Financial Signal Processing: Applications to Asset-Market Dynamics and Healthcare Finance

by

Shomesh Ernesto Chaudhuri

B.S., Harvard University (2011)
S.M., Massachusetts Institute of Technology (2014)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2018

Author .....
Department of Electrical Engineering and Computer Science
May 22, 2018

Certified by .....
Andrew W. Lo
Charles E. and Susan T. Harris Professor, Sloan School of Management
Thesis Supervisor

Accepted by .....
Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Department Committee on Graduate Students

Financial Signal Processing: Applications to Asset-Market Dynamics and Healthcare Finance

by

Shomesh Ernesto Chaudhuri

Submitted to the Department of Electrical Engineering and Computer Science
on May 22, 2018, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Electrical Engineering and Computer Science

Abstract

The seemingly random fluctuations of price and value produced by information flow and complex interactions across a diverse population of stakeholders has motivated the extensive use of stochastic processes to analyze both capital markets and the regulatory approval process in healthcare. This thesis approaches the statistical analysis of such processes through the lens of signal processing, with a particular emphasis on studying how dynamics evolve over time.

We begin with a brief introduction to financial signal processing in Part I, before turning to specific applications in the main body of the thesis. In Part II, we apply spectral analysis to understand and quantify the relationship between asset-market dynamics across multiple time horizons, and show how this framework can be used to improve portfolio and risk management. Using the Fourier transform, we decompose asset-return alphas, betas and covariances into distinct frequency components, allowing us to identify the relative importance of specific time horizons in determining each of these quantities. Our approach can be applied to any portfolio, and is particularly useful for comparing the forecast power of multiple investment strategies.

Part III addresses the growing interest from the healthcare industry, regulators and patients to include Bayesian adaptive methods in the regulatory approval process of new therapies. By applying sequential likelihood ratio tests to a Bayesian decision analysis framework that assigns asymmetric weights to false approvals and false rejections, we are able to design adaptive clinical trials that maximize the value to current and future patients and consequently, public health. We also consider the possibility that as the process unfolds, drug sponsors might stop a trial early if new information suggests market prospects are not as favorable as originally forecasted. We show that clinical trials that can be modified as data are observed are more valuable than trials without this flexibility.

Thesis Supervisor: Andrew W. Lo
Title: Charles E. and Susan T. Harris Professor, Sloan School of Management

Acknowledgments

If you want to go fast, go alone. If you want to go far, go together.

This thesis would not have been written without the support and encouragement of countless numbers of individuals.

First and foremost, I would like to thank my thesis supervisor, Professor Andrew W. Lo, whose support and guidance in both my research and life outside of the lab has been nothing short of incredible. The standards he sets and the example by which he leads have inspired me to reach for and attain goals I never thought possible. His commitment not only to research, but also to the professional development of his students is something that I will always be grateful for.

I would also like to thank my thesis committee, Professor John Guttag and Professor Leonid Kogan. Especially during the latter stages of my Ph.D. studies, the conversations I had with them really helped sharpen my thinking around some of the issues related to my research.

I also have to thank the many people behind the scenes who have been instrumental to this work. In particular, I would like to thank Jayna Cummings and Crystal Myler for everything they do to support LFE students and our research. Thank you also to my labmates and friends for making this period of my life so enjoyable.

Finally, I'm thankful for my family. A special thank you goes to my wife, Meaghan, for being a constant source of happiness in my life. I would also like to thank my brother, Sheel, for always looking out for me, and my parents Sujeet and Janet whose many sacrifices have provided me with this opportunity.

Contents

List of Figures

List of Tables

Part I

Introduction

Chapter 1

Financial Signal Processing

The field of signal processing deals with extracting information from signals produced by the behaviour or nature of some phenomenon. In finance, the most visible signal produced by a system of competing market participants is price, and signal processing can be used to extract information that may prove useful in predicting future movements. For example, the history of the daily close of a stock market index may contain information about long-term economic trends or more transient seasonal variations. This thesis explores how tools from signal processing can be used to extract information from signals produced by capital markets and the regulatory approval process in healthcare, and moreover, how systems in these areas can be designed to make efficient use of this information. Before turning to specific applications in Parts II and III, we begin by providing some historical context.

1.1 Background for Part II

Over the past 200 years, signal processing has made fundamental contributions to fields ranging from finance, communications, and neuroscience, to partial differential equations, astronomy, and geology. However, in contrast to its modern ubiquity, its origins stem from a very specific problem—modeling the orbits of celestial bodies.

In the mid-18th century, the mathematicians Leonhard Euler, Joseph-Louis Lagrange, and Alexis Clairaut each observed that orbits could be approximated as linear

combinations of trigonometric functions, i.e., sines and cosines. In fact, in order to estimate the coefficients from the data, Clairaut published the first explicit formulation of the Discrete Fourier Transform (DFT) in 1754, more than a decade before Jean-Baptiste Joseph Fourier was born. Shortly thereafter, while studying the orbit of the asteroid Pallas in 1805, Carl Friedrich Gauss discovered a computational shortcut while calculating the DFT. His calculation, which appeared posthumously as an unpublished paper in 1866, is now known as the Fast Fourier Transform (FFT), and is used in many applications in engineering, science, and mathematics. Unfortunately, Gauss' algorithm was largely forgotten until it was independently rediscovered in a more general form almost a century later by James Cooley and John Tukey in 1965.

Drawing on the progress made by pioneers such as Clairaut, spectral analysis was generalized by Fourier in his seminal paper on heat conduction. In his treatise presented to the Paris Academy in 1807, Fourier claimed that *any* arbitrary function could be represented by the superposition of trigonometric functions. This broader claim was initially received with much skepticism, but eventually the Paris Academy awarded his paper the grand prize in 1812. Despite the award, the Academy's panel of judges, which included Lagrange, Laplace and Legendre, still held reservations about the rigor of his analysis, especially in relation to the challenging question posed by convergence. Further advances by Dirichlet, Poisson, and Riemann addressed these subtle issues, and provided the foundation for today's Fourier transform, which forms the basis for analyzing signals in the frequency domain (Briggs and Henson, 1995). We will use these tools extensively in Part II of this thesis.

1.2 Background for Part III

Interestingly, it is not spectral analysis, but rather the heat equation from Fourier's treatise that has has the most influence on the field of finance. In 1900, Louis Bachelier, one of the pioneers of financial mathematics, modeled the fluctuations of stock prices as random walks in continuous time. In doing so, he connected Fourier's heat equation to option pricing, which ultimately lead to the celebrated Black-Scholes-

Merton formula. Similar to how the heat equation states that the rate of change in temperature at a point in space is proportional to the convexity of the local temperature profile, loosely speaking, the Black-Scholes-Merton partial differential equation states that the rate of change in a European call option's price at a given stock price is proportional to the convexity of the option price at that underlying stock price. This formulation captures the intuition that the value of an option is in its ability to hedge uncertainty.

Both random walks and option valuation are applied throughout Part III of this thesis. However, while the work of Bachelier and Black, Scholes and Merton focus on continuous time, we approach these problems from the perspective of discrete time. The major impetus for this choice is the result of Moore's law, i.e., technological advancements in computation have made it increasingly advantageous to represent continuous time signals in discrete time. In particular, the discrete-time domain provides greater flexibility to apply tools from signal processing on a case-by-case basis, which is especially useful in our evaluation of real options in Chapter 7.

Part II

Applications to Asset-Market Dynamics

Chapter 2

Spectral Analysis of Financial Time Series

Although spectral methods are not new to finance, current applications are sufficiently rare that a brief overview of spectral analysis may be appropriate before we turn to our own applications. We begin in Section 2.1 with a brief motivation for spectral methods, and review the literature in Section 2.2. The formulation of time-frequency analysis is provided in terms of the Discrete Fourier Transform (DFT) in Section 2.3, and the Haar Wavelet Transform in Section 2.4. We then present the main mathematical results on the co-spectrum in Section 2.5 that will be the basis of our applications to portfolio theory, and include examples using business cycle data in Section 2.6 and equity returns in 2.7. Readers familiar with spectral methods may prefer to skip this Chapter and proceed to Chapter 3.

2.1 Introduction

Although portfolio optimization models have explicitly incorporated a time dimension ever since the stochastic dynamic programming approach of Samuelson (1969) and Merton (1969, 1971, 1973), the decision-making horizon of investors has rarely been the main focus of attention. Portfolio weights are assumed either to be rebalanced continuously over time or at arbitrary but fixed discrete intervals. In both cases, the

process by which portfolio decisions are rendered is determined by dynamic optimization, yielding optimal portfolio weights that are functions of state variables evolving through time according to their laws of motion. The generality of this approach can obscure important features of the underlying process by which information is reflected in investment decisions. For example, although high-frequency trading and long-term investing can both be profitable—and both can be modeled as a dynamic optimization problem—they operate at very different frequencies using very different methods.

In this thesis, we propose a new approach to analyzing and constructing portfolios in which the frequency component is explicitly captured. Using the tools of spectral analysis—the decomposition of time series into a sum of periodic functions like sines and cosines—we show that investment strategies can differ significantly in the frequencies with which their expected returns and volatility are generated. Slower-moving strategies will exhibit more "power" at the lower frequencies while faster-moving strategies will exhibit more power at the higher frequencies. By identifying the particular frequencies that are responsible for a given strategy's expected returns and volatility, an investor will have an additional dimension with which to manage the risk/reward profile of their portfolio.

In fact, because time-domain statistics such as means, standard deviations, correlations, and beta coefficients all have frequency-domain counterparts, it is possible to apply spectral analysis to virtually all aspects of portfolio theory, linear factor models, performance and risk attribution, capital budgeting, and risk management. In each of these areas, we can decompose traditional time-series measures into the sum of frequency-specific subcomponents. For example, for a specific set of historical asset returns, we can formulate a spectral factor model to decompose the covariance matrix into the sum of high, medium, and low frequency components so that an investor can determine the best and worst sources of diversification at different frequencies and change their portfolio accordingly.

To motivate the practical relevance of frequency in the portfolio context, consider the simple market-neutral mean-reversion strategy of Lo and MacKinlay (1990). This strategy holds long positions in stocks that underperformed the average stock q days

ago and holds short positions in stocks that outperformed the average q days ago, i.e., $w_{it}(q) = -(r_{it-q} - r_{mt-q})/N$ where $r_{mt-q} = \sum_i r_{it-q}/N$ is the average stock return on date $t-q$ and N is the total number of stocks. Each lag q defines a different strategy, one intended to exploit mean reversion over a q-day horizon. Common intuition might suggest that the returns of strategies with non-overlapping horizons would be uncorrelated. Therefore, it is not surprising to learn that the correlation between the returns of the $q=1$ and $q=2$ strategies is -1.1% over the period from July 16, 1962 to December 31, 2015.[1] However, during the month of August 2007, these strategies all suffered significant losses as part of the "Quant Meltdown" (Khandani and Lo, 2007). During that month, the correlation between the $q = 1$ and $q = 2$ strategies spiked to 65.8%. Figure 2-1 provides a more dynamic view of this correlation, computed over 125-day rolling windows from 3 January 2006 through 31 December 2008. The correlation began increasing in July 2007 but the spike occurred in August 2007 and declined steadily until the correlation turned negative in the first half of 2008, only to reverse itself during the second half as the Financial Crisis unfolded.

These strange dynamics illustrate the relevance of frequency effects in financial asset returns, and spectral analysis is the most natural tool for quantifying these effects.

[1]Specifically, the strategies are implemented using data from the University of Chicago's Center for Research in Securities Prices (CRSP). Only U.S. common stocks (CRSP share code 10 and 11) are included, which eliminates REIT's, ADR's, and other types of securities, and we drop stocks with share prices below \$5 and above \$2,000.

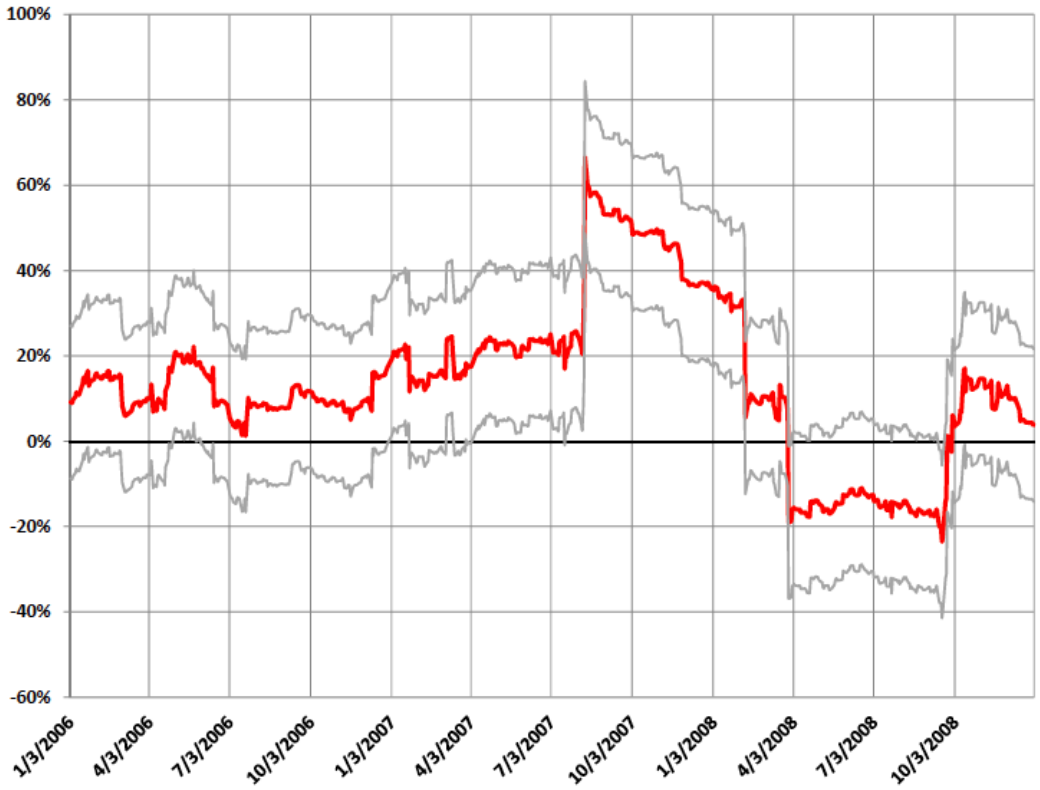

Figure 2-1: 125-day rolling-window correlation between daily mean-reversion strategies $\{w_{it}(q)\}$ with $q = 1$ and $q = 2$ where $w_{it}(q) = -(r_{it-q} - r_{mt-q})/N$ and $r_{mt-q} = \sum_i r_{it-q}/N$ is the average stock return on date $t-q$. The gray lines delineate 2-standard-deviation bands around the correlations under the null hypothesis of zero correlation.

2.2 Literature Review

The frequency domain has long been part of economics (Granger and Hatanaka, 1964; Engle, 1974; Granger and Engle, 1983; Hasbrouck and Sofianos, 1993), and the Fourier transform has been used in finance to efficiently evaluate theoretical pricing models for derivative securities (Carr and Madan, 1999). However, econometric and empirical applications of spectral analysis in economics and finance are less common, in part because economic time series are rarely considered stationary. Recently, there has been a rebirth of interest in their application to economics in response to modern advances in non-stationary signal analysis (Baxter and King, 1999; Croux et al., 2001; Ramsey, 2002; Crowley, 2007; Huang et al., 2003; Breitung and Candelon, 2006; Rua, 2010, 2012; Dew-Becker and Giglio, 2016; Bandi et al., 2017). This rebirth motivates our interest in the spectral properties of financial asset returns.

Spectral and co-spectral power, often calculated using either the Fourier or wavelet transform, provide a natural way to study the cyclical components of variance and covariance, two important measures of risk in the financial domain. Specifically, spectral power decomposes the variability of a time series resulting from fluctuations at a specific frequency, while co-spectral power decomposes the covariance between two real-valued time series, and measures the tendency for them to move together over specific time horizons. When the signals are in phase at a given frequency (i.e., their peaks and valleys coincide), the co-spectral power is positive at that frequency, and when they are out of phase, it is negative.

In a recent empirical study, Chaudhuri and Lo (2015) perform a spectral decomposition of the U.S. stock market and individual common stock returns over time. They noticed that measures related to risk and co-movement varied not only across time, but also across frequencies over time. Given that economic shocks produce distinct effects on financial assets over different time horizons, these dynamics are likely to have important implications for any theory of risk, reward, and portfolio choice. However, the traditional inputs into these analytics—means, variance, covariances, alphas, and beta—are static, and do not distinguish between the short- and long-

term components of these dynamics. The fact that the standard estimators of these statistics are invariant to how the data are ordered suggests that traditional portfolio analytics are incapable of capturing the dynamic properties of asset returns. In this thesis, we apply spectral analysis to develop dynamic, frequency-specific analogs for each of these portfolio analytics. Our exposition focuses on sample statistics of the underlying power spectrums, but we note that each equation has a population statistic analog.

Furthermore, to address the non-stationarity of financial time series, our analysis relies on the short-time Fourier transform, which applies the discrete Fourier transform (DFT) to windowed subsamples of the entire sample (Oppenheim and Schafer, 2009). Recently, wavelets (Ramsey, 2002; Crowley, 2007; Rua, 2010, 2012) and other transforms (Huang et al., 2003) have also been used to study financial data in the time-frequency domain, and depending on the specific context, these alternative techniques can provide substantial benefits in terms of implementation. For example, the sinusoids used in the short-time Fourier transform do not efficiently characterize discontinuous processes, whereas the flexibility of wavelets can be used to overcome this difficulty. Moreover, the wavelet transform provides better time resolution at high frequencies, and better frequency resolution at low frequencies, although similar results can be obtained by varying the window length used with the short-time Fourier transform. In this thesis, we consider both sets of transform, but focus primarily on the DFT for two reasons: the Fourier transform is intuitive and expositionally simple, and all our results for the Fourier transform carry over directly to the wavelet transform (albeit with greater mathematical complexity).

2.3 The Fourier Transform

One of the most structurally revealing analyses that can be performed on a time series is to express its values as a linear combination of trigonometric functions. This procedure relies on the Discrete-Time Fourier Transform (DTFT), and allows the data to be transformed to the frequency domain. Specifically, given a finite-energy

time series x_t, the DTFT is given by,

$$X(\omega) = \sum_{t=-\infty}^{\infty} x_t \, e^{-j\omega t}, \quad \omega \in [0, 2\pi) \tag{2.1}$$

where the frequency ω has units of radians per sample and j denotes the imaginary unit $\sqrt{-1}$. When x_t is real-valued, the inverse DTFT can be written in rectangular form as,

$$x_t = \frac{1}{2\pi} \int_0^{2\pi} \Big[\Re[X(\omega)] \, \cos(\omega t) - \Im[X(\omega)] \, \sin(\omega t) \Big] \, \mathrm{d}\omega, \quad t \in (-\infty, \infty) \tag{2.2}$$

or in polar form as,

$$x_t = \frac{1}{2\pi} \int_0^{2\pi} |X(\omega)| \cos(\omega t + \angle X(\omega)) \, \mathrm{d}\omega, \quad t \in (-\infty, \infty) \tag{2.3}$$

where $\Re[X(\omega)]$ and $\Im[X(\omega)]$ are the real and imaginary components of $X(\omega)$, and $|X(\omega)|$ and $\angle X(\omega)$ are its magnitude and phase, respectively.

If only a finite sample of x_t is available, or only a local portion of x_t needs to be analyzed, the DTFT reduces to the DFT. Specifically, given a sample of x_t from times $t = 0, \ldots, T-1$, the T-point DFT is given by:[2]

$$X_k = \sum_{t=0}^{T-1} x_t \, e^{-j\omega_k t}, \quad k \in [0, \, T-1] \tag{2.4}$$

where $\omega_k = 2\pi k/T$. Again, when x_t is real-valued, the inverse DFT can be written in rectangular form as,

$$x_t = \frac{1}{T} \sum_{k=0}^{T-1} \Big[\Re[X_k] \, \cos(\omega_k t) - \Im[X_k] \, \sin(\omega_k t) \Big], \quad t \in [0, \, T-1] \tag{2.5}$$

[2]In general, for finite T, $X(\omega_k) \neq X_k$ as multiplying x_t by a rectangular window results in the convolution of $X(\omega)$ with the window's DTFT in the frequency domain.

or in polar form as,

$$x_t = \frac{1}{T}\sum_{k=0}^{T-1} |X_k| \cos(\omega_k t + \angle X_k), \quad t \in [0,\, T-1]. \tag{2.6}$$

In this real-valued case, $X_k = X^*_{T-k}$, and so $|X_k|\cos(\omega_k t + \angle X_k) = |X_{T-k}|\cos(\omega_{T-k} t + \angle X_{T-k})$. Therefore, the lowest non-zero frequency occurs at $k=1$, and the highest frequency occurs at $k = \lfloor T/2 \rfloor$. The relation $h = TT_s/k$, where T_s is the time between samples and $0 \leq k \leq T/2$, can be used to convert the kth harmonic frequency to its corresponding time horizon.

As a concrete example, consider the rectangular pulse x_t for $t=0$ to $t=9$ shown in panel A of Figure 2-2. The real and imaginary components of the 10-point DFT are plotted in panels B and C, and their magnitude and phase in panels D and E. Panel F shows the reconstruction of x_t using only the constant X_0 term in (2.5), which is equivalent to the average value of the time series x_t. Panel G shows the reconstruction of x_t using both the constant term and the first non-zero low-frequency terms. These low-frequency terms are dominated by the sine term in (2.5) which has an amplitude proportional to the magnitude of the imaginary coefficients at $k = 1$ and $k = 9$ in panel C. More precisely, the amplitude of the low-frequency sinusoid can be seen in panel D, and its phase in terms of a shifted cosine in panel E. As more frequencies are included (see panel H), the output of the reconstruction begins to converge to the original time series. Ultimately, the reconstruction exactly matches the original rectangular pulse in panel A when all frequency terms are included.

Since the Fourier transform is a unitary operator that changes the basis function representation of a time series from impulses to sinusoids, Parseval's theorem states that, when represented as a vector, the Euclidean length of the time series is preserved under the transformation (with proper normalization). This observation forms the foundation of spectral decomposition, and provides a method to visualize the data in the frequency domain. This representation, known as the power spectrum, characterizes how much of the variability in the data comes from low- versus high-frequency fluctuations. We turn to this representation in Section 2.5, but first

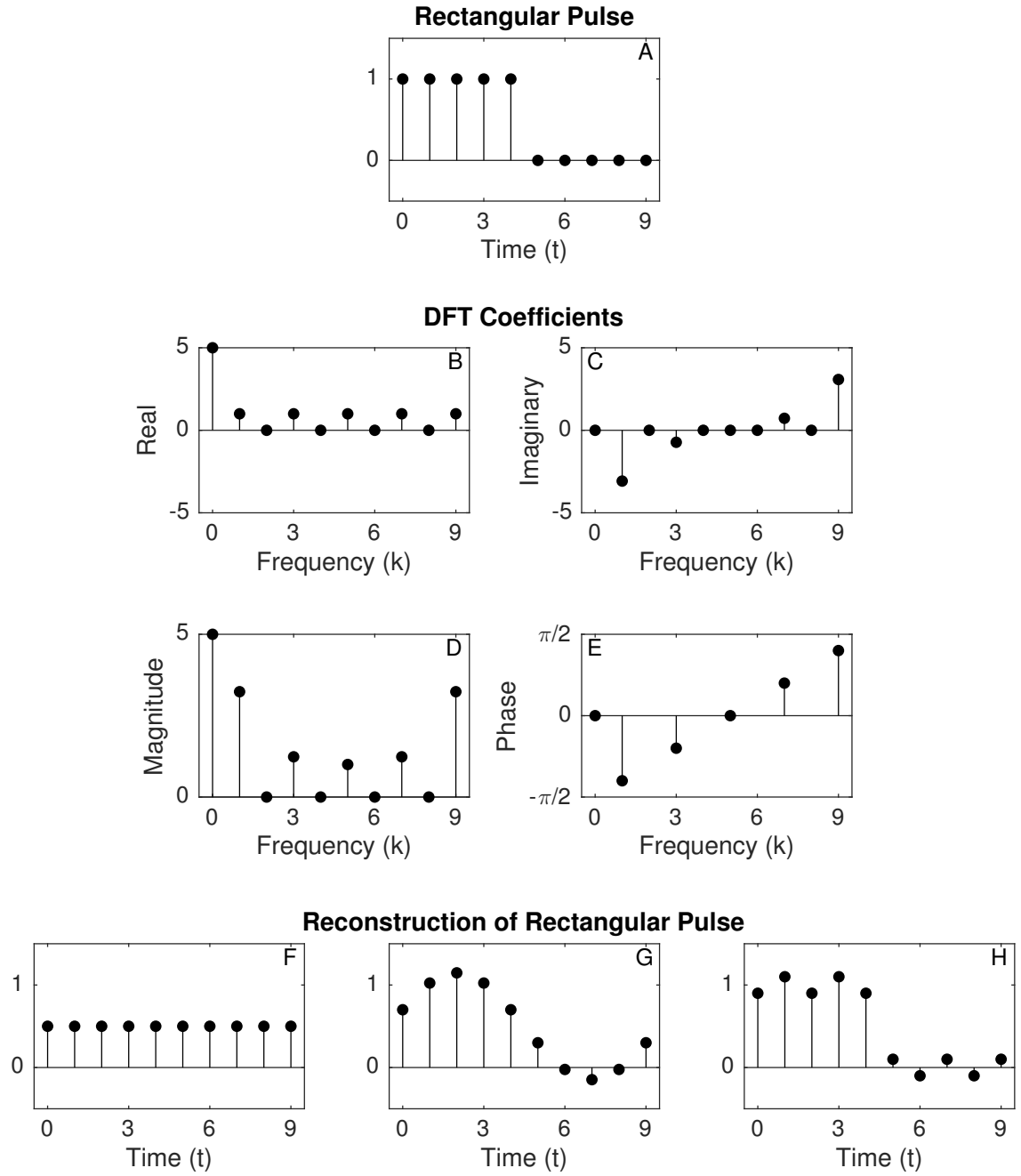

Figure 2-2: Panels B and C plot the real and imaginary components of the 10-point DFT coefficients of the rectangular pulse x_t shown in panel A. Panels D and E show the magnitude and phase of the DFT coefficients. Panels F through H show reconstructions of x_t using progressively more frequencies. The reconstruction matches the original pulse exactly when all frequencies are included.

consider time-frequency analysis using the Haar Wavelet Transform.

2.4 The Haar Wavelet Transform

In this section, we use a similar explanation to Bandi et al. (2017) to show how moving average filters in the time domain can be used to calculate the frequency components of a stationary stochastic process. Expanding the logic of Beveridge and Nelson (1981), who popularized the decomposition of a non-stationary time series into a trend component and a transient component, we can use the wavelet transform to decompose a time series into components operating over different levels of persistence.

Given a time series x_t we can construct moving averages $\overline{x}_t^{(k)}$ of window length 2^k,

$$\overline{x}_t^{(k)} = \frac{1}{2^k} \sum_{p=0}^{2^k-1} x_{t-p} \quad , \tag{2.7}$$

where $\overline{x}_t^{(0)} \equiv x_t$. Defining $\widetilde{x}_t^{(k)}$ to be the difference between moving averages of lengths 2^{k-1} and 2^k,

$$\widetilde{x}_t^{(k)} = \overline{x}_t^{(k-1)} - \overline{x}_t^{(k)} \quad . \tag{2.8}$$

we find that, $\widetilde{x}_t^{(k)}$ contains fluctuations that are not filtered out by the moving average of length 2^{k-1}, but do get filtered by the moving average of length 2^k. Therefore, the surviving fluctuations have periods in the interval $[2^k, 2^{k+1})$. Similarly, the moving average $\overline{x}_t^{(k)}$ includes fluctuations with periods that exceed length 2^{k+1}. Using (2.8), we can express x_t as,

$$x_t = \sum_{k=1}^{K} \widetilde{x}_t^{(k)} + \overline{x}_t^{(K)} \quad , \tag{2.9}$$

for any $K \geq 1$. The values $\widetilde{x}_t^{(k)}$ are known as the wavelet filter coefficients at level k, and $\overline{x}_t^{(K)}$ are the scaling filter coefficients at level K. Intuitively, (2.9) decomposes x_t into a sum of frequency-specific components, and a residual low-frequency moving average term. As K increases, $\overline{x}_t^{(K)}$ approaches a constant, long-term average. In practice, the choice of K is informed by the available sample length.

Conveniently, this decomposition can be implemented using a orthonormal matrix operator. For $K = 2$, the system of equations can be expressed as,

$$\begin{pmatrix} \overline{x}_t^{(2)} \\ \widetilde{x}_t^{(2)} \\ \widetilde{x}_t^{(1)} \\ \widetilde{x}_{t-2}^{(2)} \end{pmatrix} = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\ \frac{1}{2} & -\frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} x_t \\ x_{t-1} \\ x_{t-2} \\ x_{t-3} \end{pmatrix} \quad . \tag{2.10}$$

Letting $\mathcal{T}^{(2)}$ be the (4×4) matrix in (2.10), it can be shown that that $\mathcal{T}^{(2)}$ is orthogonal such that $\Lambda^{(2)} \equiv \mathcal{T}^{(2)}(\mathcal{T}^{(2)})^{\mathsf{T}}$ is diagonal and well-defined (Bandi et al., 2017). Therefore, by matrix inversion, one can reconstruct the original process given the filtered components,

$$\begin{pmatrix} x_t \\ x_{t-1} \\ x_{t-2} \\ x_{t-3} \end{pmatrix} = \left(\mathcal{T}^{(2)}\right)^{-1} \begin{pmatrix} \overline{x}_t^{(2)} \\ \widetilde{x}_t^{(2)} \\ \widetilde{x}_t^{(1)} \\ \widetilde{x}_{t-2}^{(2)} \end{pmatrix} \quad . \tag{2.11}$$

An extension of this procedure to any $K \geq 2$, and a recursive algorithm for the construction of the matrix $\mathcal{T}^{(K)}$ associated to any arbitrary K value is provided in Mallat (1989). The matrix $\mathcal{T}^{(K)}$ is commonly known as the Haar Wavelet Transform (HWT).

Similar to the DFT, the maximum-overlap discrete wavelet transform described in this section is an energy-preserving transform, and therefore we can use it to calculate the frequency-specific measures of volatility, alpha, and beta coefficients that are described in the rest of Part II. Alternative wavelet transforms, such as the Daubechies wavelet, can also be used. However, the structure of the HWT is useful as it makes explicit the relationship between frequency components and moving averages.

2.5 The Power Spectrum

In many situations, a time series can be modeled as the realization of a stochastic process, which can often be characterized by its first and second moments. The DTFT

of the auto- and cross-covariance functions can then be interpreted as the frequency distribution of the power contained within the variance and covariance of these time series, respectively. Similarly, the inverse DTFT can be used to find the lagged second moments as functions of the auto- and cross-power spectra.

Let $\{x_t\}$ and $\{y_t\}$ form real-valued discrete-time wide-sense stationary stochastic processes with means m_x and m_y, and cross covariance function $\gamma_{xy}[m] = \mathrm{E}[(x_{t+m} - m_x)(y_t - m_y)]$.[3] Assuming the cross-covariance function has finite energy, let $P_{xy}(\omega)$ be its DTFT,

$$P_{xy}(\omega) = \sum_{m=-\infty}^{\infty} \gamma_{xy}[m] e^{-j\omega m}. \tag{2.12}$$

The function $P_{xy}(\omega)$ is known as the cross-spectrum. Its real component, known as the co-spectrum, can be interpreted as the frequency decomposition of the covariance between x_t and y_t. Specifically, the covariance between $\{x_t\}$ and $\{y_t\}$ can be calculated using the inverse DTFT of $P_{xy}(\omega)$,

$$\mathrm{Cov}(x_t, y_t) \equiv \gamma_{xy}[0] = \frac{1}{2\pi} \int_0^{2\pi} \Re[P_{xy}(\omega)]. \tag{2.13}$$

We denote the co-spectrum[4] as $L_{xy}(\omega) \equiv \Re[P_{xy}(\omega)]$.

This calculation of the power and cross-power spectra from the auto- and cross-covariance functions assumes the first and second moments of the stochastic process are known and do not change with time; however, for practical applications, especially those in finance, the underlying distributions are often unknown and nonstationary. To address this issue, we compute the short-time Fourier transform to decompose rolling-window covariances into their frequency components. This approach uses the DFT to express windowed subsamples of x_t and y_t in the frequency domain, and then analyzes their magnitude and phase. When the time series are in phase at a given

[3]Specifically, the stochastic processes $\{x_t\}$ and $\{y_t\}$ are said to be wide-sense stationary if and only if $\mathrm{E}[x_t]$ and $\mathrm{E}[y_t]$ are constants independent of t, and $\mathrm{E}[x_{t_1} x_{t_2}]$, $\mathrm{E}[y_{t_1} y_{t_2}]$ and $\mathrm{E}[x_{t_1} y_{t_2}]$ depend only on the time difference $(t_1 - t_2)$.

[4]The co-spectrum, $L_{xy}(\omega)$, is the real part of the cross-spectrum, $P_{xy}(\omega)$. The imaginary part, $Q_{xy}(\omega)$, is called the quadrature spectrum.

frequency, the contribution that frequency makes to the sample covariance is positive; when they are out of phase, that particular frequency's contribution will be negative. Longer windows will provide better frequency resolution, but will conflict with our ability to resolve changes in the statistical properties of signals over time.

Specifically, consider a real-valued subsample of x_t and y_t from times $t = 0, \ldots, T-1$. The sample covariance over this interval can be calculated as:

$$\mathrm{Cov}\langle x_t, y_t \rangle = \frac{1}{T} \sum_{t=0}^{T-1} (x_t - \overline{x})(y_t - \overline{y}), \tag{2.14}$$

where $\overline{x}$ and $\overline{y}$ are the sample means of x_t and y_t over the same subperiod. This calculation is exactly equivalent to the one formed using the T-point DFT:

$$\mathrm{Cov}\langle x_t, y_t \rangle = \frac{1}{T} \sum_{k=1}^{T-1} \hat{L}_{xy}[k] \quad , \quad \hat{L}_{xy}[k] \equiv \frac{1}{T} \Re[X_k^* Y_k] \tag{2.15}$$

where X_k and Y_k are the T-point DFT coefficients of the subsample of x_t and y_t. Thus, the sum over $\hat{L}_{xy}[k]$ is proportional to the sample covariance of x_t and y_t. Moreover, the sum of $\hat{L}_{xy}[k]$ over a band of frequencies, $\mathrm{Cov}_K\langle x_t, y_t \rangle \,|\, K \subseteq \{1, \ldots, T-1\}$, is proportional to that band's contribution to the sample covariance. For this reason the function $\hat{L}_{xy}[k]$, called the cross-periodogram, is an estimate of the co-spectrum at the harmonic frequency ω_k, and can be interpreted as the frequency distribution of the power contained in the sample covariance. It can be shown that these estimators are asymptotically unbiased, but inconsistent. Practical implementation details, including the standard errors of these estimators, are discussed in Appendix A. Further references on the statistical properties of spectrum estimates can be found in, for example, Jenkins and Watts (1968), Hannan (1970), Anderson (1971), Priestly (1981), Brockwell and Davis (1991), Brillinger (2001), Velasco and Robinson (2001), Phillips et al. (2006, 2007), Shao and Wu (2007), Oppenheim and Schafer (2009), and Wu and Zaffaroni (2018).

Note that $k = 0$, the zero frequency, is not involved in (2.15) since adding or subtracting a constant to either time series does not change the sample covariance.

In addition, as mentioned in Section 2.3, values of k that are symmetric about $T/2$ (e.g., $k = 1$ and $k = T - 1$) have the same frequency and their contributions to the sample covariance are equivalent. Therefore, pairs of elements that correspond to the same frequency should be included together in the frequency band K to form the one-sided spectrum.[5]

As an illustrative example, suppose that,

$$x_t = \alpha_x + \beta_x F_t + u_t, \tag{2.16}$$

$$y_t = \alpha_y + \beta_y F_{t-1} + v_t, \tag{2.17}$$

where α_x, α_y, β_x and β_y are constants, and F_t, u_t and v_t are white-noise random variables that are uncorrelated at all leads and lags.

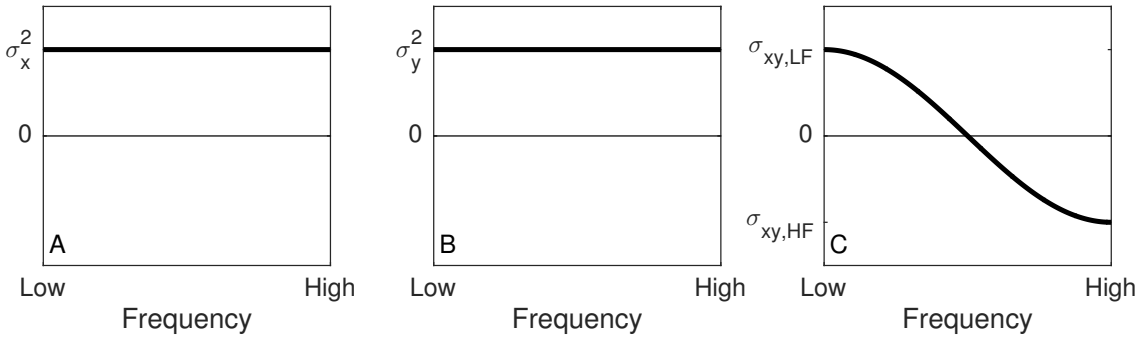

Figure 2-3: The spectral decomposition of (A) the variance of x_t, (B) the variance of y_t, and (C) the covariance of x_t and y_t.

Panels A and B display the one-sided co-spectrums, $L_{xx}(\omega)$ and $L_{yy}(\omega)$. Since x_t and y_t are serially uncorrelated at all leads and lags, their power spectrums are flat, and each frequency contributes equally to the variance. In this example, the lagged dependence of y_t on F_t relative to x_t suggests that x_t and y_t will be in phase over longer time horizons, and out of phase over shorter time horizons. As shown in panel C, this leads to a positive contribution to the covariance at low frequencies, and a negative contribution at high frequencies. Moreover, since x_t and y_t are uncorrelated, we find that $L_{xy}(\omega)$ integrates to 0.

[5]For example, see the one-sided and two-sided power spectrums in Figure 2-4. For real-valued time series, the cross-spectrum is conjugate symmetric causing the quadrature spectrum components to cancel. For this reason, we focus on the co-spectrum.

2.6 The Business Cycle

One of the most natural applications of spectral analysis is to measure the business cycle, which many studies have done (Granger and Hatanaka, 1964; King and Watson, 1996; Baxter and King, 1999). Consider U.S. real GDP from the onset of the Great Moderation in the mid-1980s to 2015. The annualized quarterly percentage change in seasonally adjusted real GDP is plotted in panel A of Figure 2-4. Notice the data exhibit longer-scale cyclical patterns in accordance with recessions and expansions, as well as high-frequency oscillations related to more transitory dynamics.

As a first step, in panel B of Figure 2-4 we subtract the mean, to view fluctuations about the long-term growth rate. We then apply the DFT to decompose this adjusted time series into its frequency components, and plot the estimated two-sided power spectrum in panel C of Figure 2-4. The horizontal axis of this graph is now frequency instead of time, and the spectrum is symmetric about the center frequency. Therefore, it is common to aggregate coupled frequencies into the one-sided power spectrum as shown in panel D of Figure 2-4. In this form, it is clear that a substantial portion of the signal's power resides in low frequencies less than 1 cycle every 5 years. These periods correspond to economic expansions and recessions, i.e., the business cycle. A reconstruction of the original time series using only these low frequencies is shown in panel E of Figure 2-4. Notice the more transitory components have been removed, and what remains features the recession of the early 1990s, the internet bubble, the Financial Crisis, and the subsequent recovery. A second reconstruction using frequencies less than or equal to 1 cycle per year, is shown in panel F of Figure 2-4, and provides a more realistic reconstruction of the original data with more transitory effects. This example demonstrates how spectral analysis can often reveal structure in a time series not immediately evident in the raw data.

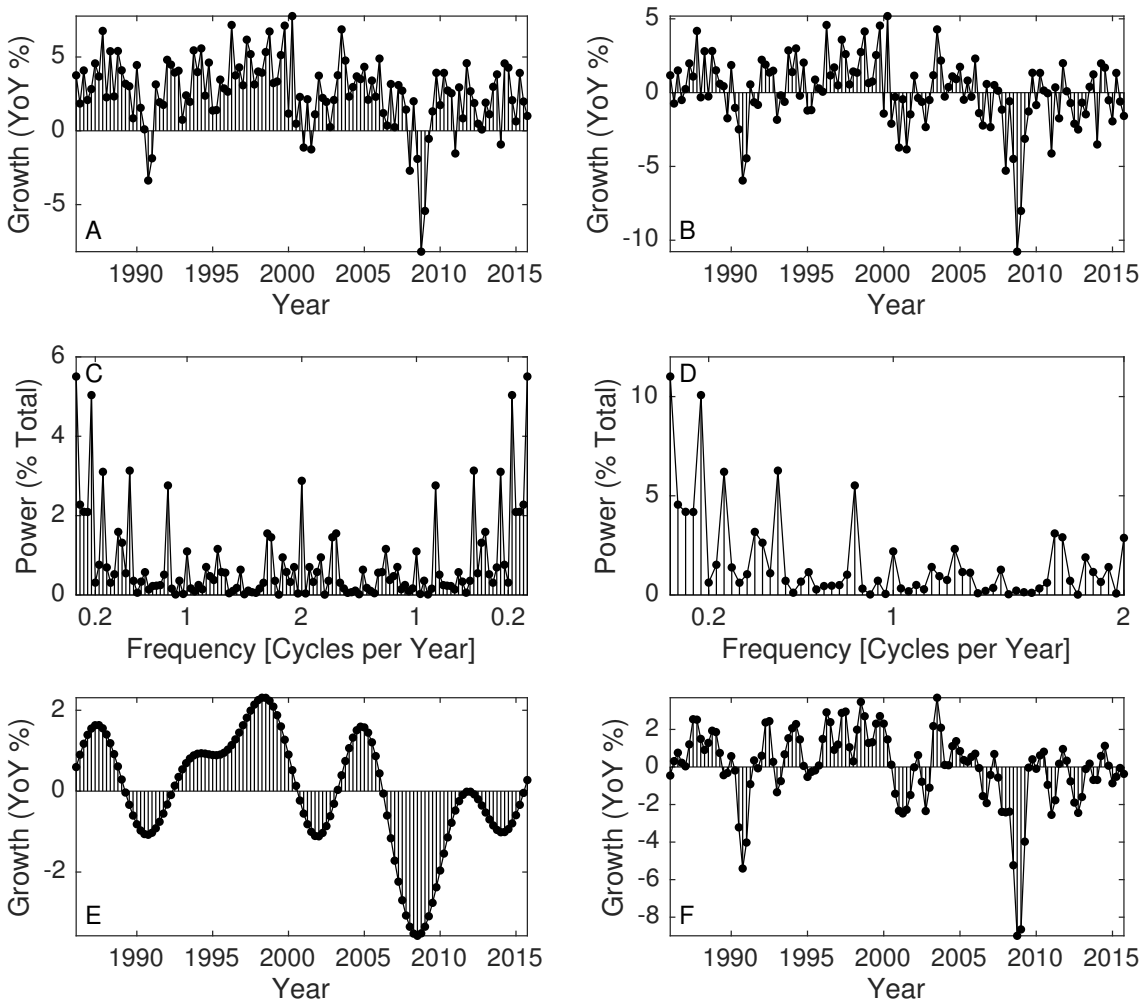

Figure 2-4: Illustration of how the DFT can be used to implement the spectral decomposition of a time series. The annualized quarterly percentage change in seasonally adjusted US real GDP from 1986 to 2015 is plotted in panel A. The same time series minus its mean is shown in panel B. Panel C shows its two-sided power spectrum after applying the DFT. Note that the horizontal axis represents frequency, and the vertical axis represents the relative contribution of each frequency to the overall variability of the time series. Panel D aggregates pairs of equivalent frequencies into the one-sided power spectrum. Panels E and F plot reconstructions of the time series using frequencies less than 1 cycle per 5 years, and less than or equal to 1 cycle per year, respectively.

2.7 Volatility

Estimating volatility is central to mean-variance portfolio management, performance attribution, and risk management. A spectral decomposition of returns allows us to measure the fraction of variability that can be attributed to fluctuations at different time scales.

Let x_t be the one-period return of a security between dates $t-1$ and t. The sample variance of returns over an interval from $t = 0, \ldots, T-1$ can be decomposed into its frequency components using (2.15):

$$\mathrm{Var}\langle x_t \rangle = \frac{1}{T} \sum_{k=1}^{T-1} \hat{L}_{xx}[k] \quad , \quad \hat{L}_{xx}[k] \equiv \frac{1}{T} |X_k|^2 \tag{2.18}$$

where X_k are the T-point DFT coefficients of the subsample of x_t. As an illustrative example, Figure 2-5 decomposes the 10-year rolling sample variance of the daily returns of the CRSP value-weighted and equal-weighted market indices from 1926 to 2015 into its low (less than 1 cycle per month), medium (between 1 cycle per month to 1 cycle per week), and high (more than 1 cycle per week) frequency components. This spectral decomposition is compared to a white noise null hypothesis where the windowed returns were rendered serially uncorrelated by generating random permutations of their order. This exercise was repeated 10,000 times from which 95% confidence intervals were formed around the flat-band, white-noise null hypothesis.

Our analysis shows that, from the mid-1960s to late-1990s, the variance of both value-weighted and equal-weighted market returns exhibited smaller fluctuations at short time scales (between 2 and 5 days), and greater fluctuations at longer time scales (greater than 1 month), than would be expected if returns were serially uncorrelated. In fact, this effect is more pronounced, and continues into the late 2000s, for the equal-weighted market returns. This low-frequency power is in agreement with the large positive serial correlation in weekly returns described in Lo and MacKinlay (1990), which would tend to shift power from high frequencies to lower frequencies. However, this spectral decomposition also shows these dynamics weakening over the subsequent

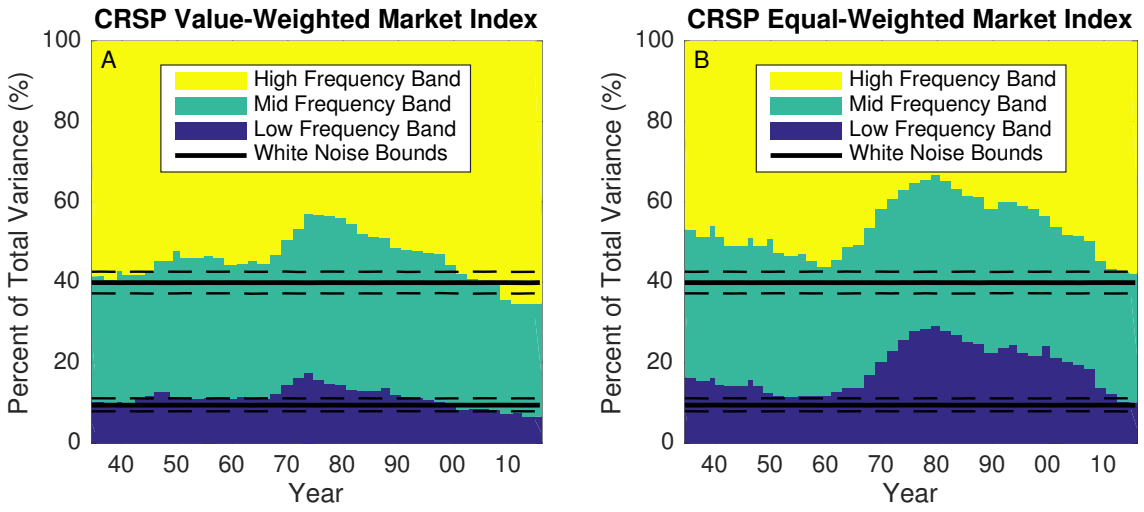

Figure 2-5: Spectral decomposition of the 10-year rolling sample variance of the daily returns of CRSP value-weighted market index (panel A), and the CRSP equal-weighted market index (panel B) from 1926 to 2015. Frequency components are grouped into 3 categories: high frequencies (more than 1 cycle per week), mid frequencies (between 1 cycle per week and 1 cycle per month), and low frequencies (less than 1 cycle per month).

decades, most likely in response to increased competitive forces and technological advances such improved telecommunications, standardized electronic information exchange protocols, and automated trading. This simple example demonstrates the usefulness of the frequency domain in visualizing complex dynamics that may exist over a wide range of leads and lags in the time domain. For example, in addition to tests in the time domain that detect local correlation between between neighboring samples, the power spectrum allows us to detect departures from white noise caused by periodic effects such as seasonal variation.

Chapter 3

Dynamic Alpha

The value added by an active investor is traditionally measured using alpha, tracking error, and the information ratio. However, these measures do not characterize the dynamic component of investor activity, nor do they consider the time horizons over which weights are changed. In this chapter, we propose a technique to measure the value of active investment that captures both the static and dynamic contributions of an investment process. This dynamic alpha is based on the decomposition of a portfolio's expected return into its frequency components using spectral analysis. The result is a static component that measures the portion of a portfolio's expected return due to passive investments and security selection, and a dynamic component that captures the manager's timing ability across a range of time horizons. Our framework can be universally applied to any portfolio, and is a useful method for comparing the forecast power of different investment processes. Several analytical and empirical examples are provided to illustrate the practical relevance of this decomposition.

3.1 Introduction

The shortest decision interval of a modern investment strategy may range from microseconds to years, a wide span of time horizons. While the legendary value investor Warren Buffett tends to change his portfolio weights rather slowly, the same cannot be said for famed day trader Steven Cohen of SAC Capital, yet both manage to gen-

erate enormous value through active investment. Although alpha, tracking error, and the information ratio are the standard tools for gauging the value-added of a portfolio manager, they can obscure important features of the underlying process by which information is reflected in investment decisions. Specifically, none of these standard performance metrics directly measure the dynamic relationship between weights and returns, which is the central focus of active investment strategies.

In this chapter, we propose a new approach to analyzing investment strategies in which the frequency component is explicitly captured. Using the tools of spectral analysis—the decomposition of time series into a sum of periodic functions like the sine and cosine functions—we show that investment strategies can differ significantly in the frequencies with which their expected returns are generated. Slower-moving strategies will exhibit more "power" at the lower frequencies, while faster-moving strategies will exhibit more power at the higher frequencies. By identifying the particular frequencies that are responsible for a given strategy's expected returns, an investor will have an additional dimension with which to manage the risk/reward profile of his portfolio.

We begin in Section 3.2 with a brief review of the financial spectral analysis literature. Our main results are contained in Sections 3.3 and 3.4, where we provide spectral decompositions for an investment strategy's forecast power. We provide numerical and empirical illustrations of these techniques in Sections 3.5–3.7, and conclude in Section 3.8.

3.2 Literature Review

In this chapter, we show that spectral analysis can be used to characterize and refine active investment strategies. The standard tools used for performance attribution originated from the Capital Asset Pricing Model (CAPM) of Sharpe (1964) and Lintner (1965). The difference between an investment's expected return and the risk-adjusted value predicted by the CAPM is referred to as alpha, and Treynor (1965), Sharpe (1966), and Jensen (1968, 1969) applied this measure to quantify the value-added of mutual-fund managers. Since then, a number of related measures have been

developed, including the Sharpe, Treynor, and information ratios. However, none of these measures explicitly depend on the relative timing of portfolio weights and returns in gauging investment skill.

In contrast, Lo (2008) proposed a novel measure of active management that quantifies the predictive power of an investment process by decomposing the expected portfolio return into the covariance between the underlying security weights and returns and the product of the average weights and average returns. In this context, a successful portfolio manager is one whose decisions induce a positive correlation between portfolio weights and returns. Since portfolio weights are a function of a manager's decision process and proprietary information, positive correlation is a direct indication of forecast power and, consequently, investment skill.

As an extension of this decomposition, we introduce the concept of dynamic alpha, which uses spectral analysis to measure the forecast power of a portfolio manager across multiple time horizons. An investment process is said to be profitable at a given frequency if there is a positive correlation between portfolio weights and returns at that frequency. When aggregated across frequencies, dynamic alpha is equivalent to Lo's (2008) active component, and therefore provides a clear indication of a manager's forecast power across time horizons. This connects spectral analysis to the standard tools of modern portfolio theory, allowing us to study the time-horizon properties of investment performance.

3.3 Dynamic Alpha

In this section, we propose an explicit measure of the value of active management—dynamic alpha—that takes into account forecast power across multiple time horizons. Expanding on the framework of the decomposition developed by Lo (2008), we use the DFT to separate the expected return of a portfolio into distinct components that depend on the correlation between portfolio weights and returns at different frequencies. The result is one component that measures the portion of a portfolio's expected return due to passive investments and active security selection, and multiple

dynamic components that capture the manager's timing ability across a range of time horizons. Our method closely parallels Hasbrouck and Sofianos (1993); however, we make a novel modification to their analysis to make it applicable to the expected returns of portfolios.

Our approach uses the DFT to express the portfolio's underlying security weights and returns in the frequency domain and then analyzes their phase. When the weights and returns are in phase at a given frequency, the contribution that frequency makes to the portfolio's expected return is positive. When they are out of phase, then that particular frequency's contribution will be negative.

If we consider a portfolio with N securities, then for $t = 0, \ldots, T-1$, the average one-period portfolio return can be calculated as,

$$\overline{r_p} = \frac{1}{T} \sum_{i=1}^{N} \sum_{t=0}^{T-1} w_{i,t} r_{i,t}, \tag{3.1}$$

where $w_{i,t}$ and $r_{i,t}$ are the realized weight and return of the ith stock at time t, respectively. Using the definition of covariance, the average portfolio return can be decomposed into a dynamic alpha component (δ_p) and a static component (ν_p) as follows,

$$\overline{r_p} = \delta_p + \nu_p, \tag{3.2}$$

$$\delta_p = \sum_{i=1}^{N} \mathrm{Cov}\langle w_{i,t}, r_{i,t} \rangle \ , \quad \nu_p = \sum_{i=1}^{N} \overline{w_{i,t}} \cdot \overline{r_{i,t}} \ . \tag{3.3}$$

The value of the static component arises from the manager's average position in a security, and can be thought of as the portion of the portfolio's return that results from collecting risk premiums, as well as the ability to select securities with favorable long-term prospects. This distinction contrasts with Lo's (2008) use of the term "passive" for the static component—in our setting, we wish to acknowledge the possibility that active management is responsible for long-term bets on specific securities, in which case a portion of a portfolio's static component may, in fact, be alpha rather than risk premia.

The value of the dynamic alpha component consists of the profitability of the portfolio manager's conscious decision to buy, sell, or avoid a security by aggregating the sample covariances between the portfolio weights, $w_{i,t}$, and security returns, $r_{i,t}$. In particular, if a manager has positive weights when security returns are positive and negative weights when returns are negative, this implies positive covariances between portfolio weights and returns, and will have a positive impact on the portfolio's average return. In effect, the covariance term captures the manager's timing ability, asset by asset.

Spectral analysis allows us to decompose this covariance term further, capturing the manager's timing ability over multiple time horizons,

$$\delta_p = \sum_{k=1}^{T-1} \delta_{p,k} \quad , \quad \delta_{p,k} = \frac{1}{T^2} \sum_{i=1}^{N} \Re[W_{i,k}^* R_{i,k}] \ , \tag{3.4}$$

where $\Re[z]$ and z^* denote the real part and complex conjugate of a complex number z, respectively, and $W_{i,k}$ and $R_{i,k}$ are the T-point DFT coefficients (see Section A.1 in the Online Supplement) of the weights and returns for stock i. In this form, the contribution to the average portfolio return by the kth harmonic frequency, where $k \in \{0, ..., T-1\}$, is clearly visible. The lowest frequency occurs at $k = 0$, and the highest frequency occurs at the value of k closest to $T/2$. Values of k that are symmetric about $T/2$ (e.g., $k = 1$ and $k = T-1$) have the same frequency, and their contributions to the average portfolio return are equivalent. The relation $h = TT_s/k$, where T_s is the time between samples and $0 \leq k \leq T/2$, can be used to convert the kth harmonic frequency to its corresponding time horizon, h. We also note that $\delta_{p,0} = \nu_p$, and it is often convenient to include $\delta_{p,0}$ when computing the DFT.

Simply put, this spectral decomposition first deconstructs the weights and returns into their various frequency components. At each frequency, if the weights and returns are in phase, then that time horizon's contribution to the average portfolio return will be positive. If the two signals are out of phase, then that particular frequency's contribution will be negative. For this reason, a value-weighted portfolio of all securities, which is traditionally considered passive, will contain no dynamic alpha across all

frequencies as long as the individual security returns are serially uncorrelated (i.e., the Random Walk Hypothesis holds for all securities). On the other hand, if returns are serially correlated, then it is possible for a buy-and-hold portfolio to yield a non-zero dynamic alpha because changes in its weights will contain information related to future returns. To distinguish between dynamically managed alpha and passive portfolios that unintentionally contain non-zero dynamic alpha, we must therefore rely on the manager's stated intentions.

In addition to quantifying the value added from active management across time horizons, we can also gauge the consistency of a portfolio manager's timing ability. Historically, the consistency of investment skill has been characterized by the volatility of the tracking error, which is a measure of the variability of the difference between the portfolio return and some benchmark return. Low tracking error volatility and a positive excess return (i.e., alpha) indicates that the manager is reliably adding value through active management. The ratio of alpha to the tracking error volatility measures the efficiency with which a manager generates excess returns and is called the *information ratio*. The higher the information ratio, the better the manager.

These measures can be incorporated into our framework by defining the *dynamic risk*, σ_δ, as the variability of the difference between the portfolio return, $r_{p,t}$, and the static component, $\nu_{p,t} = \sum_{i=1}^{N} \overline{w_{i,t}} \cdot r_{i,t}$. Specifically,

$$\sigma_\delta = \sqrt{\mathrm{Var}\langle r_{p,t} - \nu_{p,t} \rangle}, \tag{3.5}$$

where σ_δ is a measure of the risk taken by the portfolio manager in an attempt to generate higher returns by engaging in timing decisions. The dynamic information ratio, I_δ, can then be defined as,

$$I_\delta = \frac{\delta_p}{\sigma_\delta}, \tag{3.6}$$

and is a risk-adjusted measure of the dynamic alpha component. These performance metrics can be calculated for a specific range of time horizons by aggregating the frequency components of δ_p and σ_δ over the band of interest. This provides us with a risk-adjusted measure of the manager's timing ability for a specific frequency band.

Intuitively, it quantifies the manager's predictive power across a range of time horizons, but also attempts to identify the consistency of this power.

3.4 Alpha vs. Beta

To distinguish explicitly between manager outperformance and portfolio exposure to factor risk, we have to impose additional structure on the returns of the individual assets. Specifically, we consider a linear M-factor model,

$$x_{i,t} = \alpha_i + \beta_{i,1} F_{1,t} + \cdots + \beta_{i,M} F_{M,t} + \varepsilon_{i,t} \ , \tag{3.7}$$

where $x_{i,t}$ is defined to be the excess return of asset i, in excess of the risk-free rate of return, $r_{f,t}$, $F_{m,t}$ are excess factor returns, and $\mathrm{E}[\varepsilon_{i,t} \mid F_{1,t}, \ldots, F_{M,t}] = 0$. This specification is consistent with Merton's (1973a) Intertemporal Capital Asset Pricing Model and Ross's (1976) Arbitrage Pricing Theory. Since our expected-return decomposition is considerably more general than any particular asset-pricing model or linear-factor structure, we allow for an intercept, α_i, in our framework.

Under these assumptions, the portfolio's exposure to factor m is $\beta_{p,m,t} = \sum_{i=1}^{N} w_{i,t} \beta_{i,m}$. The average return of a portfolio of assets (3.2) can then be rewritten as,

$$\overline{r_p} = \underbrace{\text{Risk-Free Rate} + \text{Risk Premia} + \text{Security Selection}}_{\nu_p \equiv \text{Static Component}} + \underbrace{\text{Factor Timing} + \text{Security Timing}}_{\delta_p \equiv \text{Dynamic Component}} \tag{3.8}$$

where,

$$\text{Risk-Free Rate} \equiv \quad \overline{r_{f,t}} \tag{3.9}$$

$$\text{Risk Premia} \equiv \sum_{m=1}^{M} \overline{\beta_{p,m,t}} \cdot \overline{F_{m,t}} \tag{3.10}$$

$$\text{Security Selection} \equiv \sum_{i=1}^{N} \overline{w_{i,t}} \cdot \alpha_i \tag{3.11}$$

$$\text{Factor Timing} \equiv \sum_{m=1}^{M} \text{Cov}\langle \beta_{p,m,t}, F_{m,t} \rangle \tag{3.12}$$

$$\text{Security Timing} \equiv \sum_{i=1}^{N} \text{Cov}\langle w_{i,t}, \varepsilon_{i,t} \rangle \ . \tag{3.13}$$

Due to the structure of the linear multi-factor model, (3.8) is a more refined decomposition than (3.2). The average portfolio returns are now the sum of five components: a risk-free rate component, a risk-premia component that represents the return from the passive exposures to factor risk, a security selection component that depends on the α_i's, a factor-timing component that depends on the covariance between the portfolio's factor exposures and the underlying factors, and finally, a security-timing component that depends on the covariance between weights and the idiosyncratic component of security returns. Note that the factor- and security-timing terms can be decomposed further into their frequency components.

This factor-based decomposition demonstrates that investment expertise can manifest itself in two distinct ways: identifying cheap sources of expected return (i.e., the α_i's, which are reflected in the static component, ν_p), and creating additional expected return through factor- and security-specific timing across different time horizons (i.e., the covariance terms, which are reflected in the spectral decomposition of the dynamic component, $\delta_{p,k}$). Thus, even if all α_i's are zero, as most asset-pricing models claim, there can still be substantial value-added from active management if the investment process has the ability to time price movements over certain time horizons.

3.5 Numerical Examples

To develop further intuition for our spectral decomposition, consider the following simple numerical example of a portfolio of two assets, one that yields a monthly return that alternates between 1% and 2% (Asset 1) and the other that yields a fixed monthly return of 0.15% (Asset 2). Let the weights of this portfolio, called A1, be given by 75% in Asset 1 and 25% in Asset 2. Table 3.1 illustrates the dynamics of this portfolio over a 12-month period, where the average return of the portfolio is

1.1625% per month, all of which is due to the static component. In this case, because the weights are constant, the dynamic risk measure will also be 0%.

Month	w_1	r_1	w_2	r_2	r_p
		Strategy A1			
1	75%	1.00%	25%	0.15%	0.7875%
2	75%	2.00%	25%	0.15%	1.5375%
3	75%	1.00%	25%	0.15%	0.7875%
4	75%	2.00%	25%	0.15%	1.5375%
5	75%	1.00%	25%	0.15%	0.7875%
6	75%	2.00%	25%	0.15%	1.5375%
7	75%	1.00%	25%	0.15%	0.7875%
8	75%	2.00%	25%	0.15%	1.5375%
9	75%	1.00%	25%	0.15%	0.7875%
10	75%	2.00%	25%	0.15%	1.5375%
11	75%	1.00%	25%	0.15%	0.7875%
12	75%	2.00%	25%	0.15%	1.5375%
Mean:	75%	1.50%	25%	0.15%	**1.1625**%

Spectral decomposition of $\overline{r_p}$

ν_p	$2\delta_{p,1}$	$2\delta_{p,2}$	$2\delta_{p,3}$	$2\delta_{p,4}$	$2\delta_{p,5}$	$\delta_{p,6}$
1.1625%	0%	0%	0%	0%	0%	0%

Table 3.1: The expected return of a constant portfolio depends only on the static component.

Now consider portfolio A2, which differs from A1 only in that the portfolio weight for Asset 1 alternates between 50% and 100% in phase with Asset 1's returns which alternates between 1% and 2% (see Table 3.2). In this case, the total expected return is 1.2875% per month, of which 0.1250% is due to the positive correlation between the portfolio weight for Asset 1 and its return at the shortest-time horizon (i.e., highest frequency). In addition, the dynamic risk for this portfolio is 0.3375%, and the dynamic information ratio is about 0.37.

Finally, consider a third portfolio A3 which also has alternating weights for Asset 1, but exactly out of phase with Asset 1's returns—when the return is 1%, the portfolio weight is 100%, and when the return is 2%, the portfolio weight is 50%. Table 3.3 confirms that this is counterproductive as Portfolio A3 loses 0.1250% per month from its highest frequency component, and its total expected return is only 1.0375%. In this case, the dynamic risk is 0.3375%, and the dynamic information ratio is -0.37.

Month	w_1	r_1	w_2	r_2	r_p
			Strategy A2		
1	50%	1.00%	50%	0.15%	0.5750%
2	100%	2.00%	0%	0.15%	2.0000%
3	50%	1.00%	50%	0.15%	0.5750%
4	100%	2.00%	0%	0.15%	2.0000%
5	50%	1.00%	50%	0.15%	0.5750%
6	100%	2.00%	0%	0.15%	2.0000%
7	50%	1.00%	50%	0.15%	0.5750%
8	100%	2.00%	0%	0.15%	2.0000%
9	50%	1.00%	50%	0.15%	0.5750%
10	100%	2.00%	0%	0.15%	2.0000%
11	50%	1.00%	50%	0.15%	0.5750%
12	100%	2.00%	0%	0.15%	2.0000%
Mean:	75%	1.50%	25%	0.15%	**1.2875**%

Spectral decomposition of $\overline{r_p}$

ν_p	$2\delta_{p,1}$	$2\delta_{p,2}$	$2\delta_{p,3}$	$2\delta_{p,4}$	$2\delta_{p,5}$	$\delta_{p,6}$
1.1625%	0%	0%	0%	0%	0%	0.1250%

Table 3.2: The dynamics of the portfolio weights are positively correlated with returns at the shortest time horizon, which adds value to the portfolio and yields a positive contribution from the highest frequency ($\delta_{p,6}$).

Month	w_1	r_1	w_2	r_2	r_p
			Strategy A3		
1	100%	1.00%	0%	0.15%	1.0000%
2	50%	2.00%	50%	0.15%	1.0750%
3	100%	1.00%	0%	0.15%	1.0000%
4	50%	2.00%	50%	0.15%	1.0750%
5	100%	1.00%	0%	0.15%	1.0000%
6	50%	2.00%	50%	0.15%	1.0750%
7	100%	1.00%	0%	0.15%	1.0000%
8	50%	2.00%	50%	0.15%	1.0750%
9	100%	1.00%	0%	0.15%	1.0000%
10	50%	2.00%	50%	0.15%	1.0750%
11	100%	1.00%	0%	0.15%	1.0000%
12	50%	2.00%	50%	0.15%	1.0750%
Mean:	75%	1.50%	25%	0.15%	**1.0375**%

Spectral decomposition of $\overline{r_p}$

ν_p	$2\delta_{p,1}$	$2\delta_{p,2}$	$2\delta_{p,3}$	$2\delta_{p,4}$	$2\delta_{p,5}$	$\delta_{p,6}$
1.1625%	0%	0%	0%	0%	0%	−0.1250%

Table 3.3: The dynamics of the portfolio weights are negatively correlated with returns at the shortest time horizon, which subtracts value from the portfolio and yields a negative contribution from the highest frequency ($\delta_{p,6}$).

Note that in all three cases, the static components are identical at 1.1625% per month because the average weight for each asset is the same across all three portfolios. The only differences among A1, A2, and A3 are the dynamics of the portfolio weights at the shortest time horizon. These differences give rise to different values for the highest frequency component. As shown in (3.4), contributions from higher frequencies ($k > 0$) sum to the overall dynamic component. These higher frequency contributions can be interpreted as the portion of the dynamic component that arises from a given time horizon.

For a more realistic example, consider the long/short equity market-neutral strategy of Lo and MacKinlay (1990):

$$w_{i,t} = -\frac{1}{N}(r_{i,t-1} - r_{m,t-1}), \tag{3.14}$$

$$r_{m,t-1} = \frac{1}{N}\sum_{i=1}^{N} r_{i,t-1} \quad . \tag{3.15}$$

By buying the losers and selling the winners from date $t-1$ at the onset of each date t, this strategy actively bets on mean reversion across all N stocks, and profits from reversals that occur within the subsequent interval. For this reason, Lo and MacKinlay (1990) termed this strategy "contrarian," as it benefits from market overreaction and mean reversion, that is, when underperformance is followed by positive returns and outperformance is followed by negative returns. By construction, the weights sum to zero, and therefore the strategy is also considered a "dollar-neutral" or "arbitrage" portfolio. This implies that much of the portfolio's return should be due to active management, and that value will be added near frequencies inversely related to the mean reversion period.

Now suppose that stock returns satisfy the following simple MA(1) model,

$$r_{i,t} = \varepsilon_{i,t} + \lambda\varepsilon_{i,t-1}, \tag{3.16}$$

where the $\varepsilon_{i,t}$ are serially and cross-sectionally uncorrelated white-noise random variables with variance σ^2. In this case, the expected one-period portfolio return can be

calculated as,

$$\mathrm{E}[r_p] = -\lambda\sigma^2(1 - \frac{1}{N}) \quad . \tag{3.17}$$

We see that the expected return is proportional to the mean reversion factor, λ, and the volatility factor, σ^2. Applying our spectral decomposition (see Section A.2 in the Online Supplement), we find that,

$$\delta_{p,\omega} = -\sigma^2\Big(1 - \frac{1}{N}\Big)\Big(\lambda\cos(2\omega) + (1+\lambda^2)\cos(\omega) + \lambda\Big), \quad \omega \in [0, 2\pi). \tag{3.18}$$

The relation $h = 2\pi T_s/\omega$, where T_s is the time between samples and $\omega \in [0, \pi]$, can be used to convert frequency ω to its corresponding time horizon, h.

Column A of Figure 3-1 plots the dynamic alpha for the case of no serial correlation (λ=0). The dynamic alpha is positive at high frequencies, indicating that the weights and returns are in phase over these short time horizons. However, this added value is cancelled out since the weights and returns are out of phase at longer time horizons, resulting in zero net alpha.

Columns B and C of Figure 3-1 show the dynamic alpha for the cases of momentum ($\lambda > 0$) and mean reversion ($\lambda < 0$) in the first lag of returns, respectively. For the mean reversion case, we notice that both the lowest and highest frequencies are more profitable relative to the serially uncorrelated case. This is an intuitive result since both weights and returns now have more variability in these higher frequency fluctuations. These high-frequency components will be in phase, leading to a large positive contribution and an overall positive alpha. The momentum case is opposite in effect. Relative to the serially uncorrelated case, both the lowest and highest frequencies are less profitable, and the net contribution over all frequency components is negative.

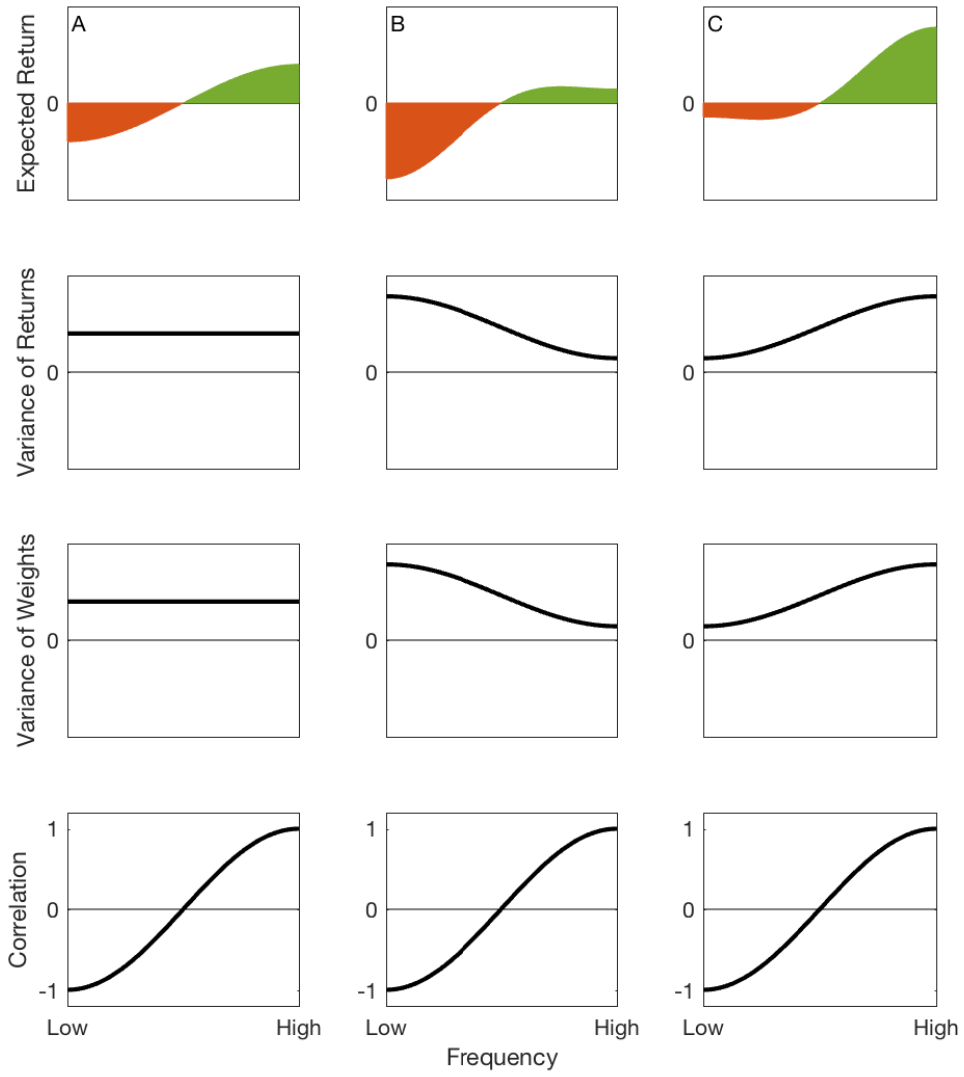

Figure 3-1: Dynamic alpha of the contrarian trading strategy applied to the serially uncorrelated (Column A), momentum (Column B), and mean reversion (Column C) implementations of (3.16).

3.6 An Empirical Example

To develop a better understanding of the characteristics of dynamic alpha, we apply our framework to Lo and MacKinlay's (1990) contrarian (mean reversion) trading strategy using historical stock market data. The fact that the weights given by (3.14) sum to zero at each date t implies very little market-beta exposure. Also, since the weights are so dynamic, much of this portfolio's return should be due to active management near frequencies inversely related to the decision period. The return for a given interval can be calculated as the profit-and-loss of the strategy's positions over that interval, divided by the capital required to support those positions. In the following analysis, we assume that Regulation T applies; therefore, the amount of capital required is one-half the total capital invested (often stated as a 2:1 leverage, or a 50% margin requirement). The unleveraged portfolio return, $r_{p,t}$ is given by:

$$r_{p,t} = \frac{\sum_{i=1}^{N} w_{i,t} r_{i,t}}{I_t} \quad , \quad I_t = \frac{1}{2} \sum_{i=1}^{N} |w_{i,t}| \ .$$

We apply (3.14) to the one-day and two-day returns of the five smallest size-decile portfolios of all NASDAQ stocks, as constructed by the University of Chicago's Center for Research in Security Prices (CRSP), from January 2, 1990 to December 29, 1995. We selected this time period purposely because of the emergence of day trading in the early 1990s, an important source of profitability for statistical arbitrage strategies. Of course, trading NASDAQ size deciles is obviously unrealistic in practice, but our purpose is to illustrate the empirical relevance of our framework, not to derive an implementable trading strategy.

Figure 3-2 illustrates the performance of the contrarian strategy for one-day and two-day mean reversion over the 1990–1995 sample period, and Table 3.4 contains summary statistics for the daily returns of the two trading strategies. For 1-day mean reversion, with an annualized average return of 31.6% and standard deviation of 7.9%, the strategy's performance is considerably better than that of a passive buy-and-hold strategy, which is one indication that active management is playing a significant role

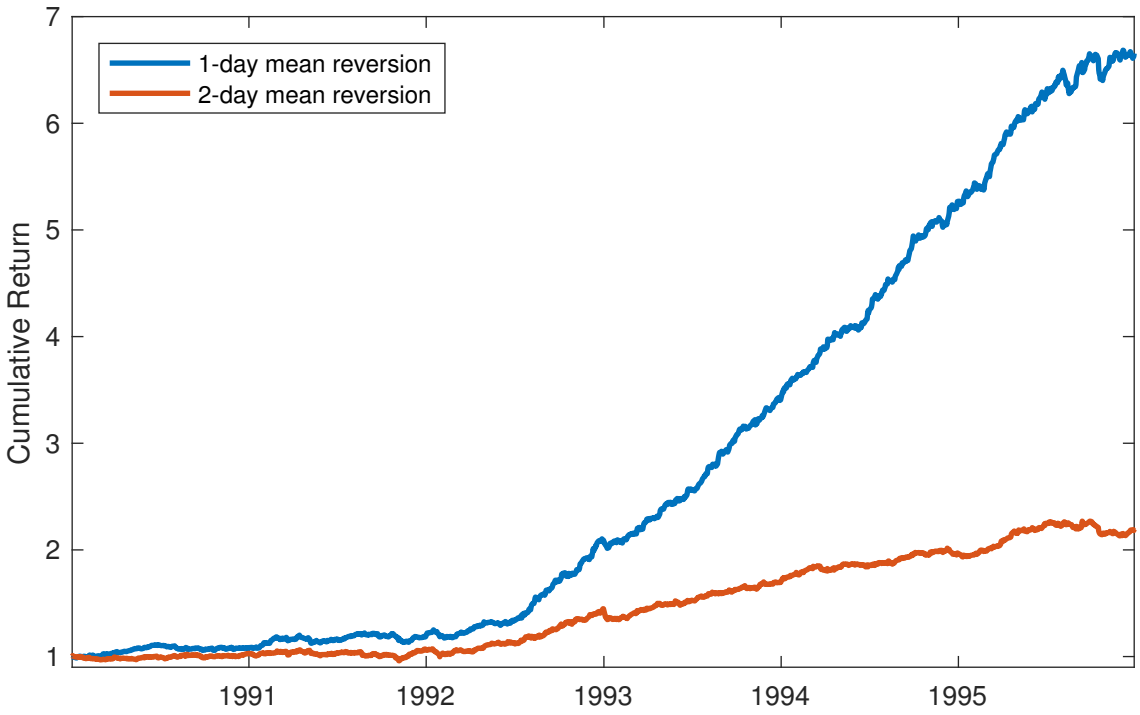

Figure 3-2: Cumulative return of a mean reversion strategy of Lo and MacKinlay (1990) over one-day and two-day returns applied to the five smallest CRSP-NASDAQ size deciles from January 2, 1990 to December 29, 1995.

in this case.

This intuition is confirmed by the decomposition of the strategy's expected return into its dynamic alpha components in Table 3.5. On an annualized basis, the dynamic component yields 32.2%, which exceeds the strategy's total expected return of 31.6%, implying a slightly negative static component. In this case, more than all of the strategy's expected return is coming from active management over a daily time horizon, and the low-frequency components are subtracting value.

The explanation for this rather unusual phenomenon was provided by Lo and MacKinlay (1990), who observed that because the contrarian strategy is, on average, long losers and short winners, it will typically be long the low-mean assets and short the high-mean assets. Therefore, the static component, i.e., the sum of average portfolio weights multiplied by average returns, will consist of positive average weights for low-mean stocks and negative average weights for high-mean stocks for this strategy—a losing proposition in the absence of mean reversion. Fortunately,

Statistic	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	1-day	2-day
Mean $\times 250$	27.5%	17.4%	13.9%	13.7%	12.8%	31.6%	13.3%
SD $\times\sqrt{250}$	12.2%	9.8%	8.9%	9.1%	9.5%	7.9%	7.8%
SR $\times\sqrt{250}$	2.25	1.77	1.56	1.51	1.35	3.98	1.69
Min	−2.9%	−2.7%	−2.7%	−3.3%	−3.5%	−2.2%	−5.2%
Median	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.0%
Max	6.7%	3.6%	2.0%	2.1%	2.3%	2.4%	1.7%
Skew	0.6	0.1	−0.5	−0.7	−0.9	−0.1	−0.8
XSKurt	5.1	2.4	2.0	3.1	3.9	1.7	8.9

Table 3.4: Summary statistics of the daily returns of the one-day and two-day mean reversion strategies of Lo and MacKinlay (1990) applied to the daily returns of the five smallest CRSP-NASDAQ size deciles, from January 2, 1990 to December 29, 1995. The Sharpe ratio (SR) is calculated relative to a 0% risk-free rate.

the positive correlation between weights and returns at high frequencies is more than sufficient to compensate for this long-term negative component.

To mitigate the loss caused by the static component, we can filter out the trend component of each size-decile portfolio before calculating the mean-reversion weights. Intuitively, the mean-reversion trading strategy will no longer place a negative bias on the weights of the smallest deciles simply because they achieve relatively large average returns. Similarly, if we perfectly filter out the low-frequency dynamics of the portfolio returns, then we can extract the profitability in the high-frequency component of returns, while not suffering the substantial losses of the low-frequency component. In other words, the mean-reversion trading strategy will be trading on the relevant high-frequency signal, and not the low-frequency "noise." Since a perfect high-pass filter cannot be implemented in practice, these low-frequency components would have to be forecasted. Therefore, rather counterintuitively, our spectral framework reveals that forecast power at low frequencies can be used to improve the overall performance of a high-frequency trading strategy.

For mean reversion over two days, with an annualized average return of 13.3% and a standard deviation of 7.8%, the strategy's performance is considerably worse than that of the one-day mean reversion strategy. Active management is playing a significant but less productive role. Here, the positive correlation between weights and returns at medium frequencies remains sufficient to compensate for the negative

Statistic	1-day	2-day
Portfolio Mean ×250	31.6%	13.3%
Static Component ×250	−0.6%	−1.0%
Dynamic Component ×250	32.2%	14.2%
Low Frequency ($h \geq$ 5d)	−44.7%	−19.1%
Med Frequency (3d $\leq h <$ 5d)	6.3%	33.7%
High Frequency ($h <$ 3d)	70.6%	−0.4%

Table 3.5: Estimates of the dynamic alpha of the daily returns of the one-day and two-day mean reversion strategies of Lo and MacKinlay (1990) applied to the five smallest CRSP-NASDAQ size-decile returns, from January 2, 1990 to December 29, 1995. Frequency components are grouped into three categories: high frequencies (more than one cycle per three days), medium frequencies (between one cycle per three days and one cycle per week), and low frequencies (less than one cycle per week).

correlation between weights and returns at the low and high frequencies.

The correlation of these two strategies' returns is only 0.26. This low correlation can be attributed to the fact that their performance is determined by market dynamics occurring in distinct and non-overlapping frequency bands. Moreover, these frequency-specific strategies can be implemented simultaneously, and can therefore be viewed as separate assets. These assets can then be combined in a portfolio to achieve diversification across multiple frequencies. In our sample period, these diversification benefits result in the Sharpe ratio being maximized when 84.6% of our capital is used to implement the one-day mean-reversion trading strategy, and the remaining capital is used to trade with mean reversion over two days. However, Table 3.5 makes it clear that both assets in our portfolio would be negatively affected by a low-frequency market shock.

3.7 Warren Buffett's Alpha

For a more realistic application of our dynamic alpha framework, we examine the returns of Warren Buffett's multinational conglomerate holding company, Berkshire Hathaway Inc., which is known for its long-term investments in public and private companies. We obtain quarterly holdings data for Berkshire Hathaway from Thomson

Reuters Institutional (13F) Holdings database (based on Berkshire's SEC filings) from 1980 to 2013, and stock return data from the CRSP Monthly Stock database.

One consequence of Warren Buffett's longer decision interval is that we are less likely to be affected by aliasing when applying our decomposition to quarterly weights and returns of his portfolio; the same cannot be said for higher-frequency trading strategies. Figure 3-3 displays the cumulative returns for Berkshire Hathaway (BRK) and a simulated reconstruction (R[BRK]) of these returns using the holdings data based on SEC filings. The correlation between these return series is 0.7, and their Sharpe ratios are 0.69 and 0.66, respectively. The high correlation and similar Sharpe ratios indicate that the reconstructed returns capture a significant fraction of Berkshire Hathaway's price dynamics. Equating the mean of the reconstructed returns with the realized returns, we use a leverage ratio of 1.41 to reconstruct Warren Buffett's levered returns (RL[BRK]). This is similar to the average leverage ratio of 1.4 estimated by Frazzini et al. (2013) using total assets to equity.

Table 3.6 contains summary statistics for the monthly returns of each time series. With an average annualized return of 22.9% over more than 30 years, Berkshire clearly has positive alpha when compared to traditional risk factors. Frazzini et al. (2013) find that Buffett's returns are due more to security selection than his effect on management, which suggests that a large component of his returns must be static alpha, i.e., high average weights on securities with large α_i's. In other words, Buffett is able to select securities that provide high average returns above and beyond the expected return resulting from passive exposures to factor risk. Moreover, if Warren Buffett has a positive long-term effect on returns due to his managerial and advisory competence, then we would also expect to find a substantial component of his returns derived from lower frequencies. Finally, Buffett is a practitioner of value investing, and so we should not expect to find a significant correlation between his portfolio weights and returns at high frequencies.

The decomposition of Berkshire Hathaway's reconstructed average portfolio return into its dynamic alpha components in Table 3.7 confirms this intuition. The static component yields an annualized return of 18.9%. In comparison, the value-weighted

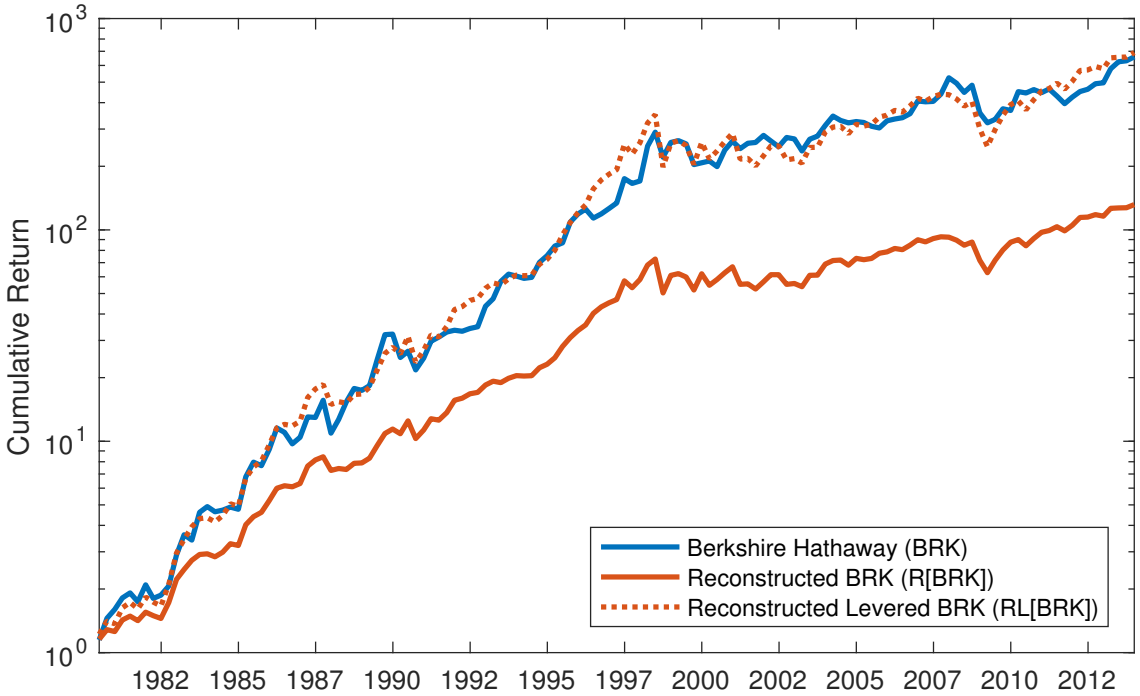

Figure 3-3: Cumulative realized returns of Berkshire Hathaway (BRK) and a simulated reconstruction (R[BRK]) using holdings data for Berkshire Hathaway from Thomson Reuters Institutional (13F) Holdings database (based on Berkshire's SEC filings) from 1980 to 2013. Equating the mean of the reconstructed returns with the realized returns, we use a leverage ratio of 1.41 to reconstruct the levered returns (RL[BRK]).

Statistic	Risk-Free	Market	BRK	R[BRK]	RL[BRK]
Mean $\times 4$	4.7%	12.8%	22.9%	16.3%	22.9%
SD $\times\sqrt{4}$	1.7%	17.4%	26.2%	17.5%	24.7%
SR $\times\sqrt{4}$	0	0.47	0.69	0.66	0.74
Min	0.0%	−23.7%	−30.1%	−30.9%	−43.6%
Median	1.2%	3.9%	4.4%	4.2%	6.0%
Max	3.8%	21.3%	46.1%	28.8%	40.7%
Skew	0.6	−0.6	0.3	−0.5	−0.5
XSKurt	0.3	0.5	0.9	1.8	1.8

Table 3.6: Summary statistics of the quarterly returns of the one-month Treasury Bill (Risk-Free) rate, the value-weighted CRSP market index (Market), Berkshire Hathaway (BRK), and a simulated reconstruction (R[BRK]) using holdings data for Berkshire Hathaway from Thomson Financial Institutional (13F) Holdings Database (based on Berkshire's SEC filings) from 1980 to 2013. Equating the means of the reconstructed returns with the realized returns we use a leverage ratio of 1.41 to reconstruct the levered returns (RL[BRK]).

CRSP market index yielded an average annualized return of 12.8% over the same interval, and the annualized risk-free interest rate (one-month Treasury Bill rate) was 4.7%. The static component of the portfolio's realized market beta over this interval using quarterly returns was 0.84, which implies a risk premium component of 6.8% and a static alpha component of 7.3%. This demonstrates that a substantial component of Berkshire Hathaway's returns results from Buffett's ability to select securities with favorable long-term prospects. The dynamic alpha component contributes an additional annualized return of 4.1% to the portfolio, most of which can be attributed to dynamics occurring at time horizons greater than 5 years. The annualized dynamic risk is 10.3%, which yields a dynamic information ratio of 0.40. This result can be attributed to Buffett's ability as a manager to improve firm performance over the long run while Berkshire maintains a position in the company, and also to his ability to time transactions based on fundamental valuations.

In contrast, the dynamics at the shortest time horizons—less than 18 months—subtract 1.2% annually from the average portfolio return. Here, the negative correlation between weights and returns can be attributed in part to transaction costs and market impact. However, the quarterly sampling frequency of the holdings data restricts our ability to study these higher frequency dynamics. By observing only quarter-end weights and cumulative returns, we have no way of inferring the profitability of dynamics occurring at these higher frequencies.

A spectral decomposition of Berkshire Hathaway's returns demonstrates conclusively that Buffett is not only a consummate long-term investor, but that the horizon of his timing ability stretches far beyond the reaches of most other portfolio managers.

Statistic	RL[BRK]
Portfolio Mean ×4	22.9%
Static Component ×4	18.9%
Risk-Free Rate	4.7%
Risk Premium	6.8%
Static Alpha	7.3%
Dynamic Component ×4	4.1%
Low Frequency ($h \geq$ 5y)	4.3%
Med Frequency (1.5y $\leq h <$ 5y)	1.1%
High Frequency ($h <$ 1.5y)	−1.2%

Table 3.7: Estimates of the static and dynamic alpha of the simulated quarterly returns of Berkshire Hathaway using holdings data for Berkshire Hathaway from Thomson Financial Institutional (13F) Holdings Database (based on Berkshire's SEC filings) from 1980 to 2013. Frequency components are grouped into three categories: high frequencies (more than one cycle per 1.5 years), medium frequencies (between one cycle per 1.5 years and one cycle per five years), and low frequencies (less than one cycle per five years). Note that table entries may not sum due to rounding.

3.8 Conclusion

In this chapter, we have applied spectral analysis to develop a dynamic measure of alpha that allows us to determine whether portfolio managers are generating alpha and over what time horizons their investment processes have forecast power. In this context, an investment process is said to be profitable at a given frequency if there is positive correlation between portfolio weights and returns at that frequency. When aggregated across frequencies, dynamic alpha is equivalent to Lo's (2008) active component, and provides a clear indication of a manager's forecast power and, consequently, active investment skill. By separating the dynamic and static components of a portfolio, it should be possible to study and improve the performance of both.

Frequency-domain representations of auto- and cross-covariances can be applied to many other financial statistics in addition to alpha. For example, dynamic versions of performance attribution, linear factor models, asset allocation models, risk management, and measures of systemic risk can all be constructed using spectral analysis.

Our framework can also be extended to other time-frequency decompositions, including the wavelet transform, to address the impact of time-varying relationships and other non-stationarities.

Chapter 4

Spectral Beta

Economic shocks can have diverse effects on financial market dynamics at different time horizons, yet traditional measures of beta do not distinguish between short- and long-term components of systematic risk. In this chapter, we apply spectral analysis to represent betas as a linear combination of frequency-specific betas, i.e., betas on components of the asset returns and components of the factor operating at different frequencies. Our spectral betas computed using the Haar Wavelet Transform can be represented as regressions on moving averages in the time domain, thereby facilitating interpretation and applicability. When applied to NYSE and AMEX stock returns from 1972 to 2016, we find that the inclusion of the frequency domain dimension can be used to select portfolios with significantly lower out-of-sample variance relative to estimators based on traditional multi-factor models.

4.1 Introduction

Spectral and co-spectral power, often calculated using either the Fourier or wavelet transform, provide a natural way to study the horizon-specific components of variance and covariance, two important measures of risk in the financial domain. Specifically, spectral power decomposes the variability of a time series resulting from fluctuations at a specific frequency, while co-spectral power decomposes the covariance between two real-valued time series, and measures the tendency for them to move together

over specific time horizons. When the signals are in phase at a given frequency (i.e., their peaks and valleys coincide), the co-spectral power is positive at that frequency, and when they are out of phase, it is negative. By identifying the particular frequencies that are responsible for a given portfolio's volatility, an investor adds another dimension with which to manage his or her risk/reward profile.

Taking the ratio of co-spectral to spectral power, we can represent the beta of an asset as a linear combination of spectral betas defined on components of asset returns and the factor operating over different frequencies. This representation, first introduced to the economic literature by Engle (1974), is such that covariance terms across frequencies do not appear, as if the components of asset returns and the factor were uncorrelated across frequencies. This is particularly helpful should one wish to investigate the contribution of individual frequencies while ignoring the impact of others. We modify Engle's (1974) band spectrum regression framework by applying the Haar Wavelet Transform (HWT), which allows us to represent the frequency components of a time series in terms of moving average filters with different window lengths. This representation of aggregation in the time domain makes the framework easily operational, while avoiding interpretational issues.

We assess the economic significance of our spectral betas by estimating the covariance matrix of stock returns using frequency-specific factor models in the context of portfolio optimization. For NYSE and AMEX stocks from 1972 to 2016, we find that modeling the frequency domain yields portfolios with lower out-of-sample variance when compared to traditional multi-factor models. Moreover, we find that both small (large) capitalization and high (low) book-to-market stocks have larger (smaller) market betas at low frequencies relative to high frequencies. This has important implications for both portfolio selection and tests of asset pricing models such as the CAPM.

In Section 4.2, we review band-spectrum regression and provide a formulation of spectral betas based on the HWT. An empirical application to portfolio optimization is contained in Section 4.3, and Section 4.4 concludes.

4.2 Spectral Beta

Linear factor models are often used in financial applications, including market model regressions, the CAPM, the APT, and the Fama-French three-factor model. The estimated beta coefficients in these models are static measures that are incapable of capturing dynamic relationships among the variables. On the other hand, band spectrum regression, proposed by Engle (1974), captures the sensitivity of the dependent variable to the fluctuations in the independent variables over different time horizons. The technique uses the Discrete Fourier Transform (DFT) to express windowed subsamples of time series in the frequency domain, and then analyzes their magnitude and phase. When the time series are in phase at a given frequency, the contribution that frequency makes to the frequency-specific beta is positive; when they are out of phase, that particular frequency's contribution will be negative.

Specifically, consider a real-valued subsample of the time series x_t and y_t from $t = 0, \ldots, T-1$. The sample covariance over this interval can be calculated as:

$$\mathrm{Cov}\langle x_t, y_t \rangle = \frac{1}{T} \sum_{t=0}^{T-1} (x_t - \overline{x})(y_t - \overline{y}) \quad , \tag{4.1}$$

where $\overline{x}$ and $\overline{y}$ are the sample means of x_t and y_t over the same subperiod. This calculation is exactly equivalent to the one formed using the T-point DFT:

$$\mathrm{Cov}\langle x_t, y_t \rangle = \frac{1}{T} \sum_{k=1}^{T-1} \hat{L}_{xy}[k] \quad , \quad \hat{L}_{xy}[k] \equiv \frac{1}{T} \Re[X_k^* Y_k] \tag{4.2}$$

where X_k and Y_k are the T-point DFT coefficients of the subsample of x_t and y_t. Thus, the sum over $\hat{L}_{xy}[k]$ is proportional to the sample covariance of x_t and y_t. Moreover, the sum of $\hat{L}_{xy}[k]$ over a band of frequencies, $\mathrm{Cov}_K\langle x_t, y_t \rangle$ where $K \subseteq \{1, \ldots, T-1\}$, is proportional to that band's contribution to the sample covariance. For this reason the function $\hat{L}_{xy}[k]$, called the cross-periodogram, is an estimate of the co-spectrum at the harmonic frequency ω_k, and can be interpreted as the frequency distribution of the power contained in the sample covariance.

For a frequency band $K \subseteq \{1, \ldots, T-1\}$, the spectral beta coefficients of an M-factor model are given by,

$$\beta_K \langle y_t; x_{1,t}, \ldots, x_{M,t} \rangle = \left(\sum_{k \in K} \hat{\mathbf{L}}_{\mathbf{xx}} \right)^{-1} \left(\sum_{k \in K} \hat{\mathbf{L}}_{\mathbf{xy}} \right) \quad , \tag{4.3}$$

where,

$$\sum_{k \in K} \hat{\mathbf{L}}_{\mathbf{xx}} \equiv \begin{pmatrix} \sum_{k \in K} \hat{L}_{x_1,x_1}[k] & \cdots & \sum_{k \in K} \hat{L}_{x_1,x_M}[k] \\ \vdots & \ddots & \vdots \\ \sum_{k \in K} \hat{L}_{x_M,x_1}[k] & \cdots & \sum_{k \in K} \hat{L}_{x_M,x_M}[k] \end{pmatrix}, \sum_{k \in K} \hat{\mathbf{L}}_{\mathbf{xy}} \equiv \begin{pmatrix} \sum_{k \in K} \hat{L}_{x_1,y}[k] \\ \vdots \\ \sum_{k \in K} \hat{L}_{x_M,y}[k] \end{pmatrix} \quad . \tag{4.4}$$

When only one factor is present, (4.3) reduces to the familiar expression,

$$\beta_K \langle y_t; x_t \rangle = \frac{\mathrm{Cov}_K \langle x_t, y_t \rangle}{\mathrm{Var}_K \langle x_t \rangle} \quad . \tag{4.5}$$

Intuitively, these calculations are computationally equivalent to estimating the beta coefficients by regressing the inverse DFT reconstruction of the time series, restricted to the frequencies specified by K. Standard errors and the F statistic for this band-spectrum regression are provided in Appendix B.

As an illustrative example, Table 4.1 tests the hypothesis that long- and short-term components of several hedge-fund style index returns are equally sensitive to market index returns across all frequencies. Specifically, we analyzed the monthly returns of the HFRI ED: Distressed/Restructuring, HFRI FOF: Market Defensive, and HFRI EH: Quantitative Directional indices relative to the monthly returns of the CRSP value-weighted market index using the spectral beta measures described above. The short-term component was assumed to include frequencies higher than 1 cycle per year, and the original series were de-meaned.

The F statistic value of 43.116 ($p < 0.001$) for the Distressed/Restructuring index rejects the hypothesis that the sensitivity of this strategy's returns to market move-

	$\hat{\beta}$ (SE)	$\hat{\beta}_{\text{LF}}$ (SE)	$\hat{\beta}_{\text{HF}}$ (SE)	$F_{1,309}$
Distressed/Restructuring	0.250	0.500	0.190	43.116***
	(0.020)	(0.067)	(0.018)	
Market Defensive	0.033	0.043	0.030	0.056
	(0.022)	(0.056)	(0.023)	
Quantitative Directional	0.704	0.709	0.703	0.009
	(0.024)	(0.064)	(0.026)	

*** $p < 0.001$

Table 4.1: All-, low-, and high-frequency beta estimates of hedge fund index monthly returns from 1990 to 2015 with the CRSP value-weighted market index returns. Frequencies are grouped into two categories: low frequencies (less than or equal to 1 cycle per year), and high frequencies (more than 1 cycle per year). F statistics are formed to compare the restricted (non-spectral) and unrestricted (spectral) regression models.

ments over this period did not differ between short- and long-term components. On the other hand, the F statistics for the Market Defensive and Quantitative Directional indices are 0.056 and 0.009, respectively. Here, the null hypothesis that the sensitivity of these indices to market movements is the same across long and short horizons cannot be rejected at the standard significance levels.

Figure 4-1 plots the cumulative returns of these hedge fund indices alongside the cumulative return of the market index. The figure illustrates that the Market Defensive index had low sensitivity to fluctuations in market returns across all frequencies ($\hat{\beta}_{LF} = 0.043$; $\hat{\beta}_{HF} = 0.030$), while the Quantitative Directional index responded strongly to all market return fluctuations ($\hat{\beta}_{LF} = 0.709$; $\hat{\beta}_{HF} = 0.703$). This observation corresponds to the relatively constant beta values across low to high frequencies listed for these strategies in Table 4.1. Conversely, the Distressed/Restructuring index appears more sensitive to long-term fluctuations, leading to significantly larger beta values at low frequencies relative to high frequencies ($\hat{\beta}_{LF} = 0.500$; $\hat{\beta}_{HF} = 0.190$). This difference may reflect the illiquidity of assets contained within the Distressed/Restructuring index, and demonstrates that a non-spectral measure of beta might underestimate the index's market risk exposure.

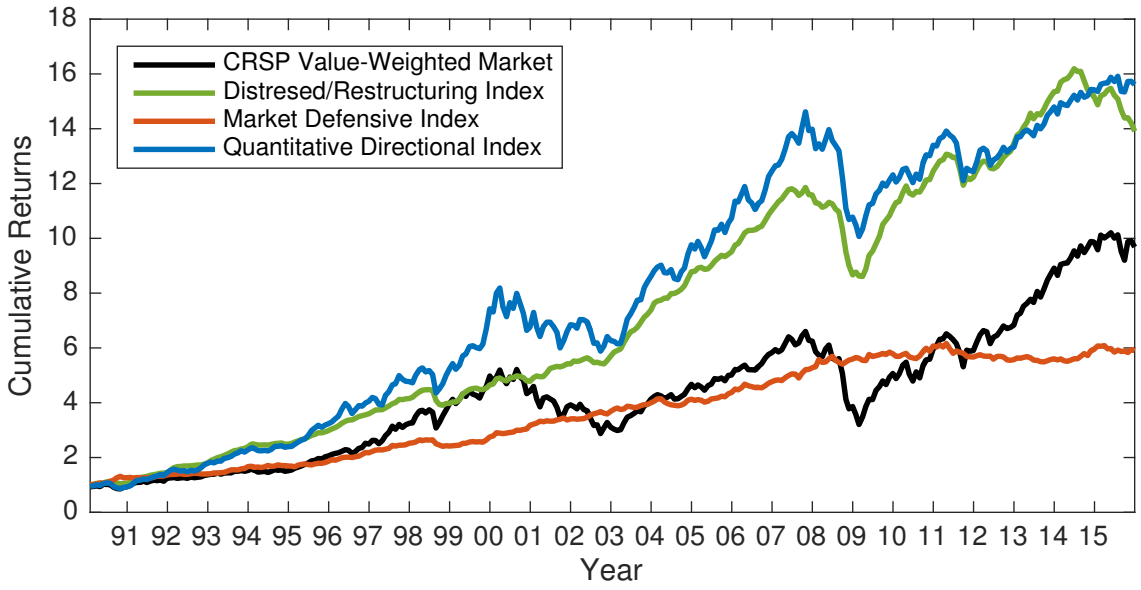

Figure 4-1: Cumulative returns of hedge fund indices alongside the CRSP value-weighted market index from 1990 to 2015.

4.3 Application to Portfolio Selection

In this section, we present empirical evidence that the inclusion of the frequency domain dimension can be used to select portfolios of stocks with lower out-of-sample variance relative to estimators based on traditional multi-factor models.

4.3.1 Portfolio selection

Markowitz's (1952) portfolio theory states that given a target value, $\tilde{\mu}$, for the expected portfolio return, the efficient portfolio weights, $\tilde{\mathbf{w}}$, are those that minimize the portfolio variance for all portfolios with expected return $\tilde{\mu}$. Mathematically, the optimization problem can be written as,

$$\tilde{\mathbf{w}} \quad = \quad \arg\min_{\mathbf{w}} \mathbf{w}^{\mathrm{T}} \boldsymbol{\Sigma} \mathbf{w} \tag{4.6}$$

subject to the constraints

$$\mathbf{w}^{\mathrm{T}} \boldsymbol{\mu} = \tilde{\mu} \quad \text{and} \quad \mathbf{w}^{\mathrm{T}} \mathbf{1} = 1 \tag{4.7}$$

where w_i is the portfolio weight on the ith security, $\mu_i = \mathrm{E}[r_i]$ and $\Sigma_{i,j} = \mathrm{Cov}(r_i, r_j)$.

An important input in this optimization problem is the covariance matrix of these securities. In practice, portfolio weights are optimized using a covariance matrix that has been estimated from historical data, and are held until the next rebalancing. The out-of-sample performance of this estimator is gauged by the variance of this optimal portfolio in the period after the portfolio has been formed. If the estimator overfits the historical data, then the out-of-sample performance can be poor, which is why imposing structure through a factor model can be beneficial.

4.3.2 Spectral Factor Model

Factor models of asset returns decompose the return on a cross-section of assets into factor-related and asset-specific components.

Let $R_{i,t}$ denote the return of asset i in a universe of N stocks. Assuming M factors, $\beta_i = (\beta_{i,1}, \ldots, \beta_{i,M})$, denotes asset i's sensitivity to the factors $f_t = (f_{1,t}, \ldots, f_{M,t})$, respectively. A factor decomposition of asset i's returns has the form,

$$R_{i,t} = \alpha_i + \beta_i f_t^{\mathsf{T}} + \varepsilon_{i,t} \quad . \tag{4.8}$$

It is commonly assumed that the asset-specific returns $\varepsilon_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{N,t})$ are cross-sectionally uncorrelated such that $\mathrm{E}[\varepsilon_t^{\mathsf{T}} \varepsilon_t] = D$, where D is a diagonal matrix. Letting B be an $N \times M$-matrix of factor betas and V be the $M \times M$ covariance matrix of the factors, the covariance matrix of returns, Σ, can expressed as

$$\Sigma = BVB^{\mathsf{T}} + D \quad . \tag{4.9}$$

Since the frequency-specific filtered components are orthogonal to one another, spectral betas can be used to decompose (4.9) into its frequency components with no

cross-frequency terms,

$$\widehat{\Sigma} = \sum_{j=1}^{J} \Sigma^{(j)} \quad , \tag{4.10}$$

$$\widehat{\Sigma}^{(j)} = B^{(j)} V B^{(j)\mathsf{T}} + D^{(j)} \quad . \tag{4.11}$$

This spectral estimate of the covariance matrix of returns models the fact that correlations of asset returns may differ across time horizons.

We apply our spectral framework to the single-index market model (CAPM) of Sharpe (1964), the Fama-French three-factor model, and a five-factor model where the factors are calculated using principal components analysis (PCA), and compare the out-of-sample performance of the non-spectral vs. spectral estimators.

4.3.3 Dataset

Our dataset and test criteria are similar to Ledoit and Wolf (2003). Stock return data are extracted from the University of Chicago's Center for Research in Securities Prices (CRSP) monthly database. Only U.S. common stocks traded on the New York Stock Exchange (NYSE) and the American Stock Exchange (AMEX) are included, which eliminates REIT's, ADR's, and other types of securities. Market returns are calculated as the equal-weighted portfolio return of this universe of stocks, and the size and book-to-market factors are extracted from Kenneth French's web site at Dartmouth[1]. For $t = 1972$ to $t = 2015$, we use an in-sample period from August of year $t - 10$ to July of year t to form an estimate of the covariance matrix of stock returns. (We rebalance on the first trading day in August because AMEX stock return data becomes available in August, 1962.) Future expected returns are estimated as the average realized return during the in-sample period. Using these expected return and covariance matrix estimates, we form the global minimum variance portfolio and the minimum variance portfolio that targets a 20% expected return. In both cases, short sales are allowed, and the universe of stocks ranges between $N = 946$ and

[1] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/index.html

$N = 1302$.

We hold these portfolios for one year, at which time they are liquidated and new optimal portfolios are formed for the next year. Our measure of performance for these competing estimators is the out-of-sample standard deviation of the estimated optimal portfolios in the period from August, 1972 to July, 2016. The goal of this empirical test is to quantify the reduction in the out-of-sample standard deviation that results from using our spectral betas under fairly practical constraints.

4.3.4 Performance

For each factor model, we compute the standard deviation of the minimum variance portfolios formed using non-spectral and spectral-based estimators of the covariance matrix. The HWT-based models are formed with 6 levels, and the DFT-based models are implemented using the corresponding frequency bands. The results including one-sided variance ratio tests for the hypothesis that the non-spectral and spectral models have equal variances are reported in table 4.2.

Model	Standard deviation (global minimum)	Variance ratio	Standard deviation ($\mathrm{E}[R] = 20\%$)	Variance ratio
CAPM	11.25 (0.35)	–	12.66 (0.39)	–
DFT-CAPM	10.77 (0.33)	1.09	11.89 (0.37)	1.13*
HWT-CAPM	10.82 (0.33)	1.08	11.92 (0.37)	1.13*
Fama-French	10.24 (0.32)	–	10.95 (0.34)	–
DFT-Fama-French	9.60 (0.30)	1.14*	10.33 (0.32)	1.12*
HWT-Fama-French	9.76 (0.30)	1.10	10.44 (0.32)	1.10
PCA	9.99 (0.31)	–	10.73 (0.33)	–
DFT-PCA	9.44 (0.29)	1.12*	10.13 (0.31)	1.12*
HWT-PCA	9.60 (0.30)	1.08	10.24 (0.32)	1.10

* $p < 0.1$

Table 4.2: Risk of minimum variance portfolios from $t = 1972$ to $t = 2015$. Standard deviation, expressed in percent, is annualized through multiplication by $\sqrt{12}$ and standard errors on these standard deviation estimates are reported in parenthesis. Ratios of the non-spectral model variance to spectral model variance are reported with their significance levels based on a one-sided F-test for equal variances.

We find that for each factor model, including the frequency domain consistently reduces the out-of-sample standard deviation by about 60 basis points, such that the

variance of the overall portfolio is lowered by about 10%. Moreover, the DFT-spectral models performs slightly better than the HWT-spectral models, but in general, their improvement over the non-spectral models is approximately equal.

To determine the economic intuition behind this improvement, we report the average frequency-specific market betas for portfolios that are sorted by size from $t = 1972$ to 2015 and book-to-market from $t = 1980$ to 2015[2] in Figure 4-2. We also report the alpha term of Sharpe's (1964) single index market model using each frequency-specific market beta as an estimate of the overall market beta.

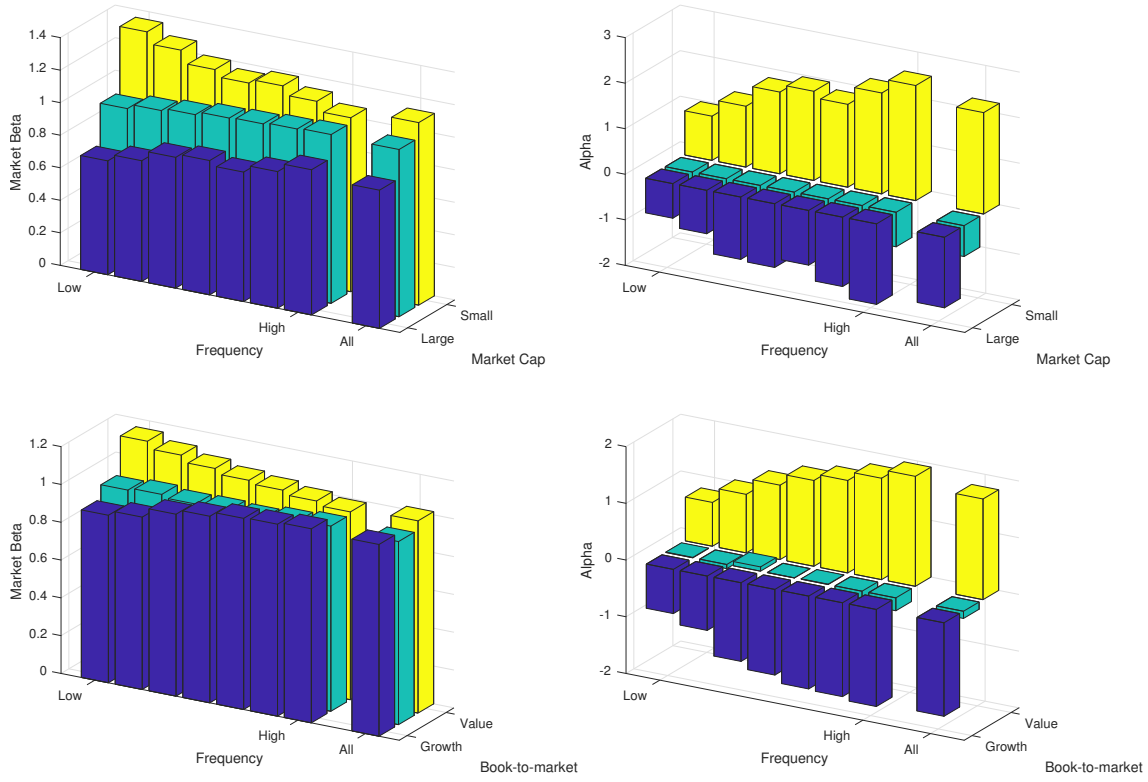

Figure 4-2: HWT-CAPM betas and corresponding alphas. Average frequency-specific market betas for portfolios that are sorted by size from $t = 1972$ to 2015 and book-to-market from $t = 1980$ to 2015. The alpha term of Sharpe's (1964) single index market model using the frequency-specific market beta as an estimate of the overall market beta is also reported.

We see that frequency-specific betas vary not only across the size and value dimensions, but also along the frequency dimension. Interestingly, small and value stocks

[2]Book value is extracted from Compustat and only becomes available to match our CRSP data around 1980. The universe of stocks for this book-to-market analysis varies from $N = 703$ to $N = 885$.

are more sensitive to market fluctuations at low frequencies, while large and growth stocks are the opposite. Economically, this suggests that small stocks and value stocks are more strongly affected by low-frequency business cycle dynamics. For example, these companies might find it more difficult to finance their operations in recessions compared to more established, non-distressed companies. Our frequency-specific beta model is able to capture these dynamics, which helps to explain the spectral models' improved out-of-sample performance.

As a result, the alpha terms in Figure 4-2 decrease as we extend the horizon. This has important implications for asset pricing models, as it suggests the single-index model is satisfied more convincingly at low frequencies. In general, we find that aggregation through the HWT provides a systematic framework to capture short- and long-term dynamics in asset returns as well as the factors. We view notions of frequency-specific risk aversion or heterogeneity in investors' investment/consumption horizons, among other channels, as being promising directions for future empirical and theoretical investigations. We hope the findings in this chapter will encourage such investigations.

4.4 Conclusion

We show that the beta of an asset can be expressed as a linear combination of frequency-specific betas, i.e., betas on components of the asset returns and components of the factor operating at different frequencies. When computed using the HWT, this beta representation can be expressed in the time domain using moving average filters, thereby facilitating interpretation and applicability. It does not involve cross-covariances between frequency-specific components of asset returns and the factor, thus permitting all frequency-specific information to be contained exclusively in frequency-specific betas.

When applied to NYSE and AMEX stock returns from 1972 to 2016, we find that the inclusion of the frequency domain can be used to estimate spectral factor models that select portfolios with significantly lower out-of-sample variance relative to esti-

mators based on traditional multi-factor models. These frequency-specific measures allow us to distinguish between short- and long-term components risk and covariances, providing additional insights into portfolio and risk management above and beyond their static counterparts. These considerations can be particularly useful when asset-dynamics differ across time horizons.

Frequency-dependent alphas, betas, variances and auto- and cross-covariances can be used to incorporate dynamics into many other financial applications. For example, spectral versions of performance attribution, linear factor models, asset-allocation models, risk management, and measures of systemic risk can all be constructed using spectral analysis.

In particular, our representation can be used to test for the impact of systematic risk at different frequencies on the pricing of assets. Our proposed framework allows for the possibility of heterogeneous pricing over different frequencies or horizons. It therefore offers a useful framework to remove the restriction of equal pricing across frequencies that is implied in the asset pricing literature.

Part III

Applications to Healthcare Finance

Chapter 5

Patient-Centered Clinical Trials

In this chapter, we review how Bayesian decision analysis can be used to incorporate patient preferences in the regulatory approval process for new therapies. By assigning weights to type I and type II errors based on patient preferences, the significance level (α) and power ($1 - \beta$) of a randomized clinical trial (RCT) for a new therapy can be optimized to maximize the value to current and future patients and consequently, to public health. As an example, we apply this technique to weight-loss devices and find that potentially effective, low-risk treatments have optimal α's larger than the traditional one-sided significance level of 5%, while potentially less effective and riskier treatments have optimal α's below 5%. Moreover, the optimal RCT design, including trial size, varies with the risk aversion, time-to-access preferences, and medical need of the target population.

5.1 Introduction

Determining the acceptable level of uncertainty associated with clinical evidence has been an important and challenging decision when regulators conduct benefit-risk assessments of novel technologies, especially for unmet medical needs. Traditional clinical trial designs typically set the one-sided significance level, i.e., the maximum allowed value for the rate of type I error (approving a device for which there is not a reasonable assurance of safety and effectiveness), at 5% regardless of the context in

which the decision is made and the public health implications of the consequences. However, the context may matter for making rational and sensible decisions with significant public health impact. In some circumstances, the consequences of making a type I error can be less important than those of type II error (not approving a device for which there is a reasonable assurance of safety and effectiveness), particularly when the device can treat a life-threatening or irreversibly debilitating disease or condition for which there are no other available treatments. Moreover, the standard value of 5% for type I error is, itself, arbitrary and not tied to any context-specific considerations.

To address this important regulatory science challenge, the Center for Devices and Radiological Health (CDRH) at the U.S. Food and Drug Administration (FDA) has used a stepwise strategy. First, CDRH has conveyed its approach to making benefit-risk assessments more robust and systematic through the release of a guidance document on benefit-risk determinations for premarket approval and De Novo classification decisions released in 2012 and updated in 2016 (US Food and Drug Administration, 2016a). The guidance document is intended to explain the FDA's thinking on the factors to take into account when making benefit-risk determinations for premarket approval of medical devices and has explicitly listed patient perspectives as one of the important factors for and CDRH staff to consider. CDRH has also made a commitment to make its regulatory decision-making more patient-centered by engaging patient stakeholders and exploring the use of quantitative methods to elicit and use patient preferences in a valid scientific manner.

To explore ways to include patient perspectives into its regulatory decision-making, CDRH sponsored a proof-of-concept pilot study in 2012 to elicit quantitative preferences on benefit-risk tradeoffs for weight-loss devices (Ho et al., 2015). Recognizing the importance of heterogeneity across the spectrum of patient preferences, the survey was designed as a discrete choice experiment, which captures not only the average, but also the distribution of patient preferences, including its variability among patients of various genders, ages, body mass indexes and previous experience with weight-loss surgeries. Moreover, the device attributes and levels considered in the study, including

benefits (amount of weight loss, weight loss duration, improvement in comorbidities), risks (side effects duration, chance of hospitalization, chance of dying from getting the device), and other device characteristics (dietary restrictions, type of operation) were selected by CDRH regulators based on a portfolio of devices that were in the development pipeline but not yet on the market. Each question in the study involved a choice between two hypothetical weight-loss devices with different attribute profiles and each attribute had varying levels. Subjects then answered eight choice questions, which revealed the relative importance of each attributes and their levels.

Since the study results have became available, they have been informing reviewers at CDRH when making their approval decisions across a wide range of weight-loss devices with differing benefit-risk profiles. However, while CDRH review staff has subjectively considered the evidence in the patient preference study, there was no objective, explicit and transparent method to directly relate the specific patient preference evidence developed in the study to the acceptable level of uncertainty associated with the submitted clinical evidence.

In recent years CDRH has achieved significant milestones to facilitate designing and conducting patient preference studies by sponsors and patient groups. In 2015 the Medical Device Innovation Consortium released a patient preference framework report sponsored by the FDA. It discusses how patient preference information can be used at various stages of the total product life cycle and also includes a catalog of existing methods for eliciting patient preferences compiled by a panel of experts. In 2016, the FDA Patient Preference Information guidance document was released, containing the following mandate (US Food and Drug Administration, 2016b):

> This guidance focuses on the specific type of patient input referred to as patient preference information [PPI], which, for the purposes of this guidance, is defined as: qualitative or quantitative assessments of the relative desirability or acceptability to patients of specified alternatives or choices among outcomes or other attributes that differ among alternative health interventions... The specific role of quantitative PPI is to provide estimates of how much different outcomes, endpoints or other attributes are

> valued by patients, and the tradeoffs that patients state or demonstrate they are willing to make among them. Such outcomes or other attributes of a device include demonstrated or posited measures of effectiveness, safety, and other device characteristics that may impact benefit-risk considerations, including (but not limited to) means of implantation, duration of effect, duration and frequency of use, and utility of the device.

However, there is a missing quantitative link between specific patient preferences and acceptable levels of uncertainty. Furthermore, methodologies that objectively, transparently and reproducibly determine these quantitative links have been elusive.

In this chapter, we use a Bayesian method proposed by Montazerhodjat et al. (2017) and Isakov et al. (2018), and quantitative patient preference data from the CDRH patient preference obesity study (Ho et al., 2015) to calculate acceptable levels of uncertainty (significance level and power) when designing pivotal clinical trials for clinical evidence required by regulatory decision-making. The Bayesian approach has long been applied to clinical trial design and analysis (Berry, 2004, 2006; Irony and Simon, 2006; Irony, 2007; US Food and Drug Administration, 2010; Bonangelino et al., 2011; Irony, 2012). Bayesian Decision Analysis (BDA), the particular method presented in this chapter, aims to optimize the balance between type I and II error rates and the severity of the consequences of making type I and type II errors based on patient preferences. The appropriate patient preference scores are scientifically elicited and estimated across safety and effectiveness, and used to construct hypothetically optimal balanced two-arm fixed-sample randomized clinical trials (RCTs) to maximize the expected the value for patients.

We take into account the fact that lengthy clinical trials provide more power but can negatively impact public health because they delay access of effective treatments to patients. In addition, we weigh the consequences of approving an ineffective treatment versus rejecting an effective intervention. If we set the significance level to be smaller and consequently more stringent, we not only reduce the chance approving an ineffective treatment but also increase the chance of rejecting an effective treatment.

We find that the BDA-optimal design is often substantially different in significance

level, power, and sample size from the conventional approach using a fixed one-sided significance level of 5%. Of course, the BDA-optimal design depends on several key assumptions. While this framework provides a systematic and quantitative method of incorporating multifaceted tradeoffs into RCT design, the usefulness of its recommendations relies on the appropriateness of these assumptions and on accurately calibrated model parameters.

Although we apply BDA to a specific medical device in this study, the framework applies more broadly to other therapeutics and this application is meant to serve as a proof-of-concept for a more general and systematic approach to incorporating patient preferences into medical device and drug approval processes.

5.2 BDA-Optimal Clinical Trial Design

In this section we use the quantitative framework from Montazerhodjat et al. (2017) and Isakov et al. (2018) to explicitly take into account patient preferences across multiple device-attributes when determining the optimal sample size and critical value of a balanced two-arm fixed-sample RCT. We first define a patient-centered value model associated with given medical device's attributes. We then assign prior probabilities to each possible combination of these attributes, and formulate the expected value of the trial. The optimal trial size ($2n$, where there are n patients in each arm of the study), and the one-sided significance level (α or critical value, λ_α) are then jointly determined to maximize the expected value of the trial. Note that maximizing the value of the trial means providing access of a safe and effective treatment to patients as soon as possible, or concluding that the treatment has not demonstrated a reasonable assurance of safety and effectiveness as soon as possible. It is equivalent to minimizing its losses, which include the consequences of incorrect decisions for all current and future patients, as well as the inefficiency of delaying access of a potentially safe and effective treatment to patients. Although the value model we now introduce is based on explicit-preference data for a specific device, these methods have been applied to a previous oncology analysis (Montazerhodjat et al., 2017; Isakov et al.,

2018) and are applicable to other therapeutics for which patient-preference data are available.

5.2.1 Patient-Centered Value Model

The CDRH weight-loss device study elicited and quantified the importance of safety, effectiveness, and other attributes of weight-loss devices into patient preference scores (Ho et al., 2015). Table 5.1 shows the preference scores estimated from the survey data for percent total body weight loss (%TBWL), which is defined as a patient's absolute amount of weight loss divided by the patient's current weight, and mortality risk. The scores are on a scale from −10 (least preferred) to +10 (most preferred), where −10 is the estimated value of a 5% mortality risk to patients. %TBWL and mortality risk were modeled as continuous variables, and their preference weights were linearly interpolated between observations.

Attribute	Level	Preference score
Average amount of weight loss (%TBWL)	0%	Reference level
	5%	+0.2
	10%	+0.6
	20%	+2.0
	30%	+4.3
Chance of dying from getting the device (mortality risk)	0%	Reference level
	1%	−3.5
	3%	−7.1
	5%	−10

Table 5.1: Estimates of preference scores by attributes and levels (Ho et al., 2015).

Patient preference scores for each attribute were then mapped directly to relative values. For example, the change in value (loss) of an increase in mortality risk from 0 to 1% can be quantified by the preference score difference between these two levels. The change in value of this increase from the patient's perspective is therefore −3.5. Similarly, the change in value (gain) of an increase in percent total body weight loss (%TBWL) from 0 to 30% is +4.3. Given both these changes, and holding other attributes constant, the net change in value would then be +0.8 and so the additional weight loss would more than compensate for the increased mortality risk according to

the patient preference information. The relative loss of value per patient, L, of using one intervention over another is then defined as this net change in value. The number of patients affected can be used to scale L to estimate a collective loss of value.

As demonstrated in Montazerhodjat et al. (2017) and Isakov et al. (2018), the value associated with a clinical trial for a superiority claim can be categorized into in-trial and post-trial value. In-trial value depends on the number of subjects in each arm of the trial, and is independent of the outcome of the trial if both arms have the same number of subjects. Post-trial value, on the other hand, is completely dependent on the outcome of the trial, and affects patients beyond the scope of the trial. In particular, we assume there is no post-trial loss in value with making a correct decision – i.e., rejecting (approving) a device that is less (more) preferred relative to the control – except for the wait time caused by the regulatory approval process. We further define the relative loss in value per person of using the investigational device under the null hypothesis ($H = 0$) as L_0, and the relative loss in value per person of foregoing the use of a the investigational device under the alternative hypothesis ($H = 1$) as L_1. If the size of the target population is N, then the aggregate loss in value of a type I or II error will be $DF_t \cdot N \cdot L_0$ and $N \cdot L_1$, respectively, where DF_t is a discount factor that decreases from 1 to 0 and accounts for the wait time, t, caused by the regulatory approval process. In other words, patients place a lower value on a treatment if it is not accessible immediately. Therefore, the aggregate loss in value caused by the length of the regulatory-approval process under the alternative hypothesis is $[1 - DF_t] \cdot N \cdot L_1$.

Finally, if the investigational device is less preferred to the control, then the n subjects in the investigational arm experience a total loss of value of $n \cdot L_0$. However, if the investigational device is preferred to the control treatment, then the n subjects in the control arm forego a better treatment and experience a total loss of value of $n \cdot L_1$. The potential losses in value associated with a fixed sample trial are tabulated in Table 5.2. Note that there is no loss in value (i.e., there is maximum value) in the hypothetically optimal scenario where the correct approval decision is made immediately and without running a trial.

	Post-trial		In-trial
	$H = 0$ Accepted	$H = 1$ Accepted	
$H = 0$	0	$DF_t \cdot N \cdot L_0$	$n \cdot L_0$
$H = 1$	$N \cdot L_1$	$(1 - DF_t) \cdot N \cdot L_1$	$n \cdot L_1$

Table 5.2: Post-trial and in-trial loss of value associated with a balanced fixed-sample RCT (Montazerhodjat et al., 2017; Isakov et al., 2018).

We additionally assume time consistent (i.e., exponential) discounting, and suggest that time-horizon preferences, which measure the capacity of patients to tolerate waiting, be elicited directly in future patient survey experiments. If the annual discount rate is r, then the discount factor is given by $DF_t = e^{-rt}$, where t is length of the regulatory-approval process. This proposed discount factor ensures that patient preferences do not change over time in such a way that they become inconsistent with one another. As in Montazerhodjat et al. (2017), the duration of the regulatory-approval process is assumed to be determined by the size of the study ($2n$), the patient accrual rate for the study (η), the time required to setup the study (s), the follow-up time of the final patient to complete the study (f), and the FDA review time (τ) such that $t = s + 2n/\eta + f + \tau$.

5.2.2 Bayesian Decision Analysis

A quantitative primary endpoint based on TBWL as a percentage of initial weight is assumed for the trial. We further assume that subjects in the treatment arm receive the investigational device and each subject's response is independent of all other responses. Diet and exercise are assumed to be administered to patients in the control arm. The response variables in the treatment arm, denoted by $\{T_1, \ldots, T_n\}$, are assumed to be independent and identically distributed, where $T_i \sim \mathcal{N}(\mu_t, \sigma_t^2)$. Similarly, the control arm responses, represented by $\{P_1, \ldots, P_n\}$, are assumed to be independently and identically distributed as $P_i \sim \mathcal{N}(\mu_p, \sigma_p^2)$. We further confine ourselves to superiority trials where the device is likely to have either a positive effect ($\mu_t > \mu_p$), or no effect ($\mu_t = \mu_p$). In such cases, the treatment effect of the device, δ, is defined as the difference between the response means in the two arms

(i.e., $\delta \equiv \mu_t - \mu_p$). In a fixed-sample trial with n subjects in each arm, we collect observations from the treatment and control arms, and form the following t-statistic:

$$T = \frac{\widehat{\mu}_t - \widehat{\mu}_p}{\sqrt{\frac{\widehat{\sigma}_t^2}{n} + \frac{\widehat{\sigma}_p^2}{n}}} \tag{5.1}$$

where $\widehat{\mu}$ and $\widehat{\sigma}$ represent the sample mean and standard deviation, respectively, and T has a noncentral t-distribution with noncentrality parameter $\delta\sqrt{\frac{n}{\sigma_t^2+\sigma_p^2}}$. Under the assumption that the two variances are equal, this distribution has $2(n-1)$ degrees of freedom. The t-statistic, T, is then compared to the critical value, λ_α. Finding that $T > \lambda_\alpha$ supports rejection of the null hypothesis (i.e., that the device has no effect). The probability of failing to reject the null hypothesis, for a device that provides a treatment effect δ with response variances σ_t^2 and σ_p^2, is therefore $P(T \leq \lambda_\alpha)$.

Assuming prior probabilities p_0 and p_1 (where $p_0 + p_1 = 1$) for the cases where the investigational device is equally effective ($H = 0$) and more effective ($H = 1$) to the control treatment, respectively, and letting V_0 and V_1 be the value created in the hypothetically optimal scenarios where the correct approval decision is made immediately and without running a trial, it is straightforward to calculate the expected value associated with an RCT design with parameters $\{n, \lambda_\alpha\}$ as

$$\mathrm{E}[\text{Value}; n, \lambda_\alpha] = p_0(V_0 - \mathrm{E}[\text{Loss} \mid H_0]) + p_1(V_1 - \mathrm{E}[\text{Loss} \mid H_1]) \tag{5.2}$$

where

$$\mathrm{E}[\text{Loss} \mid H_0] = L_0 \cdot [\alpha \cdot DF_t \cdot N + n] \quad , \tag{5.3}$$

$$\mathrm{E}[\text{Loss} \mid H_1] = L_1 \cdot [N - (1-\beta) \cdot DF_t \cdot N + n] \quad , \tag{5.4}$$

α is the significance level and $1 - \beta$ is the power of the trial. As in Montazerhodjat et al. (2017) and Isakov et al. (2018), the optimal sample size (n^*) and critical value (λ_α^*) are jointly determined such that the expected value of the trial is maximized subject to an upper bound, $\text{Power}_{\max}$, for the power level, which we set to 80% in our

simulations. This power constraint is intended to represent the practical considerations of the medical device industry. In solving the constrained optimization problem, we observe that the expected value of the trial is maximized when the expected loss, $\mathrm{E}[\mathrm{Loss}; n, \lambda_\alpha] = p_0\mathrm{E}[\mathrm{Loss} \mid H_0] + p_1\mathrm{E}[\mathrm{Loss} \mid H_1]$, is minimized.

5.3 Weight Loss Device Case Study

Using Bayesian decision analysis and the estimated preference scores, we can now formulate the BDA-optimal fixed-sample test for weight loss devices. We assume an annual discount rate of 10%, a %TBWL as a percentage of initial weight standard deviation of 25% for both arms of the study, a patient accrual rate of 100 patients per year (as well as a fixed study startup time of 6 months, final observation period of 1 year, and FDA review time of 9 months), and a target population of 100,000 patients. We also consider two separate categories of interventions, low risk and high risk, which represent devices that require non-invasive and invasive surgeries, and have mortality risks of 0.1% and 0.3%, respectively. In each case, we assume the device is either ineffective ($\mu_t = \mu_p$) or effective ($\mu_t > \mu_p$), with equal prior probability. This prior is consistent with the equipoise principle of two-arm clinical trials, which states that it is only ethical to assign the same number of patients to both arms if there is no prior information in the medical profession that favors one arm over the other (Freedman, 1987). We further subcategorize effective weight loss into low effectiveness ($\mu_t = 10\%$) and high effectiveness ($\mu_t = 20\%$). Figure 5-1 summarizes the multiple categories of investigational devices considered. Finally, the control treatment (e.g., diet and exercise) is assumed to provide moderate weight loss ($\mu_p = 2\%$) and have no additional mortality risk. In Appendix C we conduct sensitivity analyses to investigate the robustness of our analysis to perturbations in our model's key assumed parameter values.

Table 5.3 lists the optimal RCTs for the devices described above. As can be seen, the device that is assumed to be low risk, but is potentially highly effective has a relatively large BDA-optimal significance level (α) of 6.5%. This value is greater than

the traditional 5% level, reflecting the fact that patients are willing to bear increased uncertainty of receiving an ineffective device, because 1) the device is thought to be safer, and/or 2) they value the weight loss benefit and don't want to miss the opportunity of receiving an effective weight-loss device, and/or 3) they want to access a potentially effective weight-loss device sooner. The preference and ability to shorten the regulatory-approval process is especially apparent as the trial size is set to 44 patients (22 in each arm), which is approximately 6 times smaller than the trial size recommended for the low risk, low potential weight loss device.

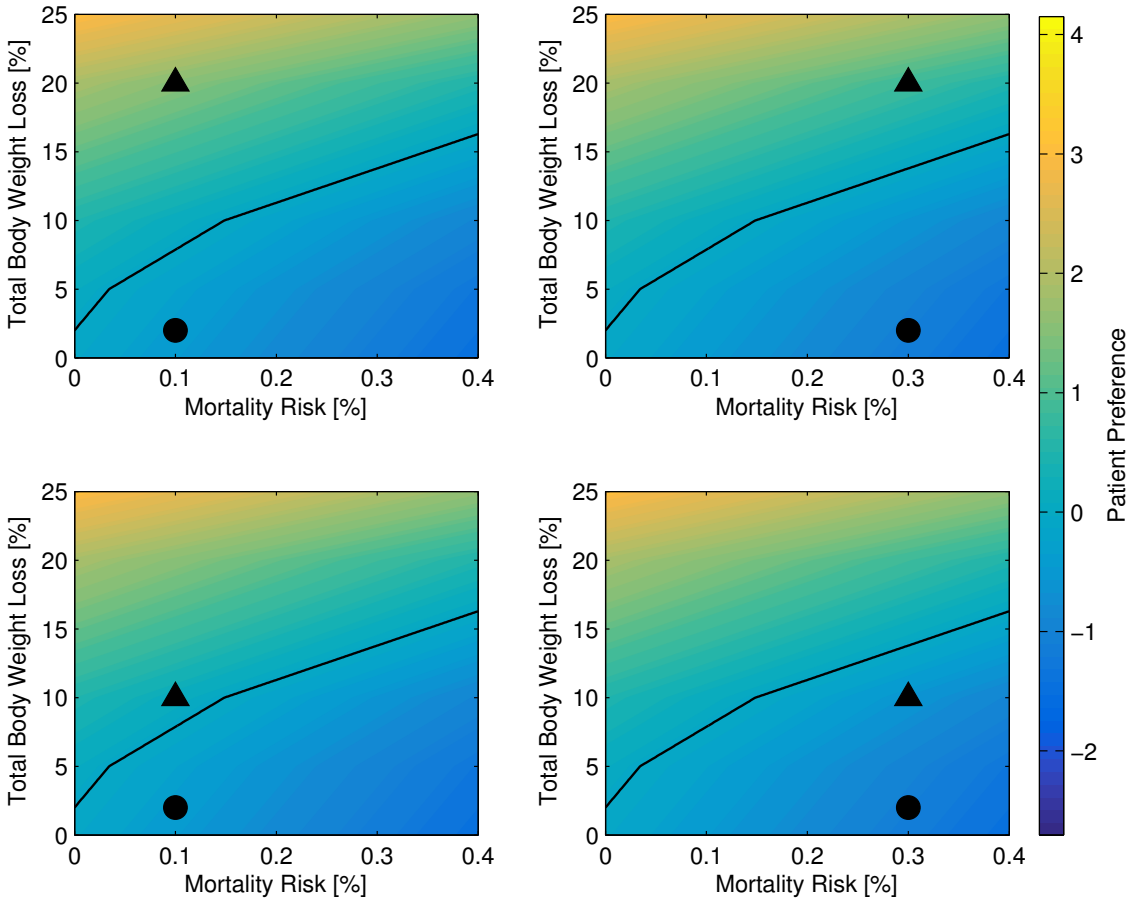

Figure 5-1: Possible device characteristics: low risk with high weight loss (top left), low risk with low weight loss (bottom left), high risk with high weight loss (top right), and high risk with low weight loss (bottom right). Circles and triangles represent the investigational device characteristics under the null hypothesis ($H = 0$) and alternative hypothesis ($H = 1$), respectively.

In contrast, the optimal significance level for the device that is assumed to have a

high mortality risk is 1.1%. The additional mortality risk causes the preference score of this device to be relatively low, and hence patients require greater evidence of clinical effectiveness. To achieve this goal, the trial size is set to 78 patients, almost twice the size of the trial recommended for the low risk device with an equivalent potential effectiveness. However, the preference to keep the trial size small remains evident, especially when compared to the 278-patient trial size recommended for the low risk, low potential weight loss device. Finally note that, because the high risk, low potential weight loss device is not preferred to the standard treatment under the alternative hypothesis (see Figure 5-1), the BDA framework recommends this device be rejected without conducting a trial.

Device characteristics under $H = 1$	Trial size $(2n)$	Critical value (λ_α)	Significance (α)	Power $(1 - \beta)$
Low risk, high %TBWL	44	1.54	6.5%	80%
Low risk, low %TBWL	278	1.83	3.5%	80%
High risk, high %TBWL	78	2.33	1.1%	80%
High risk, low %TBWL	0	–	–	–

Table 5.3: BDA-optimal RCTs for weight loss devices.

While the previous analysis studied the benefit-risk preferences of the average patient, we can also calculate BDA-optimal trial designs for more risk-tolerant early adopters. In this analysis, we derive counterpart BDA-optimal RCTs for a subset of patients who have a fraction (γ) of the risk aversion of the average survey respondent. For example, if $\gamma = 2/3$, a patient's risk aversion is one third smaller in magnitude when compared to the average patient's risk preferences seen in Table 5.1. We use this coefficient for illustrative purposes only, and recommend that preference scores from population subsets, including categorization by disease severity and other demographics, be elicited directly from survey responses. Table 5.4 compares the BDA-optimal designs for patients who are one third $(\gamma = 2/3)$ and two thirds $(\gamma = 1/3)$ less risk averse than the average patient.

The BDA-optimal RCT designs vary substantially across risk-tolerance groups. With a 1/3 decrease in risk aversion, the sample size decreases by a factor of 21–

Device characteristics under $H = 1$	Risk aversion (γ)	Trial size ($2n$)	Critical value (λ_α)	Significance (α)	Power ($1 - \beta$)
Low risk, high %TBWL	1	44	1.54	6.5%	80%
	2/3	34	1.26	10.9%	80%
	1/3	20	0.77	22.5%	80%
Low risk, low %TBWL	1	278	1.83	3.5%	80%
	2/3	190	1.36	8.7%	80%
	1/3	94	0.71	24.0%	80%
High risk, high %TBWL	1	78	2.33	1.1%	80%
	2/3	62	1.99	2.6%	80%
	1/3	44	1.54	6.5%	80%
High risk, low %TBWL	1	0	–	–	–
	2/3	0	–	–	–
	1/3	278	1.83	3.5%	80%

Table 5.4: BDA-optimal RCTs for weight loss devices for three risk-tolerance groups.

32%, and the significance level increases in magnitude by 1.5–5.2%, such that the recommended α's for low and high risk devices are on the order of 10% and 2.5%, respectively. The small number of patients and large α's for the risk tolerant subgroups relative to the average population are observed because the BDA-optimal RCT, by being less conservative, aims to grant faster access to these "early adopters." The decrease in the length of the regulatory-approval process is a consideration to offset the excess risk from the extra permissiveness in the trial, and the overall penalty – the expected harm to current and future risk-tolerant patients – may be minimized. Finally note that, if effective, the high risk, low potential weight loss device is preferred to the standard treatment by the most risk-tolerant subgroup (in fact they view the device similar to how the average population view the low risk, low potential weight loss device), and the BDA framework recommends conducting a conservative trial.

5.4 Discussion

In this chapter, we have applied a quantitative framework in which patients' preferences are the center of RCT design. We quantify the loss in value to public health associated with different actions in any fixed-sample RCT, use a BDA framework to

aggregate the value of the trial, and then determine an optimal RCT in which the expected value is maximized. We tailor this framework to weight-loss devices, using quantitative preference evidence elicited from patients through conjoint analysis, and assumptions for RCT statistics to design BDA-optimal RCTs for both average and more risk-tolerant patient populations.

Our results demonstrate that the traditional RCT design with a fixed statistical significance level does not necessarily maximize overall value (or equivalently, minimize harm) to current and future patients of an investigational treatment. For low-risk devices and risk-tolerant populations, the inefficiency is mainly caused by lengthy RCTs that are too conservative and overprotective of the type I error rate (i.e., too focused on rejecting ineffective treatments and on avoiding the harm caused by false positives.) Missed treatment opportunities do indeed harm patients, and should be considered along with the risk of approving ineffective or risky treatments.

Conversely, for some high-risk devices, such as those that require open surgery, traditional one-sided significance levels of 5% are more permissive than the BDA-optimal thresholds. These RCTs allow for a larger chance of approving ineffective or riskier treatments, such that the expected benefits are not justified by the risk to patient health. We believe that a more nuanced consideration of both significance level and power described here is instructive to the design of future clinical trials. Although we made strong assumptions here for illustrative purposes, these assumptions can be readily relaxed in future work. For example, while we have used point prevalence for simplicity in this chapter, period prevalence, incidence rates, and other epidemiological measures can be used to estimate the total population affected by the outcome of the trial. Moreover, other factors – including the time until the adverse effects of a type I error are discovered after a device is inadvertently approved, measures of disease burden, and the expected time until a new treatment is discovered that is at least as safe and effective as the investigational device – can easily be incorporated into the model (Montazerhodjat et al., 2017).

Our findings must therefore be qualified in several respects. First, many clinical trials are non-inferiority trials instead of the superiority trials we have considered.

Second, we have considered fixed sample clinical trials, when in reality clinical trials for regulatory purposes could be adaptive and might include interim analyses for early signals of effectiveness, futility, or lack of safety. Any of these possible adaptations in any given trial may alter the optimal significance level and power and appropriate modifications to our calculations are required to determine the optimal designs in these situations.

Third, the trials considered here use the percent of total body weight loss as the primary endpoint, and mortality risk as the only safety concern. For weight-loss devices, these attributes are clear and of unambiguous importance. Moreover, total body weight loss is a surrogate endpoint for morbidity and mortality; hence this trial resembles those of many other therapeutics. Other attributes may also be included, such as weight-loss duration, co-morbidities, side-effect duration, and others, which are more difficult to gauge. Study-specific definitions of type I error and type II error loss would require more nuanced treatment in these trials, but can easily be included in our BDA framework.

Fourth, we acknowledge that hypothetical patient choices such as the ones obtained in a discrete-choice experiment do not have the same clinical and emotional consequences as actual choices. However, advances in patient preference elicitation methods and best practices have helped increase the reliability of such results by ensuring that respondents are well informed, and that hypothetical biases are minimized. Despite these limitations, the estimated preference scores allow us to develop quantitative models to compare benefit-risk trade-offs across device attributes. This information is required for making patient-centered, evidence-based regulatory decisions.

Finally, we have constrained our attention to patients' medical outcomes without considering the financial cost to patients and their families, to industry, or to society. New therapies often come at a very high financial cost, which, when taken into account, may raise the bar of success for new agents, thus lowering the acceptable significance level. On the other hand, the larger the target population, the more robust the results will be to the accuracy of N and the more palatable higher finan-

cial costs might be for developing beneficial therapies. While other decision makers (insurance companies, etc.) may integrate financial costs into consideration, these concerns would not be included in regulatory decisions. On the other hand, the increased significance levels that we have proposed may lower the cost of clinical trials – which has grown to an average of $36,500 per patient as of 2013 (Battelle Technology Partnership Practice, 2015) – and reduce the risk to sponsors, which may encourage device development, lower device costs, and further accelerate clinical research.

To incorporate perspectives from the entire biomedical ecosystem, as well as the value of patient input to the device development process, CDRH has developed a patient engagement advisory committee, consisting of key stakeholder groups – patient advocacy, caregivers, physicians, medical device and biopharma executives, regulators, and policymakers. It is possible to consider whether this committee would be an appropriate forum to consider formulating explicit cost estimates for type I and type II errors. These estimates can then be incorporated into the FDA decision-making process as additional inputs to their quantitative and qualitative deliberations.

This ability of the BDA framework to systematically weigh multifaceted tradeoffs that reflect a variety of perspectives combined with its flexibility and practicality make it a potentially valuable tool for optimal RCT design. While the framework is robust, we emphasize that careful consideration must be applied to the assumptions underlying the specific models in order to produce useful recommendations. If correctly implemented, a Bayesian perspective has the potential to benefit all stakeholders. In the next chapter, we demonstrate how this framework can be incorporated into an adaptive design.

Chapter 6

Bayesian Adaptive Patient-Centered Clinical Trials

As was discussed in the previous chapter, the regulatory approval process for new medical therapies involves statistical decisions that are subject to error: approving an ineffective therapy or failing to approve an effective one. The potential harm to patients of these two types of error—known as type I and type II error, respectively—are not necessarily equal, and may also differ across patients, diseases, therapies, and other circumstances. In this chapter, we propose a patient-centered Bayesian adaptive design that applies sequential likelihood ratio tests to randomized clinical trials and incorporates patient preferences and burden-of-disease measures so as to minimize the expected harm to current and future patients. Using U.S. Burden of Disease Study 2010 data for a variety of diseases, we find that, relative to the typical balanced fixed-sample two-arm clinical trial, the expected sample size, type I, and type II error can be decreased up to 63%, 43%, and 38%, respectively in our framework.

6.1 Introduction

Adaptive protocols are becoming increasingly popular in the design of randomized clinical trials (RCTs). Unlike traditional fixed-sample balanced two-arm RCTs, adaptive trials use the accumulation of past observations to modify their progress, a

characteristic that makes Bayesian inference the ideal tool for their implementation (Anscombe, 1963; Berry, 1987; Berry and Eick, 1995; Cheng et al., 2003; Berry, 2004, 2006, 2010, 2015). Adaptive RCTs have been proposed as a way to increase the efficiency of the regulatory approval process by reducing not only time and costs, but also the probability of false positives (type I error) and false negatives (type II error). Modern RCTs use different types of adaptive designs for several purposes, including dose finding, changing the randomization proportions or accrual rate, dropping arms or doses, integrating multiple phases of development, sample size reestimation, and population enrichment (Berry, 2010, 2015; Bhatt and Mehta, 2016). Examples of adaptive RCTs include I-SPY (Harrington and Parmagiani, 2016), LUNG-MAP (Steuer et al., 2015), and GBM-AGILE (National Biomarker Development Alliance).

In this chapter, we apply sequential likelihood ratio tests to the RCT sample size of balanced two-arm clinical trials. For a given significance level and power, determined from the severity of the adverse effects of type I and type II errors on the patient (Montazerhodjat et al., 2017; Isakov et al., 2018), we find stopping boundaries for superiority and futility for the primary endpoint. This protocol allows us to end the trial early in cases where the investigational treatment shows clear signs of effectiveness or ineffectiveness on the patient, or to collect additional observations when current observations do not lead to a clear choice.

6.2 Methods

The fixed-sample RCTs considered in the previous chapter are designed so that a decision is made after a predetermined number of observations have been collected. These observations are costly to acquire, however, and introduce a delay in the regulatory decision-making process. Intuition suggests it would be preferable to make a decision after a number of observations if the resulting probability of error is small enough, and to continue collecting observations otherwise. To do so, we apply Bayesian sequential analysis to our model.

6.2.1 Bayesian Decision Analysis for Fixed-Sample Trials

As in the previous chapter, a quantitative primary endpoint is assumed for the trial. However, here we assume the response variance in each arm is known and equal to σ^2 (Isakov et al., 2018). In a fixed-sample test with n subjects in each arm, we collect observations from the treatment and control arms, and form the following Z-statistic:

$$Z_n = \frac{\sqrt{I_n}}{n} \sum_{i=1}^{n} (T_i - P_i) \ , \tag{6.1}$$

where Z_n is a normal random variable, i.e., $Z_n \sim \mathcal{N}(\delta\sqrt{I_n}, 1)$, and $I_n = \frac{n}{2\sigma^2}$ is the information in the trial (Jennison and Turnbull, 1999). The Z-statistic, Z_n, is then compared to the critical value, λ, and the therapy is approved if $Z_n > \lambda$. The probability of approving a therapy given a treatment effect δ is therefore $\Phi(\delta\sqrt{I_n} - \lambda)$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function.

Given a realization $Z_n = z_n$, it is straightforward to calculate the expected harm to patients (C), which can be measured using either patient preference information or burden of disease estimates, associated with an RCT design with parameters $\{n, \lambda\}$ as

$$\mathrm{E}[C] = \int \mathrm{E}[C \mid Z_n = z_n] \cdot \Pr(Z_n = z_n) \cdot dz_n \ . \tag{6.2}$$

Let C_{ij} be the total harm of choosing hypothesis $\widehat{H} = i$ when $H = j$. If the decision rule selects the alternative hypothesis $\widehat{H} = 1$ given $Z_n = z_n$, then the probability of error is $\Pr(H = 0 \mid Z_n = z_n)$, and the expected cost of this decision is

$$\Pr(H = 0 \mid Z_n = z_n) \cdot C_{10} + \Pr(H = 1 \mid Z_n = z_n) \cdot C_{11} \ . \tag{6.3}$$

Similarly, if the decision rule selects the null hypothesis $\widehat{H} = 0$, then the expected cost of this decision is

$$\Pr(H = 0 \mid Z_n = z_n) \cdot C_{00} + \Pr(H = 1 \mid Z_n = z_n) \cdot C_{01} \ . \tag{6.4}$$

To minimize the expected harm, also known as Bayes risk, the decision rule should

choose the hypothesis with the smallest expected cost

$$\frac{\Pr(H=0 \mid Z_n = z_n) \cdot C_{10} + \Pr(H=1 \mid Z_n = z_n) \cdot C_{11}}{\Pr(H=0 \mid Z_n = z_n) \cdot C_{00} + \Pr(H=1 \mid Z_n = z_n) \cdot C_{01}} \underset{\widehat{H}=1}{\overset{\widehat{H}=0}{\gtrless}} 1 \ . \tag{6.5}$$

Assuming prior probabilities p_0 and p_1 for $H = 0$ and $H = 1$, respectively, this equation can be written in terms of the likelihood functions using Bayes' rule

$$\frac{p_0 \cdot \Pr(Z_n = z_n \mid H=0) \cdot C_{10} + p_1 \cdot \Pr(Z_n = z_n \mid H=1) \cdot C_{11}}{p_0 \cdot \Pr(Z_n = z_n \mid H=0) \cdot C_{00} + p_1 \cdot \Pr(Z_n = z_n \mid H=1) \cdot C_{01}} \underset{\widehat{H}=1}{\overset{\widehat{H}=0}{\gtrless}} 1 \ , \tag{6.6}$$

and can be simplified to the following likelihood ratio test (LRT),

$$\Lambda(z_n) = \frac{\Pr(Z_n = z_n \mid H=0)}{\Pr(Z_n = z_n \mid H=1)} \underset{\widehat{H}=1}{\overset{\widehat{H}=0}{\gtrless}} \frac{p_1(C_{01} - C_{11})}{p_0(C_{10} - C_{00})} = \eta. \tag{6.7}$$

Taking the natural logarithm of the likelihood ratio function, $\Lambda(z_n)$, and solving for z_n allows us to express the decision rule that minimizes the expected harm for a given n as

$$z_n \underset{\widehat{H}=1}{\overset{\widehat{H}=0}{\lessgtr}} \lambda_n \ , \tag{6.8}$$

where

$$\lambda_n = \frac{\delta\sqrt{I_n}}{2} - \frac{\ln(\eta)}{\delta\sqrt{I_n}} \ . \tag{6.9}$$

As in (Isakov et al., 2018), the optimal sample size (n^*) and critical value (λ_n^*) are then jointly determined such that the expected harm of the trial is minimized subject to an upper bound for the power level, $\text{Power}_{\max}$, which we set to 90% in our simulations. This power constraint is chosen to reflect industry practice—the higher the power level, the more likely it is that a given treatment effect is detected, but at the cost of a larger sample size.

6.2.2 Sequential Analysis for Bayesian Adaptive Trials

In contrast to the fixed-sample RCTs of the previous section, we now propose to make binary decisions on the basis of a variable number of observations, choosing to

stop collecting additional observations when a sufficiently good decision can be made. Common sense tells us that it is particularly valuable to collect additional data when the current observations do not lead to a clear choice. Specifically, we have three possible choices after obtaining each observation: decide that $H = 1$, decide that $H = 0$, or continue collecting data. This can be viewed as a generalization of the standard Neyman-Pearson hypothesis test, which employs a tradeoff between the two types of error. Here, we have a three-way tradeoff between the two types of error and the time required to make a decision.

It can be shown that the sequence of log-likelihood ratios, $S_n = \ln \Lambda(Z_n)$, is a random walk under each of the two hypotheses. An appropriate decision rule for this framework is to choose $H=0$ if the sample value s_n of S_n exceeds some positive threshold, choose $H=1$ if s_n is less than some negative threshold, and to continue testing if the sample value lies in between the two thresholds. By setting these two thresholds to be $\ln(\alpha)$ and $\ln(\beta)$, the type I and type II error rates are guaranteed to be smaller than α and β, respectively (Gallager, 2014),

$$s_n \begin{cases} > \ln(1/\beta) & ; \quad \text{select } \widehat{H} = 0 \\ < \ln(\alpha) & ; \quad \text{select } \widehat{H} = 1 \\ \text{otherwise} & ; \quad \text{continue .} \end{cases} \tag{6.10}$$

Solving for z_n, the decision boundaries for superiority and futility can be rewritten as,

$$z_n \begin{cases} < \lambda_{l,n} & ; \quad \text{select } \widehat{H} = 0 \\ > \lambda_{u,n} & ; \quad \text{select } \widehat{H} = 1 \\ \text{otherwise} & ; \quad \text{continue} \end{cases} \tag{6.11}$$

where

$$\lambda_{l,n} = \frac{\delta\sqrt{I_n}}{2} - \frac{\ln(1/\beta)}{\delta\sqrt{I_n}} \quad , \quad \lambda_{u,n} = \frac{\delta\sqrt{I_n}}{2} - \frac{\ln(\alpha)}{\delta\sqrt{I_n}} \, . \tag{6.12}$$

By using the sequential decision procedure described above, for a given expected sample size, we simultaneously get the Chernoff bound error exponent for $H=1$ that

a fixed-sample trial would provide if we gave up entirely on an error exponent for $H = 0$, and vice versa (Gallager, 2014). This gain in efficiency, however, has to be weighed against the substantial increase in protocol complexity, as we shall discuss further in Section 6.4.

6.2.3 Patient Value Model

To define the target error rates α and β for our fixed-sample and Bayesian adaptive RCTs, we must first define the harm associated with regulatory delay and both types of error, or equivalently, the loss of value to the patient. Our patient-centered value model is derived from Montazerhodjat et al. (2017) and Isakov et al. (2018), but we extend their classical fixed-sample trials to the case of Bayesian adaptive trials as in Colton (1963) and Cheng et al. (2003) where there is continuous monitoring of the data and Bayesian inference is carried out during the trial. Similar to the proposal of Grieve (2015) and Isakov et al. (2018), we assign different weights to the consequences of false positives and false negatives, based on the U.S. Burden of Disease Study (US Burden of Disease Collaborators, 2013).

As in Montazerhodjat et al. (2017) and Isakov et al. (2018), to quantify the severity of the adverse effects of a type I or type II error, we estimate the years lived with disability (YLD), a measure of overall disease burden expressed as the number of years lost due to disability or ill health. A percentage point increase of burden means that a patient would be indifferent to a choice between living each year with the additional adverse effects, or losing 1% of each year if, for the rest of that year, they could live without the adverse effects. To incorporate the additional adverse effects of decreased mortality, we estimate the overall severity of a disease using the following relation:

$$s = \frac{D + \text{YLD}}{D + N} \ , \tag{6.13}$$

where D is the number of deaths caused by the disease, YLD is the years lived with disability, and N is the disease prevalence, all estimated from the U.S. Burden of Disease Study (US Burden of Disease Collaborators, 2013). As in Isakov et al.

(2018), under the alternative hypothesis, given a treatment effect $\delta > 0$, and response variance σ^2, the loss in value to the patient of forgoing the treatment, L_1, can be estimated as

$$L_1 = \min\left(\frac{\delta}{\sigma}, 1\right) s \ . \tag{6.14}$$

Alternatively, under the null hypothesis, the loss of value to the patient, L_0, of using an ineffective and possibly toxic treatment is estimated at 0.07 using the values for YLD, D, and N for the adverse effect of medical treatment in the U.S. (US Burden of Disease Collaborators, 2013; Isakov et al., 2018). This value can be made more precise by considering the specifics of the therapy under investigation, but for simplicity, we hold this value constant across all diseases in this application.

Finally we decompose the value of a clinical trial into the same in-trial and post-trial components given in Table 5.2. We also assume the rate at which patients discount treatment benefits over time, r, is linearly related to the severity of the disease, such that $r = r_0 + \gamma s_2$. Table 6.1 summarizes the parameter values used in our analysis.

Parameter	Assumed Value
Probability that treatment is effective (p_1)	50%
Treatment effect to response variability ratio ($\frac{\delta}{\sigma}$)	0.25
Excess burden caused by toxic and ineffective treatment (L_0)	0.07
Discount rate that measures baseline capacity of patients to tolerate delay (r_0)	10% per year
Discount rate factor that decreases the capacity of patients to tolerate delay in the presence of increased disease severity (γ)	10% per unit severity
Patient accrual rate (κ)	200 patients per year
Start-up time before patient enrollment	6 months
Follow-up period after enrolling the last patient	1 year
FDA review time	9 months

Table 6.1: Assumptions for RCT design.

6.3 Results

Applying the patient value model we have described, we can calculate the optimal Bayesian decision analysis (BDA) fixed-sample RCT for a therapy intended to treat a severe disease. Similar to Isakov et al. (2018), Table 6.2 ranks selected diseases from the 30 leading causes of premature mortality in the U.S. by estimated severity. As can be seen in column 2, some diseases are not as harmful as others. For example, chronic obstructive pulmonary disease (COPD) is much less severe than pancreatic cancer, which reflects the fact that pancreatic cancer has a poor prognosis, even when diagnosed early. As such, the BDA-optimal fixed-sample RCT for pancreatic cancer ($n=83$, $\alpha=37.1\%$) is much less conservative than that for COPD ($n=365$, $\alpha=1.6\%$). In general, the optimal significance level increases with the severity of the disease, and the sample size decreases, as the harm from not approving a potentially effective therapy grows larger. From the patient perspective, it is beneficial in the case of a severe disease to increase the probability of a false approval, in exchange for shorter clinical trials with potentially fewer false negatives. It should be noted that a number of the diseases in Table 6.2 are, in fact, broad collections of heterogeneous sub-diseases (e.g., breast cancer), and therefore the RCT designs here are for illustrative purposes only. The BDA framework can be readily adapted to more finely stratified categories of disease as long as information about burden of disease and survival statistics are available.

Using the estimates for statistical size and power for the optimal non-adaptive RCTs to form the stopping boundaries for superiority and futility in (6.11), we can simulate the performance of the adaptive RCT protocol. The sample size statistics (mean, standard deviation, and 25th, 50th and 75th percentiles) under both the null ($H=0$) and alternative ($H=1$) hypotheses are reported in Table 6.2 and Figures 6-1–6-2. There is a noticeable decrease in the expected and median sample sizes, especially for the larger and more conservative RCTs. The average decrease in the expected sample size across all diseases is 44% (maximum decrease = 63%) under the null hypothesis, and 37% (maximum decrease = 39%) under the alternative hypothesis.

The decrease in the expected sample size occurs because the adaptive RCT is able to stop early when a sufficiently good decision can be made.

In addition, the last two columns in Table 6.2 show that the probabilities of both type I and type II errors are consistently lower for adaptive RCTs. The average decrease in type I error across all diseases is 19% (maximum decrease = 43%), and 16% (maximum decrease = 38%) for type II error. In other words, by using the sequential test described by (6.11), we simultaneously get a reduction in the average sample size and the error rate relative to a fixed-sample RCT. As a result, the average decrease in the expected harm across all diseases is 10% (maximum decrease = 17%) under the null hypothesis, and 15% (maximum decrease = 16%) under the alternative hypothesis. While these results are striking, it should be noted that these significant reductions in error probability and sample size are accompanied by substantial increases in protocol complexity, which is an issue that we discuss further in the following section.

			Non-Adaptive			Adaptive					
						Sample Size ($H = 0$)		Sample Size ($H = 1$)			
Disease	Severity	Prevalence (000s)	Sample Size	α (%)	Power (%)	Mean (SD)	Median (IQR)	Mean (SD)	Median (IQR)	α (%)	Power (%)
Chronic kidney disease	0.04	9,919	387	1.2	88.9	142.8 (124.2)	106.0 (61.0,178.0)	246.1 (158.3)	209.0 (133.0,317.0)	0.9	89.3
Prostate cancer	0.05	3,710	381	1.3	88.9	142.6 (123.9)	103.0 (62.0,180.0)	239.6 (147.6)	208.0 (135.0,307.0)	1.1	90.5
Breast cancer	0.05	3,885	373	1.4	89.0	142.9 (122.9)	103.0 (61.0,182.0)	235.1 (146.4)	198.0 (132.0,300.0)	1.2	90.3
COPD	0.06	32,372	365	1.6	89.1	141.6 (122.4)	103.0 (60.0,180.5)	233.1 (146.3)	197.0 (128.0,302.0)	0.8	90.8
HIV/AIDS	0.10	1,160	314	3.0	89.6	142.0 (116.5)	106.0 (62.0,184.0)	195.6 (129.3)	161.0 (105.0,251.0)	2.4	90.6
Kidney cancers	0.12	329	301	3.5	89.6	140.0 (112.5)	105.0 (61.0,182.0)	184.9 (121.8)	154.0 (99.0,237.5)	3.1	91.2
Ischemic heart disease	0.12	8,896	300	3.6	89.7	142.6 (114.9)	107.0 (63.5,186.0)	183.6 (124.6)	151.0 (98.0,235.0)	3.1	90.2
Non-Hodgkin lymphoma	0.13	283	286	4.2	89.6	140.3 (112.0)	106.0 (62.0,182.0)	179.9 (125.1)	147.0 (93.0,231.0)	3.4	90.8
Colorectal cancer	0.15	799	274	4.9	89.8	139.8 (109.8)	107.0 (65.0,182.0)	168.8 (118.8)	138.0 (88.0,215.0)	3.8	91.2
Cardiomyopathy	0.17	416	264	5.5	89.9	139.7 (111.5)	107.0 (63.0,180.0)	162.6 (113.7)	133.0 (83.0,208.0)	4.5	90.9
Alzheimer's disease	0.18	5,145	260	5.8	90.0	138.4 (107.8)	108.0 (63.0,180.0)	156.9 (109.6)	128.5 (81.0,200.5)	4.2	91.6
Influenza	0.20	119	245	6.8	89.9	136.0 (103.8)	105.0 (62.0,180.0)	150.2 (106.5)	123.0 (74.0,193.0)	5.0	91.4
Leukemia	0.21	140	239	7.3	90.0	135.0 (103.0)	106.0 (63.0,173.0)	148.1 (108.6)	118.0 (72.0,191.0)	6.4	91.6
Hypertensive heart disease	0.27	185	215	9.5	90.0	131.1 (97.6)	103.0 (63.0,169.0)	131.0 (97.8)	104.0 (62.0,167.0)	7.8	91.3
CNS/PNS cancers	0.30	60	197	11.5	90.0	124.3 (91.9)	99.0 (60.0,161.0)	119.9 (90.7)	93.5 (56.0,156.5)	10.1	91.6
Liver cancer	0.44	31	147	19.4	90.0	113.4 (83.7)	90.5 (55.0,147.5)	90.8 (74.3)	68.0 (39.0,118.0)	15.8	92.2
Lung cancer	0.45	290	150	18.8	90.0	111.8 (80.9)	89.0 (55.0,145.0)	95.9 (79.9)	71.0 (40.0,126.0)	15.5	92.6
Cirrhosis of the liver	0.49	78	137	21.5	90.0	105.9 (75.3)	85.0 (52.0,137.0)	85.9 (71.8)	64.0 (36.0,112.0)	17.6	92.3
Pancreatic cancer	0.71	23	83	37.1	90.0	83.9 (65.6)	67.0 (37.0,111.0)	56.1 (55.5)	37.0 (19.0,73.0)	31.9	93.8

Table 6.2: Selected diseases from the 30 leading causes of premature mortality in the U.S. ranked with respect to their severity from lowest (top) to highest (bottom). The sample size statistics, as well as statistical size (α) and power are reported for both fixed-sample and adaptive clinical trials. SD, standard deviation; IQR, interquartile range about the median.

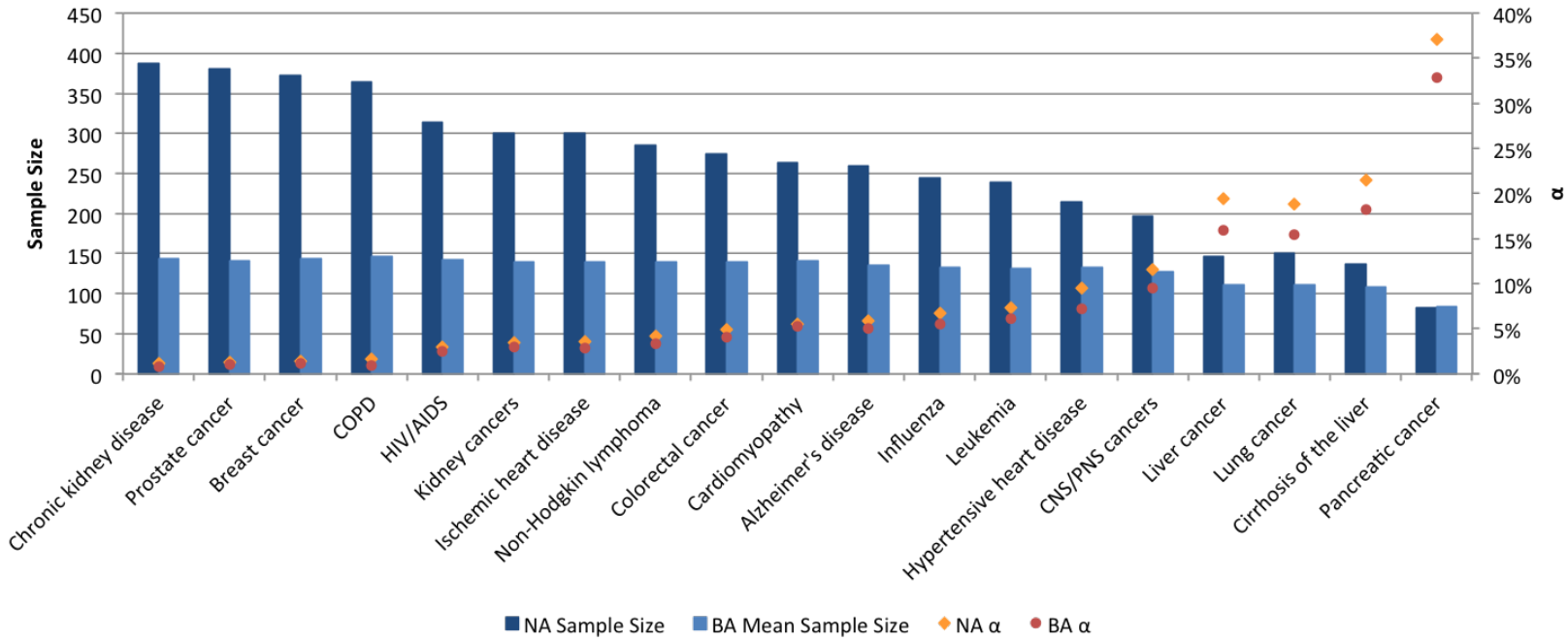

Figure 6-1: Sample sizes and type I error rates (α) for BDA-optimal non-adaptive (NA) and Bayesian adaptive (BA) randomized clinical trials under $H = 0$.

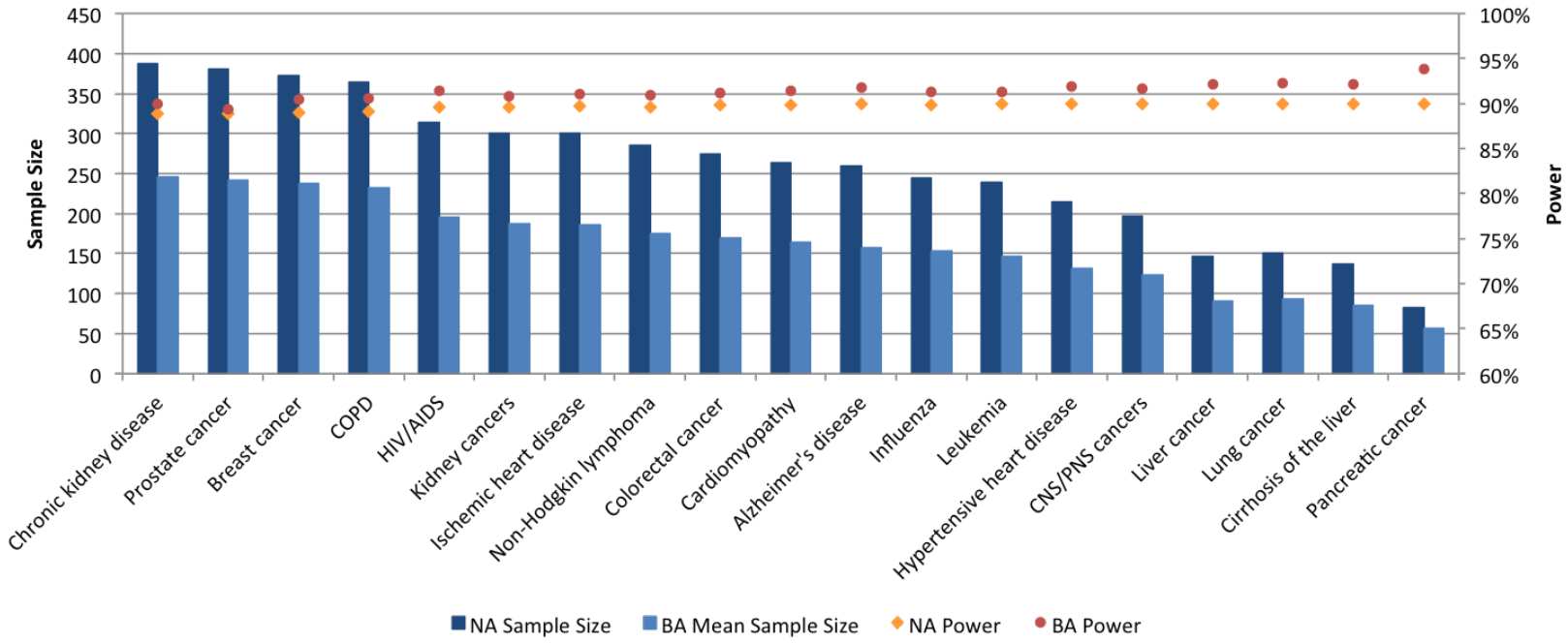

Figure 6-2: Sample sizes and power for BDA-optimal non-adaptive (NA) and Bayesian adaptive (BA) randomized clinical trials under $H = 1$.

6.4 Discussion

Our simulation results show that clinical trials with fixed sample sizes do not necessarily minimize overall harm—or equivalently, maximize overall benefit—to current and future patients of a disease. Often this is due to the fact that the drug approval process takes time and missed opportunities for early approval or rejection can harm patients. On the other hand, when current observations offer ambiguous clinical evidence, it can benefit both patients and regulators to increase the trial size to protect against both type I and type II errors. For a given significance level and power, such an adaptive strategy leads to an overall reduction in the expected trial duration and an increase in value to the patient.

There are, however, some disadvantages to adaptive designs. First, the logistics of Bayesian adaptive RCTs are more complicated than those of standard fixed-sample trials. An adaptive process requires consolidating and analyzing observations at the same time that patients are being accrued. This parallel operation adds additional overhead costs and operational complexity to an already complicated process, and increases the probability of administrative errors.

A related issue is the statistical complexity in designing futility and superiority stopping boundaries. We have considered a relatively simple example for the Z-statistic, but computing the performance characteristics for other statistics is not trivial. Moreover, non-standard adaptive designs will likely undergo additional scrutiny by regulators and institutional review boards, adding time and uncertainty to the review process.

There is also the issue of information leakage, which can diminish the credibility of a clinical trial. However, for the balanced two-arm RCTs considered here, the fact that a trial should be continued merely indicates that the data lie between the study's stopping boundaries, which is unlikely to compromise the study's integrity.

Finally, the assumption of time-invariant patient populations throughout the duration of the trial is important. While patient trends affect fixed clinical trials as well, adaptive trials are especially sensitive to changes in subject composition. Se-

quential likelihood ratio tests generally assume stationary stochastic processes, and drastic changes in the patient population would require considerably more sophisticated stopping boundaries.

6.5 Conclusion

Bayesian adaptive clinical trials are becoming more popular because they exhibit clear advantages over their fixed-sample counterparts in terms of cost, speed, and potential impact on current and future patients (Berry, 2010). In such a framework, it is straightforward to incorporate burden-of-disease measures and patient preferences so as to reflect differences in the potential harm of type I and II errors to patients. Although such processes are inherently more complex than traditional fixed-sample clinical trials, the added complexity may well be worthwhile when weighed against the savings in clinical trial costs, the number of patients exposed to potentially toxic therapies, and the time to approval of effective therapies.

Chapter 7

Financially Adaptive Clinical Trials via Option Pricing Analysis

In the previous chapter, we considered designing adaptive clinical trials from the perspective of patients. Here, we take the perspective of industry sponsors of new therapies and investors. When valuing a candidate therapy from a financial perspective, we must consider the possibility that as the regulatory approval process unfolds, industry sponsors might stop a trial early if clinical evidence suggests market prospects are not as favorable as originally forecasted. Intuition suggests that clinical trials that can be modified as new data are observed are more valuable than trials without this flexibility. In this chapter, we propose modeling the accrual of information in a clinical trial as a series of real options, which allows us to systematically design early-stopping decision boundaries that maximize the economic value to the sponsor. In an empirical analysis of selected disease areas, we find that when a therapy is ineffective, our method can decrease the expected cost incurred by the sponsor in terms of total expenditures, number of patients, and trial length by up to 46%. Moreover, by amortizing the large fixed costs associated with a clinical trial over time, financing these projects becomes less risky, resulting in lower costs of capital and larger valuations when the therapy is effective.

7.1 Introduction

When discounted cash flow (DCF) analysis is used to value a project, one must model the possibility that as the project unfolds, managers might expand the project if it goes well, or contract or even abandon the project when it does not work out as forecasted. Intuitively, projects that can be modified as new information becomes available are more valuable than projects without this flexibility. Moreover, the more uncertain the outlook, the more valuable this flexibility becomes. The ability to modify projects in the future are known as real options, and in this chapter we consider how this valuation technique can be used to design financially-optimal decision boundaries for futility in clinical trials.

Randomized clinical trials (RCTs) conducted by pharmaceutical and biotech firms proceed through a series of controlled observations. Each observation of a drug's safety and efficacy provides information about its market potential, and therefore after each observation there is an opportunity to review the data to determine whether or not the trial is worth continuing. At any point in time, the drug's sponsor knows the cost of the next observation, but do the benefits of proceeding outweigh the cost, or should the trial be stopped early for futility? Continuing the trial allows the sponsor to make future decisions about whether to continue or abandon after subsequent observations. In other words, it gives the sponsor the right, but not the obligation, to invest in future observations and stages of drug development. From a financial perspective, this can be evaluated as a series of real options.

The standard tools used to value options originate from Black and Scholes (1973) and Merton (1973b). The term "real option" was first introduced to the literature by Myers (1977), who identified that many corporate assets can be viewed as call options. Option-pricing techniques have since been used to value a variety of managerial decisions including the option to expand, contract, defer or abandon a project across a range of applications from natural resource valuation to biopharmaceutical R&D (Brennan and Schwartz, 1985; Titman, 1985; McDonald and Siegel, 1985; Trigeorgis and Mason, 1987; Ingersoll Jr and Ross, 1992; Pindyck, 1993; Moel and Tufano,

2002; Schwartz, 2004; Gunther McGrath and Nerkar, 2004; Bogdan and Villiger, 2010; Lynch and Shockley, 2016).

In biopharmaceutical R&D there are two factors that can affect the value of a drug as it navigates the regulatory approval process. The first is called scientific risk, which captures the possibility that the drug will fail to meet its clinical endpoints. This risk is generally uncorrelated with macroeconomic factors, and is therefore considered idiosyncratic. The second is market risk, which occurs when the drug is likely to be technically successful, but the sponsor learns that its economic potential is more limited than initially projected. For example, early-stage R&D projects are traditionally monetized through licensing, joint development deals, and mergers and acquisitions with other biopharma companies. Therefore, even when a drug meets its clinical endpoints, these projects may face substantial systematic risk in the form of business- and credit-cycle downturns, when investor appetite for such deals are lower (Thakor et al., 2017). As a trial progresses, clinical observations and shifting market conditions provide information on both of these risks. In this chapter, we develop a systematic valuation approach that uses the information contained in these observations to determine whether it is economically beneficial to proceed with additional observations.

7.2 The Option Value of a Clinical Trial

Management decisions during the drug development process depend on multiple parameters that may change over time as new information is collected. As time passes, and more clinical observations are collected, the sponsor will begin to develop a clearer understanding of its economic potential. This clarity can manifest itself in a number of ways including through the likelihood the drug will be able to meet its endpoints. The closer we get to the end of a clinical trial, the better informed we will be about the economic potential of the drug. For example, imagine that a drug shows early signs of toxicity, reducing our forecasted sales and increasing the probability that it will not be approved for commercial development. After revaluing the project, we

may decide to abandon the drug because it is no longer profitable. To estimate a more accurate value of a project, we must model this risk directly.

Real option valuation techniques, such as the binomial options pricing model, attempt to model this risk by forecasting the possible trajectories the economic potential of a project can follow based on both scientific and market risks. Given an estimate of the current market potential, there is uncertainty about what that estimate will be one period from now. We can model the progress of this market estimate as new information is received using a recombining binomial lattice. For a fixed-sample clinical trial with N observations, the possible trajectories of the summary statistic on the primary endpoint, Θ_n, can be represented as sample paths in this framework. A new clinical observation of the effectiveness or safety of a new therapy can either be positive or negative, reflected by the left (L) and right (R) steps of the binomial lattice in Figure 7-1. Similarly, the state of the market, S_n, can independently improve (U) or deteriorate (D) with a certain probability.

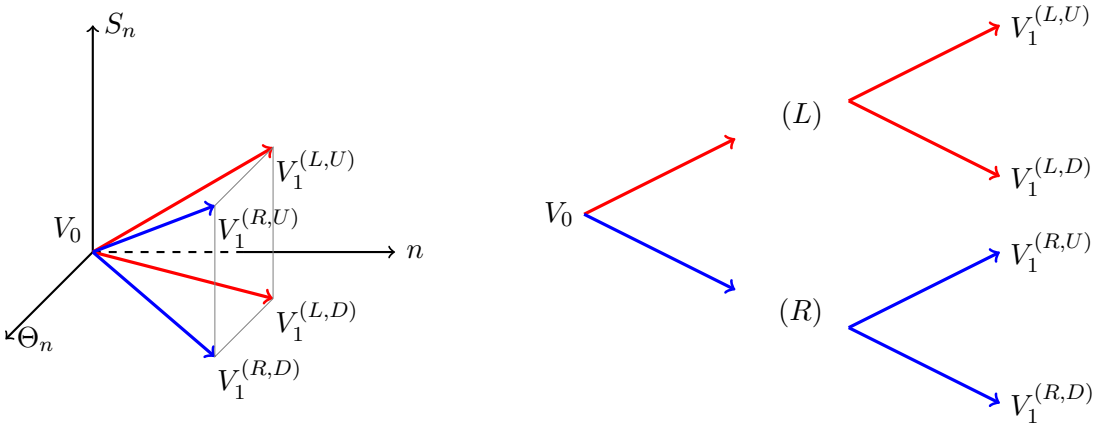

Figure 7-1: A single step of the binomial lattice. The scientific risk, Θ_n, can either step left (L) or right (R) depending on the outcome of the most recent clinical evidence, and the market risk, S_n, can either step up (U) or down (D) depending on evolving market conditions. The right panel separates these 2 independent sources of risk into a 2-step process.

The right panel of Figure 7-1 illustrates how this 2-dimensional risk can be visualized as a binomial tree. First the summary statistic of the clinical trial is updated, followed by a revision to the market conditions. Now suppose that at the start of the

period, the cost of the next clinical observation is K, and the sponsor has the option to continue or abandon the clinical. In this scenario, the value of drug at the start of the period is given by,

$$V_0 = \max\left(0,\ \frac{\mathrm{E}_0^Q[V_1]}{e^{r_f \Delta t}} - K\right) \quad , \tag{7.1}$$

where the risk-neutral probability measure is denoted by Q, the risk-free rate by r_f, and the length of the period by Δt. Since we have assumed the scientific risk is independent of the market risk, its risk-neutral probabilities are equal to its physical probabilities, and therefore (7.1) can be expressed as,

$$V_0 = \max\left(0,\ \frac{\mathrm{E}_0^P\left[\mathrm{E}_0^Q[V_1 \mid \Theta_1]\right]}{e^{r_f \Delta t}} - K\right) \quad , \tag{7.2}$$

where the physical probability measure is denoted by P. The calculation of a project's value then becomes straightforward. Once all possible scenarios with their corresponding cash flows and various risk-neutral probabilities have been defined, we multiply the future values by their probabilities to calculate the expected future value under the risk-neutral measure, and then discount this expected cash flow at the risk-free rate back one layer. After we subtract the cost of continuing the trial, if the value at any node is negative, we abandon the project, and set the value of the project at that node to \$0. In these states, the project value increases from a negative value to \$0. This difference in value is directly linked to the option to abandon.

The farther we proceed into the future, the less certain we are about the state of the market or the clinical trial, and the tree therefore branches out as we move forward in time (see Figure 7-2). The leaves of the tree represent the possible market states at the end of the clinical trial, which for pivotal phase 3 trials precedes product launch, and for early-stage trials precedes a later-stage trial. Working from the leaves of the tree backwards to its root, we can calculate the value of the drug at each node, one layer at a time. The present value of the project then corresponds to the value

calculated at the root node,

$$\begin{aligned} V_N &= \max\Big(0,\ f(\Theta_N, S_N) - I\Big) \\ V_{N-1} &= \max\left(0,\ \frac{\mathrm{E}^P_{N-1}\big[\mathrm{E}^Q_{N-1}[V_N \mid \Theta_N]\big]}{e^{r_f \Delta t}} - K\right) \\ &\vdots \\ V_0 &= \max\left(0,\ \frac{\mathrm{E}^P_0\big[\mathrm{E}^Q_0[V_1 \mid \Theta_1]\big]}{e^{r_f \Delta t}} - K\right) \ , \end{aligned} \tag{7.3}$$

where $f(\cdot)$ is a function that defines the economic value of a drug at the end of the clinical trial for a given realization of (Θ_n, S_n), and I is the investment required for the next stage of development.

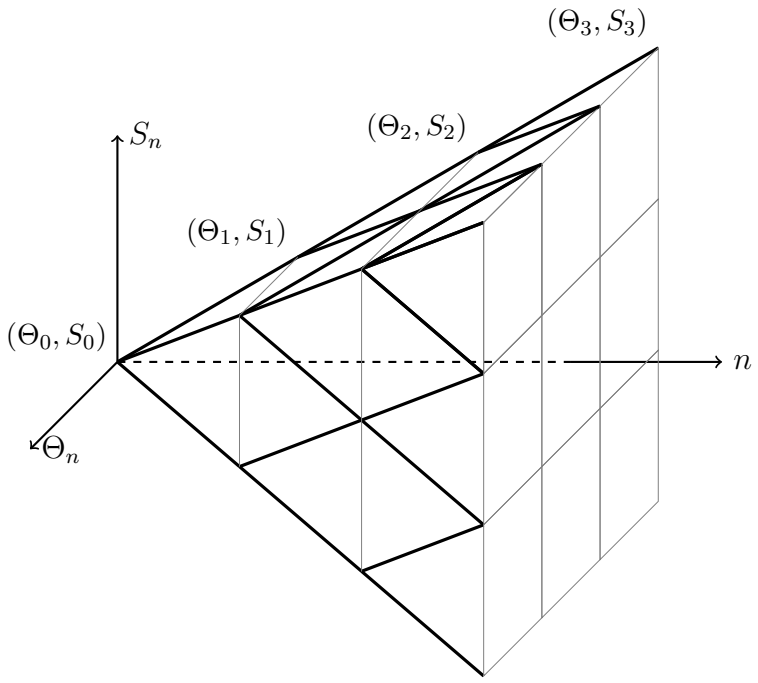

Figure 7-2: Binomial lattice model. The farther we proceed into the future, the less certain we are about the state of the market or the clinical trial, and the tree therefore branches out as we move forward in time.

The economic value of a drug upon commercial launch is often estimated by forecasting the drug's peak annual sales and sales curve over time given market conditions

and the drug's attributes. This uncertainty is eventually resolved with the progression of the drug through the clinical trial as its safety and effectiveness becomes known. In addition to these factors, future sales will be influenced by the growth rate of the target market, the marketing power of the company, other treatments available, the price elasticity for the disease, and so on. These cash flows must then be discounted at an appropriate cost of capital which is often estimated as the expected rate of return investors will demand from similarly risky projects. While the appropriate discount rate will depend on factors such as investor risk aversion and the correlation of the project's cash flows with other investments, in general, the cost of capital will be higher in the earlier stages of drug development relative to the period after a successful commercial launch.

7.3 Defining the Scientific and Market Risk Processes

To develop a better understanding of our valuation framework, consider a balanced two-arm RCT that uses the Z-statistic as a measure of its primary quantitative endpoint. As in the previous chapter, given an observation $Z_n = z_n$, we define the likelihood ratio of z_n, $\Lambda(z_n)$, as

$$\Lambda(z_n) = \frac{\Phi(Z_n = z_n \mid H = 0)}{\Phi(Z_n = z_n \mid H = 1)} \quad . \tag{7.4}$$

Taking the natural logarithm of the likelihood ratio function and using (6.1) to substitute in for Z_n, the log-likelihood function, $\ln \Lambda(Z_n) \equiv \Theta_n$, can be expressed as a random walk,

$$\Theta_n = \sum_{i=1}^{n} X_i \quad , \quad \text{where} \quad X_i = \frac{\delta^2 - 2\delta(T_i - P_i)}{4\sigma^2} \quad , \tag{7.5}$$

and Θ_n is defined to be the scientific risk process. The mean displacement between Θ_n and Θ_{n+1} under the null and alternative hypothesis is then given by,

$$\mu_X = \begin{cases} \dfrac{\delta^2}{4\sigma^2}\ , & H = 0 \\ \\ -\dfrac{\delta^2}{4\sigma^2}\ , & H = 1 \end{cases}\ , \tag{7.6}$$

and its variance is,

$$\sigma_X^2 = \frac{\delta^2}{2\sigma^2}\ . \tag{7.7}$$

The right (R) and left (L) additive factors for Θ_n in the binomial lattice can then be modeled as $\pm\sigma_X$, respectively, and their physical probabilities as,

$$p_R = 1 - p_L = \frac{\mu_x - L}{R - L}\ . \tag{7.8}$$

These factors represent clinical evidence of the effectiveness of the drug. Adding L for each leftward step, and R for each rightward step to Θ_n, models a possible evolution of the summary statistic of the clinical trial as observations get collected.[1]

Similarly, if we model the market risk process (S_n) as geometric brownian motion, the up (U) and down (D) multiplicative factors in the binomial lattice are given by $e^{\pm\sigma\sqrt{\Delta t}}$, respectively, where σ is the underlying market risk volatility. Their physical probabilities are then given by,

$$p_U = 1 - p_D = \frac{e^{r\Delta t} - D}{U - D}\ , \tag{7.9}$$

where r is the underlying cost of capital, and their risk-neutral probabilities are given

[1]Note that while we have considered normally distributed response variables for expositional purposes, as long as the observations are conditionally independent and identically distributed, the log-likelihood function will follow a random walk. By the central limit theorem, for large enough n (e.g., $n > 50$), we can approximate each X_i as $\mathcal{N}(\mu_X, \sigma_X^2)$. (This approximation is exact for normally distributed response variables.)

by,

$$q_U = 1 - q_D = \frac{e^{r_f \Delta t} - D}{U - D} \quad , \tag{7.10}$$

where r_f is the risk-free rate. Here, U and D represent changes to the economic potential of the drug due to external shocks in market conditions. The binomial lattice is then formed by multiplying S_n by U for each upward step, and by D for each downward step, which models the evolution of market conditions over the course of the clinical trial.

7.4 Optimal Decision Boundaries for Futility

Using the risk processes defined in the previous section, we can now design an optimal decision boundary that, when crossed, informs the sponsor that the clinical trial should be stopped early for futility. Setting $H = 1$ in (7.6), we can determine which nodes, (Θ_n, S_n), result in the project being abandoned in (7.3). This design choice mitigates the chance of a false early rejection when the drug is effective, yet preserves its ability to stop early when the drug is ineffective ($H = 0$).

The decision boundary formed by this methodology can be demonstrated by considering a numerical example. We consider a balanced two-arm RCT that uses the Z-statistic, Z_n in (6.1), as a measure of its primary quantitative endpoint. Traditionally, Z_n is compared to a critical value, λ, and the drug progresses to the next stage of development if $Z_n > \lambda$. The probability of approving a therapy given a treatment effect δ is therefore $\Phi(\delta\sqrt{I_n} - \lambda)$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function. Here, we assume $\lambda = 1.64$ for phase 2 and $\lambda = 1.96$ for phase 3, such that the probability of a false approval given the drug has no effect is 5% and 2.5%, respectively. A complete list of assumptions for this numerical example are provided in Table 7.1.

To visualize the early-stopping boundary defined by applying our valuation tech-

nique, note that Z_n is related to Θ_n by the following equation

$$Z_n = \frac{\delta\sqrt{I_n}}{2} - \frac{\Theta_n}{\delta\sqrt{I_n}} \quad . \tag{7.11}$$

Figure 7-3 plots this decision boundary as a function of the number of clinical observations, n, for this hypothetical example. Initially, when n is small, the decision boundary is very conservative. Common sense tells us that it is valuable to continue collecting data in this scenario because the small sample size restricts the quality of decision that can be made. Only if the drug has clearly underperformed does it make economic sense to stop the trial for futility at this early stage. As n increases, the decision boundary becomes less conservative because a drug that has performed poorly up to this point will have little chance of meeting its primary endpoint. Finally, as n approaches its target accrual of 276 patients for phase 2 and 1,052 patients for phase 3, the decision boundary approaches $Z_n = 1.64$ and $Z_n = 1.96$, respectively, in accordance with the approval criteria.

The second factor that affects the decision boundary over time is unanticipated shocks, either positive or negative, in economic conditions as modeled by the market volatility parameter. If market conditions deteriorate, then continuation of the trial at earlier stages becomes less valuable. In this scenario, the optimal decision is to use a less conservative threshold for futility, which results in abandoning the trial at an earlier stage. On the other hand, if market conditions improve, then the drug becomes more valuable, and the optimal decision is to collect more evidence before we can reject the hypothesis that the drug is effective.

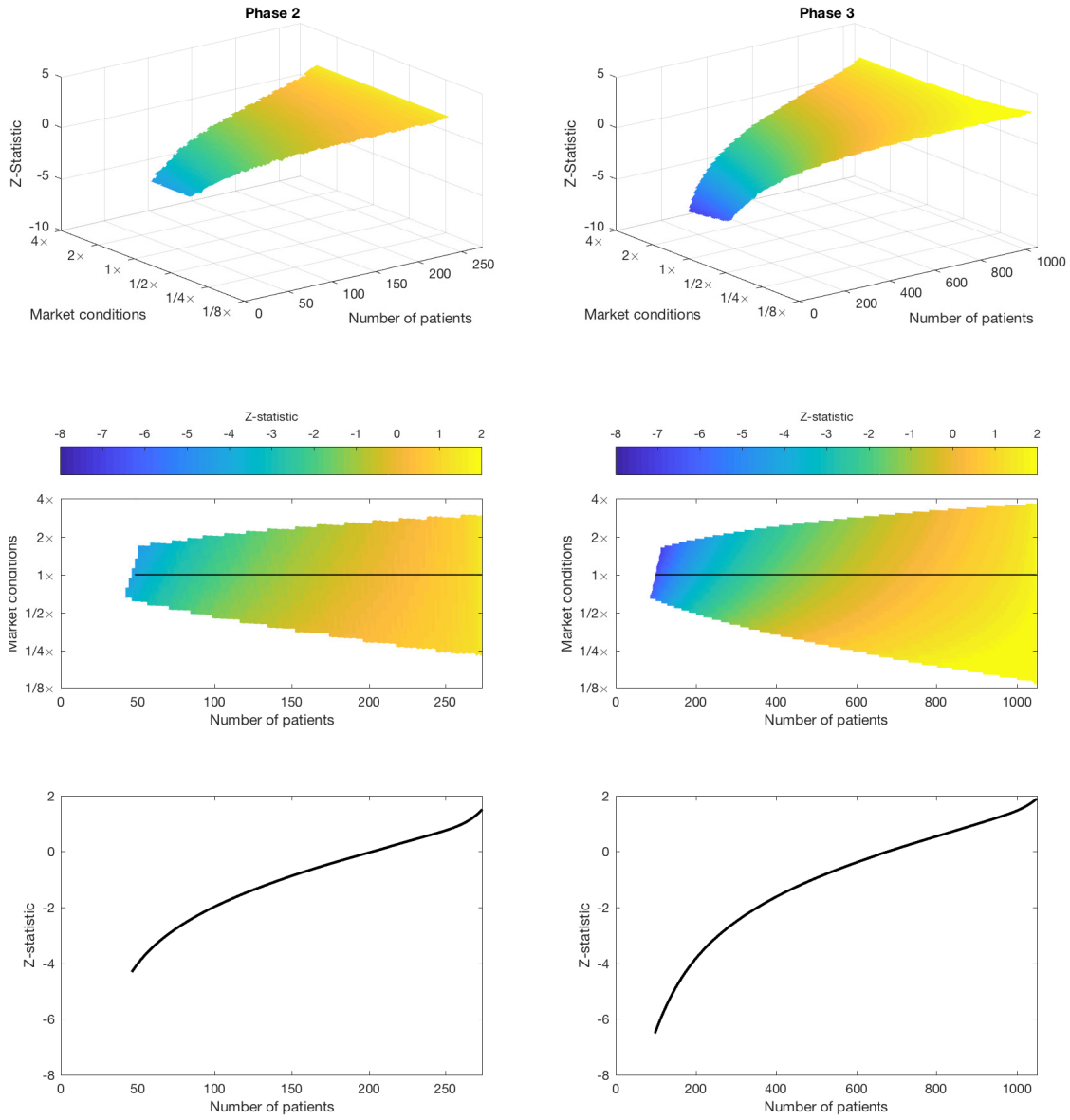

Figure 7-3: Financially-optimal futility boundaries. The left and right columns show the decision boundaries for the phase 2 and phase 3 clinical trials described by Table 7.1. The top row provides a 3-dimensional visualization of the decision boundary surface as a function of the market conditions and number of patients accrued to the study. The middle rows provide a top-down perspective of the decision boundary surface, and the bottom rows provide a cross-section of the surface given stable market conditions.

Parameter	Phase 2	Phase 3	Comments
Significance level (α)	5%	2.5%	Probability of a false approval under $H = 0$.
Statistical power ($1 - \beta$)	80%	90%	Probability of a correct approval under $H = 1$.
Standardized difference (δ/σ)	0.3	0.2	Average treatment effect under $H = 1$ in units of standard deviations of the response variable.
Target accrual ($2N$)	276	$1,052$	Total number of patients in the trial if run to completion.
Cost per patient ($K/2$)	$\$40,000$	$\$42,000$	The cost of clinical trials varies across disease groups and depends on multiple factors. On average, clinical trials have been estimated to cost $40,000 and $42,000 per patient for phase 2 and phase 3 trials, respectively Battelle Technology Partnership Practice (2015).
Trial length (T)	2 years	3 years	T/N defines the time between 2 observations, Δt, assuming uniform patient accrual.
Median annual sales	–	$300MM	The drug is forecasted to generate $300 million per year in sales if it meets its primary endpoint.
Net margin	–	20%	Percentage of revenues remaining as profit after all operating, interest, and tax expenses have been deducted from annual sales. In this case, the expected annual profit is $60 MM per year.
Years of exclusivity	–	13	Revenues from a successful therapy are expected to be generated for a 13-year period of exclusivity after FDA approval before patent expiration.
Launch costs	–	$50MM	launch-related investment during the year a new therapy enters the market.
Probability of success	58.3%	59.0%	Average estimates for the probability of a successful transition from phase 2 to phase 3, and phase 3 to approval across therapeutic areas Wong et al. (2018). These values are used to estimate the *a priori* probability of $H = 1$.
Annual market volatility	50%	50%	Market risk that affects the drug's economic potential independent of its clinical trial results (i.e., scientific risk). Factors include business- and credit-cycle risk, regulation, and competition.
Annualized cost of capital (r)	20%	10%	Conventional estimates for early and late-stage clinical trials.
Risk-free rate (r_f)	3%	3%	Annual yield on a US 10-year Treasury Note.

Table 7.1: Parameter assumptions for the phase 2 and phase 3 clinical trials.

7.5 Empirical Results

The financial value of these boundaries can be exhibited by applying our methodology to a set of common disease areas. In this section, we vary the cost per patient, median annual sales, and probability of a successful phase 3 trial to determine the optimal futility boundaries across disease groups. Table 7.2 reports the present values of a candidate drug for each disease group under both the null ($H = 0$) and alternative ($H = 1$) hypotheses for clinical trials with (adaptive) and without (non-adaptive) the option to end the trial early for futility.

These table entries show that disease groups with greater costs per patient, lower median sales, and lower *a priori* probabilities of success have relatively less conservative decision boundaries. When a drug is ineffective, this results in greater savings in terms of total expenditures, number of patients, and trial length. Here, managerial flexibility allows the sponsor to avoid significant clinical trial costs when the economic potential of the drug is learned to be poor. For example, under the null hypothesis, phase 3 clinical trials for oncology, which have the higher cost per patient, were stopped approximately 42.1% early, increasing the NPV for this scenario from −$67.5 million to −$35.7 million. In general, the average increase in NPV from using the futility boundaries was $3.3 million for phase 2, and $14.3 million for phase 3. Moreover, trials that used the optimal futility boundaries were on average 31% smaller in terms of total number of patients and overall trial length. Figure 7-4 illustrates these savings by phase and disease area.

On the other hand, if the drug is effective, the likelihood of crossing the early-stopping boundary decreases, and so the option to abandon the trial early for futility is not exercised. In this case, the option expires worthless, but at no extra cost to the sponsor, causing the NPV for both the adaptive and non-adaptive trial designs to converge. The option to abandon early therefore allows the sponsor to hedge the risk of downside scenarios, while maintaining their ability to fully extract the benefits of positive outcomes. It is this nonlinear payoff structure that makes this adaptive design so valuable.

Phase 2			Non-Adaptive				Adaptive					
			$H = 0$	$H = 1$			$H = 0$			$H = 1$		
Disease	Cost per patient ($M)	$POS_{3,APP}$ (%)	NPV ($MM)	NPV ($MM)	2N	T	NPV ($MM)	E[2N] (SD)	E[T] (SD)	NPV ($MM)	E[2N] (SD)	E[T] (SD)
Cardiovascular	25.0	62.2	−6.0	22.3	276	2	−3.4	181.0 (59.6)	1.31 (0.43)	22.4	268.7 (25.7)	1.95 (0.19)
Central Nervous System	39.5	51.1	−8.3	88.8	276	2	−5.0	197.4 (51.6)	1.43 (0.37)	89.0	270.8 (18.3)	1.96 (0.13)
Metabolic	18.5	51.6	−2.7	89.8	276	2	−1.4	207.2 (46.2)	1.50 (0.34)	89.9	271.8 (14.7)	1.97 (0.11)
Hematology	30.0	59.0	−6.2	70.9	276	2	−3.7	198.1 (51.2)	1.44 (0.37)	71.1	270.9 (18.0)	1.96 (0.13)
Infectious	17.5	75.3	−2.5	89.1	276	2	−1.2	207.8 (45.9)	1.51 (0.33)	89.1	271.8 (14.5)	1.97 (0.11)
Oncology	67.5	35.5	−17.2	19.2	276	2	−7.9	154.3 (70.9)	1.12 (0.51)	20.1	262.9 (42.6)	1.91 (0.31)
Respiratory	30.5	59.0	−7.0	39.7	276	2	−4.2	188.4 (56.1)	1.37 (0.41)	39.9	269.7 (22.1)	1.95 (0.16)

Phase 3			Non-Adaptive				Adaptive					
			$H = 0$	$H = 1$			$H = 0$			$H = 1$		
Disease	Cost per patient ($M)	Annual sales ($MM)	NPV ($MM)	NPV ($MM)	2N	T	NPV ($MM)	E[2N] (SD)	E[T] (SD)	NPV ($MM)	E[2N] (SD)	E[T] (SD)
Cardiovascular	26.0	145	−25.6	83.2	1052	3	−12.6	569.0 (211.0)	1.62 (0.60)	84.8	996.0 (148.1)	2.84 (0.42)
Central Nervous System	40.5	422	−38.3	343.5	1052	3	−22.9	682.3 (190.1)	1.95 (0.54)	344.2	1036.3 (67.0)	2.96 (0.19)
Metabolic	19.0	371	−17.0	314.1	1052	3	−10.5	717.2 (176.2)	2.05 (0.50)	314.3	1038.9 (55.3)	2.96 (0.16)
Hematology	31.0	302	−29.6	233.1	1052	3	−17.2	663.9 (193.1)	1.89 (0.55)	233.8	1031.2 (79.4)	2.94 (0.23)
Infectious	18.0	265	−16.8	209.3	1052	3	−10.1	686.6 (183.8)	1.96 (0.52)	209.7	1032.8 (71.5)	2.95 (0.20)
Oncology	69.0	344	−67.5	236.8	1052	3	−35.7	608.8 (212.3)	1.74 (0.61)	238.9	1023.7 (106.3)	2.92 (0.30)
Respiratory	31.0	213	−30.2	144.7	1052	3	−16.2	616.3 (203.9)	1.76 (0.58)	145.8	1017.9 (110.2)	2.90 (0.31)

Table 7.2: Phase 2 and phase 3 clinical trial statistics for selected disease areas. Abbreviations: $POS_{3,APP}$, probability of success of a phase 3 to approval transition; NPV, net present value; 2N, total number of patients in the trial; M, thousands; MM, millions; SD, standard deviation. Cost per patient estimates are from Battelle Technology Partnership Practice (2015), $POS_{3,APP}$ estimates are from Wong et al. (2018) where we have used the overall success rates for hematologic and respiratory diseases, and annual sales estimates are from Bogdan and Villiger (2010).

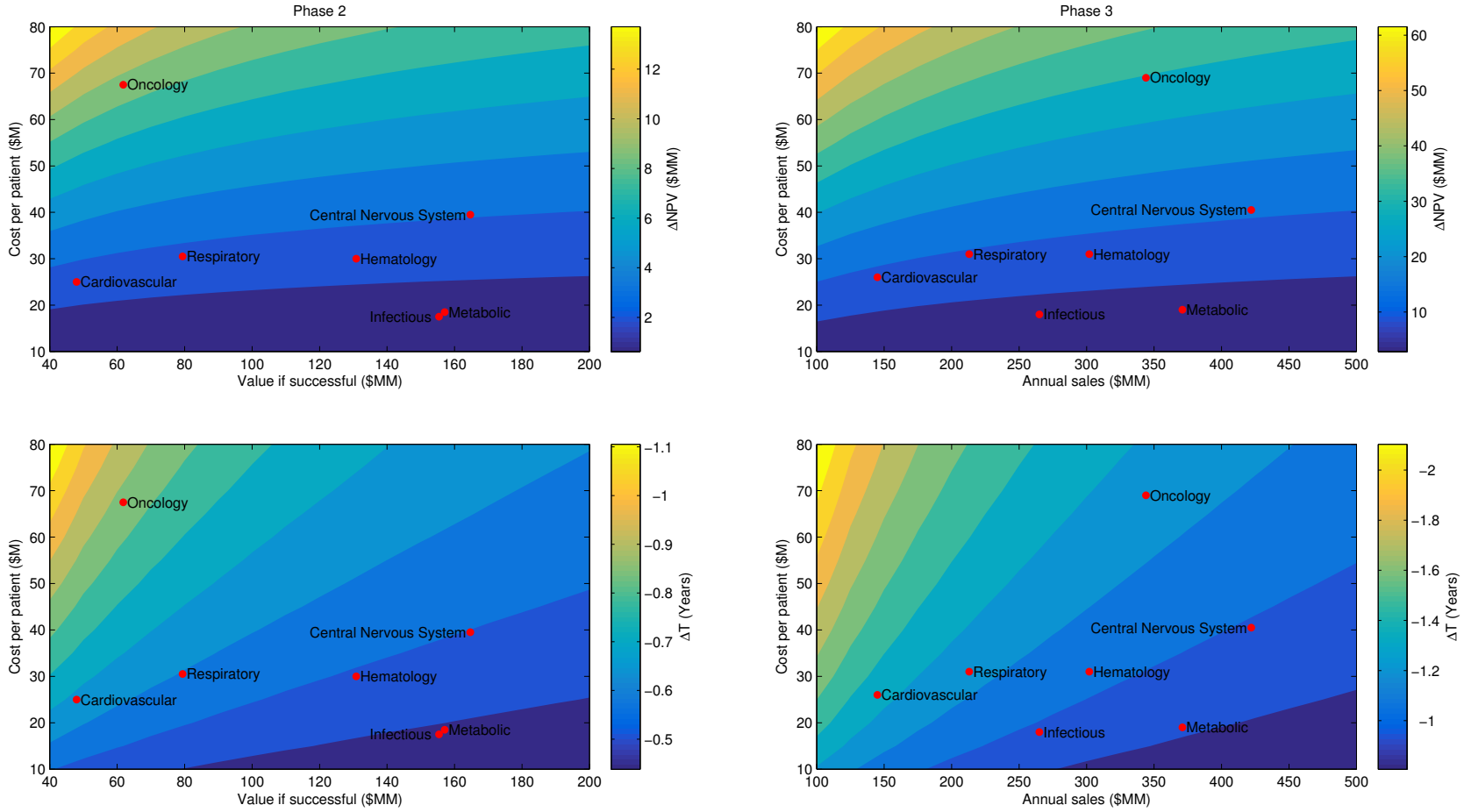

Figure 7-4: Increases in value (ΔNPV) and decreases in clinical trial length (ΔT) from using the optimal futility boundaries in the case of an ineffective drug ($H = 0$).

The previous analysis assumed that the cost of capital for both trials with and without the option to end the trial early for futility were the same. However, as described by others Myers and Howe (1997); Thakor et al. (2017), because of financial leverage, maintaining a large, fixed commitment to biopharmaceutical R&D is difficult in the face of business- and credit-cycle downturns. By providing sponsors and investors with more frequent and systematic "exit options" to cut their losses in the face of a market downturn, the systematic risk component of their investment will be reduced. Moreover, these adaptive clinical trials can be funded according to more granular milestones reducing the amount of financial leverage inherent to the project. Since biotechnology firms have a greater clinical trial costs relative to their size, a reduction in the leverage effect should be more beneficial for them, thus reducing their exposure to systematic risk and consequently their cost of capital.

Figure 7-5 provides a sensitivity analysis that investigates the effect of a change in the cost of capital. We find that if the cost of capital of an early-phase trial were reduced by half, from 20% per year to 10% per year (similar to the cost of capital of late-stage trials), then the value of an effective therapy increases on average by \$15.2 million. Similarly, a 5% per year reduction in the cost of capital from 10% to 5% increases the average value of a phase 3 clinical trial under $H = 1$ by \$152.8 million. Moreover, the added value is most prominent for disease groups that have the greatest economic potential given a successful therapy. This acute sensitivity suggests that even modest reductions in the systematic risk component faced by investors can have substantial benefits in terms of larger clinical trial valuations and increased funding for biomedical R&D.

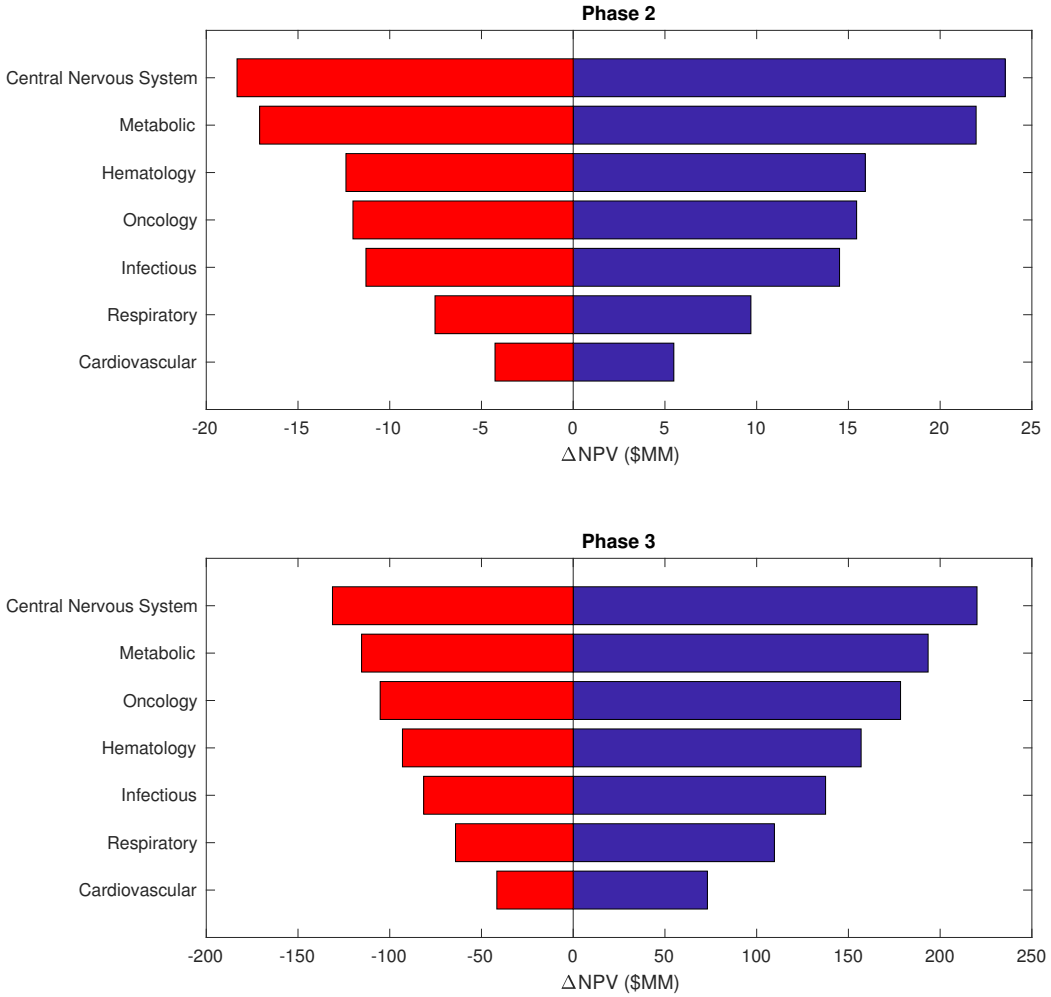

Figure 7-5: Sensitivity of the value (ΔNPV) of an effective drug ($H = 1$) to the cost of capital for selected disease areas. From the left to right of each bar, the three number summary corresponds to [1.5, 1, 0.5] times the cost of capital proposed in Table 7.1.

7.6 Discussion

Our empirical results show that fixed-sample RCTs without the option to abandon early fail to maximize the economic value of candidate drug therapies. Often this is because of missed opportunities to stop the trial early when clinical evidence suggests lackluster future market prospects. In contrast, adaptive clinical trials that take advantage of this optionality mitigate downside risk and result in an overall increase in value to the sponsor and investors.

The framework that we have described is a simplified version of reality. The resolution of a drug's economic potential depends on multiple factors including the trajectory of disease incidence that the drug is intended to treat, the rate of population growth over time, income growth, reimbursement rates, and so forth. For example, imagine that a competing drug shows outstanding clinical results, reducing our forecasted sales. After revaluing the project, we may decide to abandon the drug because it is no longer profitable. In our empirical results, these factors have been modeled as a general brownian motion process, but to estimate a more accurate value of a project, we should model these other factors that affect the market risk process directly. Moreover, we could expand the leaves of the binomial pricing model's tree to include a wider range of economic potential at launch, perhaps with a blockbuster scenario in addition to other, more refined, intermediate outcomes that may not follow a log-normal distribution. In general, clinical trials can also involve complicated fee structures with various kinds of triggers for contract clauses. Therefore our state-space representation can take on many forms, since each trial has its own peculiarities and unique scenarios.

Since R&D programs and clinical trials are complex and uncertain, modeling them can quickly become unmanageable. If we are not careful, however, the added complexity can convolute our analysis to the point where they are no longer useful to guide decision-making. Therefore, models must be pruned to the point where they show us the most important links between present and future decisions. Our primary goal in this paper was not to develop a detailed representation of the regulatory

approval process, but rather to demonstrate how a clinical trial with both scientific and market risk can be valued as a series of real options. Nevertheless, while we have focused on a binomial recombining lattice representation, the principles behind our valuation technique can easily be extended to more general distributions that provide more realistic assumptions about the nature of the uncertainty. In these cases, simulations and sophisticated numerical analysis will be required to estimate early-stopping decision boundaries.

7.7 Conclusion

Clinical trials with financially-optimal futility boundaries exhibit clear economic advantages over their fixed-sample, non-adaptive counterparts. In particular, the traditional fixed-sample trials are inflexible, resulting in missed opportunities to stop the trial early for futility. Conversely, financially adaptive trials add economic value by conditionally funding future stages of a trial only when a drug shows commercial potential. The ability of our framework to systematically design decision boundaries that inform the sponsor when to stop a trial early for futility make it a potentially valuable tool for capital budgeting. While our framework can be generalized, we emphasize that careful consideration must be applied to the assumptions underlying the specific models in order to produce useful recommendations.

Part IV

Conclusion

Chapter 8

Summary of Findings

This thesis approached the statistical analysis of stochastic processes in capital markets and regulatory decision-making in healthcare through the lens of signal processing. We explored how information could be extracted from the evolution of signals such as price and clinical observations, and moreover, how systems in these areas can be designed to make efficient use of this information.

The main body of the thesis was separated into two parts. Part II focused on the role of the frequency domain in investing. We presented the main mathematical results of the power spectrum in Chapter 2, which formed the basis of our applications to portfolio theory. In Chapter 3, we applied spectral analysis to develop a dynamic measure of alpha that allows us to determine whether portfolio managers are capturing alpha and over what time horizons their investment processes have forecast power. In this context, an investment process was said to be profitable at a given frequency if there was positive correlation between portfolio weights and returns at that frequency. When aggregated across frequencies, the dynamic alpha was shown to be equivalent to Lo's (2008) Active/Passive decomposition, and provided a clear indication of a manager's forecast power and, consequently, active investment skill.

In Chapter 4, we extended the frequency domain representations of auto- and cross-covariances to a financial measure of systematic risk, beta. Applying Engle's (1974) band-spectrum regression, we demonstrated that the beta of an asset can be expressed as a linear combination of frequency-specific betas, i.e., betas on components

of the asset returns and components of the factor operating at different frequencies. When applied to NYSE and AMEX stock returns from 1972 to 2016, we found that the inclusion of the frequency domain can be used to estimate spectral factor models that select portfolios with significantly lower out-of-sample variance relative to estimators based on traditional multi-factor models. These frequency-specific measures were shown to distinguish between short- and long-term components risk and covariances, providing additional insights into portfolio and risk management above and beyond their static counterparts.

Part III focused on efficient designs for randomized clinical trials (RCTs) in the development and approval process of new therapies. In Chapter 5, we applied the Bayesian decision analysis (BDA) framework of Montazerhodjat et al. (2017) and formed a quantitative model in which patients' preferences are the center of RCT design. We quantified the loss in value to public health associated with different actions in any fixed-sample RCT, used a BDA framework to aggregate the value of the trial, and then determined an optimal RCT in which the expected value to patients is maximized. We tailored this framework to weight-loss devices, and our results demonstrated that traditional RCT designs with a fixed statistical significance level do not necessarily maximize overall value (or equivalently, minimize harm) to current and future patients. For low-risk devices and risk-tolerant populations, the inefficiency was mainly caused by lengthy RCTs that were too conservative and overprotective of the type I error rate. Conversely, for some high-risk devices, such as those that require open surgery, traditional one-sided significance levels of 5% were more permissive than the BDA-optimal thresholds.

Chapter 6 extended this framework to Bayesian adaptive clinical trials through the use of sequential likelihood ratio tests. We found that adaptive trials exhibited clear advantages over their fixed-sample counterparts in terms of cost, speed, and potential impact on current and future patients. We concluded that although such processes are inherently more complex than traditional fixed-sample clinical trials, the added complexity may well be worthwhile when weighed against the savings in clinical trial costs, the number of patients exposed to potentially toxic therapies, and

the time to approval of effective therapies.

Chapters 5 and 6 focused on patient value, and did not include a consideration of a clinical trial's economic costs. In Chapter 7, we modeled a clinical trial as a series of real options, which allowed us to systematically design early-stopping decision boundaries that maximize the economic value to the industry sponsor. Our numerical simulations demonstrated that these financially adaptive clinical trials exhibit clear economic advantages over their fixed-sample, non-adaptive counterparts. In particular, fixed-sample trials were inflexible, resulting in missed opportunities to stop the trial early for futility. Conversely, financially adaptive trials added economic value by conditionally funding future stages of a trial only when a drug showed commercial potential. The ability of our framework to systematically design decision boundaries that inform the sponsor when to stop a trial early for futility make it a potentially valuable tool for capital budgeting.

Part V

Appendices

Appendix A

Chapter 2 Supplement

In this appendix, we derive statistical properties of the main estimators in Chapter 2 that are required for conducting standard inferences such as hypothesis tests and significance-level calculations.

A.1 General Moment Properties of the Power Spectrum

Consider the real-valued discrete-time wide-sense stationary stochastic processes $\{x_t\}$ and $\{y_t\}$ with means m_x and m_y, and cross-covariance function $\gamma_{xy}[m] = \mathrm{E}[(x_{t+m} - m_x)(y_t - m_y)]$. Assuming the cross-covariance function has finite energy, let $P_{xy}(\omega)$ be its Discrete-Time Fourier Transform (DTFT) such that,

$$P_{xy}(\omega) = \sum_{m=-\infty}^{\infty} \gamma_{xy}[m] e^{-j\omega m}, \tag{A.1}$$

$$\gamma_{xy}[m] = \frac{1}{2\pi} \int_0^{2\pi} P_{xy}(\omega) e^{j\omega m} \mathrm{d}\omega. \tag{A.2}$$

The function $P_{xy}(\omega)$ is known as the cross spectrum, and can be interpreted as the frequency distribution of the power contained in the covariance between x_t and y_t. A rectangular window of length T can be used to select a finite-length subsample of x_t and y_t. Forming the cross spectrum estimate from the DFT of this finite subsample

we find that $\mathrm{E}[C_{xy}[k]]$ is not generally equal to $P_{xy}(\omega_k)$, where $\omega_k = 2\pi k/T$, and is therefore a biased estimator. The bias results from the convolution of the true power spectrum, $P_{xy}(\omega)$, with the DTFT of the aperiodic autocorrelation function of the window. As the window length increases, its DTFT approaches a Dirac delta function, and so the bias approaches 0. Thus, $\mathrm{E}[C_{xy}[k]]$ is an asymptotically unbiased estimator of $P_{xy}(\omega_k)$ (Oppenheim and Schafer, 2009). Moreover, over a wide range of conditions, it can be shown that,

$$\mathrm{Var}[L_{xy}[k]] \approx \frac{1}{2}(P_{xx}(\omega_k)P_{yy}(\omega_k) + \Lambda_{xy}^2(\omega_k) - \Psi_{xy}^2(\omega_k)), \tag{A.3}$$

where $\Lambda_{xy}(\omega)$ and $\Psi_{xy}(\omega)$ are the theoretical co-spectrum and quadrature spectrum between x_t and y_t, respectively. At the harmonic frequencies, which are separated in frequency by $1/T$, these frequency-specific estimators of the co-spectrum are approximately uncorrelated (Jenkins and Watts, 1968). This property can be used to estimate the variance of the sum of co-spectrum estimators, $L_{xy}[k]$.

A few important implementation details still remain. Notice that the variance of the co-spectrum estimates are not consistent as they do not asymptotically approach 0 as T increases. Averaging the co-spectrum estimates calculated over overlapping time intervals can reduce the variance of the spectral estimates at the expense of introducing bias. In addition, windowing procedures (e.g., multiplication by a Hamming window) can be applied to the data before calculating the DFT. This procedure will generally decrease spectral leakage at the expense of reducing spectral resolution. An estimate of the co-spectrum can also be calculated from the Fourier transform of the estimated cross-covariance function. Finally, if x_t and y_t are sampled at a low frequency relative to the rate at which their properties change, then the decomposition will be biased due to a phenomenon known as aliasing. See Oppenheim and Schafer (2009) and Jenkins and Watts (1968) for a more detailed discussion of these advanced implementation techniques.

Appendix B

Chapter 4 Supplement

This Appendix summarizes the relevant band-spectrum regression properties from Engle (1974).

B.1 Standard Errors and F-Tests for Spectral Betas

If we define $\tilde{y}$ to be a $T \times 1$ column vector, and $\tilde{x}$ to be a $T \times M$ matrix, where the kth rows are the DFT coefficients of Y_k and $[X_{1,k}, \ldots, X_{M,k}]$, respectively, then an M-factor band-spectrum regression can be specified as:

$$A\tilde{y} = A\tilde{x}\beta + A\tilde{\varepsilon}, \tag{B.1}$$

where A is a $T \times T$ matrix with ones on the diagonals corresponding to included frequencies and zeros elsewhere. Using this form, (4.3) can be rewritten as:

$$\beta_K = \big((A\tilde{x})^\dagger (A\tilde{x})\big)^{-1} \big((A\tilde{x})^\dagger (A\tilde{y})\big) \tag{B.2}$$

$$\mathrm{Var}(\beta_K) = \big((A\tilde{x})^\dagger (A\tilde{x})\big)^{-1} \sigma^2 \ , \tag{B.3}$$

where "†" denotes the conjugate transpose, and an estimator of σ^2 is given by:

$$A\tilde{u} = A\tilde{y} - A\tilde{x}\beta_K, \tag{B.4}$$

$$\hat{\sigma}^2 = \frac{(A\tilde{u})^\dagger (A\tilde{u})}{T_K - M} \ . \tag{B.5}$$

From these equations, the standard errors of β_K can be estimated. If a regression model that forces the β's to fit all frequencies (the restricted model) holds, then a regression model that allows the β's to differ across frequency bands (the unrestricted model) will be relatively inefficient. In this case, Fisher's F-test can be used to determine if the unrestricted model yields a significantly better fit. Assuming the same T' frequencies are used for both models, and letting $\tilde{u}$ and $\tilde{v}$ be the unrestricted model's and restricted model's residuals, respectively, the F-statistic is given by,

$$F = \frac{\left(\frac{v^\dagger v - u^\dagger u}{n_u - n_v}\right)}{\left(\frac{u^\dagger u}{T' - n_u}\right)}, \tag{B.6}$$

where the unrestricted model has n_u parameters, and the restricted model has n_v parameters. Under the null hypothesis that the unrestricted model does not provide a significantly better fit than the restricted model, F will have an F distribution with $(n_u - n_v, T' - n_u)$ degrees of freedom.

Appendix C

Chapter 5 Supplement

In this Appendix, we investigate the robustness of our results in Chapter 5 to the parameter assumptions in our model.

C.1 Sensitivity Analysis

We calculate the optimal balanced two-arm fixed sample RCT for each of the 4 device types in our study as we vary the power constraint ($\text{Power}_{\max}$), the time horizon discount rate (r), and the control device efficacy (μ_p). The optimal α and sample size values associated with the perturbed parameters are given in Tables C.1, C.2, and C.3.

In Table C.1, we find that both the trial size and α decrease as the power constraint decreases. By allowing a larger type II error rate, the optimization can afford to decrease both the trial length and the type I error rate. On the other hand, in Table C.2, the wait time induced by the regulatory-approval process becomes less harmful as the discount rate decreases, and so the optimization can afford to increase the sample size to reduce both the type I and type II error rates. Furthermore, as the treatment effect of the control device increases in C.3, it becomes relatively more difficult to detect a difference between the control and investigational devices, especially for the low efficacy devices. As such, the sample size must increase to maintain accuracy. Finally, in addition to providing specific BDA-optimal RCT recommendations,

this sensitivity analysis highlights the need for carefully considered assumptions and accurately calibrated preference models when implementing the BDA-framework.

Device characteristics under $H = 1$	Constraint ($\text{Power}_{\max}$)	Trial size ($2n$)	Critical value (λ_α)	Significance (α)	Power ($1 - \beta$)
Low risk, high %TBWL	100%	78	1.11	13.5%	98.1%
	90%	54	1.36	9.0%	90%
	80%	44	1.54	6.5%	80%
Low risk, low %TBWL	100%	338	1.72	4.4%	89.0%
	90%	338	1.72	4.4%	89.0%
	80%	278	1.83	3.5%	80%
High risk, high %TBWL	100%	124	2.05	2.1%	97.4%
	90%	92	2.16	1.7%	90%
	80%	78	2.33	1.1%	80%
High risk, low %TBWL	100%	0	–	–	–
	90%	0	–	–	–
	80%	0	–	–	–

Table C.1: BDA-optimal RCTs for weight loss devices for decreasing power constraints.

Device characteristics under $H = 1$	Discount rate (r)	Trial size ($2n$)	Critical value (λ_α)	Significance (α)	Power ($1 - \beta$)
Low risk, high %TBWL	20%	30	1.13	13.4%	80%
	15%	36	1.32	9.8%	80%
	10%	44	1.54	6.5%	80%
	5%	56	1.85	3.5%	80%
	1%	88	2.53	0.7%	80%
Low risk, low %TBWL	20%	220	1.53	6.4%	80%
	15%	244	1.66	4.9%	80%
	10%	278	1.83	3.5%	80%
	5%	338	2.10	1.8%	80%
	1%	472	2.63	0.4%	80%
High risk, high %TBWL	20%	64	2.03	2.3%	80%
	15%	68	2.12	1.9%	80%
	10%	78	2.33	1.1%	80%
	5%	88	2.53	0.7%	80%
	1%	118	3.06	0.1%	80%
High risk, low %TBWL	20%	0	–	–	–
	15%	0	–	–	–
	10%	0	–	–	–
	5%	0	–	–	–
	1%	0	–	–	–

Table C.2: BDA-optimal RCTs for weight loss devices for decreasing discount rates.

Device characteristics under $H = 1$	Control %TBWL (μ_p)	Trial size ($2n$)	Critical value (λ_α)	Significance (α)	Power ($1 - \beta$)
Low risk, high %TBWL	0%	38	1.62	5.7%	80%
	2%	44	1.54	6.5%	80%
	4%	50	1.42	8.1%	80%
Low risk, low %TBWL	0%	180	1.84	3.4%	80%
	2%	278	1.83	3.5%	80%
	4%	508	1.86	3.2%	80%
High risk, high %TBWL	0%	64	2.35	1.1%	80%
	2%	78	2.33	1.1%	80%
	4%	92	2.22	1.4%	80%
High risk, low %TBWL	0%	0	–	–	–
	2%	0	–	–	–
	4%	0	–	–	–

Table C.3: BDA-optimal RCTs for weight loss devices for increasing control device efficacy.

Bibliography

Anderson, T. 1971. *The Statistical Analysis of Time Series*. New York, NY: John Wiley & Sons, Inc.

Anscombe, F. 1963. Sequential medical trials. *J Am Stat Assoc* 58:365–383.

Bandi, F., B. Perron, A. Tamoni, and C. Tebaldi. 2017. The scale of predictability. Accessed on May 16, 2018 at `https://ssrn.com/abstract=2184260`.

Battelle Technology Partnership Practice. 2015. Biopharmaceutical Industry-Sponsored Clinical Trials: Impact on State Economies. Accessed January 25, 2018 at `http://phrma-docs.phrma.org/sites/default/files/pdf/biopharmaceutical-industry-sponsored-clinical-trials-impact-on-state-economies.pdf`.

Baxter, M., and R. King. 1999. Measuring business cycles: Approximate band-pass filters for economic time series. *Review of Economics and Statistics* 81:575–593.

Berry, D. 1987. Interim Analysis in Clinical Trials: The Role of the Likelihood Principle. *Am Stat* 41:117–122.

Berry, D. 2004. Bayesian statistics and the efficiency and ethics of clinical trials. *Stat Sci* 19:175–187.

Berry, D. 2006. Bayesian clinical trials. *Nat Rev Drug Discov* 5:27–36.

Berry, D. 2010. Adaptive clinical trials: the promise and the caution. *J Clin Oncol* 29:606–609.

Berry, D. 2015. The Brave New World of clinical cancer research: Adaptive biomarker driven trials integrating clinical practice with clinical research. *Mol Oncol* 9:951–959.

Berry, D., and S. Eick. 1995. Adaptive assignment versus balanced randomization in clinical trials: a decision analysis. *Stat Med* 14:231–246.

Beveridge, S., and C. R. Nelson. 1981. A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle'. *Journal of Monetary economics* 7:151–174.

Bhatt, D., and C. Mehta. 2016. Adaptive designs for clinical trials. *N Engl J Med* 375:65–74.

Black, F., and M. Scholes. 1973. The pricing of options and corporate liabilities. *Journal of political economy* 81:637–654.

Bogdan, B., and R. Villiger. 2010. *Valuation in life sciences.* Springer.

Bonangelino, P., T. Irony, S. Liang, X. Li, V. Mukhi, S. Ruan, Y. Xu, X. Yang, and C. Wang. 2011. Bayesian approaches in medical device clinical trials: a discussion with examples in the regulatory setting. *Journal of biopharmaceutical statistics* 21:938–953.

Breitung, J., and B. Candelon. 2006. Testing for short- and long-run causality: A frequency-domain approach. *Journal of Econometrics* 132:363–378.

Brennan, M. J., and E. S. Schwartz. 1985. Evaluating natural resource investments. *Journal of business* pp. 135–157.

Briggs, W., and V. Henson. 1995. *The DFT: An Owners' Manual for the Discrete Fourier Transform.* Philadelphia, PA: Society for Industrial and Applied Mathematics.

Brillinger, D. 2001. *Time Series: Data Analysis and Theory.* Philadelphia, PA: Society for Industrial and Applied Mathematics.

Brockwell, P., and R. Davis. 1991. *Time Series: Theory and Methods.* 2nd ed. New York: Springer-Verlag.

Carr, P., and D. Madan. 1999. Option valuation using the fast Fourier transform. *Journal of Computational Finance* 2:61–73.

Chaudhuri, S., and A. Lo. 2015. Spectral analysis of stock-return volatility, correlation, and beta. *IEEE Signal Processing and Signal Processing Education Workshop (SP/SPE)* pp. 232–236.

Cheng, Y., F. Su, and D. Berry. 2003. Choosing sample size for a clinical trial using decision analysis. *Biometrika* 90:923–936.

Colton, T. 1963. A model for selecting one of two medical treatments. *J Am Stat Assoc* 58:388–400.

Croux, C., M. Forni, and L. Reichlin. 2001. A measure of comovement for economic variables: Theory and empirics. *Review of Economics and Statistics* 83:232–241.

Crowley, P. 2007. A guide to wavelets for economists. *Journal of Economic Surveys* 21:207–267.

Dew-Becker, I., and S. Giglio. 2016. Asset pricing in the frequency domain: theory and empirics. *The Review of Financial Studies* 29:2029–2068.

Engle, R. 1974. Band Spectrum Regression. *International Economic Review* 15:1–11.

Frazzini, A., D. Kabiller, and L. Pedersen. 2013. Buffett's alpha. Tech. rep., National Bureau of Economic Research.

Freedman, B. 1987. Equipoise and the ethics of clinical research. *New England Journal of Medicine* 317:141–145.

Gallager, R. 2014. *Stochastic Processes: Theory for Applications.* Cambridge University Press.

Granger, C., and R. Engle. 1983. Applications of Spectral Analysis in Econometrics. *Handbook of Statistics* 3:93–109.

Granger, C., and M. Hatanaka. 1964. *Spectral Analysis of Economic Time Series*. Princeton, NJ: Princeton University Press.

Grieve, A. 2015. How to test hypotheses if you must. *Pharm Stat* 14:139–150.

Gunther McGrath, R., and A. Nerkar. 2004. Real options reasoning and a new look at the R&D investment strategies of pharmaceutical firms. *Strategic Management Journal* 25:1–21.

Hannan, E. 1970. *Multiple Time Series*. New York, NY: John Wiley & Sons, Inc.

Harrington, D., and G. Parmagiani. 2016. I-SPY 2-A Glimpse of the Future of Phase 2 Drug Development? *New England Journal of Medicine* 375:7–9.

Hasbrouck, J., and G. Sofianos. 1993. The trades of market makers: An empirical analysis of NYSE specialists. *The Journal of Finance* 48:1565–1593.

Ho, M. P., J. M. Gonzalez, H. P. Lerner, C. Y. Neuland, J. M. Whang, M. McMurry-Heath, A. B. Hauber, and T. Irony. 2015. Incorporating patient-preference evidence into regulatory decision making. *Surgical endoscopy* 29:2984–2993.

Huang, N., M. Wu, W. Qu, S. Long, S. Shen, and J. Zhang. 2003. Applications of Hilbert-Huang transform to non-stationary financial time series analysis. *Applied Stochastic Models in Business and Industry* 19:245–268.

Ingersoll Jr, J. E., and S. A. Ross. 1992. Waiting to invest: Investment and uncertainty. *Journal of Business* pp. 1–29.

Irony, T. Z. 2007. Evolving methods: Evaluating medical device interventions in a rapid state of flux. In L. Olsen, D. Aisner, and J. McGinnis (eds.), *The learning healthcare system: Workshop summary*, pp. 93–99. The National Academies Press.

Irony, T. Z. 2012. Statistical methods in clinical trials. In F. Faltin, R. Kennet, and F. Ruggeri (eds.), *Statistical methods in healthcare*, pp. 22–54. Wiley.

Irony, T. Z., and R. Simon. 2006. Applications of Bayesian methods to medical device trials. In K. Becker and J. Whyte (eds.), *Clinical Evaluation of Medical Devices, Principles and Case Studies*, pp. 99–116. Humana Press.

Isakov, L., V. Montazerhodjat, and A. Lo. 2018. Is the FDA too conservative or too aggressive?: A Bayesian decision analysis of clinical trial design. *Journal of Econometrics* p. to appear.

Jenkins, G. M., and D. G. Watts. 1968. *Spectral Analysis and Its Applications*. San Francisco, CA: Holden-Day.

Jennison, C., and B. Turnbull. 1999. *Group Sequential Methods with Applications to Clinical Trials*. Chapman and Hall.

Jensen, M. 1968. The performance of mutual funds in the period 1945–1964. *The Journal of Finance* 23:389–416.

Khandani, A., and A. Lo. 2007. What happened to the quants in August 2007? *Journal of Investment Management* 5:5–54.

King, R., and M. Watson. 1996. Money, prices, interest rates and the business cycle. *Review of Economics and Statistics* 78:35–53.

Ledoit, O., and M. Wolf. 2003. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of empirical finance* 10:603–621.

Lintner, J. 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics* 47:13–37.

Lo, A. 2008. Where do alphas come from?: A new measure of the value of active investment management. *Journal of Investment Management* 6:1–29.

Lo, A., and A. MacKinlay. 1990. When are contrarian profits due to stock market overreaction? *Review of Financial Studies* 3:175–205.

Lynch, J., and R. Shockley. 2016. Valuation of a Developmental Drug as a Real Option. *Journal of Applied Corporate Finance* 28:118–126.

Mallat, S. G. 1989. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE transactions on pattern analysis and machine intelligence* 11:674–693.

McDonald, R. L., and D. R. Siegel. 1985. Investment and the valuation of firms when there is an option to shut down. *International economic review* pp. 331–349.

Merton, R. 1969. Lifetime Portfolio Selection under Uncertainty: The Continuous Time Case. *Review of Economics and Statistics* 51:247–257.

Merton, R. 1973a. An Intertemporal Capital Asset Pricing Model. *Econometrica* 41:867–887.

Merton, R. C. 1973b. Theory of rational option pricing. *The Bell Journal of economics and management science* pp. 141–183.

Moel, A., and P. Tufano. 2002. When are real options exercised? An empirical study of mine closings. *The Review of Financial Studies* 15:35–64.

Montazerhodjat, V., S. Chaudhuri, D. Sargent, and A. Lo. 2017. Use of Bayesian Decision Analysis to Minimize Harm in Patient-Centered Randomized Clinical Trials in Oncology. *JAMA Oncology* .

Myers, S. C. 1977. Determinants of corporate borrowing. *Journal of financial economics* 5:147–175.

Myers, S. C., and C. D. Howe. 1997. *A Life-cycle Financial Model of Pharmaceutical R & D.* Program on the Pharmaceutical Industry, Sloan School of Management, Massachusetts Institute of Technology.

National Biomarker Development Alliance. 2017. GBM AGILE. Accessed September 7, 2017 at `http://nbdabiomarkers.org/gbm-agile`.

Oppenheim, A., and R. Schafer. 2009. *Discrete-Time Signal Processing.* 3rd ed. Upper Saddle River, NJ: Prentice Hall.

Phillips, P., Y. Sun, and S. Jin. 2006. Spectral Density Estimation and Robust Hypothesis Testing Using Steep Origin Kernels without Truncation. *International Economic Review* 47:837–894.

Pindyck, R. S. 1993. Investments of uncertain cost. *Journal of financial Economics* 34:53–76.

Priestly, M. 1981. *Spectral Analysis and Time Series.* London-New York: Academic Press.

Ramsey, J. 2002. Wavelets in economics and finance: Past and future. *Studies in Nonlinear Dynamics & Econometrics* 6:1–27.

Ross, S. 1976. The arbitrage theory of capital asset pricing. *Journal of economic theory* 13:341–360.

Rua, A. 2010. Measuring comovement in the time-frequency space. *Journal of Macroeconomics* 32:685–691.

Rua, A. 2012. Wavelets in economics. *Economic Bulletin and Financial Stability Report Articles* pp. 71–79.

Samuelson, P. A. 1969. Lifetime Portfolio Selection by Dynamic Stochastic Programming. *Review of Economics and Statistics* 51:239–246.

Schwartz, E. S. 2004. Patents and R&D as real options. *Economic Notes* 33:23–54.

Shao, X., and W. Wu. 2007. Asymptotic spectral theory for nonlinear time series. *The Annals of Statistics* 35:1773–1801.

Sharpe, W. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance* 19:425–442.

Sharpe, W. 1966. Mutual fund performance. *The Journal of Business* 39:119–138.

Steuer, C., V. Papadimitrakopoulou, R. Herbst, M. Redman, H. F., P. Mack, S. Ramalingam, and D. Gandara. 2015. Innovative Clinical Trials: The LUNG-MAP Study. *Clin. Pharmacol. Ther.* 97:488–491.

Thakor, R. T., N. Anaya, Y. Zhang, C. Vilanilam, K. W. Siah, C. H. Wong, and A. W. Lo. 2017. Just how good an investment is the biopharmaceutical sector? *Nature biotechnology* 35:1149.

Titman, S. 1985. Urban land prices under uncertainty. *The American Economic Review* 75:505–514.

Treynor, J. 1965. How to rate management of investment funds. *Harvard Business Review* 43:63–75.

Trigeorgis, L., and S. P. Mason. 1987. Valuing managerial flexibility. *Midland corporate finance journal* 5:14–21.

US Burden of Disease Collaborators. 2013. The state of us health, 1990-2010: Burden of diseases, injuries, and risk factors. *JAMA* 310:591–606. URL `http://dx.doi.org/10.1001/jama.2013.13805`.

US Food and Drug Administration. 2010. Guidance for the use of Bayesian statistics in medical device clinical trials. Accessed January 25, 2018 at `https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071121.pdf`.

US Food and Drug Administration. 2016a. Factors to consider when making benefit-risk determinations in medical device premarket approval and de novo classifications. Accessed January 25, 2018 at `https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm506679.pdf`.

US Food and Drug Administration. 2016b. Patient preference information – Voluntary submission, review in premarket approval applications, humanitarian device exemption applications, and de novo requests, and inclusion in decision summaries and device labeling. Accessed January 25, 2018 at `http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM446680.pdf`.

Velasco, C., and P. Robinson. 2001. Edgeworth Expansions for Spectral Density Estimates and Studentized Sample Mean. *Econometric Theory* 17:497–539.

Wong, C. H., K. W. Siah, and A. W. Lo. 2018. Estimation of clinical trial success rates and related parameters. *Biostatistics* p. to appear.

Wu, W., and P. Zaffaroni. 2018. Asymptotic Theory for Spectral Density Estimates of General Multivariate Time Series. *Econometric Theory* 34:1–22.