux Besnainou, Isabelle, and Roland Portait, 1997, The numeraire portfolio: A new perspective on financial theory, *The European Journal of Finance* 3, 291–309.

Balduzzi, Pierluigi, and Anthony W. Lynch, 2000, Predictability and transaction costs: The impact on rebalancing rules and behavior, *The Journal of Finance* 55, 2285–2309.

Baltagi, Badi H, Georges Bresson, and Alain Pirotte, 2007, Panel unit root tests and spatial dependence, *Journal of Applied Econometrics* 22, 339–360.

Baltagi, Badi H, and Chihwa Kao, 2001, *Nonstationary panels, cointegration in panels and dynamic panels: A survey*, volume 15 (Emerald Group Publishing Limited).

Balvers, Ronald, Yangru Wu, and Erik Gilliland, 2000, Mean reversion across national stock markets and parametric contrarian investment strategies, *The Journal of Finance* 55, 745–772.

Banerjee, Anindya, 1999, Panel data unit roots and cointegration: an overview, *Oxford Bulletin of economics and Statistics* 61, 607–629.

Baumol, William J., 1952, The transactions demand for cash: An inventory theoretic approach, *Quarterly Journal of Economics* 66, 545–556.

Blum, Avrim, and Adam Kalai, 1999, Universal portfolios with and without transaction costs, *Machine Learning* 35, 193–205.

Bogen, Jules Irwin, and Herman Edward Krooss, 1960, *Security credit: Its economic role and regulation* (Prentice-Hall).

Borodin, Andrei N., and Paavo Salminen, 2002, *Handbook of Brownian motion: facts and formulae* (Springer).

Brandt, Michael W., 1999, Estimating portfolio and consumption choice: A conditional Euler equations approach, *The Journal of Finance* 54, 1609–1645.

Brandt, Michael W., 2003, Hedging demands in hedging contingent claims, *Review of Economics and Statistics* 85, 119–140.

Brandt, Michael W., 2009, Portfolio choice problems, *Handbook of Financial Econometrics* 1, 269–336.

Breiman, L, 1961, Optimal gambling systems for favorable games, *The Kelly Capital Growth Investment Criterion* 47.

Breitung, Jörg, and M Hashem Pesaran, 2008, Unit roots and cointegration in panels, in László Mátyás, and Patrick Sevestre, eds., *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice* (Springer).

Brennan, Michael J., Eduardo S. Schwartz, and Ronald Lagnado, 1997, Strategic asset allocation, *Journal of Economic Dynamics and Control* 21, 1377–1403.

Buss, Adrian, and Bernard Dumas, 2013, Financial-market equilibrium with friction, Technical report, National Bureau of Economic Research.

Campbell, John Y., and Samuel B. Thompson, 2008, Predicting excess stock returns out of sample: Can anything beat the historical average?, *Review of Financial Studies* 21, 1509–1531.

Chang, Yoosoon, 2002, Nonlinear IV unit root tests in panels with cross-sectional dependency, *Journal of Econometrics* 110, 261–292.

Chang, Yoosoon, and Joon Y Park, 2002, On the asymptotics of ADF tests for unit roots, *Econometric Reviews* 21, 431–447.

Choi, In, 2001, Unit root tests for panel data, *Journal of international money and Finance* 20, 249–272.

Choi, In, 2006, *Combination unit root tests for cross-sectionally correlated panels* (Econometric Theory and Practice: Frontiers of Analysis and Applied Research: Essays in Honor of Peter CB Phillips. Cambridge University Press).

Christensen, Morten Mosegaard, 2005, On the history of the growth optimal portfolio, *Preprint, University of Southern Denmark* 389.

Cocco, Joo F., Francisco J. Gomes, and Pascal J. Maenhout, 2005, Consumption and portfolio choice over the life cycle, *Review of Financial Studies* 18, 491–533.

Constantinides, George M., 1986, Capital market equilibrium with transaction costs, *Journal of Political Economy* 94, 842–862.

Cont, Rama, 2001, Empirical properties of asset returns: stylized facts and statistical issues, *Quantitative Finance* 1.

Cover, Thomas M., 1991, Universal portfolios, *Mathematical Finance* 1, 1–29.

De Jong, RM, 1995, Laws of large numbers for dependent heterogeneous processes, *Econometric Theory* 11, 347–358.

Demetrescu, Matei, Uwe Hassler, and Adina-Ioana Tarcolea, 2006, Combining significance of correlated statistics with application to panel data, *Oxford Bulletin of Economics and Statistics* 68, 647–663.

Dickey, David A, and Wayne A Fuller, 1981, Likelihood ratio statistics for autoregressive time series with a unit root, *Econometrica* 1057–1072.

Dufour, Jean-Marie, and Abdeljelil Farhat, 2001, Exact nonparametric two-sample homogeneity tests for possibly discrete distributions., *Working paper* .

Dumas, Bernard, and Elisa Luciano, 1991, An exact solution to a dynamic portfolio choice problem under transactions costs, *The Journal of Finance* 46, 577–595.

Fama, Eugene F, and Kenneth R French, 2002, The equity premium, *The Journal of Finance* 57, 637–659.

Fan, Jianqing, et al., 2005, A selective overview of nonparametric methods in financial econometrics, *Statistical Science* 20, 317–337.

Fan, Jianqing, Chunming Zhang, and Jian Zhang, 2001, Generalized likelihood ratio statistics and wilks phenomenon, *Annals of statistics* 153–193.

Feller, William, 1971, An introduction to probability and its applications Vol. ii, *John Wiley, New York* .

Fisher, R.A., 1932, *Statistical Methods for Research Workers* (Oliver and Boyd, Edinburgh, 4th Edition).

Fortune, Peter, 2000, Margin requirements, margin loans, and margin rates: Practice and principles, *New England Economic Review* 19–44.

Fortune, Peter, 2001, Margin lending and stock market volatility, *New England Economic Review* 3–26.

Gabaix, Xavier, 2012, Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance, *The Quarterly Journal of Economics* 127, 645–700.

Gao, Jiti, and Maxwell King, 2004, Adaptive testing in continuous-time diffusion models, *Econometric Theory* 20, 844–882.

Gerber, Hans U, and Elias SW Shiu, 1994, Option pricing by Esscher transforms, *Transactions of the Society of Actuaries* 46, 140.

Gourieroux, Christian, and Joann Jasiak, 2001, *Financial econometrics: Problems, models, and methods*, volume 1 (Princeton University Press Princeton, NJ).

Gyorfi, Laszlo, and Istvan Vajda, 2008, Growth optimal investment with transaction costs, in *Algorithmic Learning Theory*, 108–122, Springer.

Hamilton, James Douglas, 1994, *Time series analysis*, volume 2 (Princeton university press Princeton).

Hansen, Bruce E, 1991, Strong laws for dependent heterogeneous processes, *Econometric Theory* 7, 213–221.

Hardouvelis, Gikas A., and Panayiotis Theodossiou, 2002, The asymmetric relation between initial margin requirements and stock market volatility across bull and bear markets, *Review of Financial Studies* 15, 1525–1559.

Harris, Richard ID, 1992, Testing for unit roots using the augmented Dickey-Fuller test: Some issues relating to the size, power and the lag structure of the test, *Economics letters* 38, 381–386.

Higham, Nicholas J, 1988, Computing a nearest symmetric positive semidefinite matrix, *Linear algebra and its applications* 103, 103–118.

Hong, Yongmiao, and Haitao Li, 2005, Nonparametric specification testing for continuous-time models with applications to term structure of interest rates, *Review of Financial Studies* 18, 37–84.

Horváth, Márk, and András Urbán, 2011, Growth optimal portfolio selection with short selling and leverage, *Machine Learning for Financial Engineering, eds. L. Györfi, G. Ottucsák, H. Walk, Imperial College Press* 151–176.

Hurlin, Christophe, Valérie Mignon, et al., 2007, Second generation panel unit root tests, *Working paper, https://halshs.archives-ouvertes.fr/halshs-00159842* .

Im, Kyung So, M Hashem Pesaran, and Yongcheol Shin, 2003, Testing for unit roots in heterogeneous panels, *Journal of econometrics* 115, 53–74.

James, Barry, Kang James, and Yongcheng Qi, 2008, Limit theorems for correlated Bernoulli random variables, *Statistics & Probability Letters* 78, 2339–2345.

Karlin, Samuel, and Howard M. Taylor, 1981, *A second course in stochastic processes*, volume 2 (Gulf Professional Publishing).

Karlin, Samuel, and Howard M. Taylor, 1998, *An Introduction to Stochastic Modeling* (Academic Press).

Kelly, John L., 1956, A new interpretation of information rate, *Information Theory, IRE Transactions on* 2, 185–189.

Kirk, Donald E, 2004, *Optimal control theory: an introduction* (Courier Dover Publications).

Kolmogoroff, Andrei, 1931, Über die analytischen methoden in der wahrscheinlichkeitsrechnung, *Mathematische Annalen* 104, 415–458.

Latane, Henry Allen, 1959, Criteria for choice among risky ventures, *The Journal of Political Economy* 144–155.

Levin, Andrew, and Chien-Fu Lin, 1992, Unit root test in panel data: asymptotic and finite-sample properties, *Discussion Paper No. 96-56, University of California at San Diego* .

Levin, Andrew, Chien-Fu Lin, and Chia-Shang James Chu, 2002, Unit root tests in panel data: asymptotic and finite-sample properties, *Journal of econometrics* 108, 1–24.

Lin, Sheldon X., 1998, Double barrier hitting time distributions with applications to exotic options, *Insurance: Mathematics and Economics* 23, 45–58.

Linetsky, Vadim, 2004, Computing hitting time densities for CIR and OU diffusions: Applications to mean-reverting models, *Journal of Computational Finance* 7, 1–22.

Liu, Hong, and Mark Loewenstein, 2002, Optimal portfolio selection with transaction costs and finite horizons, *Review of Financial Studies* 15, 805–835.

Lo, Andrew W., and A.C. MacKinlay, 1999, *A Non-Random Walk Down Wall Street* (Princeton).

Lo, Andrew W, and Archie Craig MacKinlay, 1988, Stock market prices do not follow random walks: Evidence from a simple specification test, *Review of financial studies* 1, 41–66.

Long, John, 1990, The numeraire portfolio, *Journal of Financial Economics* 26, 29–69.

MacLean, Leonard C., Edward O. Thorp, and William T. Ziemba, 2011, *The Kelly capital growth investment criterion: Theory and practice*, volume 3 (World Scientific).

Maddala, Gangadharrao S, and Shaowen Wu, 1999, A comparative study of unit root tests with panel data and a new simple test, *Oxford Bulletin of Economics and statistics* 61, 631–652.

Maenhout, Pascal J., 2004, Robust portfolio rules and asset pricing, *Review of Financial Studies* 17, 951–983.

Malkiel, Burton G., and Eugene F. Fama, 1970, Efficient capital markets: A review of theory and empirical work, *The Journal of Finance* 25, 383–417.

Markowitz, Harry, 1952, Portfolio selection, *The Journal of Finance* 7, 77–91.

Martellini, Lionel, and Philippe Priaulet, 2002, Competing methods for option hedging in the presence of transaction costs, *The Journal of Derivatives* 9, 26–38.

Mcleish, Donald L., 2002, Highs and lows: Some properties of the extremes of a diffusion and applications in finance, *Canadian Journal of Statistics* 30, 243–267.

Medvegyev, Péter, 2007, *Stochastic integration theory*, Oxford Graduate Texts in Mathematics (Oxford University Press).

Mehra, Rajnish, and Edward C Prescott, 1985, The equity premium: A puzzle, *Journal of monetary Economics* 15, 145–161.

Mehra, Rajnish, and Edward C Prescott, 2003, The equity premium in retrospect, *Handbook of the Economics of Finance* 1, 889–938.

Merton, Robert C., 1969, Lifetime portfolio selection under uncertainty: The continuous-time case, *Review of Economics and Statistics* 51, 247–257.

Merton, Robert C., 1971, Optimum consumption and portfolio rules in a continuous-time model, *Journal of Economic Theory* 3, 373–413.

Merton, Robert C., and P. A. Samuelson, 1992, *Continuous-time finance* (Blackwell Boston).

Moon, Hyungsik Roger, and Benoit Perron, 2004, Testing for a unit root in panels with dynamic factors, *Journal of econometrics* 122, 81–126.

Pesaran, M Hashem, 2004, General diagnostic tests for cross section dependence in panels, *Cambridge Working Papers in Economics, https://www.repository.cam.ac.uk/handle/1810/446* .

Pesaran, M Hashem, 2007, A simple panel unit root test in the presence of cross-section dependence, *Journal of Applied Econometrics* 22, 265–312.

Peters, Ole, 2011, Optimal leverage from non-ergodicity, *Quantitative Finance* 11, 1593–1602.

Phillips, Peter CB, and Pierre Perron, 1988, Testing for a unit root in time series regression, *Biometrika* 75, 335–346.

Phillips, Peter CB, and Zhijie Xiao, 1998, A primer on unit root testing, *Journal of Economic Surveys* 12, 423–470.

Platen, Eckhard, 2006, A benchmark approach to finance, *Mathematical Finance* 16, 131–151.

Poundstone, William, 2010, *Fortune's Formula: The untold story of the scientific betting system that beat the casinos and wall street* (Macmillan).

Pritsker, Matt, 1998, Nonparametric density estimation and tests of continuous time interest rate models, *Review of Financial Studies* 11, 449–487.

Redner, Sidney, 2001, *A guide to first-passage processes* (Cambridge University Press).

Said, Said E, and David A Dickey, 1984, Testing for unit roots in autoregressive-moving average models of unknown order, *Biometrika* 71, 599–607.

Samuelson, P. A., 1969, Lifetime portfolio selection by dynamic stochastic programming, *Review of Economics and Statistics* 239–246.

Santa-Clara, Pedro, and Shu Yan, 2010, Crashes, volatility, and the equity premium: Lessons from S&P 500 options, *The Review of Economics and Statistics* 92, 435–451.

Saretto, Alessio, and Heather E. Tookes, 2013, Corporate leverage, debt maturity, and credit supply: The role of credit default swaps, *Review of Financial Studies* 26, 1190–1247.

Shiller, Robert J., 2005, *Irrational exuberance* (Random House LLC).

Song, Zhaogang, 2011, A martingale approach for testing diffusion models based on infinitesimal operator, *Journal of Econometrics* 162, 189–212.

Szimayer, A, and R Maller, 2004, Testing for mean reversion in processes of Ornstein - Uhlenbeck type, *Statistical inference for stochastic processes* 7, 95–113.

Teräsvirta, Timo, and Zhenfang Zhao, 2011, Stylized facts of return series, robust estimates and three popular models of volatility, *Applied financial economics* 21, 67–94.

Tippett, Leonard Henry Caleb, et al., 1931, *The methods of statistics.* (London: Williams & Norgate Ltd.).

Tobin, James, 1956, *The interest-elasticity of transactions demand for cash* (Cowles Foundation for Research in Economics at Yale University).

Valenti, Davide, Bernardo Spagnolo, and Giovanni Bonanno, 2007, Hitting time distributions in financial markets, *Physica A: Statistical Mechanics and its Applications* 382, 311–320.

Valov, Angel Veskov, 2009, *First passage times: Integral equations, randomization and analytical approximations*, Ph.D. thesis, University of Toronto.

Wachter, Jessica A., and Motohiro Yogo, 2010, Why do household portfolio shares rise in wealth?, *Review of Financial Studies* 23, 3929–3965.

Wilmott, Paul, 1998, *Derivatives* (John Willey & Sons, New York).

Wolf, Fredric M, 1986, *Meta-analysis: Quantitative methods for research synthesis*, volume 59 (Sage).

Zhao, Zhibiao, et al., 2008, Parametric and nonparametric models and methods in financial econometrics, *Statistics Surveys* 2, 1–42.