

Time Scale Macroeconometrics

Clark Lundberg

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2014

Reading Committee:

Eric Zivot, Chair

Yu-chin Chen

Don Percival

Program Authorized to Offer Degree:
Department of Economics

©Copyright 2014

Clark Lundberg

University of Washington

Abstract

Time Scale Macroeconometrics

Clark Lundberg

Chair of the Supervisory Committee:

Professor Eric Zivot

Department of Economics

The central focus of this dissertation is to develop robust econometric methods to identify and learn about the effects of time horizon on economic relationships. The first chapter provides a broad overview of some of the disparate methods used in the literature to identify such effects and, where appropriate, attempts to interpret them in a unified filtering framework. The second chapter briefly summarizes wavelet methods – a central tool used in the framework I develop throughout the rest of the dissertation. In chapter two I also establish some useful asymptotic properties of wavelet transforms. The third chapter introduces a multiresolution regression (MRR) model based on nonparametric wavelet methods. I derive statistical properties and asymptotic behavior of the MRR estimator – establishing that the MRR is robust to a broad class of scale misspecification. I conclude the third chapter by using a simple MRR model to demonstrate that common assumptions of high frequency noise in portfolio return models are inappropriate. The final chapter expands on the MRR framework by introducing dynamic MRR models in the context of cross sectional asset pricing. I find considerable evidence for scale structure in the cross section of portfolio returns and show that a single financial variable can aggregate multiple sources of risk over different horizons, yet the market prices these risks associated with different time horizons separately. By directly incorporating scale structure into asset pricing models, asset pricing performance dramatically improves while offering inference on the time horizon dynamics of risk channels.

TABLE OF CONTENTS

	Page
List of Figures	iii
List of Tables	iv
Chapter 1: Time Scale Econometrics	1
1.1 Introduction	1
1.2 Scale-based Methods – An Overview	2
1.3 Conclusion	8
Chapter 2: Wavelets	11
2.1 Introduction	11
2.2 Multiresolution Decomposition and Wavelet Filters	12
2.3 Asymptotic Behavior of the Multiresolution Decomposition	19
2.4 Some Useful Lemmas and Corollaries	20
2.5 Simulation	21
2.6 Conclusion	22
Chapter 3: A Multiresolution Regression Framework	27
3.1 Introduction	27
3.2 The Multiresolution Regression (MRR) Model	29
3.3 Properties of the MRR Estimator	31
3.4 Asymptotic Distribution	35
3.5 Simulation	41
3.6 A Test for High Frequency Noise in Portfolio Returns	44
3.7 Conclusion	46
Chapter 4: Time Scale Structure in the Cross Section of Portfolio Returns	52
4.1 Introduction	52
4.2 Multiresolution Asset Pricing	55
4.3 Methodology	58

4.4	Data	67
4.5	Results	67
4.6	Conclusion	69
Appendix A: Chapter 2 Proofs: Wavelets		85
A.1	Proof of Lemma 1	86
A.2	Proof of Corollary 1	90
A.3	Proof of Lemma 2	90
A.4	Proof of Corollary 2	90
Appendix B: Chapter 3 Proofs: A Multiresolution Regression Framework		92
B.1	Proof of Lemma 3	93
B.2	Proof of Lemma 4	94
B.3	Proof of Theorem 1	95
B.4	Proof of Theorem 2	96
B.5	Proof of Corollary 3	96
B.6	Proof of Lemma 5	97
B.7	Proof of Corollary 4	99
B.8	Proof of Lemma 6	99
B.9	Proof of Lemma 7	100
B.10	Proof of Corollary 5	101
B.11	Proof of Theorem 3	101

LIST OF FIGURES

Figure Number	Page
1.1 Long Horizon Squared Gain Function – Amplification and Attenuation	9
1.2 Long Horizon Squared Gain Functions and Filter Length	10
2.1 Asymptotic Behavior of Daubechies-class of Wavelet Filters	23
2.2 Asymptotic Variance Behavior	24
2.3 Determinant of Wavelet Correlation Matrix	25
2.4 Determinant of Details Correlation Matrix	26
3.1 Asymptotic Standard Error Behavior	47
4.1 25 Size/B-M Portfolios Full Sample (1926-2012)	70
4.2 25 Size/B-M Portfolios Short Sample (1970-2010)	71
4.3 100 Size/B-M Portfolios Full Sample (1926-2012)	72
4.4 48 Industry Portfolios Full Sample (1926-2012)	72
4.5 10 Momentum Portfolios Full Sample (1926-2012)	73

LIST OF TABLES

Table Number	Page
3.1 Correct Specification: IID Errors – $L = 2, n = 64$	47
3.2 Correct Specification: IID Errors – $L = 256, n = 4096$	48
3.3 IID Errors – $L = 16$	48
3.4 $AR(1)$ Errors – $L = 16$	48
3.5 Heteroskedastic Errors – $L = 16$	48
3.6 IID Errors – $L = 128$	49
3.7 $AR(1)$ Errors – $L = 128$	49
3.8 First Quintile of Size Sorted Portfolios	49
3.9 Second Quintile of Size Sorted Portfolios	50
3.10 Third Quintile of Size Sorted Portfolios	50
3.11 Fourth Quintile of Size Sorted Portfolios	51
3.12 Fifth Quintile of Size Sorted Portfolios	51
4.1 Consumption Growth and Market Excess Returns by Scale	70
4.2 Performance Measures of Asset Pricing Models	71
4.3 Scale Homogeneity Test: First Quintile Size Portfolios by B-M Quintile	73
4.4 Scale Homogeneity Test: Second Quintile Size Portfolios by B-M Quintile	74
4.5 Scale Homogeneity Test: Third Quintile Size Portfolios by B-M Quintile	74
4.6 Scale Homogeneity Test: Fourth Quintile Size Portfolios by B-M Quintile	74
4.7 Scale Homogeneity Test: Fifth Quintile Size Portfolios by B-M Quintile	75
4.8 Scale Homogeneity Test: 48 Industry Sorted Portfolios (1 of 6)	75
4.9 Scale Homogeneity Test: 48 Industry Sorted Portfolios (2 of 6)	75
4.10 Scale Homogeneity Test: 48 Industry Sorted Portfolios (3 of 6)	76
4.11 Scale Homogeneity Test: 48 Industry Sorted Portfolios (4 of 6)	76
4.12 Scale Homogeneity Test: 48 Industry Sorted Portfolios (5 of 6)	76
4.13 Scale Homogeneity Test: 48 Industry Sorted Portfolios (6 of 6)	77
4.14 Scale Homogeneity Test: 10 Momentum Sorted Portfolios (1 of 2)	77
4.15 Scale Homogeneity Test: 10 Momentum Sorted Portfolios (2 of 2)	77

Chapter 1

TIME SCALE ECONOMETRICS

1.1 Introduction

Time horizons play an integral role in macroeconomics and finance. Certainly, economic and financial time series exhibit different properties over different horizons; for example, Fama and French (1988) and Poterba and Summers (1988) both document negative serial correlation in financial returns over multi-year horizons, while Lo and MacKinley (1988) document positive serial correlation in returns over daily, weekly, and monthly horizons. These different scale-based properties can be economically meaningful – Campbell, Lo, and MacKinley (1997) speculate that many macroeconomic effects – like the business cycle – likely manifest in financial markets over these longer horizons. Likewise, predictability has been of central interest in finance and economics for decades. Various tests of predictability show up in asset pricing, international finance, and more broadly in macroeconomics, often with similar mixed results. A phenomenon that has received continued attention has been the apparent predictability of financial variables over long horizons for which inconclusive or negative evidence exists over short horizons. A popular technique for identifying these longer horizon characteristics is the so-called long-horizon regression which is typically implemented using rolling sums of variables sampled at higher frequencies. These long-horizon regressions often find significant evidence for predictability – hence the attention they have received. However, the fundamental approach of long-horizon regressions – aggregating a time series via rolling sums – destroys the higher frequency information; rolling sums are, in effect, a crude low-pass filter¹. As a result, long-horizon regressions impose an *a priori* assumption that the higher-frequency components of the relationship under investigation are not of interest. Indeed, the high frequency components in the regression are

¹This is clear when it is noted that a rolling sum is simply a convolution of the time series with a set of filter coefficients equal to 1. The low-pass characteristics of this filter can be verified by looking at the squared gain function of this so-called “boxcar” filter.

often thought of as uninformative noise and that the low-pass filtering operation embedded into long-horizon regressions strengthens the signal in the data while reducing the noise. Valkanov (2003) summarizes an early heuristic argument in the literature that such a denoising operation is responsible for the statistically significant findings for predictability in the finance literature – i.e. returns are predictable but short-horizon regressions suffer from noise that obscures predictability. Indeed, even contemporary research operates under this assumption. Campbell and Yogo (2006) argue that predictive regressions generally have low power because returns are extremely noisy. However, the model they present is formulated under the null that (high-frequency) innovations to returns and the predictor variables are correlated. In this case, the interpretation of such high-frequency innovations as noise no longer seems appropriate.

The central problem in addressing apparent scale-dependent economic relationships seems to be a general lack of a unifying empirical framework for identifying and estimating these effects. As a result, there seems to be a feedback loop of sorts – a consensus about the scale-based stylized facts and empirical findings has yet to emerge and inform theory, which has, in turn, failed to inform empirical work. That said, there is a considerable body of theoretical finance literature that at least indirectly proposes horizon-based effects (e.g. Merton (1973), Jagannathan and Wang (1996), Campbell and Cochrane (1999)). There is also a growing body of applied econometric work modeling time-horizon components more directly. The remainder of this chapter presents a brief overview of some of the empirical work directly addressing time horizon effects in econometric models. Section 1.2 provides a brief overview of some existing scale-based econometric methods including long horizon regression, state space models, scale-based volatility models, mixed-sampling frequency methods, and wavelet methods while Section 1.3 concludes.

1.2 Scale-based Methods – An Overview

A well-adopted methodology for investigating economic relationship over longer scales is the so-called long horizon regression. The long-horizon approach constructs new variables that capture the “long-horizon” behavior of a time series via a simple rolling sum. Valkanov (2003) identifies two approaches to long-horizon regressions – 1) regressing an aggregated

dependent variable on a lagged regressor, and 2) regressing an aggregated dependent variable on an aggregated regressor (or lagged aggregated regressor). The rolling sum operation is easily identified as a convolution and hence, a filtering operation. Consider a covariance stationary time series $\{X_t\}$ with a bounded spectral density function and a rolling sum $\{Q_t^L\}$ with width L :

$$Q_t^L = \sum_{l=0}^{L-1} X_{t+l}$$

The series $\{Q_t^L\}$ is obtained by convolving the series $\{X_t\}$ with the vector $\{a_l\} = \mathbf{1}_L$, i.e. $a_l = 1$ for $l = 0, \dots, L - 1$. Frequency domain filters are often designed such that they preserve the variance of the original series over the desired pass band, however the rolling sum convolution introduces an amplification operation; that is, the variance of the series $\{X_t\}$ over the pass band is *magnified* by this operation. Figure 1.1 illustrates this amplification phenomenon with a long horizon filter of length $L = 8$. Figure 1.2 demonstrates how this effect is compounded by increasing the filter length. The normalized version of this type of filter is an $MA(L)$ filter, i.e. by dividing the impulse response sequence by the filter length, the filter then preserves the variance of $\{X_t\}$ in the lower frequency range while attenuating it at the higher frequency ranges. If we define an $MA(L)$ impulse response sequence as $\{b_l\}$, then the long horizon impulse response sequence $\{a_l\}$ is equivalent to $L\{b_l\}$ and the squared gain function associated with the long horizon filter is $L^2|B(f)|^2$ where $B(f)$ is the Fourier transform of the $MA(L)$ impulse response sequence. The variance of the $MA(L)$ -filtered sequence shrinks at a rate of $1/L$ while the amplification term for the long horizon-filtered sequence grows at a rate of L^2 , hence variance of filtered variables using long horizon filter grows linearly in filter length. Incidentally, this frequency domain interpretation of the long horizon filter leads to a frequency domain understanding of $I(1)$ processes. We could construct a “random walk filter” as the limit of a long horizon filter as $L \rightarrow \infty$. The straightforward observation that random walks do not have finite variance is clear in the frequency domain from the observation above.

Much of the contemporary work in long horizon regressions has focused on correcting spurious inference due to the imposed persistence in long horizon regressors. The rolling sum operation embedded in long-horizon methods induces persistence in the regressors that

leads to a breakdown in traditional first-order asymptotic theory and potentially spurious test results. Valkanov (2003) derives an asymptotic theory for long-horizon regressions under the local-to-unity framework. Both Valkanov (2003) and Campbell and Yogo (2006) work with a rescaled t-tests and a Bonferroni method for addressing the local-to-unity nuisance parameter in t tests. Hjalmarsson (2011) follows a similar approach. A more recent contribution to the long horizon regression literature is the introduction of the extended instrumental variables (IVX) approach by Magdalinos and Phillips (2009) that avoids the problem of induced persistent regressors.

Although these contemporary econometric approaches address the issue of spurious inference in long-horizon regressions, the role of the filtered high-frequency components in the data remains to be addressed. There is an *a priori* assumption of high frequency noise common to all long horizon approaches which may or may not be appropriate depending on the application.

A potential approach to relax the noise assumptions embedded in long horizon methods is the state space framework. However, directly modeling scale behavior can be challenging task using state space methods. There is a long history of unobserved components-based (UC) decompositions of non stationary time series into transitory and permanent components (e.g. Watson (1986), Clark (1987)) which certainly captures a degree of scale effects in the the time series. Harvey and Trimbur (2003) introduce stochastic cycles into the state space framework, thereby allowing a more thorough and robust approach to identifying and estimating scale effects. Harvey, Trimbur and Van Dijk (2006) and Creal, Koopman, and Zivot (2010) expand on this approach. For the most part, these state space models are formulated and interpreted in the time domain exclusively – even the stochastic cycle component is typically specified in the time domain. The notable exception to this is Creal, Koopman, and Zivot (2010) who interpret the dynamic parameters of the stochastic cycle as a time varying band pass filter, however they do not explicitly present the frequency domain features of their estimated filter. Nevertheless, the highly parameterized nature of these state space scale models leads to two primary drawbacks: 1) they can be difficult or impossible to implement, often requiring complicated Bayesian methods that themselves have steep learning curves, and 2) the highly parameterized nature creates a high degree

of sensitivity to model specification. Creal, Koopman, and Zivot (2010) in particular find that different specifications of the variance processes of their scale components lead to quite different scale component estimates. However, despite these challenges, state space models can be incredibly flexible and missing data and/or mixed sampling frequencies can be incorporated as well.

A contemporary strand of literature addresses scale issues through mixed sampling frequency data – including the so-called mixed data sampling (MIDAS) regression of Ghysels, Santa-Clara, and Valkanov (2003) as well as the generalized autoregressive distributed lag (GADL) model of Chen and Tsay (2011). In both cases, the econometric approach is not centrally concerned with directly modeling scale effects, but rather implementing a simple reduced form approach to addressing mixed sampling frequencies as an alternative to the far more complicated state space methods mentioned above. For example, fitting a predictive regression for macroeconomic indicator variables sampled monthly or quarterly using high(er) frequency (e.g. daily) financial variables. In general, a MIDAS regression can be represented:

$$y_t = \beta_0 + \beta_1 B(L^{1/m})x_{t-1}^{(m)} + \epsilon_t^{(m)} \quad (1.1)$$

where

$$B(L^{1/m}) = \sum_{j=0}^{j_{max}} b_j L^{j/m}$$

and for the identification of β_1 :

$$\sum_{j=0}^{j_{max}} b_j = 1$$

Essentially, a sequence $\{x_s\}$ is sampled m times more frequently than $\{y_t\}$. The standard MIDAS notation $\{y_t\}$ has integer indices while $\{x_s\}$ is indexed fractionally, hence is observable at $s = 1/m, 2/m, 3/m, 4/m, \dots, 1, 1 + 1/m, 1 + 2/m, \dots$. The sequence used as a regressor $\{x_t^{(m)}\}$ is constructed via a weighted average of $x_t, x_{t-1/m}, x_{t-2/m}, \dots, x_{t-j_{max}/m}$ where j_{max} is the maximum lag included in the model – often set to $m - 1$. The basic approach of the GADL model is quite similar to the MIDAS approach but avoids some of the complications in estimation through clever reparameterization using a Vandermonde

matrix with Almon lag polynomials which allows for linear estimation methods without imposing the restriction on lag weights to identify β_1 . In both approaches the lag weights are estimated with a least squares optimization procedure. Although neither MIDAS nor GADL methods directly model scale effects, they both implicitly include filtering and down-sampling operations that are inherently scale related – the lag weights equivalent to filter coefficients and the temporal aggregation scheme a downsampling operation.

To illustrate this point, consider a series $\{z_t\}$ that is a convolution of the lag weights $\{b_j\}$ with the series $\{x_t\}$ for $t = 1/m, 2/m, 3/m, 4/m, \dots, 1 + 1/m, 1 + 2/m, \dots$:

$$z_t = \sum_{j=0}^{j_{max}} b_j x_{t-j/m}$$

It is worth noting that $x_t^{(m)}$ in equation (1.1) is indexed such that $x_t^{(m)}$ is m periods ahead of $x_{t-1}^{(m)}$. The regressor $\{B(L^{1/m})x_{t-1}^{(m)}\}$ in equation (1.1) is simply the integer time indices $t = 1, 2, 3, \dots$ of the filtered sequence $\{z_t\}$ or, equivalently, the filtered sequence $\{z_t\}$ *down-sampled* at rate m .

The frequency domain characteristics of MIDAS/GADL filters depend on the lag order – which affects the degree of resolution in the frequency domain – and the data itself. Because MIDAS/GADL filters are the solutions to a least squares optimization problem, they effectively represent a filter that best captures the frequency domain characteristics of the regression relationship under investigation – the strongest signal to noise ratio is extracted subject to the downsampling condition and the filter length.

I illustrate the frequency domain characteristics of the MIDAS filter with the following example in which, for ease of exposition, I consider a single regressor that is sampled twice as frequently as the dependent variable. Let us index the high-frequency sampling by s and impose a maximum lag order of two for simplicity. Then the lag parameterization is equivalent to a filtered variable z_s constructed by:

$$z_s^{(2)} = B(L)x_s^{(2)} = b_0 x_s + b_1 x_{s-1}$$

Following the convention in the MIDAS literature, the sum of the lag weights is constrained

to unity so the above MA filter can be reparameterized to:

$$z_s^{(2)} = B(L)x_s^{(2)} = b_0x_s + (1 - b_0)x_{s-1}$$

The frequency response of the filter then depends on the b_0 parameter: $b_0 > 1$ corresponds to a high-pass filter, $0 < b_0 < 1$ to a low-pass filter. Note that in a higher-order lag parameterization the number of parameters allows for band-pass filter characteristics as well. The regressor z_t is then obtained by downsampling the filtered series $z_s^{(2)}$ by rate m (here $m = 2$):

$$z_t = z_s^{(2)} \Big|_{\downarrow 2}$$

In contrast to other related filtering/downsampling operations like the discrete wavelet transform (DWT), the MIDAS impulse response sequence (lag weights) is constructed in the time domain based on a least squares optimization procedure. The filter construction process is driven by the data and the filter coefficients (or lag weights) will vary depending on the sample. It is typical in the MIDAS literature to use the less frequently sampled series – the dependent variable – as the reference sample frequency and fractionally index the higher-frequency series.

A separate literature that is increasingly incorporating direct scale modeling is the volatility modeling literature. Adrian and Rosenberg (2008) parametrically model scale-based volatility factors using two components of volatility, each with a different rate of mean reversion. Engle, Ghysels, and Sohn (2009) develop a GARCH extension in the MIDAS framework and model short-horizon volatility with a high-frequency GARCH process while long-horizon dynamics are captured with a MIDAS polynomial. Engle and Rangel (2008) develop a spline-GARCH model to describe short-horizon and long-horizon components of volatility. Rangel and Engle (2009) extend this model in a multivariate GARCH framework using dynamic conditional correlations to capture short- and long-horizon components of equity correlations. In the spline-GARCH framework the mean-reversion characteristics of the GARCH process is a function that slowly varies over time. This slowly varying function captures long-horizon dynamics while the short-horizon dynamics are described by the

GARCH process itself.

Finally, wavelet-related methods offer an alternative (or in many cases, a complement) to traditional time series econometric methods. Ramsey and Lampart (1998) was one of the early introductions of wavelets to the economics literature but much of the contemporary work – characterized by Gençay, Selçuk, and Whitcher (2001a, 2001b, 2003, 2005) and In and Kim (2012) – centers around so-called multiscale regressions. That is, the variables in a time series regression model are each decomposed into components that describe the time series over particular time scales. The regression model is then estimated separately for each scale. Certainly this approach allows identification of scale effects in regression relationships but the manner in which identification occurs makes interpretation of the models challenging. Testing and more broadly integrating the approach with mainstream econometric methods can also be a problem with these approaches.

1.3 Conclusion

The organizing theme of scale methods in econometrics is a decided lack thereof. However, there is a clear growth across the econometrics literature in models that attempt to explicitly incorporate scale effects. This chapter has presented a very brief overview of some of the major strands in the literature modeling scale effects in economic relationships. Where possible, I have tried to interpret these different approaches in a unified filtering framework.

Figure 1.1: Long Horizon Squared Gain Function – Amplification and Attenuation

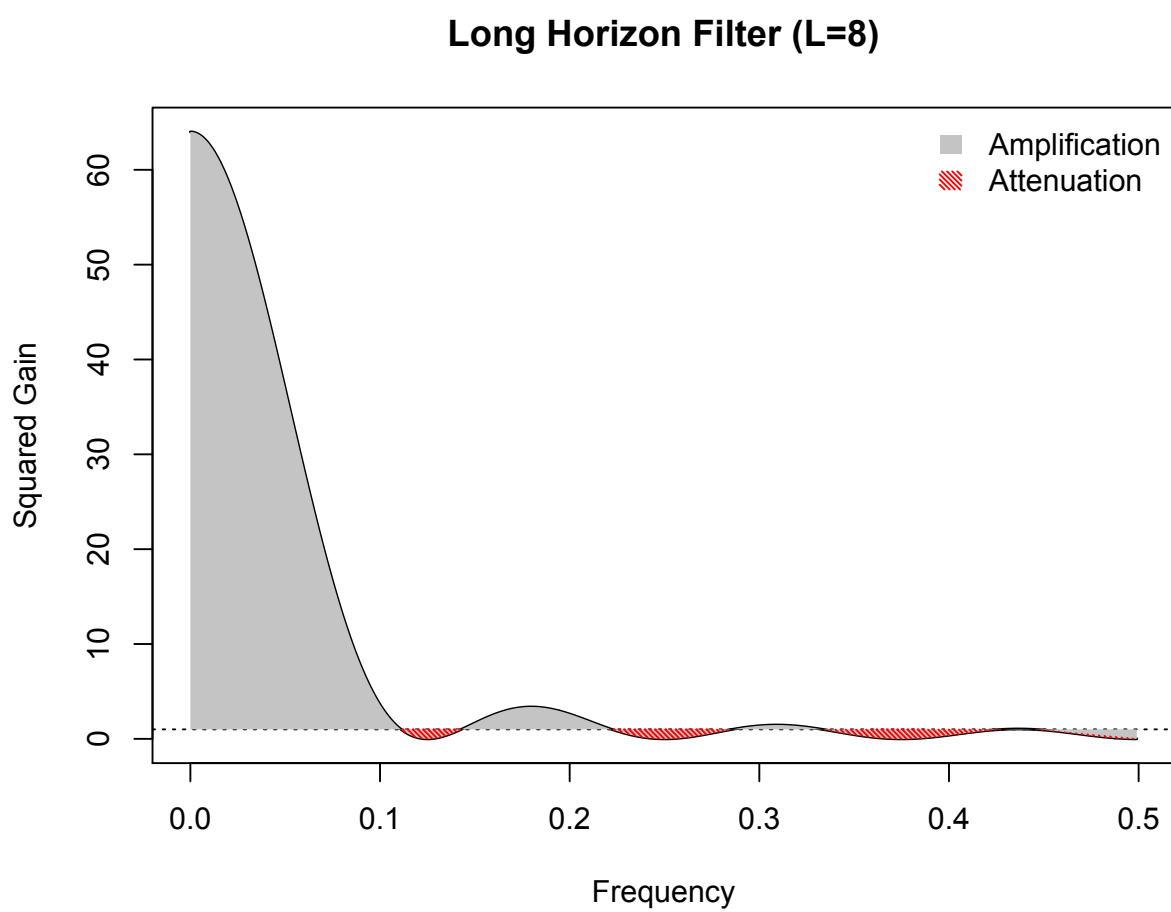
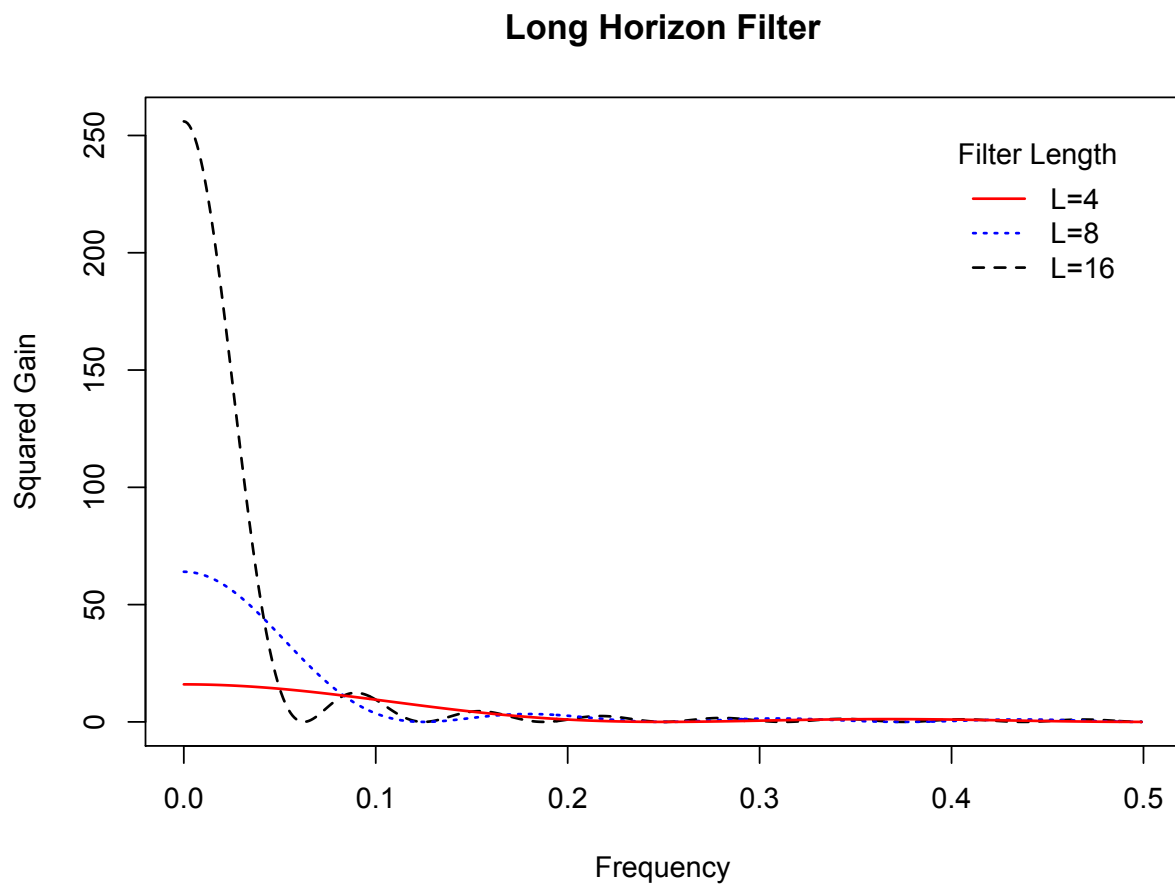


Figure 1.2: Long Horizon Squared Gain Functions and Filter Length



Chapter 2

WAVELETS

2.1 Introduction

The popularity of wavelet methods has grown tremendously over the last several decades with the bulk of growth concentrated in the signals processing, engineering, and geosciences literature. Roughly speaking, wavelets decompose the characteristics of a time series over different time horizons (scales). The central method for doing so are the “little waves” themselves – these wavelets are able to localize both time and frequency components of a time series. While the theory developed in this dissertation focuses on weakly stationary time series, a major attraction of wavelet methods is their flexibility in addressing a variety of non stationary characteristics in time series. Because of this, wavelets offer very promising avenues for nonparametric inference on economic times series of questionable stationarity.

As mentioned, wavelets capture the time characteristics of a time series over different horizons and thus constitute a “mixed-domain” analysis technique that simultaneously describes the time series in the time domain as well as the frequency domain. Strictly speaking, the mixed wavelet domain is the time-scale domain, in contrast to the time-frequency domain of the short-time Fourier Transform (STFT), the Wigner-ville distribution, and associated time-frequency distributions. However, this distinction is a subtle one – scales are defined over frequency bands – so in a loose sense, time-scale methods like wavelet transforms can easily be interpreted in the time-frequency domain and throughout this chapter and those that follow I interpret results in this manner. Because of their mixed-domain nature, wavelets allow for a flexible, hybrid approach to synthesizing traditional time domain methods (which econometricians have overwhelmingly focused on) with Fourier analysis of the frequency domain. The central tool in pure frequency domain (spectral) analysis is the Fourier transform. This chapter proceeds under the assumption that readers are at least modestly familiar with spectral methods and Fourier analysis. For a comprehensive

introduction that is notationally consistent with this chapter and those that follow, I refer readers to Percival and Walden (1993, 2000). The remainder of this chapter is organized as follows: Section 2.2 briefly introduces wavelet methods, while Section 2.3 establishes some useful asymptotic properties of wavelet transforms. Section 2.4 formalizes these properties in several useful lemmas and corollaries. Section 2.5 illustrates some properties established in Section 2.3 through a brief simulation exercise. Section 2.6 concludes.

2.2 Multiresolution Decomposition and Wavelet Filters

Wavelet transforms, and their associated multiresolution analysis are convenient tools for both stationary and nonstationary time series analysis. At their most basic, the means with which wavelet transforms localize time series characteristics in both time and scale is via sets of differencing and local averaging operations. There are a plethora of wavelet transforms that have been developed but I restrict my attention to the orthogonal discrete wavelet transform (DWT) and its non-decimated nonorthogonal variety the maximal overlap discrete wavelet transform (MODWT). Both the DWT and MODWT decompose the time series into what are typically referred to as “approximately” uncorrelated sequences. The DWT is a convenient introduction to wavelet transforms and a powerful tool in its own right but suffers from a number of undesirable characteristics. The MODWT, in contrast, is slightly more sophisticated than the DWT but mitigates many of the shortcomings of the DWT. I introduce each below.

2.2.1 The Discrete Wavelet Transform

The DWT decomposes a time series into wavelet and scaling coefficients different *levels* – which are typically indexed by j – where each level corresponds to the dyadic physical scale $2^j \Delta t$ for sampling interval Δt . This scale corresponds to a nominal pass band of $(1/2^{j+1}, 1/2^j]$ – i.e. the level j wavelet coefficients describe the time domain characteristics of the original time series (approximately) over the nominal pass band. The wavelet coefficients describe how the time series changes over time at that scale while the scaling coefficients describe a local average of the time series over an equivalent scale. This is done using a so-called mother wavelet filter whose impulse response sequence satisfies the following

properties:

$$\sum_{l=0}^{L-1} h_l = 0 \quad (2.1)$$

$$\sum_{l=0}^{L-1} h_l^2 = 1 \quad (2.2)$$

$$\sum_{l=0}^{L-1} h_l h_{l+2n} = 0 \quad \forall n \neq 0, n \in \mathbb{Z} \quad (2.3)$$

The associated scaling filters g_l are defined via the quadrature mirror relationship of the above wavelet filters:

$$g_l = (-1)^{l+1} h_{L-1-l} \quad l = 0, 1, 2, \dots, L-1$$

Properties (2.1) - (2.3) establish the wavelet filter as an energy preserving series of differencing operations. Higher levels of the wavelet decomposition are obtained by stretching (or dilating) the original mother wavelet filter such that the differencing operation becomes a more sophisticated differencing of local averages. These dilated versions of the mother wavelet are typically denoted by indexing the level of the scale decomposition j , ie. the filter coefficients associated with level j wavelet coefficients would be denoted $h_{j,l}$ with $l = 0, 1, 2, \dots, L_j - 1$. The level j dilation stretches out the mother wavelet of length L to a length of $L_j = (2^j - 1)(L - 1) + 1$. As an illustration, consider the Haar wavelet filter of length $L = 2$. The level 2 Haar wavelet filter is of length $(2^2 - 1)(2 - 1) = 4$ while the the level 3 Haar wavelet filter will be of length $L_j = (2^3 - 1)(L - 1) + 1 = 8$. The higher levels of the wavelet filters must also satisfy properties (2.1) - (2.3).

The filtering operation can also be interpreted as a linear transform of the time series. In the case of the orthogonal DWT, this transform can be represented as:

$$\mathbf{W}_{N \times 1} = \mathcal{W}_{N \times N} \mathbf{X}_{N \times 1} \quad (2.4)$$

where X is a time series, \mathbf{W} is the vector of wavelet and scaling transform coefficients that describe the time series characteristics at various scales ($N/2^j$ coefficients for each scale

2^{j-1}), and \mathcal{W} is the transformation matrix associated with the set of filtering operations described above. However, the DWT has a number of drawbacks: it can only deal with time series of length 2^{J_0} for some integer J_0 and it involves a downsampling operation such that for each level j , the time series at scale 2^{j-1} is only described by a series of $N/2^j$ wavelet coefficients resulting in a certain loss of resolution in the time domain. This is reflected by partitioning the \mathbf{W} matrix as follows:

$$\mathbf{W}_{N \times 1} = \begin{bmatrix} \mathbf{W}'_1 & \mathbf{W}'_2 & \cdots & \mathbf{W}'_{J_0} & \mathbf{V}'_{J_0} \end{bmatrix}'$$

$\begin{matrix} 1 \times N/2^1 & 1 \times N/2^2 & & 1 \times 1 & 1 \times 1 \end{matrix}$

where \mathbf{W}_j are the level j wavelet coefficients and \mathbf{V}_{J_0} is the level J_0 scaling coefficient. The \mathcal{W} matrix can be similarly partitioned:

$$\mathcal{W}_{N \times N} = \begin{bmatrix} \mathcal{W}'_1 & \mathcal{W}'_2 & \cdots & \mathcal{W}'_{J_0} & \mathcal{V}'_{J_0} \end{bmatrix}'$$

$\begin{matrix} N \times N/2^1 & N \times N/2^2 & & N \times 1 & N \times 1 \end{matrix}$

The first row of the $N/2^j \times N$ matrix \mathcal{W}_j is $h_{j,1}$, $h_{j,0}$, and $N - L_j$ zeros followed by the remaining filter coefficients in reverse order. This is equivalent to reversing and then circularly shifting to the right by two the sequence $\{h_{j,l}\}$ followed by $N - L_j$. Subsequent rows of this matrix contain this first row shifted (circularly) to the right by two, so that the second row begins with two zeros, followed by $\{h_{j,l}\}$, followed by $N - L_j - 2$ zeros; while the last row begins with $N - L_j$ zeroes followed by $\{h_{j,l}\}$.

A natural result of partitioning the \mathbf{W} and \mathcal{W} matrices by level defines a so-called *multiresolution analysis* (MRA) which is a set of time series *details* $\mathcal{D}_j = \mathcal{W}'_j \mathbf{W}_j$ for $j = 1, 2, \dots, J_0$ and *smooth* $\mathcal{S}_{J_0} = \mathcal{V}'_{J_0} \mathbf{V}_{J_0}$ that capture the behavior of the original times series at different scales. Note that although \mathcal{D}_j is of length N , there is still some loss of frequency domain (scale domain) resolution due to the compression (downsampling) embedded in the original DWT.

The DWT and associated details and smooth have a number of convenient properties. The first is that the details and smooth constitute an additive decomposition of the time

series across scales. For a level J_0 DWT, the following property holds:

$$X = \mathcal{D}_1 + \mathcal{D}_2 + \cdots + \mathcal{D}_{J_0} + \mathcal{S}_{J_0} \quad (2.5)$$

This decomposition is referred to in the wavelet literature as a multiresolution analysis (MRA). Likewise, the wavelet and scaling coefficients decompose the variance of the times series across scales. For a level J_0 DWT, the following property holds:

$$\|X\|^2 = \|W_1\|^2 + \|W_2\|^2 + \cdots + \|W_{J_0}\|^2 + \|V_{J_0}\|^2 \quad (2.6)$$

where $\|\cdot\|$ denotes the Euclidean norm. This property establishes that the wavelet coefficients decompose the sample variance of the time series across scales:

$$Var(X) = \sum_{j=1}^{J_0} Var(W_j) \quad (2.7)$$

In fact, the norm of the wavelet (scaling) coefficients and the norm of the details (smooth) are equivalent in the DWT, ie. $\|W_j\|^2 = \|\mathcal{D}_j\|^2$ for $j = 1, 2, \dots, J_0$ and $\|V_j\|^2 = \|\mathcal{S}_j\|^2$. Hence the DWT details and smooth can be thought of as a variance decomposition as well as an additive decomposition of the time series.

2.2.2 The Maximal Overlap Discrete Wavelet Transform

The maximal overlap discrete wavelet transform (MODWT) is a redundant version of the DWT that avoids the downsampling operation embedded in the DWT. A standing convention in the wavelet literature is to distinguish MODWT filters and coefficients from that of the DWT with a tilde, ie. if \mathbf{W}_j denotes the level j wavelet coefficients from the DWT, the level j MODWT wavelet coefficients would be denoted $\widetilde{\mathbf{W}}_j$. In contrast to the linear transform of the DWT, the MODWT can be thought of as a *set* of linear transforms, one for each level $j = 1, 2, \dots, J_0$ plus one additional level J_0 scaling filter transform:

$$\widetilde{\mathbf{W}}_j = \widetilde{\mathcal{W}}_j \mathbf{X} \quad (2.8)$$

$\begin{matrix} N \times 1 & N \times N & N \times 1 \end{matrix}$

hence the MODWT results in a series of wavelet coefficients at each level j that is of the same length as the original time series, effectively giving a very accurate, time-localized description of the time series characteristics at that particular scale. The MODWT filters are simply rescaled versions of DWT filters h_l :

$$\tilde{h}_l = h_l/\sqrt{2}$$

Since all subsequent discussion will focus on the MODWT rather than the DWT, I forgo the tilde notation and instead all wavelet-related variables will denote MODWT versions.

Although the MODWT operation is not an orthogonal transformation, like the DWT it does preserve the energy of the time series in the decomposition; that is, the wavelet coefficients decompose the energy of the time series across scales in the following way:

$$\|X\|^2 = \sum_{j=1}^{J_0} \|W_j\|^2 + \|V_{J_0}\|^2 \quad (2.9)$$

where $\|\cdot\|$ denotes the Euclidean norm. This property establishes that – like the DWT – the wavelet coefficients decompose the variance of the time series across scales:

$$Var(X) = \sum_{j=1}^{J_0} Var(W_j) + Var(V_{J_0}) \quad (2.10)$$

The MODWT also allows for an analogue to the DWT additive decomposition of the time series via the MRA *Details* and *Smooth*. For a level J_0 MODWT, the following property holds:

$$X = \mathcal{D}_1 + \mathcal{D}_2 + \cdots + \mathcal{D}_{J_0} + \mathcal{S}_{J_0} \quad (2.11)$$

The filtering interpretation of wavelet transforms is most evident in MODWT and related transforms. The level j wavelet coefficients are the result of a (circular)¹ convolution of the

¹There are a variety of methods for relaxing the circularity assumption including a reflection operation in which the time series is split into two parts and reflected about the endpoints $t = 0$ and $t = N - 1$. This approach minimizes any artifacts of the circular convolution without introducing any new data into the operation and is appropriate for a variety of stationary and non stationary processes.

filter coefficients with the original time series X :

$$W_{j,t} = \sum_{l=0}^{L_j-1} h_{j,l} X_{t-l \bmod N}$$

The DWT wavelet coefficients are obtained by down sampling these MODWT coefficients or equivalently using the tilde notation to distinguish DWT and MODWT coefficients:

$$W_{j,t} = 2^{j/2} \widetilde{W}_{j,2t+1}$$

where the downsampling is embedded in the indexation of the MODWT wavelet coefficients and the $2^{j/2}$ is from the rescaling factor in the MODWT filters.

Having considered the filtering interpretation of the MODWT, the frequency domain characteristics of the MODWT details and smooth can now be considered by regarding the derivation of the details and smooth as a two-step filtering operation. In the first step, the level j wavelet coefficients are obtained as the output of a filter with transfer function $H_j(f)$ applied to the time series X . The level j details are then the output of filtering the wavelet coefficients with a transfer function that is the complex conjugate of $H_j(f)$ (usually denoted $H_j^*(f)$)². This establishes the level j details as the output of a filter with a transfer function equal to $H_j^*(f)H_j(f) = |H_j(f)|^2$ applied to X . Because the transfer function for the MODWT details (and smooth) is real valued, the details (and smooth) constitute the output of a zero-phase filter. This relationship between the MODWT wavelet and scaling filters and the MODWT details and smooth filters establishes the fact that for the MODWT, $\|\mathcal{D}_j\|^2 \leq \|W_j\|^2$ with a strict inequality for $L < \infty$. This leads to a distinction between an additive decomposition of the series via the details and smooth and an additive decomposition of the variance via the wavelet and scaling coefficients. The MODWT, therefore, requires both the MRA components as well as the wavelet and scaling coefficients to capture both decomposition properties for finite filter lengths.

This is an admittedly limited introduction to wavelet transforms. For more detail on

²Filtering with the complex conjugate of the transfer function is equivalent to taking the matrix transpose of the transform matrix W_j . See Percival and Walden (2000) for more details.

the properties of wavelets and wavelet-based transforms we refer the reader to Percival and Walden (2000), Gençay, Selçuk, and Whitcher (2001a), or In and Kim (2012).

2.2.3 Pass Band Aggregation

As outlined above, the details and smooth (and their associated wavelet and scaling coefficients) correspond to particular frequency bands. It is a simple procedure to aggregate these pass bands to create wider band-pass filters. Following Percival et al. (2011) the aggregation procedure can be initiated using the details and smooth by simple addition:

$$\mathcal{D}_{j_{min},j_{max}} = \sum_{i=j_{min}}^{j_{max}} \mathcal{D}_i$$

Intuitively it makes sense to aggregate adjacent frequency bands to create wider pass band components but in practice more complicated pass bands like notch-like filters could be created; again, this is implemented through a simple summation of the details and smooth.

Using the details preserves the additive decomposition of the time series, however in order to accurately preserve the variance decomposition in finite samples, associated wavelet and scaling coefficients are required. Again, following Percival et al. (2011) these coefficients can be obtained via the frequency domain by adding the squared gain functions associated with the aggregating frequency bands:

$$\mathcal{A}(f)_{j_{min},j_{max}} = \sum_{i=j_{min}}^{j_{max}} \mathcal{H}_i(f)$$

where $\mathcal{H}_i(f)$ is the squared gain function associated with the level i wavelet coefficients and details. The wavelet coefficients $C_{j_{min},j_{max}}$ of the aggregated details are then obtained via an inverse discrete Fourier transform:

$$C_{j_{min},j_{max}} = iDFT \left\{ \left(\mathcal{A}\left(\frac{k}{N}\right)_{j_{min},j_{max}} \right)^{\frac{1}{2}} \mathcal{X}_k \right\}$$

where \mathcal{X}_k indicates the discrete Fourier transform of the original time series.

In practice, this pass band aggregation can be an important tool for inference about time scale relationships over economically meaningful frequency bands. For example, in chapter 4 I aggregate levels four through seven of a MODWT on monthly sampled data which corresponds to physical time scales equal to 14 - 128 months – the shortest and longest NBER dated recessions – to implement a wavelet-based business cycle decomposition. Yogo (2008) similarly uses biorthogonal wavelets to create a wavelet-based business cycle filter.

2.3 *Asymptotic Behavior of the Multiresolution Decomposition*

I introduce the asymptotic properties (as the filter length $L \rightarrow \infty$) of the MODWT by considering a popular class of orthogonal wavelet filters – the Daubechies-class wavelet filters. The Daubechies wavelet filters of length L do not have a closed form expression and instead are characterized by the squared gain function of the scaling filter $\mathcal{G}(f)$:

$$\mathcal{G}(f) = 2\cos^L(\pi f) \sum_{l=0}^{\frac{L}{2}-1} \binom{\frac{L}{2}-1+l}{l} \sin^{2l}(\pi f)$$

Filter coefficients are obtained by factoring the squared gain functions in the frequency domain to obtain transfer functions and taking an inverse Fourier transform of the associated transfer functions. Various factorization criteria lead to different Daubechies-class filter coefficients, however the transfer functions for these filters differ only in their phase functions. For the purposes of exposition I limit my attention to Daubechies’ so-called least-asymmetric (LA) wavelet filters that have approximately linear phase. Regardless of the factorization scheme, the squared gain functions of all Daubechies-class scaling filters converge to ideal low-pass filters – the so called “brick wall filter” – as L increases (Lai 1995). This property extends to wavelet filters as well – as L increases the wavelet filters converge to ideal high-pass and band-pass filters. This leads to a perfect partitioning of the frequency domain, meaning that the wavelet/scaling coefficients (and details/smooth) are perfectly uncorrelated with one another. Figure 2.1 illustrates this convergence for Daubechies wavelet filters. Notice the frequency overlaps that occur for small L ; this is the mechanism that leads to spectral leakage in the wavelet transform and an “approximate”

uncorrelation across levels.

2.4 Some Useful Lemmas and Corollaries

In this section I formalize some useful asymptotic properties of the wavelet/scaling coefficients and details/smooth. The following lemma formalizes the vanishing correlation property:

Lemma 1 *Let $\{X_t\}$ be a stationary stochastic process with a bounded absolutely continuous spectrum. Then any Daubechies-class MODWT wavelet and scaling coefficients for $\{X_t\}$ are asymptotically uncorrelated across scales, that is:*

$$\lim_{\substack{L, N \rightarrow \infty \\ L/N \rightarrow 0}} \{Cov[W_{i,t}, W_{j,u}]\} = 0 \quad \forall i \neq j \text{ and } i, j \leq J_0$$

with an equivalent statement holding for the scaling coefficients V_{J_0} .

Because the details and smooth are linear functions of the wavelet and scaling coefficients, the corollary below follows naturally from Lemma 1:

Corollary 1 *Let $\{X_t\}$ be a stationary stochastic process with a bounded absolutely continuous spectrum. Then any Daubechies-class MODWT details and smooth for $\{X_t\}$ are asymptotically uncorrelated across scales, that is:*

$$\lim_{\substack{L, N \rightarrow \infty \\ L/N \rightarrow 0}} \{Cov[\mathcal{D}_i, \mathcal{D}_j]\} = 0 \quad \forall i \neq j \text{ and } i, j \leq J_0$$

with an equivalent statement holding for the smooth \mathcal{S}_{J_0} .

For the MODWT, there is a distinction between the additive decomposition of a time series via the details and smooth and the variance decomposition via the wavelet and scaling coefficients. However, this distinction asymptotically vanishes; the following lemma establishes that the energy of the level j details (smooth) is asymptotically equivalent to the energy of the level j wavelet (scaling) coefficients:

Lemma 2 *The energy in the wavelet coefficients is asymptotically equal to the energy in the MRA components. That is:*

$$\lim_{\substack{L, N \rightarrow \infty \\ L/N \rightarrow 0}} \left\{ \|W_j\|^2 - \|\mathcal{D}_j\|^2 \right\} = 0$$

It follows that the the MODWT-based MRA is an asymptotic variance decomposition:

Corollary 2 *The MRA decomposition*

$$X = \sum_{j=1}^{J_0} \mathcal{D}_j + \mathcal{S}_j$$

is an asymptotic variance decomposition that is equivalent to the variance decomposition given by the MODWT wavelet coefficients:

$$\text{Var}[X] = \sum_{j=1}^{J_0} \text{Var}[\mathcal{D}_j] + \text{Var}[\mathcal{S}_{J_0}]$$

where

$$\text{Var}[\mathcal{D}_j] = \text{Var}[W_j]$$

2.5 Simulation

As a brief illustration of the Lemmas and Corollaries established above, I simulate a length $n = 1024$ sequence of i.i.d. standard normal random variables $\{X_t\}$ and implement a level $J_0 = 4$ MODWT and MODWT-based MRA using Daubechies extremal phase filters of length $L = 2, 4, 6, 8, 10, 12, 14, 16, 18, 20$ and calculate the absolute difference in the variance of the level j wavelet (scaling) coefficients and the level j details (smooth). Figure 2.2 plots the results, clearly showing a decline in the absolute difference as filter lengths increase, illustrating properties established in Lemma 2 and Corollary 2. To illustrate the properties established in Lemma 1 and Corollary 1, Figures 2.3 and 2.4 plot the determinant of the correlation matrix for the details and smooth. Note that the correlation matrix for perfectly uncorrelated random variables will be the identity matrix and hence the determinant will be

unity. Figure 2.4 clearly shows the correlation in details and smooth declining as the filter size increases. Figure 2.3 shows the so-called “approximate” uncorrelation between wavelet coefficients and a good deal of the variation in the determinant is likely due to numerical error, however a pattern of convergence does emerge from $L = 12$ and up.

2.6 Conclusion

In this chapter I briefly summarize the discrete wavelet transform (DWT), the maximal overlap discrete wavelet transform (MODWT) and their associated multiresolution analyses (MRA). I establish some asymptotic properties of the MODWT wavelet/scaling coefficients as well as the MRA details/smooth, formalizing these properties in several Lemmas and Corollaries. I then concluded with a simulation illustrating these asymptotic properties applied to a Gaussian white noise process.

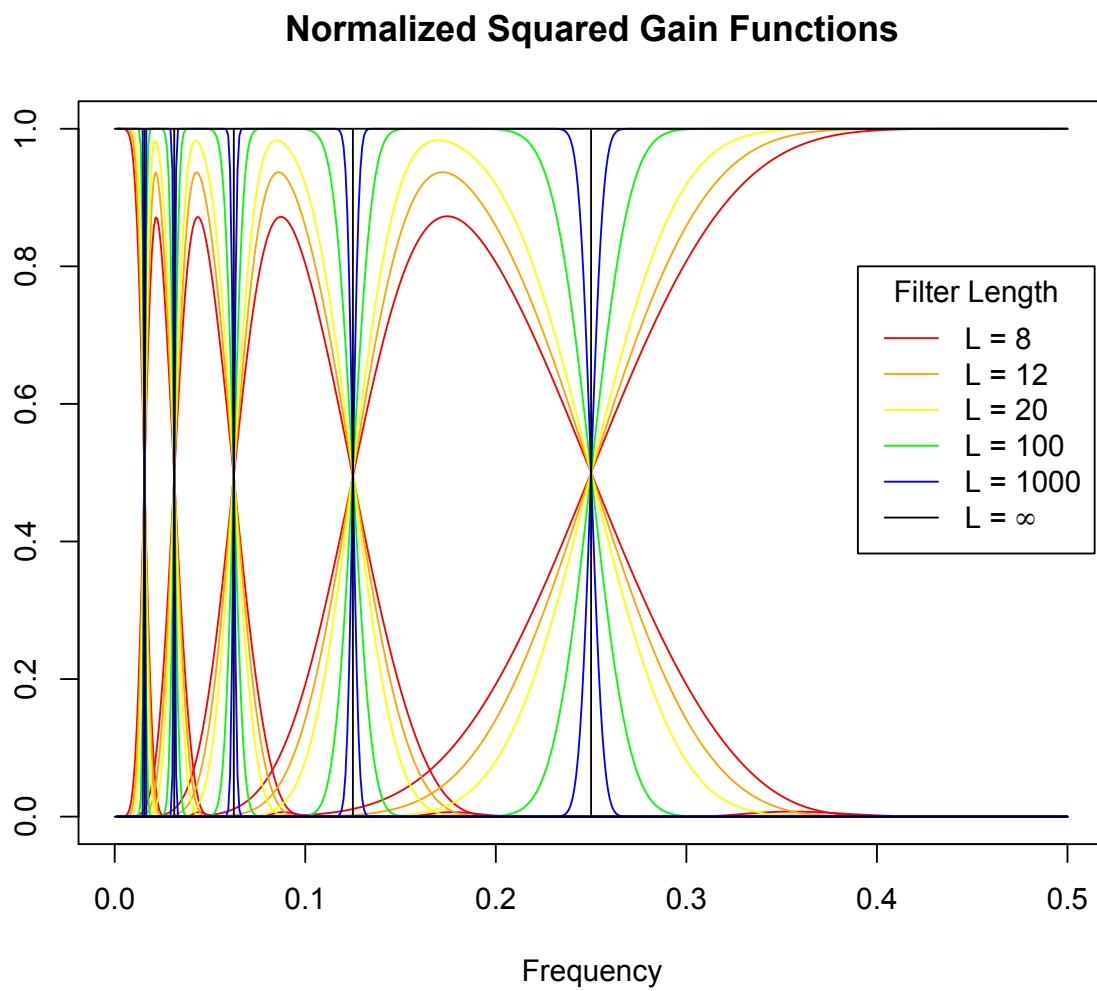


Figure 2.1: Asymptotic Behavior of Daubechies-class of Wavelet Filters

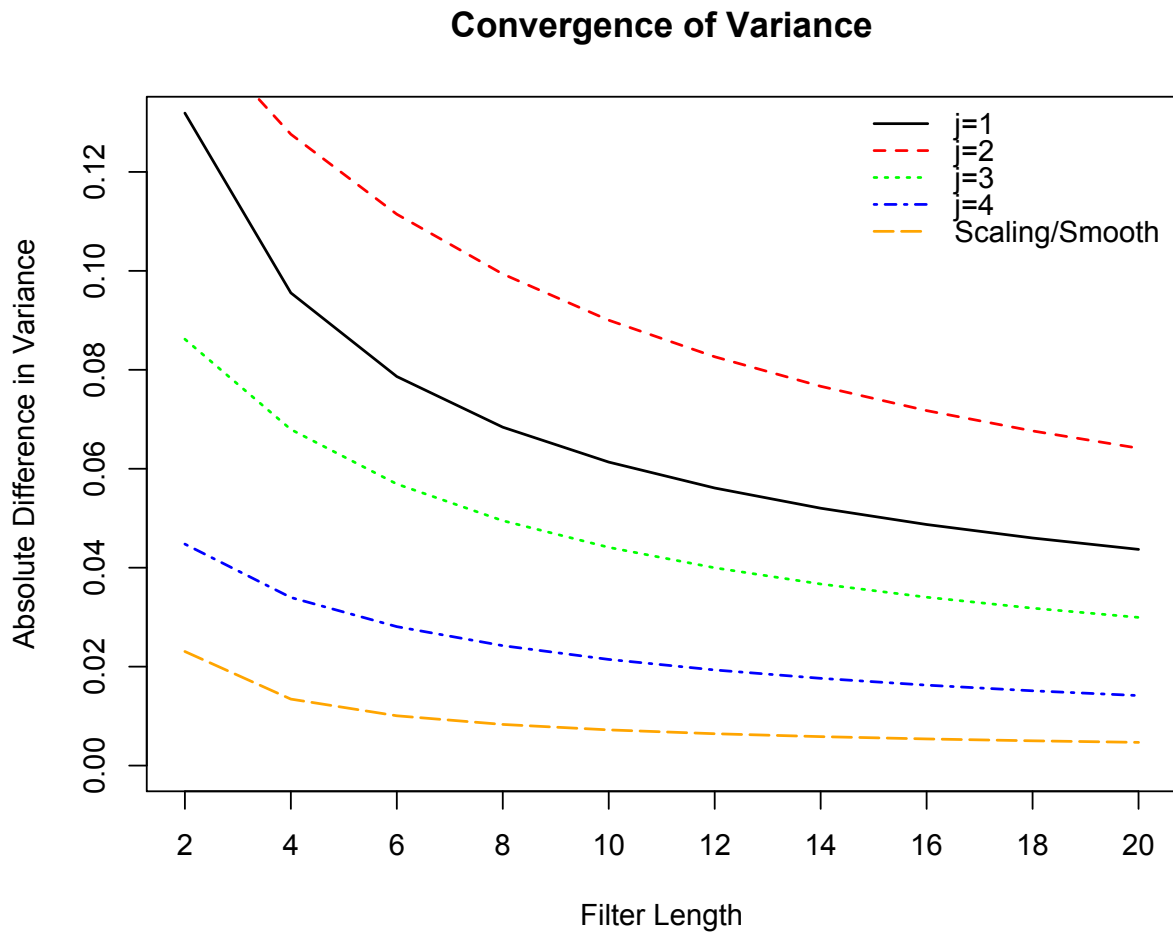


Figure 2.2: Asymptotic Variance Behavior

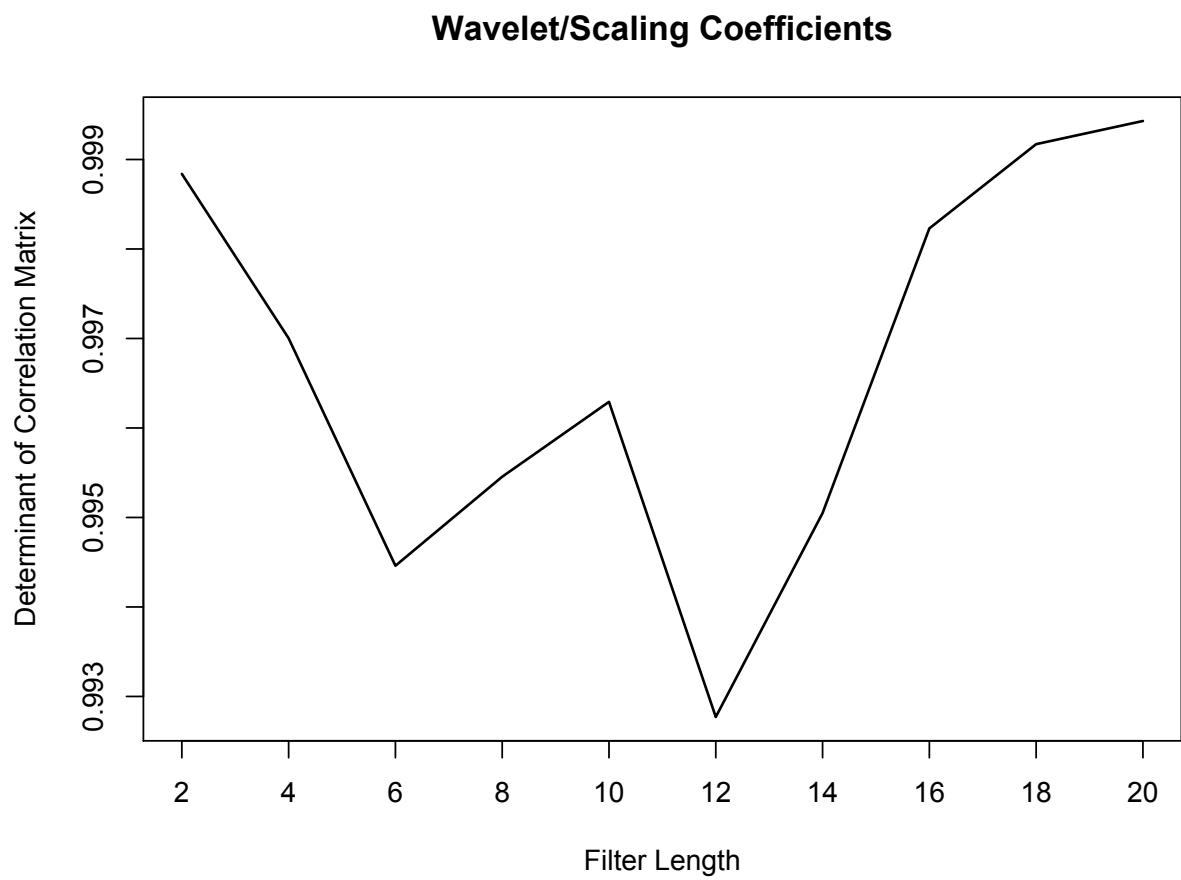


Figure 2.3: Determinant of Wavelet Correlation Matrix

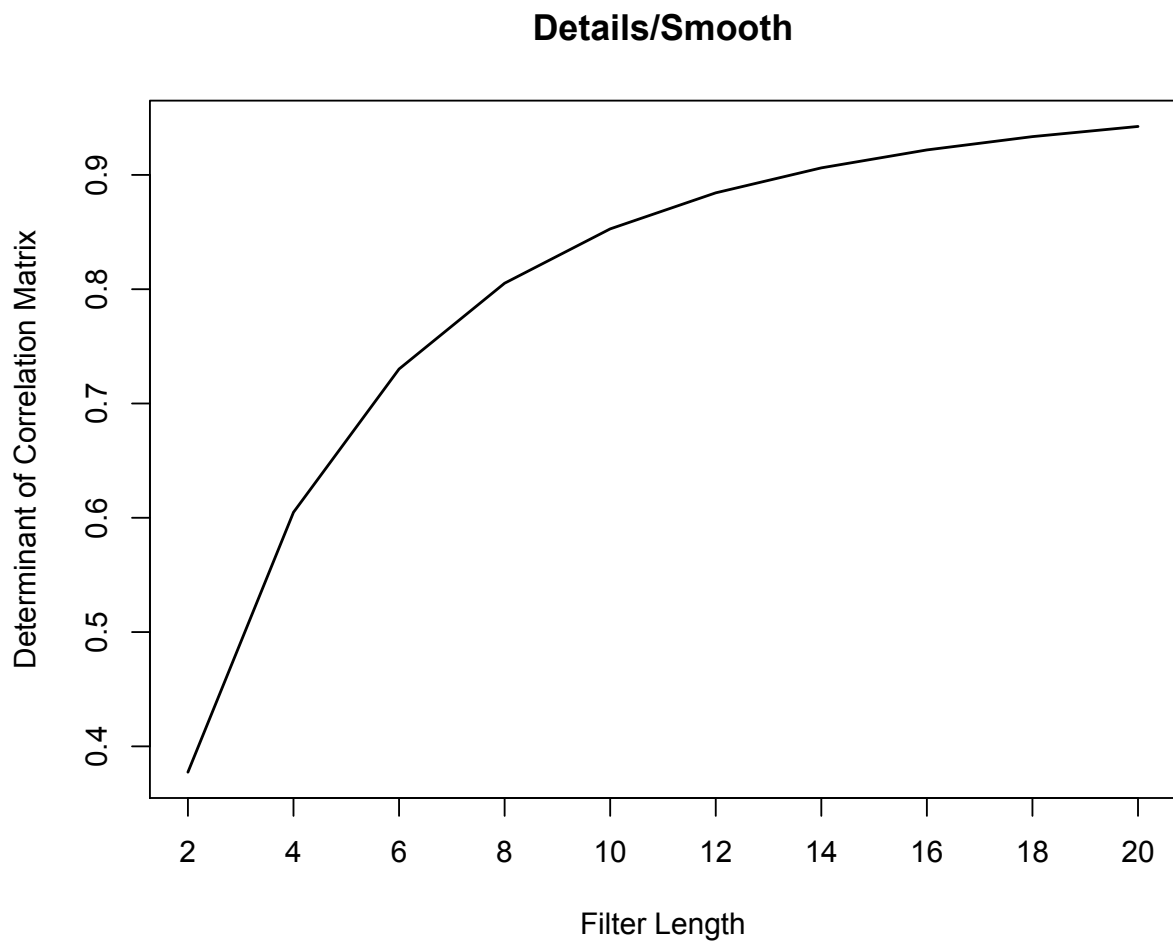


Figure 2.4: Determinant of Details Correlation Matrix

Chapter 3

A MULTIREOLUTION REGRESSION FRAMEWORK

3.1 Introduction

Scale-based econometric methods are an important and growing subset of the literature. In this chapter I develop a robust and easy to implement time scale regression model based on wavelet multiresolution components. I propose an estimator for this class of multiresolution regressions (MRR) and derive its asymptotic properties. For ease of exposition, I present the MRR framework using a single regressor but note that the approach generalizes easily to multiple regression models. Centrally, the MRR framework I present here is semi-parametric in nature. A scale-based decomposition of an independent regression variable is implemented through nonparametric wavelet transforms and the decomposed components are then utilized as new regressors in a parametric linear model. Multiresolution regression has a number of advantages over existing scale-based econometric methods. Arguably the dominant advantage over fully parametric approaches is robustness to scale specification. Model based scale decompositions are often very sensitive to specification of scale dynamics leading to very different results under different specifications. The framework I develop below is also incredibly simple to implement – arguably a huge advantage over complicated, highly parameterized state space models. The framework developed below also has a number of advantages over the long horizon regression approach as well. MRR models include both long-horizon components and orthogonal, high-frequency components commonly thought of as noise. A common assumption in the finance and economics literature is the presence of noise – often high frequency noise – that obscures the signal of interest. This is one of the central motivations for the long-horizon regression approach in the finance literature. From a pure signal-to-noise perspective, MRR offer advantages over other filter-based (parametric or nonparametric) approaches because the decomposed scale-based components of the regressor are all included as instruments. A hypothesis of high frequency noise can then be

rigorously tested using simple t or chi-squared tests. If instead, the high-frequency components *are* correlated with the dependent variable, but perhaps to a weaker degree than longer-horizon components, then the MRR estimates will capture this signal feature.

Quite aside from the above statistical advantages with respect to noise, MRR methods allow unique inference about the scale-based characteristics of economic relationships. By decomposing regressors into approximately uncorrelated scale-based components, I allow the regressor to affect the dependent variable differently over different time horizons. These differential scale-based effects can actually capture very different economic relationships. As an expository example, consider the cross-sectional CAPM in which monthly portfolio excess returns are regressed on monthly excess market returns to measure the portfolios' exposure to the market which is treated as the single relevant risk factor that drives cross-sectional asset price variation. Suppose now that excess market returns are decomposed into short-, medium-, and long-horizon components corresponding to horizons shorter, equal to, and longer than business cycles, respectively. Intuitively, the risks described by covariation of portfolio returns with these components are different¹. Similarly, regressions of uncovered interest parity (UIP) related puzzles (forward premium puzzle, delayed overshooting puzzle, etc.) have intuitive scale-based characteristics – and findings that UIP holds better over longer horizons strongly support this argument. In overshooting and delayed overshooting frameworks (e.g. Dornbush (1976), Eichenbaum and Evans (1995), etc.), explicit scale-based dynamics can emerge and failing to incorporate these dynamics into time series regressions can limit inferential ability. However, incorporating these dynamics correctly is an inherently challenging task. As in the state space framework above, the VAR framework favored by many in the international finance literature (e.g. Cushman and Zha (1997), Faust and Rogers (2003), Kim (2005) and Scholl and Uhlig (2008)) suffers from model specification and identification assumptions leading to often contradictory results.

Finally, as a brief comment, the MRR framework also offers potential advantages over the MIDAS/GADL framework as well. MIDAS/GADL filters inherently restrict the frequency domain characteristics of the filtered regressor. The central identification scheme used in

¹Chapter 4 covers just this topic

MIDAS/GADL models is the downsampling of the filtered regressor, however the frequency characteristics of this filtered regressor have already been truncated when the downsampling operation is performed. Although I do not explicitly develop an MRR analogue to the MIDAS/GADL approach in this chapter, the scale decomposition used in the MRR approach affords the opportunity to downsample each of the scale components separately (possibly stochastically). This procedure is positioned to dramatically improve the aliasing problem present in simple downsampling operations.

For now, however, I limit my attention to developing a simple MRR framework and establish properties and asymptotic behavior of the MRR estimators. The remainder of this chapter is organized as follows: Section 3.2 introduces the MRR model and the MRR estimator. Section 3.3 establishes properties of the MRR estimator while Section 3.4 derives the asymptotic distribution for it. Section 3.5 illustrates these properties via Monte Carlo simulation. Section 3.6 implements a test for high frequency noise in portfolio returns using an MRR model. Section 3.7 concludes.

3.2 The Multiresolution Regression (MRR) Model

For expository purposes, I develop the MRR framework through an extension to the following simple predictive linear regression model:

$$Y_{t+m} = \gamma_0 + \gamma_1 X_t + u_t \quad t = 1, 2, \dots, n, \quad (3.1)$$

which can be written in vector formulation as

$$Y = X\gamma + u$$

The simplicity of the model will allow easier illustration and intuition about the approach while easily generalizing to multivariate regression models. I introduce the MRR framework in the context of predictive regression for two reasons: 1) much of the attention to scale based econometric methods has been in the context of predictive regressions, and 2) the theory developed easily accommodates the case where $m = 0$.

The central departure from this standard linear regression model is the application of a level J_0 MODWT-based MRA decomposition of the regressor X where $J_0 \leq \log_2(n)$ is fixed. The model is then demeaned and the marginal effects of X on Y are allowed to vary across scales:

$$Y_{t+m} - \mu_Y = \sum_{j=1}^{J_0} \beta_j \mathcal{D}_{j,t} + \beta_{J_0+1} \mathcal{S}_{J_0,t} + \nu_t \quad t = 1, 2, \dots, n \quad (3.2)$$

$$Y = \mathcal{D}\beta + \nu$$

where \mathcal{D} is a $n \times (J_0 + 1)$ matrix of scale-based MRA decomposition components²:

$$\mathcal{D} \equiv \begin{bmatrix} \mathcal{D}_1 & \mathcal{D}_2 & \dots & \mathcal{D}_{J_0} & \mathcal{S}_{J_0} \end{bmatrix}$$

where

$$X_t - \mu_X = \sum_{j=1}^{J_0} \mathcal{D}_{j,t} + \mathcal{S}_{J_0,t}$$

In general, the MRA decomposition used to construct the matrix \mathcal{D} will follow from a MODWT filter of length L where $L < \infty$ (finite impulse response). The finite impulse response (FIR) MRR estimator then is simply the OLS estimator of equation (3.2):

$$\hat{\beta} = (\mathcal{D}'\mathcal{D})^{-1}\mathcal{D}'Y \quad (3.3)$$

Although the columns of \mathcal{D} are approximately orthogonal to one another, the small amount of spectral leakage resulting from finite length filters results in a small – but nonzero – amount of correlation between columns. However, this remaining correlation vanishes as the filter length increases. Consequently, the FIR MRR takes the form above but will collapse into a simpler, component-by-component OLS estimator in the limit. This asymptotic result is explored further below.

A brief comment about the demeaning of the model is in order. The demeaning is done primarily for technical reasons – certain spectral estimators utilized in the asymptotic

² \mathcal{D} is allowed to be composed of pass-band aggregated components, in which case the number of columns would be equal to $k + 1$ where k is the number of components remaining after aggregation.

theory developed below are well-behaved when the mean of the series is known (or without loss of generality, assumed to be zero). However, the central goal of the MRR framework is inference about differential marginal effects across time horizons – estimates of the regression coefficients will be identical to those from a model that has not been demeaned but spectral estimates will inherit well-behaved properties under the know mean of zero. Nevertheless, if an intercept estimate is required, one can be backed out of the model. By construction, the level j details \mathcal{D}_j will have a mean of zero for $j = 1, 2, \dots, J_0$ while the smooth \mathcal{S}_{J_0} will have a mean of μ_X . An ad hoc estimate of an intercept parameter is then given by $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_{J_0+1}\bar{X}$.

3.3 Properties of the MRR Estimator

It is also worth noting that equation (3.1) is an econometric model – not a description of the underlying data generating process (DGP). Consequently, I proceed by making the following very general assumption about X :

Assumption 1 $\{X_t, t \in \mathbb{Z}\}$ is a real-valued linear processes given by:

$$X_t - \mu_X = \sum_{l=0}^{\infty} \psi_l \epsilon_{t-l} \quad (3.4)$$

The innovations $\{\epsilon_t\}$ is an ergodic martingale difference sequence with $E[\epsilon_t^2] = \sigma_\epsilon^2 < \infty$, $E[|\epsilon_t|^4] < \infty$, and $\sum_{l \in \mathbb{Z}} |\psi_l| < \infty$.

I consider the properties of the MRR estimator in equation (3.3) under three data generating processes. The first DGP I consider describes a correctly specified model, that is, the true DGP is given by:

$$Y = \mathcal{D}\beta + \nu \quad (3.5)$$

where \mathcal{D} is an $n \times (J_0 + 2)$ matrix of a constant and $J_0 + 1$ scale-based components that partition the frequency domain. The second DGP I consider lacks any scale characteristics at all and is equivalent to a standard linear regression model:

$$Y = X\gamma + u \tag{3.6}$$

where X is an $n \times 2$ matrix of a constant and a single, scale-independent regressor. The final DGP I consider has scale characteristics but the frequency domain partitioning is assumed to be different than that in equation (3.5):

$$Y = \mathcal{B}\theta + \zeta \tag{3.7}$$

where \mathcal{B} is an $n \times (k + 1)$ matrix of a constant and k scale-based components that orthogonally partition the frequency domain but with bandpass characteristics different from that of \mathcal{D} , i.e. \mathcal{B} is a different frequency domain partitioning. In each DGP above, the error terms are assumed to be (weakly) stationary and ergodic. Specifically, the errors are assumed to be covariance stationary processes whose Wold representation innovations are martingale difference sequences (MDS) with finite fourth moments and absolutely convergent MA coefficients.

3.3.1 Unbiased and Consistent

Unbiasedness and consistency are particularly desirable properties of statistical estimators. In truth, unbiasedness can almost solely be shown to hold under correct model specification and, indeed, this is also the case for the MRR estimator. However, an additional assumption must be made for the MRR estimator to be unbiased under DGP (3.5) (i.e. correct model specification). The MRR estimator in equation (3.3) is an OLS estimator and hence, requires the exogeneity of the regressors in order for unbiasedness to hold. However, because the regressors \mathcal{D} are scale-based components of a single regressor X , the ordinary exogenous regressor assumption becomes slightly more sophisticated, as detailed below.

Assumption 2 *The regressors \mathcal{D} are strictly exogenous, i.e. $E[\nu|\mathcal{D}] = 0$. Since \mathcal{D} is a frequency domain partitioning of the regressor X this exogeneity assumption can be categorized as a “scale exogeneity” assumption on X . That is, X is assumed to be uncorrelated with ν over all frequency bands in \mathcal{D} .*

The “scale exogeneity” assumption above is a stricter form of exogeneity than is typically assumed. The typical scale-independent exogeneity assumption implies $Cov[X, \nu] = 0$ but the MRA decomposition of X implies that this is a necessary, but insufficient condition for unbiased and consistent estimates. For simplicity, consider an MRA decomposition in which $J_0 = 1$, then $X_t = \mathcal{D}_{1,t} + \mathcal{S}_{1,t}$. An error term ϵ_t is “scale exogenous” if $Cov[\mathcal{D}_{1,t}, \epsilon_t] = 0$ and $Cov[\mathcal{S}_{1,t}, \epsilon_t] = 0$. The typical exogeneity assumption $Cov[X_t, \epsilon_t] = 0$ is implied by this condition but the converse is not true.

The unbiasedness of the MRR estimator under a correctly specified model follows immediately from the assumption above, however establishing consistency requires the following lemmas:

Lemma 3 *Let $\{X_t\}$ be defined as in Assumption 1 with $\{\mathcal{D}_{j,t}\}$ be the level j details from a Daubechies-class MODWT-based filtering operation on $\{X_t\}$ with filter length L . Let $\{U_t\}$ be a stationary linear process with a purely continuous spectrum with spectral density function $S_U(f)$ and an absolutely summable cross covariance sequence with $\{X_t\}$: $\sum_{\tau \in \mathbb{Z}} s_{XU,\tau} < \infty$. If $\{\mathcal{Q}_t\}$ is a band-limited realization (i.e. filtered version) of $\{U_t\}$ such that the filtering operation is asymptotically (in n) a zero-phase ideal bandpass filter with passband $\mathcal{F}_{\mathcal{Q}} \in [-1/2, 1/2]$, then under the condition $L, n \rightarrow \infty$ and $L/n \rightarrow 0$:*

$$\frac{1}{n} \sum_{t=1}^n \mathcal{D}_{j,t} \mathcal{Q}_t \xrightarrow{p} 2 \int_{\mathcal{F}_{\mathcal{Q}\mathcal{D}_j}} C_{XU}(f) df$$

where $\mathcal{F}_{\mathcal{Q}\mathcal{D}_j} \equiv \mathcal{F}_{\mathcal{Q}} \cap (1/2^{j+1}, 1/2^j]$ and $C_{XU}(f)$ is the real part of the cross spectrum – or cospectrum – between $\{X_t\}$ and $\{U_t\}$. In particular, if $\{\mathcal{Q}_t\}$ is a band-limited realization of $\{X_t\}$, then $\frac{1}{n} \sum_{t=1}^n \mathcal{D}_{j,t} \mathcal{Q}_t \xrightarrow{p} 2 \int_{\mathcal{F}_{\mathcal{Q}\mathcal{D}_j}} S_X(f) df$. An equivalent statement holds for the smooth $\{S_{J_0,t}\}$ with $\mathcal{F}_{\mathcal{Q}S_{J_0}} \equiv \mathcal{F}_{\mathcal{Q}} \cap [0, 1/2^{J_0}]$.

Lemma 4 *Under the condition $L, n \rightarrow \infty$ and $L/n \rightarrow 0$, the matrix $n^{-1} \mathcal{D}' \mathcal{D}$ converges in probability to a diagonal matrix with the partitioned auto spectrum of X along the diagonal as in lemma 3. Similarly, the matrix $\mathcal{D}' Y$ converges in probability to a $(J_0 + 1) \times 1$ matrix whose elements partition the cospectrum, again as in lemma 1. This defines an asymptotic*

– or infinite impulse response (IIR) – MRR estimator:

$$\hat{\beta}_j = \frac{\sum_{t=1}^n \mathcal{D}_{j,t} Y_t}{\sum_{t=1}^n \mathcal{D}_{j,t}^2}$$

With these lemmas, the consistency and unbiasedness of the MRR estimator can be established for correctly specified models in the following theorem:

Theorem 1 *Let $\{X_t\}$ be defined as in Assumption 1 and $\{Y_t\}$ be defined as in DGP (3.5) with ν a stationary and ergodic MDS with $E[\nu|\mathcal{D}] = 0$. Then the MRR estimator in equation (3.3) is an unbiased and consistent estimate of β in equation (3.5).*

A particularly attractive property of the MRR estimator is that it is also asymptotically robust to certain classes of model misspecification. In the typical time series context, asymptotic results emerge as the sample size approaches infinity $n \rightarrow \infty$, however MRR regressors are the output of filtering operations and the notion of asymptotics that I use here is slightly different than that addressed in most of the econometric literature – here I use the term asymptotic to refer to both the sample size n and the filter length L going to infinity with the condition that the sample size grows faster than the filter length. The following theorem formalizes the robust consistency of the MRR estimator under this asymptotic framework:

Theorem 2 *Let $\{X_t\}$ be defined as in Assumption 1 and $\{Y_t\}$ be defined as in DGP (3.7) with ζ a stationary and ergodic MDS satisfying $E[\zeta|\mathcal{D}] = 0$. Then the MRR estimator $\hat{\beta}_j$ from equation (3.3) is a consistent estimate of θ_k in equation (3.7) under the condition $L, n \rightarrow \infty$ with $L/n \rightarrow 0$, provided the frequency band of the k th component of \mathcal{B} in equation (3.7) covers the frequency band of \mathcal{D}_j .*

This is a powerful result – in particular, the above theorem covers the case in DGP (3.6) where the “frequency band” of the regressor under the DGP is the full frequency domain. Clearly the full domain covers any frequency band of \mathcal{D}_j , leading to the following corollary:

Corollary 3 *Theorem 2 implies that under the condition $L, n \rightarrow \infty$ with $L/n \rightarrow 0$ the MRR estimator $\hat{\beta}_j$ converges to γ_1 under the DGP in equation (3.6) provided that $E[\nu|\mathcal{D}] = 0$.*

Theorem 2 and Corollary 3 establish the MRR estimator as asymptotically robust to a broad class of model misspecification – inappropriate application of the scale based decomposition still leads to asymptotically valid parameter estimates. This is particularly encouraging as, in practice, the presence – and exact structure – of scale-based economic relationships is unlikely to be known a priori. If, for example, an MRR model is inappropriately applied to a scale-independent regression relationship, the parameter estimates will still converge to the same (true) value. As mentioned above, specification issues in parametric scale models can lead to very different parameter estimates – not only are MRR methods robust to how scale components are modeled, they are robust to the types of scale misspecification above as well.

Theorem 2 and Corollary 3 establish *asymptotic* robustness to certain types of misspecification. However, in practice, the MRR estimator is also *approximately* unbiased for the coefficient vectors γ and θ in equations (3.6) and (3.7) (in the sense described above). Strictly speaking the MRR estimator in equation (3.3) is in fact biased for the coefficient vectors γ and θ ; however, because this bias comes from the fact that $(\mathcal{D}'\mathcal{D})^{-1}\mathcal{D}'X \neq I$ (likewise with $(\mathcal{D}'\mathcal{D})^{-1}\mathcal{D}'\mathcal{B}$), it is quite small. Recall that the columns of \mathcal{D} are *approximately* uncorrelated with one another and are filtered versions of X . As a result, $\mathcal{D}'\mathcal{D}$ is *approximately* diagonal and \mathcal{D}'_jX is *approximately* $\mathcal{D}'_j\mathcal{D}_j$. Consequently, there is bias under model misspecification but in practice, it is quite small. Figure 3.3 illustrates this phenomenon for a filter length of $L = 16$.

3.4 Asymptotic Distribution

I derive the asymptotic distribution of the MRR estimator in a particular way – building it up with rather general lemmas. I do this primarily for three reasons: 1) the results involve a general multivariate asymptotic distribution for the wavelet variance and wavelet covariance jointly – a result that contributes directly to the wavelet literature, 2) the asymptotic behavior of the MRR estimator is significantly simplified by directly accounting for the limiting behavior of the wavelet filters, and finally 3) deriving the asymptotic distribution in this way explicitly controls for serial correlation and heteroskedasticity – even of the conditional variety. In order to proceed in this manner, I extend Assumption 1 about $\{X_t\}$

to $\{Y_t\}$ with the following assumption.

Assumption 3 $\{X_t, t \in \mathbb{Z}\}$ (as defined as in Assumption 1) and $\{Y_t, t \in \mathbb{Z}\}$ are two real-valued linear processes given by:

$$X_t - \mu_X = \sum_{l=0}^{\infty} \psi_l \epsilon_{t-l} \quad (3.8)$$

$$Y_t - \mu_Y = \sum_{k=0}^{\infty} \phi_k \eta_{t-k} \quad (3.9)$$

The vector of innovations $[\epsilon_t \ \eta_t]'$ is an ergodic martingale difference sequence:

$$\begin{bmatrix} \epsilon_t \\ \eta_t \end{bmatrix} \sim MDS \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\epsilon^2 & \Delta \\ \Delta & \sigma_\eta^2 \end{bmatrix} \right)$$

with $E[\epsilon^4] < \infty$ with $\sum_{l \in \mathbb{Z}} |\psi_l| < \infty$ and $E[\eta^4] < \infty$ with $\sum_{l \in \mathbb{Z}} |\phi_l| < \infty$.

In general, the Wold representation theorem guarantees any covariance stationary $\{Y_t\}$ can be represented by (3.9) with square summable MA coefficients and white noise (but not necessarily i.i.d.) innovations. It is a small step to make the additional assumptions above. Indeed, Assumption 3 is satisfied under all DGPs considered in this chapter. To illustrate this point, consider DGP (3.6) and without loss of generality, let $\mu_X = \mu_Y = 0$. Assumption 1 implies that under DGP (3.6), Y_t is given by:

$$Y_t = \beta \sum_{l=0}^{\infty} \psi_l \epsilon_{t-l} + u_t$$

and since $\{u_t\}$ is assumed to be covariance stationary:

$$Y_t = \beta \sum_{l=0}^{\infty} \psi_l \epsilon_{t-l} + \sum_{k=0}^{\infty} \xi_k v_{t-k} = \sum_{j=0}^{\infty} \beta \psi_j \epsilon_{t-j} + \xi_j v_{t-j}$$

which implies that η_t in Assumption 3 is a composite error term. It is worth noting that I could proceed using the multivariate Wold representation of $[Y_t \ X_t]'$ based on a DGP for Y_t , however I aim to derive the asymptotic distribution for the MRR estimator under all of

the DGP considered, rather than under a particular DGP. Consequently, I proceed under Assumption 3 – which is quite general in the context of linear DGP.

3.4.1 Central Limit Theorems

Define a statistic Z_m as follows:

$$n\bar{Z}_m = \begin{bmatrix} n\bar{Z}_{1,m} \\ n\bar{Z}_2 \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^n Z_{1,m,t} \\ \sum_{t=1}^n Z_{2,t} \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^n X_t Y_{t+m} \\ \sum_{t=1}^n X_t^2 \end{bmatrix} \quad (3.10)$$

Lemma 5 *Let $\{X_t\}$ and $\{Y_t\}$ be defined as in assumption 3. Without loss of generality let $\mu_X = \mu_Y = 0$. Then for Z_m defined above:*

$$\sqrt{n}(\bar{Z}_m - \bar{Z}_m^*) \xrightarrow{d} N(\mathbf{0}, \Omega) \quad (3.11)$$

where

$$\bar{Z}_m^* = [\text{Cov}(X_t, Y_{t+m}) \quad \text{Var}(X_t)]'$$

and

$$\Omega = \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{bmatrix}$$

and

$$\omega_{11} = \sum_{k \in \mathbb{Z}} s_{X,k} s_{Y,k} + \sum_{k \in \mathbb{Z}} s_{XY,k} s_{XY,2m-k} + \left(\frac{E[\epsilon_0^2 \eta_0^2]}{\Delta^2} - \frac{\sigma_\epsilon^2 \sigma_\eta^2}{\Delta^2} - 2 \right) s_{XY,m}^2$$

$$\omega_{12} = \omega_{21} = 2 \sum_{k \in \mathbb{Z}} s_{XY,k+m} s_{X,k} + \left(\frac{E[\epsilon_0^3 \eta_0]}{\Delta \sigma_\epsilon^2} - 3 \right) s_{XY,m} s_{X,0}^2$$

$$\omega_{22} = 2 \sum_{k \in \mathbb{Z}} s_{X,k}^2 + \left(\frac{E[\epsilon_0^4]}{\sigma_\epsilon^4} - 3 \right) s_{X,0}^2$$

Corollary 4 *The statistic $\sqrt{n}(\bar{Z}_{1,m}/\bar{Z}_2 - \text{Cov}[X_t, Y_{t+m}]/\text{Var}[X_t])$ is asymptotically normal with mean zero and variance:*

$$\xi_m^2 = \nabla f(\bar{Z}_m^*)' \Omega \nabla f(\bar{Z}_m^*) = \frac{\omega_{11}}{\text{Var}[X_t]^2} - \frac{2\omega_{12}\text{Cov}[X_t, Y_{t+m}]}{\text{Var}[X_t]^3} + \frac{\omega_{22}\text{Cov}[X_t, Y_{t+m}]^2}{\text{Var}[X_t]^4}$$

where $f(\bar{Z}_m) = \bar{Z}_{1,m}/\bar{Z}_2$.

In general, Ω (and hence ξ_m^2) is not in an estimable form. Serroukh and Walden (2000) and Serroukh, Walden, and Percival (2000) utilize spectral estimators for ω_{11} and ω_{22} . It remains to establish that ω_{12} can also be estimated with spectral methods; the following lemma will be a useful step towards establishing this.

Lemma 6 *Let $\{X_t\}$ and $\{Y_t\}$ be defined as in Assumption 3. Without loss of generality, let $\mu_X = \mu_Y = 0$. Then $\sum_{k \in \mathbb{Z}} |s_{XY,k}| < \infty$ and consequently:*

$$\lim_{n \rightarrow \infty} n \text{Cov}[\bar{X}, \bar{Y}] = \sum_{k \in \mathbb{Z}} s_{XY,k}$$

The above lemma describes the asymptotic covariance between two sample means – a critical step in understanding the properties of the asymptotic variance covariance matrix Ω in equation (3.11). Specifically, the ω_{12} term is an asymptotic covariance between the sample mean for $Z_{1,m}$ and the sample mean for Z_2 . The following lemma establishes the absolute summability of the cross covariance function for $Z_{1,m}$ and Z_2 which will in turn allow use of spectral estimation methods for the Ω matrix.

Lemma 7 *Let $\{X_t\}$ and $\{Y_t\}$ be defined as in assumption 3 and without loss of generality let $\mu_X = \mu_Y = 0$. Then for $Z_{1,m}$ and Z_2 defined as in equation (3.10), the cross covariance sequence $\{s_{Z_1 Z_2, k}\}$ is absolutely summable:*

$$\sum_{k \in \mathbb{Z}} |s_{Z_1 Z_2, k}| < \infty$$

where $s_{Z_1 Z_2, k} = \text{Cov}[Z_{1,t,m}, Z_{2,t+k}]$

Absolute summability of an auto (cross) covariance sequence is a sufficient condition to establish the (cross) spectral density function as a Fourier pair with the auto (cross)

covariance sequence. Now consider the auto covariance matrix for $Z_m = [Z_{1,m} \ Z_2]'$:

$$\mathbf{s}_{Z,k} = \begin{bmatrix} s_{Z_1,k} & s_{Z_1 Z_2,k} \\ s_{Z_2 Z_1,k} & s_{Z_2,k} \end{bmatrix}$$

Serroukh and Walden (2000) show that Assumption 3 is sufficient for $\sum_{k \in \mathbb{Z}} |s_{Z_1,k}| < \infty$ and $\sum_{k \in \mathbb{Z}} |s_{Z_2,k}| < \infty$. This finding, in conjunction with Lemma 7, establishes the spectral density matrix of $Z_m = [Z_{1,m} \ Z_2]'$ as the Fourier transform of $\mathbf{s}_{Z,k}$:

$$S_Z(f) = \begin{bmatrix} \sum_{k \in \mathbb{Z}} s_{Z_1,k} e^{-i2\pi f k} & \sum_{k \in \mathbb{Z}} s_{Z_1 Z_2,k} e^{-i2\pi f k} \\ \sum_{k \in \mathbb{Z}} s_{Z_2 Z_1,k} e^{-i2\pi f k} & \sum_{k \in \mathbb{Z}} s_{Z_2,k} e^{-i2\pi f k} \end{bmatrix} = \begin{bmatrix} S_{Z_1,m}(f) & S_{Z_1 Z_2}(f) \\ S_{Z_1 Z_2}(f) & S_{Z_2}(f) \end{bmatrix}$$

which, evaluated at $f = 0$, leads to

$$S_Z(0) = \begin{bmatrix} \sum_{k \in \mathbb{Z}} s_{Z_1,k} & \sum_{k \in \mathbb{Z}} s_{Z_1 Z_2,k} \\ \sum_{k \in \mathbb{Z}} s_{Z_2 Z_1,k} & \sum_{k \in \mathbb{Z}} s_{Z_2,k} \end{bmatrix}$$

This property, along with Lemmas 6 and 7, and Serroukh and Walden (2000), establish that $S_Z(0)$ above is a valid (and computationally feasible) estimator for Ω in equation (3.11). The following corollary extends this fact to the distribution of the statistic $\bar{Z}_{1,m}/\bar{Z}_2$:

Corollary 5 *The asymptotic variance of the statistic $\bar{Z}_{1,m}/\bar{Z}_2$ is given by:*

$$\xi_m^2 = \frac{S_{Z_1,m}(0)}{\text{Var}[X_t]^2} - \frac{2S_{Z_1 Z_2,m}(0)\text{Cov}[X_t, Y_{t+m}]}{\text{Var}[X_t]^3} + \frac{S_{Z_2}(0)\text{Cov}[X_t, Y_{t+m}]^2}{\text{Var}[X_t]^4}$$

3.4.2 Asymptotic Distribution of MRR Estimator

Finally, with these preliminaries concluded, the asymptotic distribution of the MRR estimator is established with the following theorem:

Theorem 3 *Let $\{X_t\}$ and $\{Y_t\}$ be defined as in Assumption 3. Without loss of generality let $\mu_X = \mu_Y = 0$ and let $\hat{\beta} = (\mathcal{D}'\mathcal{D})^{-1}\mathcal{D}'Y$ be the MRR estimator. Let β_o be the (possibly redundant) length $J_0 + 2$ vector of limits to the MRR estimator such that as $L, n \rightarrow \infty$ while $L/n \rightarrow 0$, $\hat{\beta}_j \rightarrow \beta_{o,j}$. Then $\sqrt{n}(\hat{\beta} - \beta_o)$ is asymptotically (as $L, n \rightarrow \infty$ and $L/n \rightarrow 0$)*

normal with mean 0 and variance equal to the diagonal matrix $S_\beta(0)$, the j th diagonal of which is given by:

$$\frac{S_{\mathcal{D}_j Y, m}(0)}{\left(\int_{1/2^{j+1}}^{1/2^j} S_X(f) df\right)^2} - \frac{2S_{\mathcal{D}_j Y, \mathcal{D}_j^2, m}(0) \int_{1/2^{j+1}}^{1/2^j} C_{XY, m}(f) df}{\left(\int_{1/2^{j+1}}^{1/2^j} S_X(f) df\right)^3} + \frac{S_{\mathcal{D}_j^2}(0) \left(\int_{1/2^{j+1}}^{1/2^j} C_{XY, m}(f) df\right)^2}{\left(\int_{1/2^{j+1}}^{1/2^j} S_X(f) df\right)^4}$$

where $S_{(\cdot)}(f)$ denotes the power spectral density function and $C_{(\cdot)}(f)$ denotes the cospectrum. An equivalent statement holds for the last column in \mathcal{D} composed of the MRA smooth.

Because this variance is formulated from zero-frequency spectral estimators, it directly accounts for any serial correlation or heteroskedasticity; any estimate thereof will be a HAC-type variance estimator. Additionally, the regressors in an MRR model are generated regressors which OLS standard errors fail to account for. However, the variance expression above directly accounts for variation in the generated regressors using spectral estimators. Although this asymptotic variance might look convoluted, it is actually quite computationally compact with a well established methodology for estimating in a reliable manner – the subject of section 3.4.3.

3.4.3 Spectral Density Estimation

As detailed above, computing asymptotic standard errors for coefficient estimates requires estimation of several spectral density functions at zero frequency. A popular and widely adopted methodology in the statistics and physical sciences literature is the multitaper approach (e.g. Serroukh, Walden, and Percival (2000), Serroukh and Walden (2000), and Serroukh (2012)) which I briefly introduce below.

The benchmark spectral density estimator – the periodogram – suffers from bias due to spectral leakage. Solutions to this problem called modified periodograms or *direct spectral estimates* involve the use of data tapers – sequences of real-valued constants $\{a_t\}$ that are used in conjunction with a time series $\{X_t\}$ to form a new sequence $\{a_t X_t\}$. The spectral estimate of the sequence $\{a_t X_t\}$ is then called a direct spectral estimate $\hat{S}_X^{(d)}(f)$ of the spectral density function of X_t . The tapering procedure is interpreted in the frequency domain

as reducing the spectral leakage properties (and hence bias) inherent in the periodogram. The central idea behind multitaper spectral estimation is the averaging of multiple direct spectral estimates formed from orthogonal data tapers, thereby reducing both bias and variance in the spectral estimate, ie.

$$\hat{S}_X^{(mt)}(f) = \frac{1}{K} \sum_{k=0}^{K-1} \hat{S}_{X,k}^{(d)}(f)$$

Percival and Walden (1993) observe that a direct spectral estimate at zero frequency is distributed $\hat{S}^{(d)}(0) \sim S(f)\chi_1^2$, while a multitaper estimate is distributed $\hat{S}^{(mt)}(0) \sim S(f)\chi_K^2$ for orthogonal tapers when the mean of the time series is known. A popular class of data taper used in the multitaper approach is the Slepian tapers – tapers formed from discrete prolate spheroidal sequences (dpss's). Following Serroukh, Walden, and Percival (2000), Serroukh and Walden (2000), and Serroukh (2012) I estimate the spectral density functions above using multitaper estimates based on $K = 5$ Slepian tapers with resolution bandwidth $2W = 9/n$. For more details on multitaper methods and Slepian tapers in particular, the reader is referred to Percival and Walden (1993).

3.5 Simulation

To illustrate the statistical properties and asymptotic results derived above I simulate data under a variety of DGP. Specifically, I simulate data under DGPs (3.5) and (3.6) under i.i.d. error processes, $AR(1)$ error processes, and heteroskedastic error processes. In order to investigate both the behavior of the MRR estimator as well as the proposed estimator for the asymptotic variance thereof, I implement the following Monte Carlo procedure:

1. Generate a length n $AR(1)$ sequence $\{X_t\}$ with autoregressive parameter 0.8 and $N(0, 2)$ innovations.
2. Simulate a sequence $\{Y_t\}$ parametrically under DGP (3.5) and (3.6) with the three types of error processes considered. γ in DGP (3.6) is chosen to be 1 and β in DGP (3.5) is chosen to be $[3 \ 9 \ 1 \ -3 \ 1/2]'$. The i.i.d. errors are generated $N(0, 2)$. The

$AR(1)$ errors are generated with an autoregressive parameter of 0.7 and $N(0, 1)$ innovations. The heteroskedastic errors are generated $N(0, 2\{\mathbf{1}_{(t \leq n/4)}\} + 1\{\mathbf{1}_{(n/4 < t \leq n/2)}\} + 3\{\mathbf{1}_{(1/2 < t \leq 3n/4)}\} + 1\{\mathbf{1}_{(3n/4 < t \leq n)}\})$.

3. Calculate the MRR coefficients and standard errors for a level 4 MRR model with filter length L .
4. Repeat steps 1-3 500 times.

The procedure above results in simulated distributions for both the MRR coefficients as well as the estimated standard errors. The unbiasedness and/or consistency of the MRR estimators is observable via the average of the simulated coefficient distribution, while the performance of the asymptotic standard errors is observable by comparing the standard deviation of the simulated coefficient distributions to the average of the estimated standard errors.

Table 3.1 presents the results under a correctly specified MRR model (i.e. Y_t is generated as in DGP (3.5)) with a relatively small sample size of $n = 64$ and the shortest Daubechies-class filter of $L = 2$. Clearly, even with short filters and small sample sizes, the MRR estimator is unbiased for correctly specified models. The primary shortcoming highlighted by table 3.1 is the relatively poor performance of the estimated asymptotic standard errors. But indeed this is entirely expected as the estimated standard errors are explicitly *asymptotic* standard errors. Table 3.2 illustrates the improvement in estimated standard error performance as the filter length and sample size increase ($L = 256$ and $n = 4096$).

Of potentially greater interest is the performance of the MRR estimator under *incorrectly* specified scale models. As established in Theorem 2, the MRR estimator is still consistent under certain types of misspecification – particularly the type represented by fitting of an MRR model to DGP (3.6) (i.e. no scale structure in the data). However, simulation of the asymptotic behavior as defined in this paper ($L, n \rightarrow \infty$ while $L/n \rightarrow 0$) requires the ability to generate Daubechies-class filters of arbitrarily large size; unfortunately large L Daubechies wavelet filters can be troublesome to compute. Because the filter coefficients themselves do not, in general, have closed form expressions, the coefficients are typically

computed from different solution criteria for solving the complex polynomial function defined by factoring the (closed form) squared gain function. Most major software packages implementing wavelet filters limit functionality to predetermined filter length (typically $L = 20$). However, because the MRR approach works with the MODWT-based details and smooth, rather than the wavelet coefficients themselves, a convenient workaround to the Daubechies factorization problem is available. Recall that the transfer function for the level j MODWT-based details is equal to the *squared gain* function of the level j wavelet filter, i.e. closed form expressions for the Daubechies *details* and *smooth* filters are available. In order to compute the level $j = 1, 2, \dots, J_0$ details and smooth based on longer filter lengths ($L > 20$) required for asymptotic evaluation of the MRR estimator, I utilize the following steps:

1. Evaluate the Daubechies squared gain function for filter length L at the Fourier frequencies $\{k/n\}$ for $k = 0, 1, 2, \dots, n - 1$.
2. Take the inverse Fourier transform of the squared gain function to obtain the filter coefficients associated with a length L details (smooth) filter.³
3. To apply the details and smooth filters, the original time series X_t is circularly filtered with the impulse response sequence (filter coefficients) derived above.

This approach is similar to that taken in Percival et al. (2011) who create composite MODWT-based details filters in the frequency domain and obtain impulse response sequences via inverse Fourier transforms. Tables 3.3 - 3.5 illustrate that even for a relatively shorter filter length of $L = 16$ and a moderate sample size $n = 1024$, the small sample misspecification bias has already all but vanished and the estimated asymptotic standard errors perform quite well. Tables 3.6 - 3.7 illustrate the same for filter length $L = 128$. The asymptotic behavior of the estimated standard errors under different error structures is summarized in Figure 3.1 which plots the sum over all coefficients ($j = 1, 2, \dots, J_0 + 1$)

³In fact the inverse Fourier transform results in an impulse response sequence of length n , though $L - n$ of the coefficients will be numerically zero within reasonable bounds of computational rounding error.

of the absolute difference between the standard deviation of the simulated coefficient distribution and the average of the estimated asymptotic standard error, i.e. if $\hat{\beta}^{(m)}$ is the simulated distribution of the MRR coefficients and $\hat{se}^{(m)}(\hat{\beta})$ is the simulated distribution of the MRR standard errors, then the value plotted in Figure 3.1 is:

$$\sum_{j=1}^{J_0+1} \left| \text{Var}[\hat{\beta}_j^{(m)}]^{1/2} - E[\hat{se}^{(m)}(\hat{\beta})] \right|$$

Figure 3.1 clearly shows that the estimated standard errors are converging to the (true) standard deviations of the coefficient distributions.

3.6 A Test for High Frequency Noise in Portfolio Returns

Having developed an asymptotic theory and distributional properties for the MRR model and validated these properties through Monte Carlo simulation, I now turn my attention to an empirical application in financial economics. Many empirical investigations into financial returns often make *a priori* assumptions about the presence of high frequency noise – i.e.. economically uninformative components of a financial signal. However, it is not enough to simply implement (for example) long horizon regressions, observe that predictability has improved⁴, and proceed. Nor is it sufficient to simply take a particular model proposing high frequency noise as given and proceed. In order to validate models and assumptions that postulate high frequency noise, a rigorous test is in order.

I investigate this phenomenon in the context of the simple CAPM model in an MRR framework with Fama and French's (1993) 25 size and book-to-market sorted portfolios available from Kenneth French's website⁵. Data sampling frequency is monthly and the sample period considered is January 1926 to December 2012. – a sample size of 1038. Both portfolio returns and market returns are transformed to excess returns by subtracting off the risk free rate. An MRR model is fit to the basic CAPM regression with $J_0 = 7$ levels using Daubechies filters of length $L = 16$. In this framework, high frequency

⁴Much of the contemporary work on long horizon regressions has focused on correcting spurious findings based on incorrect asymptotic theory (e.g. Valkanov (2000), Hjalmarsen (2010), Phillips and Lee (2013))

⁵Data accessible at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

noise corresponds to statistically insignificant β_j coefficients at the lower levels (i.e. higher frequency bands) which is straightforward to test. In addition to identification of the high frequency components, the MRR framework is particularly well suited to financial data as the standard errors are robust to heteroskedasticity and autocorrelation due to their formulation in the frequency domain.

Tables 3.8 - 3.12 present the results of this MRR model with coefficient estimates reported for each level as well as t-statistics for the null hypothesis that the coefficient is equal to zero with significance at the .001, .01, .05, and .1 levels indicated by ***, **, *, and . respectively. Before discussing these results, a few preliminary remarks on MRR t-tests are in order. Percival and Walden (1993) observe that direct spectral estimators (and hence multitaper estimates) do not converge to the true spectral density functions (say $S(f)$) as $n \rightarrow \infty$ but instead are asymptotically distributed $S(f)\chi_{df}^2/df$ where $df = 2$ for $0 < f < 1/2$ and $df = 1$ for $f = 0$. This implies that multitaper spectral estimates at zero frequency are asymptotically distributed $S(0)\chi_K^2$ where K is the number of tapers used in the estimate. As a result, the t-statistics from an estimated MRR model will be asymptotically t random variables with K degrees of freedom (they will not be asymptotically normal). The statistical significance reported in 3.8 - 3.12 reflect this property – since 5 orthogonal tapers were used to construct the standard errors the critical values used in the tests are quantiles of the t_5 distribution.

All 25 size-BM sorted portfolios exhibit coefficients that are significantly different that zero at lower levels (higher frequencies) at the one percent level of significance or lower (i.e. stronger results), indicating that the high frequency movements in the market portfolio are correlated with movements in every one of the portfolios. This pattern fails to pass even the most basic criteria for noise – the high frequency components of the market portfolio are informative in describing portfolio returns. These findings suggest both a careful reappraisal of *a priori* noise assumptions in empirical finance as well as a bevy of questions about the underlying economic mechanisms leading to such strong empirical results.

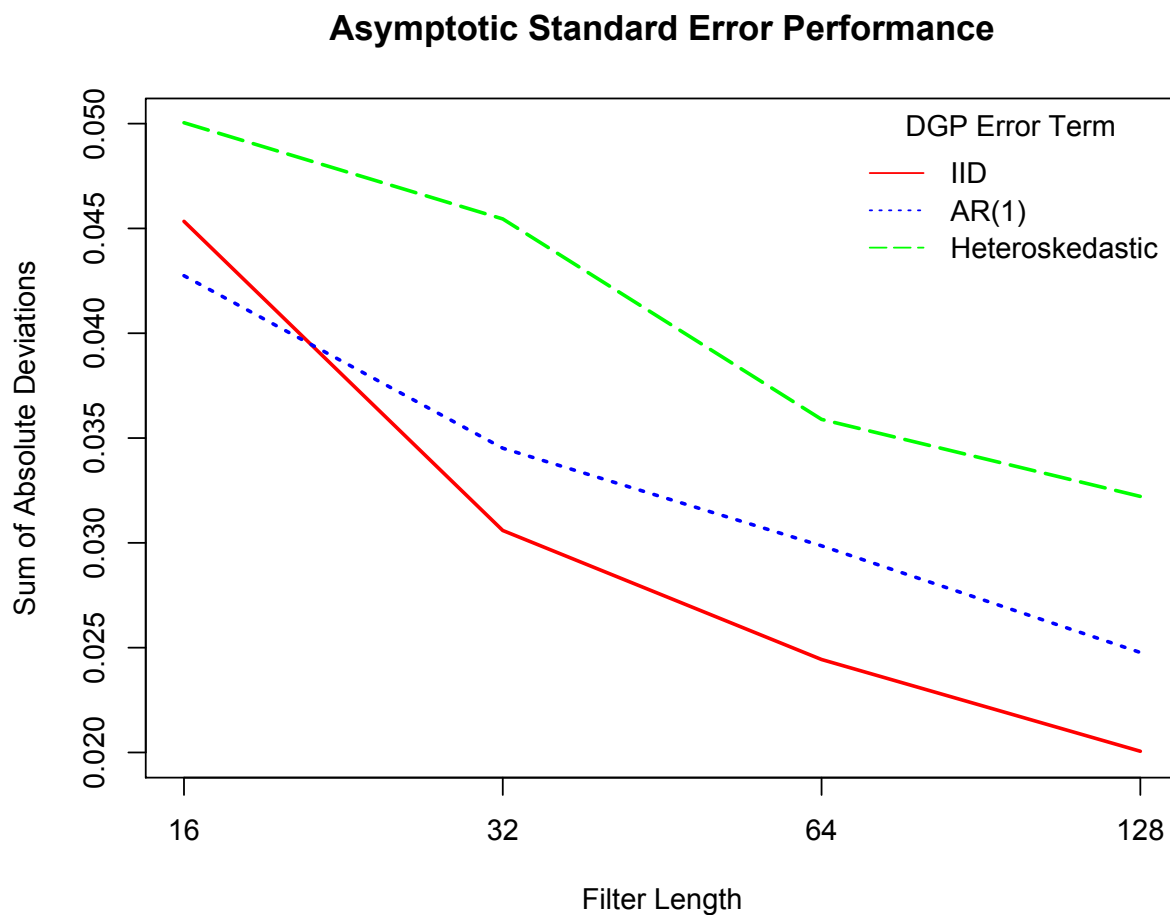
Table 3.1: Correct Specification: IID Errors – $L = 2$, $n = 64$

		Mean	Mean	Std Dev
	β	$\hat{\beta}$	$\widehat{se}(\beta)$	$\hat{\beta}$
β_1	3.000	3.027	0.543	0.404
β_2	9.000	8.968	0.484	0.609
β_3	1.000	1.005	0.342	0.607
β_4	-3.000	-2.987	0.223	0.550
β_5	0.500	0.499	0.154	0.339

3.7 Conclusion

In this chapter I present a scale-based regression model and established properties and asymptotic behavior of the proposed estimator. My approach is based on wavelet-based scale decompositions and leverages a mature wavelet literature in other disciplines to establish these properties. This depth and breadth of work in wavelet methods in other disciplines also offers many potential extensions to the MRR framework using well-developed methods from other fields. Monte Carlo simulation demonstrates that even in finite samples with relatively short filter lengths, my proposed MRR estimation scheme is quite robust to scale misspecification – a powerful result when scale structure is not known *a priori* (which is nearly always). I apply my MRR model to extend the standard CAPM model for Fama and French’s (1993) 25 size/book-to-market-sorted portfolios. In contrast to the conventional assumptions made in the finance literature, I find no evidence of high frequency portfolio noise – or rather, I find strong evidence that the high frequency components of the market portfolio have significant descriptive power for individual portfolio returns. As mentioned, these findings are strongly counter to the standard assumptions and highlight the importance of including a scale dimension in many time series regressions.

Figure 3.1: Asymptotic Standard Error Behavior

Table 3.2: Correct Specification: IID Errors – $L = 256$, $n = 4096$

		Mean	Mean	Std Dev
	β	$\hat{\beta}$	$\widehat{se}(\hat{\beta})$	$\hat{\beta}$
β_1	3.000	3.009	0.180	0.151
β_2	9.000	8.989	0.205	0.204
β_3	1.000	1.002	0.177	0.165
β_4	-3.000	-2.999	0.125	0.159
β_5	0.500	0.499	0.118	0.083

Table 3.3: IID Errors – $L = 16$

		Mean	Mean	Std Dev
	β	$\hat{\beta}$	$\widehat{se}(\beta)$	$\hat{\beta}$
β_1	1.000	1.003	0.040	0.056
β_2	1.000	0.998	0.035	0.053
β_3	1.000	1.003	0.037	0.041
β_4	1.000	0.998	0.041	0.035
β_5	1.000	0.999	0.022	0.021

Table 3.4: AR(1) Errors – $L = 16$

		Mean	Mean	Std Dev
	β	$\hat{\beta}$	$\widehat{se}(\beta)$	$\hat{\beta}$
β_1	1.000	1.000	0.022	0.026
β_2	1.000	0.999	0.031	0.041
β_3	1.000	1.001	0.042	0.051
β_4	1.000	0.996	0.052	0.063
β_5	1.000	1.001	0.041	0.049

Table 3.5: Heteroskedastic Errors – $L = 16$

		Mean	Mean	Std Dev
	β	$\hat{\beta}$	$\widehat{se}(\beta)$	$\hat{\beta}$
β_1	1.000	1.001	0.057	0.071
β_2	1.000	1.001	0.047	0.069
β_3	1.000	0.998	0.043	0.054
β_4	1.000	1.001	0.046	0.045
β_5	1.000	1.001	0.028	0.030

Table 3.6: IID Errors – $L = 128$

		Mean	Mean	Std Dev
	β	$\hat{\beta}$	$\widehat{se}(\beta)$	$\hat{\beta}$
β_1	1.000	0.999	0.038	0.044
β_2	1.000	1.001	0.033	0.041
β_3	1.000	0.998	0.027	0.032
β_4	1.000	0.998	0.027	0.027
β_5	1.000	1.000	0.017	0.018

Table 3.7: $AR(1)$ Errors – $L = 128$

		Mean	Mean	Std Dev
	β	$\hat{\beta}$	$\widehat{se(\beta)}$	$\hat{\beta}$
β_1	1.000	0.999	0.018	0.022
β_2	1.000	0.999	0.027	0.031
β_3	1.000	1.001	0.035	0.040
β_4	1.000	0.998	0.043	0.050
β_5	1.000	0.997	0.036	0.040

Table 3.8: First Quintile of Size Sorted Portfolios

	S_1BM_1	S_1BM_2	S_1BM_3	S_1BM_4	S_1BM_5
β_1	1.5064 (9.164)***	1.2007 (9.682)***	1.3025 (7.35)**	1.0983 (7.725)**	1.1633 (6.507)**
β_2	1.5582 (8.985)***	1.4957 (8.686)***	1.2241 (7.58)**	1.2435 (7.5)**	1.3893 (7.126)**
β_3	2.181 (11.264)***	2.1368 (8.975)***	1.9787 (8.8)***	1.9361 (8.996)***	2.0113 (9.175)***
β_4	1.4132 (8.286)***	1.4231 (6.707)**	1.2836 (5.524)**	1.2563 (5.303)**	1.2646 (5.97)**
β_5	1.9446 (11.033)***	1.6845 (9.342)***	1.5335 (8.382)***	1.4827 (7.261)**	1.7368 (7.28)**
β_6	1.4751 (7.147)**	1.2664 (5.5)**	1.2222 (5.195)**	1.3453 (5.154)**	1.4296 (6.146)**
β_7	1.5109 (5.447)**	0.8105 (3.615)*	1.2306 (5.67)**	1.1095 (5.332)**	0.7599 (3.66)*
β_8	0.1252 (0.316)	0.2931 (1.058)	0.3723 (0.849)	0.2644 (0.755)	0.3304 (0.753)

Table 3.9: Second Quintile of Size Sorted Portfolios

	S_2BM_1	S_2BM_2	S_2BM_3	S_2BM_4	S_2BM_5
β_1	1.2372 (16.638)***	1.1279 (12.236)***	1.0556 (11.295)***	1.0969 (9.596)***	1.1955 (8.175)***
β_2	1.1036 (14.624)***	1.3157 (10.779)***	1.1631 (10.542)***	1.2219 (9.379)***	1.3997 (8.822)***
β_3	1.513 (16.261)***	1.6503 (11.873)***	1.6339 (10.363)***	1.6486 (9.278)***	1.8475 (9.393)***
β_4	1.3822 (12.759)***	1.1179 (9.103)***	1.1023 (7.073)**	1.1046 (6.49)**	1.2743 (5.936)**
β_5	1.2193 (11.124)***	1.3011 (10.667)***	1.3061 (9.141)***	1.3371 (8.316)***	1.3907 (6.323)**
β_6	1.2732 (8.361)***	1.1681 (6.651)**	1.0835 (5.94)**	1.1688 (7.109)**	1.2081 (6.532)**
β_7	1.5824 (6.819)**	1.1893 (5.719)**	0.8313 (4.661)*	0.6336 (4.294)*	0.7412 (4.344)*
β_8	0.493 (1.333)	0.448 (1.372)	0.4061 (2.025).	0.3445 (1.171)	0.4738 (1.62)

Table 3.10: Third Quintile of Size Sorted Portfolios

	S_3BM_1	S_3BM_2	S_3BM_3	S_3BM_4	S_3BM_5
β_1	1.1638 (18.437)***	1.1021 (16.041)***	1.0689 (11.663)***	1.0189 (12.373)***	1.2784 (7.646)**
β_2	1.279 (14.142)***	1.0921 (13.819)***	1.1568 (10.728)***	1.1542 (10.442)***	1.4359 (8.572)***
β_3	1.597 (14.045)***	1.2738 (15.524)***	1.4092 (10.986)***	1.4138 (10.46)***	1.7055 (8.113)***
β_4	1.2618 (10.751)***	1.1435 (10.391)***	1.0839 (7.214)**	1.0933 (9.444)***	1.114 (4.769)*
β_5	1.2761 (9.71)***	1.1659 (10.573)***	1.1389 (9.644)***	1.2277 (10.889)***	1.5955 (7.097)**
β_6	1.188 (10.709)***	1.113 (8.785)***	0.9918 (7.489)**	1.0061 (6.262)**	0.9611 (5.403)**
β_7	1.1233 (7.31)**	0.988 (5.866)**	1.0355 (9.041)***	0.8592 (5.61)**	0.9653 (6.478)**
β_8	0.7086 (3.408)*	0.4597 (1.874)	0.4521 (1.934)	0.4611 (3.052)*	0.3737 (1.428)

Table 3.11: Fourth Quintile of Size Sorted Portfolios

	S_4BM_1	S_4BM_2	S_4BM_3	S_4BM_4	S_4BM_5
β_1	1.0817 (31.796)***	1.0474 (21.048)***	1.0686 (18.467)***	1.1349 (11.132)***	1.3489 (7.779)**
β_2	1.0426 (23.762)***	1.125 (15.17)***	1.0809 (13.532)***	1.1515 (9.237)***	1.4608 (7.663)**
β_3	1.0812 (21.615)***	1.2144 (12.366)***	1.2014 (12.193)***	1.3676 (9.004)***	1.7224 (7.434)**
β_4	1.1614 (18.961)***	1.1378 (9.475)***	1.1301 (12.411)***	1.0482 (7.086)**	1.3051 (5.549)**
β_5	1.0571 (14.915)***	1.035 (12.251)***	1.1327 (10.586)***	1.3719 (10.066)***	1.6285 (7.228)**
β_6	0.9574 (12.733)***	1.0347 (9.016)***	1.1532 (9.826)***	0.9529 (7.471)**	1.0049 (8.317)***
β_7	0.9782 (7.093)**	0.7517 (6.422)**	0.8012 (8.036)***	0.3718 (4.422)*	0.2604 (2.003)
β_8	0.6109 (3.34)*	0.7219 (3.095)*	0.6561 (3.97)*	0.2369 (1.705)	0.7614 (2.785)*

Table 3.12: Fifth Quintile of Size Sorted Portfolios

	S_5BM_1	S_5BM_2	S_5BM_3	S_5BM_4	S_5BM_5
β_1	0.9808 (65.78)***	0.947 (40.457)***	0.9388 (28.177)***	1.0489 (11.122)***	1.0846 (7.979)***
β_2	0.9655 (26.453)***	0.937 (25.239)***	1.0407 (13.261)***	1.1781 (8.701)***	0.8929 (7.477)**
β_3	0.9454 (18.059)***	0.7955 (20.553)***	0.9364 (8.942)***	1.2743 (7.398)**	1.7089 (11.663)***
β_4	0.9619 (16.075)***	0.9505 (13.561)***	0.8766 (8.556)***	0.9692 (6.545)**	-0.8629 (-3.206)*
β_5	0.8815 (23.834)***	0.9004 (12.26)***	1.0963 (8.949)***	1.3229 (8.282)***	3.46 (6.034)**
β_6	0.9701 (21.001)***	0.915 (17.566)***	1.0299 (12.268)***	0.9518 (11.635)***	5.0591 (7.666)**
β_7	1.0896 (6.84)**	0.8366 (11.274)***	0.8309 (8.222)***	1.0394 (20.092)***	7.282 (6.618)**
β_8	0.8535 (5.61)**	0.627 (3.436)*	0.9977 (6.349)**	0.7803 (3.921)*	1.7879 (1.349)

Chapter 4

**TIME SCALE STRUCTURE IN THE CROSS SECTION OF
PORTFOLIO RETURNS****4.1 Introduction**

The study of macroeconomic relationships without incorporating asset markets leaves critical gaps in our understanding. The opposite is true as well – understanding how macroeconomic risks are transmitted to asset markets and, consequently, priced, remains a central question in macrofinance. However, well performing asset pricing models that utilize macroeconomic risk factors continue to be elusive. Chen, Roll, and Ross (1986) find that some macroeconomic factors have explanatory power on asset returns, however they curiously find that aggregate consumption does not have explanatory power for asset returns – a somewhat vexing result given the consumption-based formulation of most theoretical asset pricing models. This result is especially troubling in light of Cochrane’s (2001) emphasis that any asset pricing factor model is an extension of the basic consumption-based model. The poor performance of aggregate consumption as an asset pricing factor underpins our lack of understanding about links between asset markets and the macroeconomy.

I propose a scale-based asset pricing framework – based on a modified MRR model developed in chapter 3 – that offers unique insights into the relationship between risk and time scale. I propose that this relationship is central to linkages between macroeconomics and financial markets. My framework helps make sense of the apparent consumption contradiction, and allows for a more refined understanding of risks associated with traditional asset pricing factors. I argue that scale effects can dramatically impact asset pricing relationships via a phenomenon I style “risk aggregation.” The basic premise is that a single measured factor can aggregate multiple sources of risk over different scales. Section 4.2 below outlines this risk aggregation phenomenon in the context of excess market returns. Conversely, a single factor might only describe risk over a particular scale – I find aggregate

consumption to exhibit this phenomenon – the longer time scale components of aggregate consumption have explanatory power on asset returns while higher frequency components of consumption do not. These often complex interactions between scale effects and risk factors certainly complicates economic interpretation of traditional asset pricing models – both the economics interpretation of scale-based risk factors, as well as the economic interpretation of the relative success or failure of macroeconomic factors in pricing portfolios.

Although this is the first (to my knowledge) direct investigation of the linkages between general asset pricing factors and time scale, scale effects in and of themselves are not novel in theoretical or empirical asset pricing – with the greatest attention being given to the business cycle. Merton's (1973) ICAPM features state variable that describe time-varying investment opportunities which clearly vary over the business cycle while Campbell and Cochrane's (1999) external habit models introduces a state variable that tracks the business cycle even more explicitly. Again in 2001 Cochrane highlights the importance of recession state variables. Changing market exposure over the business cycle was one of the main motivating considerations in Jagannathan and Wang's (1996) conditional CAPM model in which they utilize a yield spread that strongly forecasts the business cycle as an instrument for the conditional market risk premium. A variety of authors have also experimented with sampling frequencies to capture various scale effects in asset pricing models; see for example Andersen et al. (2005, 2006), Harvey (1989), Cochrane (1996) and Jagannathan and Wang (1996) among others. A newer theme in the literature involves explicitly modeling multiple scale factors separately as in Adrian and Rosenberg (2008), Engle and Rangel (2009), Engle, Gysels, and Sohn (2009) and Rangel and Engle (2012) – who all model volatility risk factors at separate scales – and Cenesizoglu and Reeves (2013) who utilize mixed frequency methods to construct separate scale components of beta.

A separate but related literature utilizes wavelet transforms to investigate asset pricing relationships at different scales. This rather sparse literature has primarily focused on scale-by-scale factor pricing models with the majority applying a simple scale-by-scale CAPM – that is, both portfolio returns and market returns are decomposed via wavelet transforms and individual time-series regressions are run for each scale; see Gençay, Selçuk, and Whitcher (2001b, 2003, 2005) and In and Kim (2012) who revisit the CAPM across scales

and investigate the long-run performance of the Fama-French three factor model via wavelet-based multiresolution decomposition. In and Kim (2012) differ from Gençay, Selçuk, and Whitcher (2001b, 2003, 2005) in that they attempt to leverage scale-based methods to explain the cross-section of returns, a step that previous wavelet-based asset pricing work has stopped short of taking.

In contrast to the existing wavelet asset pricing literature, my MRR approach more easily integrates with the general asset pricing literature and allows for a meaningful interpretation of the economics behind such scale effects. The MRR model that I fit aggregates groups of MODWT-based details and smooth into short-, medium-, and long-run components. However, in an extension to the MRR framework developed in Chapter 3, I modify the conditional wavelet variance methodology proposed in Percival et al. (2011) to estimate conditional betas for each scale component which I treat as separate risk factors in estimating risk premia from conditional cross-sectional regressions. Because this approach utilizes a fixed-length filter, an appeal to the asymptotic properties the MRR estimator is a tenuous prospect. Consequently, I provide a bootstrapping scheme in which statistical properties of these modified MRR estimators can be considered. I view this approach as an explicit treatment of the scale-dependent beta alluded to in Jagannathan and Wang (1996) and a natural integration of the wavelet asset pricing literature into the main stream asset pricing literature.

My approach is consistent with non-wavelet-based scale factor models introduced by Adrian and Rosenberg (2008), Engle and Rangel (2009), Engle, Gysels, and Sohn (2009), Rangel and Engle (2012), and Cenesizoglu and Reeves (2013) which all exploit scale structure but do not do so on a scale-by-scale basis. My model most closely aligns with Cenesizoglu and Reeves' (2013) semiparametric approach in that the time scale components are constructed nonparametrically while factor pricing relationships are estimated using linear models. However, my approach differs from theirs on a number of fronts. Most notably, I relate these scale effects back to general factor models through a "risk aggregation" phenomenon and, by extension, to implications to factor models utilizing macroeconomic variables. I also provide an economic interpretation of these scale factors – highlighting the risk aggregation phenomenon for market excess returns.

I demonstrate the appropriateness of my MRR model by introducing a finite sample statistical test for differential scale effects in linear models, finding strong evidence in the majority of the canonical Fama-French portfolios under investigation. I compare the performance of MRR model in pricing alternative sets of Fama-French portfolio sortings and find a strong relationship between asset pricing performance and scale structure, suggesting that the relative success of the MRR model comes from directly exploiting scale structures present in the data. To this end, I note that my model performs quite well in explaining the cross-section of excess portfolio returns across all sets of portfolio sorting schemes considered. My model wildly outperforms the benchmark Fama-French (1993, 1996) three-factor model and significantly improves pricing accuracy over the similar three-component-beta model of Censizoglu and Reeves (2013) over an identical sample period. My model also significantly improves accuracy (based on SSPE and RMSPE) over all model specifications presented in Adrian and Rosenberg (2008) over a similar sample period.

The remainder of this chapter is organized as follows: Section 4.2 outlines the intuition behind multiresolution asset pricing. Section 4.3 introduces the proposed methodology. Section 4.4 introduces the data set used, followed by asset pricing results in section 4.5. Section 4.6 concludes.

4.2 *Multiresolution Asset Pricing*

My multiresolution approach is motivated by the curious observation that asset pricing factors do not necessarily describe a single source of risk – a single “risk factor” might proxy very different sources of risk at different time scales. I believe this aggregation of multi-scale risk is one of the reasons that macroeconomic factors have performed relatively poorly in asset pricing applications. My goals with this approach are two-fold: to understand the relationships between scale and risk, and to begin to relate such relationships to macroeconomic sources of risk. Existing multiresolution investigations of asset pricing relationships (e.g. Gençay, Selçuk, and Whitcher (2001b, 2003, 2005) and In and Kim (2012)) have been largely focused on identifying heterogeneity across scales without identifying the sources of scale effects or attempting to orient such scale-based approaches in the broader asset pricing or macro-finance literature.

Scale effects and risk aggregation are potential problems in any factor pricing model but to clearly and concisely investigate the issue I propose a simple modified CAPM-style asset pricing model in which market returns are decomposed by scale and treated as separate risk factors – a factor associated with short-run movements in market returns (under 1.3 years), a factor associated with medium-run movements or the traditional business cycle (1.3 - 10.6 years), and a long-run factor (over 10.6 years). The economic interpretation of these risk factors leads to unique insights into the effects of time scale in factor pricing models. The medium-run factor is clearly tied to the business cycle; business cycle components of consumption and market returns are well documented. During recessions the market tends to be down, as does aggregate consumption. So a portfolio that moves with the market over the business cycle is a poor hedge against consumption risk at that scale hence I style this component a “business-cycle risk factor.”

The intuition behind a short-run risk factor is a bit more subtle, however. Why would consumers care about market returns month to month? If average annual market returns are around 8%, why would a consumer care about a 5% loss one month if they know that, on average, they will tend to make it up over the next year? I argue that the risk associated with short-run movements in the market corresponds to downside liquidity risk. On occasion, consumers receive negative income shocks that could come in the form of job loss, pay cuts, or other asset-based cash flow shocks (rental income, etc.). Alternatively, I could formulate an argument based on expenditure shocks like unanticipated medical costs, automobile costs, home repair, etc. to the same effect. Traditional consumption based asset pricing models predict that consumer’s will keep some of their wealth in risk-free liquid assets. However, if the funds required to meet the income (or expenditure) shock exceeds the consumer’s liquid holdings, they will need to liquidate part of their risky portfolio holdings to weather the shock. If such income shocks are positively correlated with higher-frequency returns (e.g. monthly), then the consumer faces potential situations in which they must sell off part of their portfolio when it is doing badly. The effect on portfolio wealth compounds over time leading to a change in wealth that is much larger than the amount liquidated. For an infinitely lived consumer, this effect is unbounded. If instead, income shocks are negatively correlated or uncorrelated with returns, the downside risk of having to liquidate portfolio

holdings when the portfolio is performing poorly is greatly reduced.

One might be tempted to conflate these two sources of risk, as they are both driven by consumption, however the two ideas are indeed separate. Consider a counter-cyclical stock that tends to see higher returns during recessions. In general, this stock would be very desirable. However, suppose monthly returns on this stock are perfectly correlated with income (expenditure) shocks to a representative consumer. There is a nontrivial probability that the consumer will face liquidity constraints and have to sell part of their holdings of this stock – and they must do so when the stock is doing badly. This does not imply that the consumer will refuse to hold the stock, however, but they will demand a risk premium above and beyond a similar countercyclical stock whose monthly returns are negatively correlated with income shocks.

An interesting implication of this interpretation is that the short-run downside risk factor is related to future expected consumption via the wealth effect but does not relate directly to contemporaneous consumption (except via intertemporal optimization). As Cochrane (2001) emphasizes, any factor pricing model is an extension of the basic consumption-based model and factors are simply proxies for consumption. As a cursory check of the intuition above I look at how log consumption growth responds to market returns at different scales. If the downside liquidity risk interpretation is correct, then there should not be a clear relationship between short-run market returns and log consumption growth. Using monthly aggregate consumption data and monthly market excess returns I estimate the following model:

$$\Delta c_t = \alpha + \beta_1 \mathcal{R}_{short,t}^m + \beta_2 \mathcal{R}_{medium,t}^m + \beta_3 \mathcal{R}_{long,t}^m + \varepsilon_t \quad (4.1)$$

where c_t is log consumption and \mathcal{R}^m are the scale components of market excess returns. Indeed, table 4.1 demonstrates that the relationship between consumption growth and the short-run component of market returns is not statistically different than zero. That is, short run movements in portfolio wealth are not transmitted to consumption growth. This finding is consistent with the proposed “downside liquidity risk” interpretation described above.

The intuition behind a long-run market risk factor can again be tied back to consumption by looking at table 4.1. An increase in the long-run component of market excess returns,

which I interpret as a proxy for the long-run growth rate of wealth, leads to a decrease in contemporaneous consumption. If consumption and asset wealth are cointegrated, as argued in Lettau and Ludvigson (2001), then a fall in the current long-run growth rate of wealth implies a corresponding expected *future* increase as wealth adjusts back to the common stochastic trend, provided expected consumption growth is not too volatile. This drives up the expected present discounted value of future wealth and hence, current consumption. This interpretation is consistent with findings in table 4.1. An alternative argument follows from Campbell and Cochrane (1999) and Constantinides (1990) habit formation models. Consumption growth is high when marginal utility is high, i.e. when consumption is close to the habit. Consumption will be low when the expected present discounted value of future wealth is low, i.e. when the long-run growth rate of wealth is low. Again, I have a theoretical framework that is consistent with findings in table 4.1. In both stories, I tie the long-run component to risk associated with the expected present discounted value of future wealth.

These findings and intuition illuminate potential ways forward in developing an understanding of linkages between the macroeconomy and financial markets. My interpretation of the above scale-based risk factors imply relationships to certain macroeconomic variables. Downside liquidity risk is related to liquidity demand, interest rates, and correlation of income shocks with asset returns while business cycle risk is clearly related to many aggregate macroeconomic variables with business cycle components. However, this business cycle risk factor is only relevant over particular time horizons. For example, aggregate consumption growth might be a candidate macroeconomic risk factor but the risk that this factor is capturing is only relevant over the medium and long run. For now, I limit my attention to developing the scale-based framework to address these questions and leave the task of exploring scale-based macroeconomic risk factors for future work.

4.3 Methodology

As described above, my basic model explores the scale effects and risk aggregation phenomenon in a simple time-varying CAPM model in a MRR framework. This MRR approach is quite flexible and can be easily implemented in multifactor asset pricing models or with different scale components; however, motivated by my intuition about composite risk fac-

tors above I opt to investigate scale effects with a single traditional risk factor – exposure to the market. As Jagannathan and Wang (1996) observe, market exposure changes over time – over the business cycle at the very least. Thus I consider the following (conditional) CAPM-style model:

$$R_t^i - R_t^f = \alpha_i + \beta_{i,t}(R_t^m - R_t^f) + \varepsilon_{i,t} \quad (4.2)$$

where R_t^i are returns to portfolio i at time t , R_t^m are returns to the market, and R_t^f is the risk-free rate. I then extend the basic MRR model to allow for dynamic parameters in the following way:

$$R_t^i - R_t^f = \beta_{1,i,t}\mathcal{R}_{short,t}^m + \beta_{2,i,t}\mathcal{R}_{medium,t}^m + \beta_{3,i,t}\mathcal{R}_{long,t}^m + \nu_{i,t} \quad (4.3)$$

where $\beta_{j,i,t}$ is the exposure to the j th component of the market at time t . Note that I do not include an intercept in equation (4.3). The choice to omit an intercept parameter is motivated by two observations: one, a *static* intercept would simply recenter the long-run component \mathcal{R}_{long}^m about zero and would have no substantive effect on how the long-run component behaves over time and hence would have no effect on the risk premia estimates from equation (4.4); and two, a *dynamic* intercept term would not be separately identifiable from the long-run component, nor would the long-run component and conditional intercept be individually economically interpretable if identification were possible. I do not miss any economically meaningful inference by omitting the intercept parameter. Consequently, I demean both portfolio excess returns as well as market excess returns and omit an intercept parameter in model (4.3). Allowing time variation in the scale-based coefficients from equation (4.3) factors is not only consistent with a time-varying beta in equation (4.2), but also with the interpretation of these scale-based risk factors discussed above and potential time variation thereof.

Having estimated the scale-based risk factors above, I then estimate the conditional risk premia for these factors by monthly cross-sectional OLS à la Fama and Macbeth (1973):

$$R_t^i - R_t^f = \gamma_{0,t} + \gamma_{1,t}\beta_{1,i,t} + \gamma_{2,t}\beta_{2,i,t} + \gamma_{3,t}\beta_{3,i,t} + v_{i,t} \quad (4.4)$$

resulting in conditional risk premia $\gamma_{j,t}^i$ for each scale factor $j = 1, 2, 3$. My economic interpretation of these scale risk factors also suggests time variation in the gamma parameters – for example, if consumers tend to increase their liquid holdings during periods of increased income volatility then the short-run downside liquidity risk factor will command a lower premium as consumers are less likely to face downside liquidity constraints.

As mentioned above, multiresolution methods are not entirely novel in asset pricing but they have been limited to a scale-by-scale investigation of asset pricing relationships (e.g. Gençay, Selçuk, and Whitcher (2001b, 2003, 2005) and In and Kim (2012)). For the moment, I set aside estimation of the conditional betas and gammas and observe that the multiresolution model in equation (4.3) can be thought of as an extension of previous multiresolution CAPM investigations via a reconstruction of portfolio returns. Consider a three-component decomposition of returns to the i th portfolio:

$$R_t^i - R_t^f = \mathcal{R}_{short,t}^i + \mathcal{R}_{medium,t}^i + \mathcal{R}_{long,t}^i \quad (4.5)$$

and a scale-by-scale estimation of portfolio market exposure akin to Gençay, Selçuk and Whitcher (2001, 2003, 2005) and In and Kim (2012):

$$\mathcal{R}_{j,t}^i = \alpha_j^i + \beta_{i,j} \mathcal{R}_{j,t}^m + \varepsilon_{j,t}^i \quad j = short, medium, long \quad (4.6)$$

Equations (4.5) and (4.6) imply

$$R_t^i - R_t^f = \alpha_1^i + \alpha_2^i + \alpha_3^i + \beta_{i,1} \mathcal{R}_{short,t}^m + \beta_{i,2} \mathcal{R}_{medium,t}^m + \beta_{i,3} \mathcal{R}_{long,t}^m + \varepsilon_{1,t}^i + \varepsilon_{2,t}^i + \varepsilon_{3,t}^i \quad (4.7)$$

which is equivalent to equation (4.3) with $\alpha^i = \alpha_1^i + \alpha_2^i + \alpha_3^i$ and $\varepsilon_t^i = \varepsilon_{1,t}^i + \varepsilon_{2,t}^i + \varepsilon_{3,t}^i$. I observe that, by construction, the short- and medium-run components are mean-zero thus the α_1^i and α_2^i will be approximately zero. As mentioned above, inclusion of an intercept term in equation (4.7) (α_3^i) is equivalent to demeaning the long-run component $\mathcal{R}_{long,t}^m$.

Thus my approach can be viewed as a bridge between the traditional, scale-independent asset pricing literature and the growing wavelet-based asset pricing literature.

It should be noted that the MRR framework and associated wavelet filtering methods

share a characteristic common to nearly all frequency domain filters – namely, the two-sided nature of the convolutions that are associated with the filtering operations. A wavelet centered at time t_0 has support both to the left ($t < t_0$) and to the right ($t > t_0$), meaning that “future” information is utilized to obtain the multiresolution components at time t_0 .

This is a potential shortcoming of wavelet methods; however, a “two-sided” approach is not without precedent or merit. I view time scale decomposition in the MRR framework as a nonparametric analogue of smoothed estimates of state variables in state space models. If forecasting or on-line filtering is the primary objective, then smoothed state estimates aren’t required; however if the actual *level* of the unobserved state is of interest, as it is here, then smoothed estimates – based on the entire sample – are often used in lieu of informationally conditional estimates (e.g. Creal, Koopman, and Zivot (2010)). Thus I view the MRR regressors as optimal (i.e. using all relevant information as in smoothed states) estimates of unobserved components akin to those obtained in state space models.

4.3.1 Scale Aggregation

Using monthly data, I define the short-run component such that it corresponds to levels 1 - 4 (1.3 years and under)¹ and the medium-run component such that it corresponds to levels 5 - 7 (between 1.3 and 10.6 years)². The medium-run component is thus defined over the shortest and longest NBER dated recessions and hence captures business cycle fluctuation in the time series. The long-run component then corresponds to the level 7 smooth. Yogo (2008) uses a similar approach in aggregating details with biorthogonal wavelets to construct a business cycle filter. For other uses of wavelet transforms in business cycle filtering, see, for example, Aguiar-Conraria and Soares (2011), Addo, Billio and Guegan (2012), and Raihan, Wen and Zeng (2005).

¹The level 4 details correspond to a physical scale of $2^4 = 16$ months, or 1.3 years

²The level 7 details correspond to a physical scale of $2^7 = 128$ months, or 10.6 years

4.3.2 A Static Test for Heterogenous Scale Effects

Gençay, Selçuk, and Witcher (2001a) find evidence of heterogeneity across scales in foreign exchange markets while both Gençay, Selçuk, and Witcher (2003, 2005) and In and Kim (2013) find evidence for heterogeneity in market exposure across scales but stop short of formalizing a statistical test for heterogeneity. As a specification test for my MRR model in equations (4.3) and (4.4), I introduce a test for unconditional scale homogeneity. Although I model conditional parameters in equations (4.3) and (4.4), I formulate an unconditional test rather than a conditional test for two reasons: one, computational and analytic simplicity, and two, conditional homogeneity across scales implies unconditional homogeneity – if unconditional homogeneity is rejected, then conditional homogeneity is by extension also rejected. Clearly Chapter 3 develops asymptotically valid testing procedures for MRR models, however the fixed length filters used here suggest that an asymptotic test might be less appropriate. As a result, I propose a finite impulse response (FIR) test for scale specification.

As mentioned, the test is unconditional in nature. Substituting the scale decomposition of market returns from the MRR model into the original linear model in equation (4.2) and dropping the time subscripts on the parameters, I see that equation (4.2) is equivalent to:

$$R_t^i - R_t^f = \alpha_i + \beta_i (\mathcal{R}_{short,t}^m + \mathcal{R}_{medium,t}^m + \mathcal{R}_{long,t}^m) + \varepsilon_{i,t} \quad (4.8)$$

If the data in (4.8) are also demeaned, then this model nests inside the following unconditional version of equation (4.3) where I allow heterogeneity of betas across scale components:

$$R_t^i - R_t^f = \beta_{1,i} \mathcal{R}_{short,t}^m + \beta_{2,i} \mathcal{R}_{medium,t}^m + \beta_{3,i} \mathcal{R}_{long,t}^m + \nu_{i,t} \quad (4.9)$$

If the marginal effect of market returns on portfolio returns is homogenous across time scales, then $\beta_{j,i} = \beta_i$ for all scale components $j = 1, 2, 3$. This joint restriction *could* be tested using the following Wald statistic:

$$w = N \left(R \hat{\beta} \right)' \left(R \hat{V} R' \right)^{-1} \left(R \hat{\beta} \right)' \quad (4.10)$$

where \hat{V} is the estimated asymptotic variance-covariance matrix for $\hat{\beta}$. A naïve estimator of V is:

$$\hat{V} = \left(N^{-1} \sum_{t=1}^N \mathcal{R}'_t \mathcal{R}_t \right)^{-1} \left(N^{-1} \sum_{t=1}^N \hat{v}_t^2 \mathcal{R}'_t \mathcal{R}_t \right) \left(N^{-1} \sum_{t=1}^N \mathcal{R}'_t \mathcal{R}_t \right)^{-1} \quad (4.11)$$

where N is the sample size, \mathcal{R} is the $N \times 3$ matrix of scale components, \hat{v} are the residuals from regression (4.9), and R in (4.10) is the 2×3 restriction matrix:

$$R = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}.$$

A *more* appropriate, and robust, estimator of V is the diagonal asymptotic variance matrix derived in Chapter 3. However, I also propose another heteroskedastic and autocorrelation consistent (HAC) statistic based on Kiefer, Vogelsang, and Bunzel's (2000) robust fixed-b variation of an F-test. The test statistic \mathcal{F}^* is robust to general forms of autocorrelation and heteroskedasticity and computed as follows:

$$\mathcal{F}^* = N \left(R \hat{\beta} \right)' \left(R \hat{B} R' \right)^{-1} \left(R \hat{\beta} \right) / 2 \quad (4.12)$$

where

$$\hat{B} = \left(N^{-1} \sum_{t=1}^N \mathcal{R}'_t \mathcal{R}_t \right)^{-1} \left(N^{-2} \sum_{t=1}^N \hat{S}_t \hat{S}'_t \right) \left(N^{-1} \sum_{t=1}^N \mathcal{R}'_t \mathcal{R}_t \right)^{-1}$$

and \hat{S}_t is the partial sum:

$$\hat{S}_t = \sum_{j=1}^t \mathcal{R}'_j \hat{v}_j$$

Kiefer, Vogelsang, and Bunzel (2000) show that this statistic has the following limiting distribution:

$$\mathcal{F}^* \stackrel{a}{\sim} W_2(1)' \left(\int_0^1 V_2(r) V_2(r)' dr \right)^{-1} W_2(1) / 2 \quad (4.13)$$

where $W_2(1)$ is a two-dimensional Wiener process and $V_2(r)$ is a two-dimensional Brownian bridge. Critical values and p-values for this HAC scale test statistic can be calculated by simulating this distribution.

4.3.3 Dynamic Parameter Estimation

Having established a test to validate the MRR model in equation (4.3), I now return to the problem of estimating the conditional parameters in (4.3) and (4.4). As outlined in Chapters 2 and 3, the MRR regressors are approximately uncorrelated. This implies that the betas in equation (4.3) can be estimated using the IIR MRR estimator:

$$\hat{\beta}_{j,i} = \frac{Cov(R^i - R^f, \mathcal{R}_j^m)}{Var(\mathcal{R}_j^m)} \quad (4.14)$$

where \mathcal{R}_j^m is the j th component of excess market returns. The dynamic version of this being:

$$\hat{\beta}_{j,i,t} = \frac{Cov_t[R^i - R^f, \mathcal{R}_j^m]}{Var_t[\mathcal{R}_j^m]} \quad (4.15)$$

Percival et al. (2011) introduce a technique for estimating the conditional variance associated with a grouping of multiresolution components as I have proposed above. Following their method, I use a width $m + 1$ Gaussian sliding window $w_{\tau,m}$ to estimate the conditional wavelet variance:

$$Var_t[\mathcal{R}_j^m] = \sum_{\tau=t-\frac{m}{2}}^{t+\frac{m}{2}} w_{\tau,m} C_{j,\tau}^2 \quad j = short, medium, long \quad (4.16)$$

where C_j are the transform coefficients associated with the aggregated multiresolution component \mathcal{R}_j^m . The scale-based attributes of these components naturally lead to a scale-based window length selection rule. I propose a window length that is approximately half the physical bandwidth of the component³. For the short-term component – which aggregates scales 1 - 4 – I choose a window length of $\frac{1}{2}(2^4) + 1 = 9$. For the medium-term component – which aggregates scales 5 - 7 – I choose a window length of $\frac{1}{2}(2^7 - 2^4) + 1 = 57$. For the long-term component – which captures all scales greater than 7 – I choose a window length of $\lfloor \frac{1}{2}(N - 2^7) \rfloor + 1$ where N is the length of the time series and $\lfloor \cdot \rfloor$ indicates the floor function. Although these window length selections are somewhat arbitrary, they do

³The length is “approximately” half because I use an odd length window such that the window is symmetric about t

roughly capture the length of time over which the different components will vary; that is, short-term components will vary over shorter time periods while medium- and long-term components will vary over longer time periods by construction. The choice of using the transform coefficients (related to the wavelet/scaling coefficients) rather than the aggregated multiresolution components (related to the details/smooth) is due to the fact that the fixed bandwidth nature of the estimation procedure does not allow an appeal to the asymptotic properties of the details/smooth components established in Chapter 2 – namely the asymptotic variance decomposition. As a result, the variance decomposition property of the wavelet transform is unique to the transform coefficients.

Equation (4.16) provides a way of estimating the denominator in equation (4.15); I propose a similar modification to denominator of the IIR MRR estimator from Chapter 3:

$$Cov_t[R^i - R^f, \mathcal{R}_j^m] = \sum_{\tau=t-\frac{m}{2}}^{t+\frac{m}{2}} w_{\tau,m}(R_\tau^i - R_\tau^f)C_{j,\tau} \quad (4.17)$$

This defines the following estimator for the betas in equation (4.3):

$$\hat{\beta}_{j,i,t} = \frac{\sum_{\tau=t-\frac{m}{2}}^{t+\frac{m}{2}} w_{\tau,m}(R_\tau^i - R_\tau^f)C_{j,\tau}}{\sum_{\tau=t-\frac{m}{2}}^{t+\frac{m}{2}} w_{\tau,m}C_{j,\tau}^2} \quad (4.18)$$

4.3.4 *Bootstrapping Conditional Parameters*

I derive the asymptotic distribution for the MRR estimator in Chapter 3, however the modifications introduced above change the distributional behavior of estimator – the finite filter length, the use of the transform coefficients, and most importantly the fixed bandwidth procedure. Furthermore, there is not a clear closed form distribution for the risk premia in equation (4.4) due to the doubly generated β regressors. Consequently, I propose a bootstrapping framework to obtain standard errors and confidence intervals for the betas and risk premia in the model. Bootstrapping procedures rely on repeatedly resampling the data and recalculating the statistic of interest based on the resampled data. This leads to an empirical distribution of the statistic that approximates the true distribution of the original statistic.

The challenge in bootstrapping the statistics of interest lies in preserving the scale structure present in the data. To address this issue, I propose a variation on the wild bootstrapping scheme – originally proposed by Wu (1986) – that preserves the time-dependent features of the betas and the risk premia. The resampling scheme is as follows:

1. Simulate the b th bootstrap sample of portfolio excess returns using the fitted values with the addition of randomly sampled pricing errors $\hat{v}_{i,t}^*$ and randomly generated standard normal variables $\xi_{i,t}$:

$$\widehat{(R_t^i - R_t^f)}_{(b)} = \hat{\gamma}_{0,t} + \hat{\gamma}_{1,t}\hat{\beta}_{1,i,t} + \hat{\gamma}_{2,t}\hat{\beta}_{2,i,t} + \hat{\gamma}_{3,t}\hat{\beta}_{3,i,t} + \hat{v}_{i,t}^* \xi_{i,t} \quad (4.19)$$

2. Recalculate the conditional betas and risk premia parameters based on the b th bootstrap sample of portfolio excess returns and the original components of market returns.
3. Repeat above a statistically significant number of times B .
4. Calculate the standard errors and 95% confidence intervals for the betas and gammas using the resulting bootstrap distribution.

A residual-based bootstrapping scheme would be simpler, however that approach relies on a certain degree of pricing error homogeneity across portfolios and across time. I verify that a wild bootstrap scheme is more appropriate than a simple residual-based approach by testing homogeneity of cross-sectional pricing error distributions using the k -sample Anderson-Darling test of Scholz and Stephens (1987) which rejects the null hypothesis (at any reasonable level) that the cross-sectional pricing errors all come from a common distribution. The unconditional (across time and portfolio) residuals then exhibit heteroskedastic properties of unknown form but remain symmetrically distributed around zero; hence I can use a wild bootstrap to address the heteroskedasticity.

4.4 Data

I consider monthly returns on Fama and French's (1993) 25 size and book-to-market sorted portfolios available from Kenneth French's website⁴. I consider two sample periods in my analysis. The first covers the full sample of the data set: January 1926 to December 2012. The second covers January 1970 to December 2010. I consider this second sample period to allow direct comparison to the performance measurements detailed in Cenesizoglu and Reeves (2013) which considers the static CAPM, the Fama and French (1993, 1996) three-factor model, and a three-component beta model. As a robustness check I also consider 100 size and book-to-market sorted portfolios, 48 industry sorted portfolios, and 10 momentum sorted portfolios, all available on Kenneth French's website.

4.5 Results

Table 4.2 presents several performance metrics comparing the fit of the MRR model to Cenesizoglu and Reeves (2013) three-component-beta model, the Fama-French (1993, 1996) three-factor model with constant factor loadings, a state space time-varying parameter CAPM with filtered and smoothed betas, and the CAPM with constant beta on the 25 size and book-to-market sorted portfolios from January 1970 - December 2010. The state-space model for the time-varying CAPM is:

$$R_t^i - R_t^f = \alpha + \beta_{i,t}(R_t^m - R_t^f) + \varepsilon_{i,t} \quad (4.20)$$

$$\beta_{i,t} = \xi\beta_{i,t-1} + \nu_{i,t} \quad (4.21)$$

The time-varying CAPM utilizes Kalman filtered coefficients $\beta_{i|t}$ while the smoothed time-varying CAPM utilizes Kalman smoothed coefficients with information from the full sample $\beta_{i|T}$. Various treatments of the α parameter ($\alpha = 0$, constant α , time-varying α_t) in the state space model do not substantively affect results. SSPE is the sum of squared pricing errors – calculated as the squared sum over the cross-section of average pricing errors – while RMSPE is the root mean squared pricing error – calculated as $\sqrt{SSPE/N}$ where N is the

⁴Data accessible at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

size of the cross-section. Clearly the MRR model outperforms the other models according to all three metrics. Of particular note is the large improvement over the time-varying CAPM model. The impressive performance of the multiresolution beta model is not a result of allowing time-variation in the betas, but rather the incorporation of scale effects in fitting the model. Furthermore, the comparative performance of the smoothed time-varying parameter CAPM over the non-smoothed version is also compelling. As mentioned above, wavelet transforms are based on two-sided filtering operations which are not informationally conditional. The smoothed time-varying CAPM model performance shows only a very small improvement over the informationally conditional version, suggesting that the performance of the multiresolution model is not a result of any type of “look-ahead” effects due to the two-sided wavelet filtering operations. Tables 4.3 - 4.7 provide further evidence for scale effects driving the multiresolution beta model, illustrating the results from the scale homogeneity tests on the 25 size and book-to-market sorted portfolios. All portfolios exhibit statistically significant differential scale effects with respect to the market portfolio with the exception of three large-size portfolios (Table 4.7). Most portfolios exhibit similar unconditional beta sizes across the short- and medium-term components with a markedly lower long-term beta. These differential scale effects suggest that the MRR approach is appropriate in describing portfolio returns. I also note that my model outperforms all model specifications presented in Adrian and Rosenberg (2008) over similar – though not identical – sample periods according to SSPE and RMSPE criterion.

Figure 4.1 plots realized versus fitted returns from the MRR model on the full sample period (January 1926 - December 2012) while figure 4.2 plots realized versus fitted returns for my model from the shorter sample period considered in table 4.2. Particularly worth noting is relatively good cross-sectional performance – no single portfolio is badly mispriced. This includes the small-growth portfolio (bottom left point on figure 4.2) which is typically more difficult to price and which the other asset pricing models considered, including the three-component beta model, are unable to accurately price.

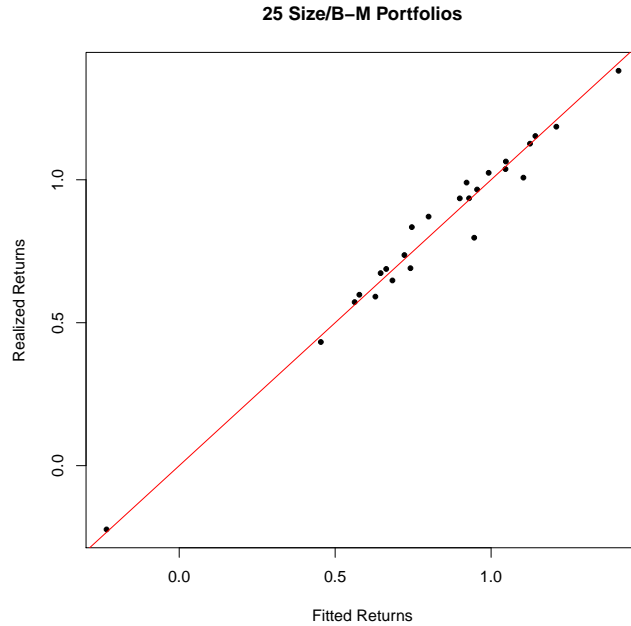
I also consider realized versus fitted returns on three other sets of sorted portfolios: 100 size and book-to-market sorted portfolios, 48 industry sorted portfolios, and 10 momentum sorted portfolios, plotted in figures 4.3, 4.4, and 4.5 and respectively. The fit on all three

sets of portfolios is quite good, however the fit on the industry portfolios is not quite as tight as on the size and book-to-market portfolios or the momentum portfolios. A look at tables 4.8 - 4.13 – which show the results from scale homogeneity tests on the 48 industry sorted portfolios – suggests why this might be. There is far less scale structure in the industry sorted portfolios than in the size and book-to-market portfolios. Indeed, less than half of the industry portfolios reject the null of scale homogeneity at the 10-percent level. In contrast, tables 4.14 and 4.15 confirm that there is evidence of scale structure present in 8 out of 10 momentum sorted portfolios at the 10-percent level. The smaller scale effects in the industry sorted portfolios is intuitive based on my economic interpretation the short run risk factor. The Fama-French size/book-to-market portfolios are constructed using market capitalization as one of the primary criterion for sorting assets. Market capitalization is in turn loosely related to liquidity – large cap stocks tend to be marginally more liquid than small cap stocks. This relation between size and liquidity is captured in the short-run component of market returns which I interpret as being related to downside liquidity risk. In contrast, industry sorted portfolios do not exhibit the same degree of exposure to liquidity risks as size/book-to-market sorted portfolios because a portfolio might include both large and small cap stocks. The results strongly suggest that the relative increase in pricing performance from the multiresolution model is coming directly from exploiting scale structure that is present in the data – the more pronounced the scale structure, the more pronounced the gains in pricing performance from exploiting it.

4.6 Conclusion

In this chapter I introduce an extension to the MRR framework developed in Chapter 3 for asset pricing applications that allows us to address aggregation of risk factors over multiple time scales. I find that traditional risk factors often aggregate multiple sources of risk over different scales and a scale-based treatment can significantly refine our understanding of priced risk. My MRR asset pricing model performs quite well in the cross-section, outperforming the benchmark Fama-French (1993,1996) three-factor model as well as newer scale-based conditional factor models like Adrian and Rosenberg (2008) and Cenesizoglu and Reeves (2013). Although the MRR model offers a rather significant improvement in

Figure 4.1: 25 Size/B-M Portfolios Full Sample (1926-2012)



asset pricing performance, I feel that the true contribution of such a model is in refining our understanding of the relationship between risk and time scale. I demonstrate via a scale-based inquiry into the relationship between aggregate consumption and market returns that a multiresolution approach offers a promising avenue for exploring asset pricing in a broader macroeconomic context. Further work will focus on exploring these macroeconomic links in greater detail.

Table 4.1: Consumption Growth and Market Excess Returns by Scale

	Estimate	Std. Error	t value	Pr(> t)
Intercept	0.56	0.02	25.41	0.00
Short-run Market Returns	0.00	0.01	0.90	0.37
Medium-run Market Returns	0.05	0.02	2.45	0.01
Long-run Market Returns	-0.29	0.08	-3.49	0.00

Table 4.2: Performance Measures of Asset Pricing Models

	RMSPE	SSPE	Adj. R^2
MRR Beta	0.0123	0.0740	0.9460
Three Comp. Beta ($\beta_{12m,d}$ & $\beta_{60m,d}$ & $\beta_{120m,m}$)	0.1162	0.3376	0.7536
Fama-French Three Factor Model	0.1253	0.3927	0.7134
TVP CAPM Smooth	0.1779	0.7909	0.6710
TVP CAPM	0.1797	0.8071	0.6643
CAPM with constant β	0.2102	1.1047	0.1972

Note: Performance measures report performance on 25 size/B-M sorted portfolios from January 1970 - December 2010. Measures on the three-component beta, the Fama-French, and the CAPM with constant β models are from Cenesizoglu and Reeves (2013). TVP CAPM is a state space time-varying parameter treatment of the CAPM.

Figure 4.2: 25 Size/B-M Portfolios Short Sample (1970-2010)

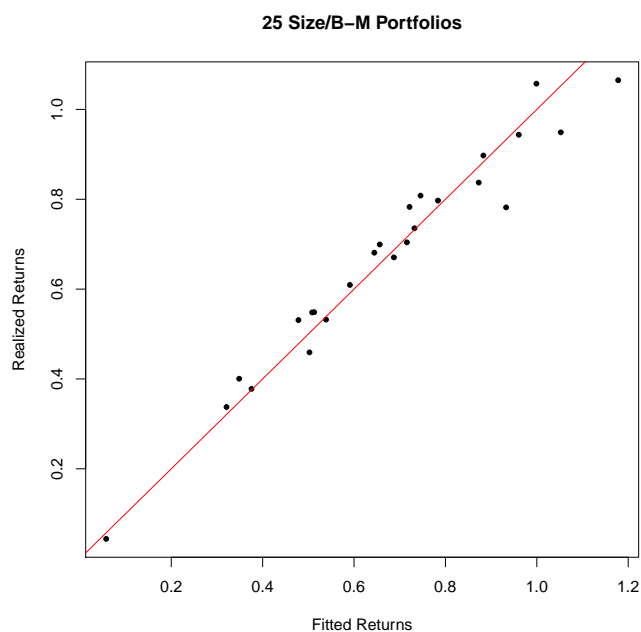


Figure 4.3: 100 Size/B-M Portfolios Full Sample (1926-2012)

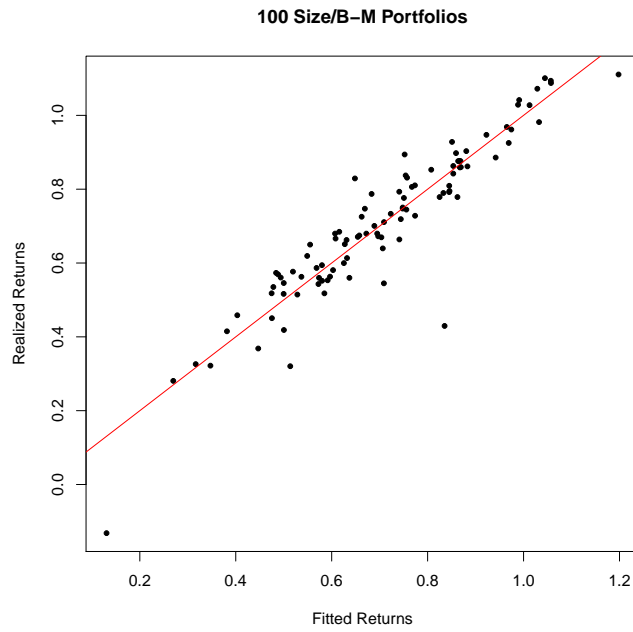


Figure 4.4: 48 Industry Portfolios Full Sample (1926-2012)

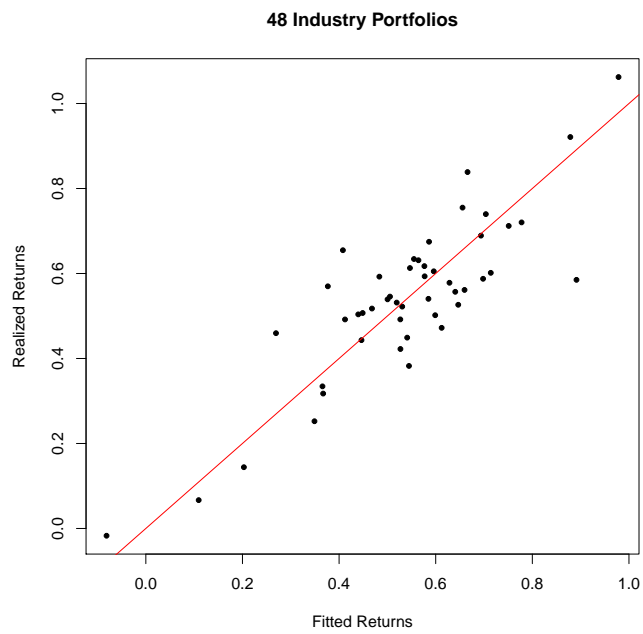


Figure 4.5: 10 Momentum Portfolios Full Sample (1926-2012)

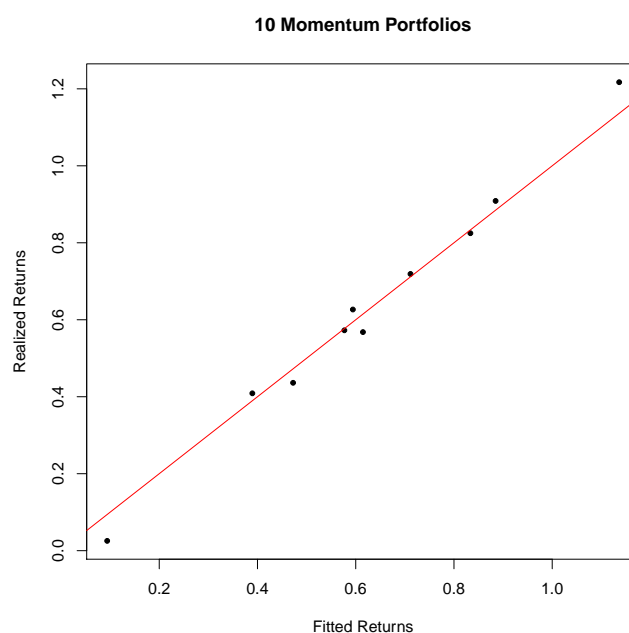


Table 4.3: Scale Homogeneity Test: First Quintile Size Portfolios by B-M Quintile

	SmallLow	Small2	Small3	Small4	SmallHigh
Short-term	1.6286 (0.0503)	1.4624 (0.0417)	1.3842 (0.0318)	1.2895 (0.0298)	1.3809 (0.035)
Medium-term	1.7035 (0.193)	1.4697 (0.1599)	1.3842 (0.1221)	1.4087 (0.1142)	1.5484 (0.1342)
Long-term	0.0916 (0.7393)	0.2561 (0.6122)	0.3538 (0.4677)	0.2268 (0.4374)	0.2896 (0.5139)
Scale Test Statistic	4.4467*	3.8672*	4.8312*	6.8948**	5.9366*
Robust Scale Test Statistic	100.75***	116.86***	66.76***	67.75***	75.59***
Significance level:	*** = .999	** = .99	* = .95	· = .9	

Table 4.4: Scale Homogeneity Test: Second Quintile Size Portfolios by B-M Quintile

	TwoLow	Two2	Two3	Two4	TwoHigh
Short-term	1.2439 (0.0252)	1.2736 (0.0227)	1.1872 (0.021)	1.2278 (0.022)	1.3734 (0.0277)
Medium-term	1.2842 (0.0968)	1.2243 (0.087)	1.1821 (0.0804)	1.2156 (0.0842)	1.26 (0.106)
Long-term	0.5553 (0.3705)	0.4411 (0.3332)	0.3835 (0.3078)	0.3119 (0.3225)	0.4578 (0.4061)
Scale Test Statistic	3.6027	6.4964*	6.788**	8.0452**	6.0878*
Robust Scale Test Statistic	23.36**	86.02***	102.32***	89.31***	68.10***
Significance level:	*** = .999	** = .99	* = .95	. = .9	

Table 4.5: Scale Homogeneity Test: Third Quintile Size Portfolios by B-M Quintile

	ThreeLow	Three2	Three3	Three4	ThreeHigh
Short-term	1.2768 (0.0194)	1.1287 (0.0146)	1.154 (0.0155)	1.1322 (0.0177)	1.3935 (0.0253)
Medium-term	1.2352 (0.0745)	1.1372 (0.056)	1.0647 (0.0595)	1.1127 (0.0681)	1.2604 (0.0969)
Long-term	0.7071 (0.2853)	0.4534 (0.2145)	0.4353 (0.2278)	0.4537 (0.2606)	0.366 (0.3712)
Scale Test Statistic	4.245*	9.8936**	11.9322***	6.8167**	9.3246**
Robust Scale Test Statistic	57.61***	93.15***	128.60***	115.98***	196.06***
Significance level:	*** = .999	** = .99	* = .95	. = .9	

Table 4.6: Scale Homogeneity Test: Fourth Quintile Size Portfolios by B-M Quintile

	FourLow	Four2	Four3	Four4	FourHigh
Short-term	1.0734 (0.0135)	1.1051 (0.0119)	1.0974 (0.0139)	1.1754 (0.0176)	1.4473 (0.0265)
Medium-term	1.0281 (0.0519)	1.0219 (0.0456)	1.126 (0.0533)	1.1265 (0.0674)	1.2725 (0.1016)
Long-term	0.593 (0.1989)	0.6942 (0.1745)	0.6402 (0.204)	0.1882 (0.2582)	0.7227 (0.389)
Scale Test Statistic	6.4835*	8.5411**	5.2752*	15.0091***	6.1415*
Robust Scale Test Statistic	56.99***	28.57**	50.42***	62.17***	19.26*
Significance level:	*** = .999	** = .99	* = .95	. = .9	

Table 4.7: Scale Homogeneity Test: Fifth Quintile Size Portfolios by B-M Quintile

	BigLow	Big2	Big3	Big4	BigHigh
Short-term	0.9685 (0.0093)	0.9198 (0.0093)	0.9693 (0.0131)	1.1252 (0.019)	1.0004 (0.0638)
Medium-term	0.9295 (0.0355)	0.9062 (0.0359)	1.0436 (0.0503)	1.1388 (0.0729)	4.1017 (0.2445)
Long-term	0.8348 (0.136)	0.6014 (0.1373)	0.9871 (0.1927)	0.7668 (0.2791)	2.2789 (0.9365)
Scale Test Statistic	2.0592	5.4778*	2.0032	1.6744	149.0144***
Robust Scale Test Statistic	16.98*	28.5**	4.32	13.91.	16.03*
Significance level:	*** = .999	** = .99	* = .95	.	. = .9

Table 4.8: Scale Homogeneity Test: 48 Industry Sorted Portfolios (1 of 6)

	Agric	Food	Soda	Beer	Smoke	Toys
Short-term	0.9008 (0.0508)	0.6781 (0.0329)	0.8454 (0.0551)	0.753 (0.041)	0.686 (0.0548)	1.1798 (0.0478)
Medium-term	0.2916 (0.2063)	0.4974 (0.1337)	0.7846 (0.2236)	0.5783 (0.1662)	0.3068 (0.2225)	0.9005 (0.194)
Long-term	0.5997 (0.7026)	1.4079 (0.4553)	0.6441 (0.7613)	2.0883 (0.5661)	0.4579 (0.7576)	1.5223 (0.6606)
Scale Test Statistic	6.21**	5.35*	0.167	8.79**	2.856	2.101
Significance level:	*** = .99	** = .95	*	. = .9		

Table 4.9: Scale Homogeneity Test: 48 Industry Sorted Portfolios (2 of 6)

	Fun	Books	Hshld	Clths	Hlth	MedEq
Short-term	1.3613 (0.0491)	1.0468 (0.0338)	0.8025 (0.0303)	1.1246 (0.0432)	1.1737 (0.0656)	0.9085 (0.0345)
Medium-term	1.3504 (0.1992)	1.1551 (0.1372)	0.7184 (0.1232)	1.1283 (0.1753)	0.6071 (0.2663)	0.66 (0.1398)
Long-term	0.6293 (0.6785)	1.474 (0.4674)	1.192 (0.4194)	1.0936 (0.597)	1.3336 (0.9068)	0.8873 (0.4762)
Scale Test Statistic	1.198	1.121	1.868	0.003	3.907	2.624
Significance level:	*** = .99	** = .95	*	. = .9		

Table 4.10: Scale Homogeneity Test: 48 Industry Sorted Portfolios (3 of 6)

	Drugs	Chems	Rubbr	Txtls	BldMt	Cnstr
Short-term	0.7843 (0.0359)	1.0504 (0.0314)	1.0689 (0.0366)	1.1454 (0.0532)	1.1846 (0.0339)	1.3272 (0.0434)
Medium-term	0.6524 (0.1459)	0.8222 (0.1276)	0.8944 (0.1484)	0.8025 (0.2158)	0.871 (0.1376)	0.9104 (0.1761)
Long-term	1.3587 (0.4968)	0.6192 (0.4346)	1.3293 (0.5055)	1.2047 (0.7348)	1.1335 (0.4687)	0.8515 (0.5998)
Scale Test Statistic	2.402	3.9	1.444	1.827	3.73	5.973*
Significance level:	**** = .99	*** = .95	** = .9			

Table 4.11: Scale Homogeneity Test: 48 Industry Sorted Portfolios (4 of 6)

	Steel	FabPr	Mach	ElcEq	Autos	Aero
Short-term	1.3202 (0.0477)	1.171 (0.0532)	1.2508 (0.0308)	1.2069 (0.0323)	1.1208 (0.0476)	1.125 (0.0437)
Medium-term	0.9728 (0.1934)	0.8713 (0.2159)	0.9047 (0.1248)	1.1457 (0.131)	1.24 (0.193)	1.2178 (0.1774)
Long-term	0.3829 (0.6586)	0.5029 (0.7351)	0.2866 (0.425)	1.1901 (0.446)	1.1244 (0.6572)	0.9216 (0.6042)
Scale Test Statistic	4.686*	2.361	12.89***	0.176	0.294	0.327
Significance level:	**** = .99	*** = .95	** = .90			

Table 4.12: Scale Homogeneity Test: 48 Industry Sorted Portfolios (5 of 6)

	Ships	Guns	Gold	Mines	Coal	Oil
Short-term	1.1252 (0.0544)	0.8521 (0.0546)	0.6727 (0.101)	1.1297 (0.0536)	1.253 (0.0878)	0.7937 (0.0421)
Medium-term	0.6305 (0.2209)	0.3178 (0.2216)	0.2104 (0.4101)	0.7424 (0.2175)	0.3383 (0.3564)	0.7301 (0.1708)
Long-term	-0.6834 (0.7524)	0.2789 (0.7548)	-0.3851 (1.3965)	-0.4028 (0.7406)	-2.0556 (1.2136)	0.4235 (0.5817)
Scale Test Statistic	11.624***	6.059**	1.848	9.088**	13.938***	0.602
Significance level:	**** = .99	*** = .95	** = .9			

Table 4.13: Scale Homogeneity Test: 48 Industry Sorted Portfolios (6 of 6)

	Util	Telcm	PerSv	BusSv	Comps	Chips
Short-term	0.5077 (0.0332)	0.766 (0.0316)	1.132 (0.0461)	1.3157 (0.0305)	1.2257 (0.0456)	1.419 (0.0412)
Medium-term	0.6684 (0.1346)	0.9116 (0.1282)	0.5356 (0.1871)	1.1091 (0.1238)	1.3664 (0.1849)	1.2983 (0.1673)
Long-term	0.7235 (0.4585)	1.1875 (0.4367)	2.2497 (0.6373)	1.8736 (0.4214)	0.8682 (0.6298)	1.5179 (0.5698)
Scale Test Statistic	1.125	1.918	10.434***	4.756*	0.97	0.476
Significance level:	*** = .99	** = .95	* = .9			

Table 4.14: Scale Homogeneity Test: 10 Momentum Sorted Portfolios (1 of 2)

	Low	2nd	3rd	4th	5th
Short-term	1.5662 (0.0303)	1.3502 (0.0227)	1.1825 (0.0183)	1.1009 (0.015)	1.0421 (0.0128)
Medium-term	1.4099 (0.1164)	1.1468 (0.0873)	1.0775 (0.0701)	1.0618 (0.0576)	0.8991 (0.0491)
Long-term	0.7851 (0.4404)	0.8054 (0.3304)	0.7269 (0.2653)	0.7217 (0.218)	0.7418 (0.1856)
Scale Test Statistic	4.271	4.252	4.028	4.298	5.921*
Significance level:	*** = .99	** = .95	* = .9		

Table 4.15: Scale Homogeneity Test: 10 Momentum Sorted Portfolios (2 of 2)

	6th	7th	8th	9th	High
Short-term	1.0264 (0.0106)	0.9736 (0.0108)	0.9341 (0.0112)	0.9629 (0.0133)	1.0189 (0.021)
Medium-term	0.9999 (0.0408)	0.9455 (0.0414)	0.934 (0.0431)	1.0275 (0.0511)	1.0706 (0.0807)
Long-term	0.7597 (0.1544)	0.8955 (0.1566)	0.8379 (0.1629)	0.5999 (0.1932)	0.5428 (0.3054)
Scale Test Statistic	3.85	0.53	0.44	6.089**	4.006
Significance level:	*** = .99	** = .95	* = .9		

BIBLIOGRAPHY

- Addo, Peter Martey, Monica Billio, and Dominique Guegan.** 2012. “Alternative Methodology for Turning-Point Detection in Business Cycle : A Wavelet Approach.” HAL Universit Paris1 Panthon-Sorbonne (Post-Print and Working Papers) halshs-00694420.
- Adrian, Tobias, and Joshua Rosenberg.** 2008. “Stock Returns and Volatility: Pricing the Short-Run and Long-Run Components of Market Risk.” *Journal of Finance*, 63(6): 2997–3030.
- Aguiar-Conraria, Luís, and Maria Joana Soares.** 2011. “Business cycle synchronization and the Euro: A wavelet analysis.” *Journal of Macroeconomics*, 33(3): 477–489.
- Campbell, John Y., and John Cochrane.** 1999. “Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior.” *Journal of Political Economy*, 107(2): 205–251.
- Campbell, John Y., and Motohiro Yogo.** 2006. “Efficient tests of stock return predictability.” *Journal of Financial Economics*, 81(1): 27–60.
- Campbell, John Y, and W Andrew.** 1997. *Lo, and A. Craig MacKinlay, 1997, The Econometrics of Financial Markets*. Princeton University Press.
- Caraiani, Petre.** 2012. “Stylized facts of business cycles in a transition economy in time and frequency.” *Economic Modelling*, 29(6): 2163–2173.
- Cenesizoglu, Tolga, and Jonathan J. Reeves.** 2013. “CAPM, Components of Beta and the Cross Section of Expected Returns.” CIRANO CIRANO Working Papers 2013s-09.
- Chen, Nai-Fu, Richard Roll, and Stephen A Ross.** 1986. “Economic Forces and the Stock Market.” *Journal of Business*, 383–403.

- Chen, Yu-chin, and Wen-Jen Tsay.** 2011. "Forecasting Commodity Prices with Mixed-Frequency Data: An OLS-Based Generalized ADL Approach."
- Clark, Peter K.** 1987. "The Cyclical Component of U.S. Economic Activity." *The Quarterly Journal of Economics*, 102(4): 797–814.
- Cochrane, John H.** 1996. "A Cross-Sectional Test of an Investment-Based Asset Pricing Model." *Journal of Political Economy*, 104(3): 572–621.
- Cochrane, John Howland.** 2001. *Asset pricing*. Princeton [u.a.]:Princeton Univ. Press.
- Constantinides, George M.** 1990. "Habit formation: A Resolution of the Equity Premium Puzzle." *Journal of Political Economy*, 519–543.
- Creal, Drew, Siem Jan Koopman, and Eric Zivot.** 2010. "Extracting a robust US business cycle using a time-varying multivariate model-based bandpass filter." *Journal of Applied Econometrics*, 25(4): 695–719.
- Cushman, David O., and Tao Zha.** 1997. "Identifying Monetary Policy in a Small Open Economy Under Flexible Exchange Rates." *Journal of Monetary Economics*, 39(3): 433–448.
- Dornbusch, Rudiger.** 1976. "Expectations and Exchange Rate Dynamics." *The Journal of Political Economy*, 1161–1176.
- Eichenbaum, Martin, and Charles L Evans.** 1995. "Some Empirical Evidence on the Effects of Shocks to Monetary Policy on Exchange Rates." *The Quarterly Journal of Economics*, 110(4): 975–1009.
- Engle, Robert F., and José Gonzalo Rangel.** 2009. "High and Low Frequency Correlations in Global Equity Markets." Banco de México Working Papers 2009-17.
- Engle, Robert F, Eric Ghysels, and Bumjean Sohn.** 2009. "Stock Market Volatility and Macroeconomic Fundamentals." *Review of Economics and Statistics*, 0(0).
- Fama, Eugene F, and James D MacBeth.** 1973. "Risk, Return, and Equilibrium: Empirical Tests." *Journal of Political Economy*, 81(3): 607–36.

- Fama, Eugene F, and Kenneth R French.** 1988. "Permanent and Temporary Components of Stock Prices." *Journal of Political Economy*, 96(2): 246–73.
- Fama, Eugene F, and Kenneth R French.** 1993. "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics*, 33(1): 3–56.
- Fama, Eugene F, and Kenneth R French.** 1996. "Multifactor Explanations of Asset Pricing Anomalies." *The Journal of Finance*, 51(1): 55–84.
- Faust, Jon, and John H. Rogers.** 2003. "Monetary Policy's Role in Exchange Rate Behavior." *Journal of Monetary Economics*, 50(7): 1403–1424.
- Gençay, Ramazan, Faruk Selçuk, and Brandon J Whitcher.** 2001a. *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*. Academic press.
- Gençay, Ramazan, Faruk Selçuk, and Brandon Whitcher.** 2001b. "Scaling Properties of Foreign Exchange Volatility." *Physica A: Statistical Mechanics and its Applications*, 289(1): 249–266.
- Gençay, Ramazan, Faruk Selçuk, and Brandon Whitcher.** 2003. "Systematic Risk and Timescales." *Quantitative Finance*, 3(2): 108–116.
- Gencay, Ramazan, Faruk Selcuk, and Brandon Whitcher.** 2005. "Multiscale systematic risk." *Journal of International Money and Finance*, 24(1): 55–70.
- Ghysels, Eric, Pedro Santa-Clara, and Rossen Valkanov.** 2004. "The MIDAS Touch: Mixed Data Sampling Regression Models." Anderson Graduate School of Management, UCLA University of California at Los Angeles, Anderson Graduate School of Management qt9mf223rs.
- Harvey, Andrew C., and Thomas M. Trimbur.** 2003. "General Model-Based Filters for Extracting Cycles and Trends in Economic Time Series." *The Review of Economics and Statistics*, 85(2): 244–255.

- Harvey, Andrew C., Thomas M. Trimbur, and Herman K. Van Dijk.** 2007. "Trends and cycles in economic time series: A Bayesian approach." *Journal of Econometrics*, 140(2): 618–649.
- Harvey, Campbell R.** 1989. "Time-varying conditional covariances in tests of asset pricing models." *Journal of Financial Economics*, 24(2): 289–317.
- Hjalmarsson, Erik.** 2011. "New Methods for Inference in Long-Horizon Regressions." *Journal of Financial and Quantitative Analysis*, 46(03): 815–839.
- In, Francis, and Sangbae Kim.** 2012. *An Introduction to Wavelet Theory in Finance: A Wavelet Multiscale Approach*. World Scientific Publishing Company.
- Jagannathan, Ravi, and Zhenyu Wang.** 1996. "The Conditional CAPM and the Cross-Section of Expected Returns." *Journal of Finance*, 51(1): 3–53.
- Kiefer, Nicholas M, Timothy J Vogelsang, and Helle Bunzel.** 2000. "Simple Robust Testing of Regression Hypotheses." *Econometrica*, 68(3): 695–714.
- Kim, Soyoung.** 2005. "Monetary Policy, Foreign Exchange Policy, and Delayed Overshooting." *Journal of Money, Credit and Banking*, 37(4): 775–82.
- Lai, Ming-Jun.** 1995. "On the Digital Filter Associated with Daubechies' Wavelets." *IEEE Transactions on Signal Processing*, 43(9): 2203–2205.
- Lettau, Martin, and Sydney Ludvigson.** 2001. "Consumption, Aggregate Wealth, and Expected Stock Returns." *Journal of Finance*, 56(3): 815–849.
- Lewellen, Jonathan, Stefan Nagel, and Jay Shanken.** 2010. "A skeptical appraisal of asset pricing tests." *Journal of Financial Economics*, 96(2): 175–194.
- Lo, Andrew W, and Archie Craig MacKinlay.** 1988. "Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test." *Review of Financial Studies*, 1(1): 41–66.
- Merton, Robert C.** 1973. "An Intertemporal Capital Asset Pricing Model." *Econometrica*, 41(5): 867–87.

- Percival, DB, SM Lennox, Y-G Wang, and RE Darnell.** 2011. “Wavelet-based Multiresolution Analysis of Wivenhoe Dam Water Temperatures.” *Water Resources Research*, 47(5): W05552–1.
- Percival, Donald B.** 1995. “On Estimation of the Wavelet Variance.” *Biometrika*, 82(3): 619–631.
- Percival, Donald B, and Andrew T Walden.** 1993. *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques*. Cambridge Univ. Press, New York.
- Percival, Donald B, and Andrew T Walden.** 2000. *Wavelet Methods for Time Series Analysis*. Vol. 4, Cambridge University Press.
- Phillips, Peter C.B., and Ji Hyung Lee.** 2013. “Predictive regression under various degrees of persistence and robust long-horizon regression.” *Journal of Econometrics*, 177(2): 250–264.
- Phillips, Peter C.B., and Tassos Magdalinos.** 2009. “Econometric Inference in the Vicinity of Unity.” Sim Kee Boon Institute for Financial Economics Working Papers CoFie-06-2009.
- Poterba, James M, and Lawrence H Summers.** 1988. “Mean Reversion in Stock Prices: Evidence and Implications.” *Journal of Financial Economics*, 22(1): 27–59.
- Priestley, Maurice Bertram.** 1981. *Spectral Analysis and Time Series*. Academic Press.
- Raihan, Sharif Md, Yi Wen, and Bing Zeng.** 2005. “Wavelet: A new tool for business cycle analysis.” *Federal Reserve Bank of St. Louis Working Paper Series*.
- Ramsey, James B., and Camille Lampart.** 1998. “The Decomposition of Economic Relationships by Time Scale Using Wavelets: Expenditure and Income.” *Studies in Non-linear Dynamics & Econometrics*, 3(1): 2.

- Rangel, Jos Gonzalo, and Robert F. Engle.** 2012. "The Factor–Spline–GARCH Model for High and Low Frequency Correlations." *Journal of Business & Economic Statistics*, 30(1): 109–124.
- Scholl, Almuth, and Harald Uhlig.** 2008. "New Evidence on the Puzzles: Results from Agnostic Identification on Monetary Policy and Exchange Rates." *Journal of International Economics*, 76(1): 1–13.
- Scholz, FW, and MA Stephens.** 1987. "K-sample Anderson–Darling tests." *Journal of the American Statistical Association*, 82(399): 918–924.
- Serroukh, A, and AT Walden.** 2000. "Wavelet scale analysis of bivariate time series ii: statistical properties for linear processes." *Journal of Nonparametric Statistics*, 13(1): 37–56.
- Serroukh, Abdeslam.** 2012. "Wavelet Coefficients Cross-correlation Analysis of Times Series." *Electronic Journal of Applied Statistical Analysis*, 5(2): 289–296.
- Serroukh, Abdeslam, AT Walden, and DB Percival.** 2000. "Statistical properties and uses of the wavelet variance estimator for the scale analysis of time series." *Journal of the American Statistical Association*, 95(449): 184–196.
- Tonn, Victor Lux, H.C. Li, and Joseph McCarthy.** 2010. "Wavelet domain correlation between the futures prices of natural gas and oil." *The Quarterly Review of Economics and Finance*, 50(4): 408–414.
- Valkanov, Rossen.** 2003. "Long-horizon regressions: theoretical results and applications." *Journal of Financial Economics*, 68(2): 201–232.
- Watson, Mark W.** 1986. "Univariate detrending methods with stochastic trends." *Journal of Monetary Economics*, 18(1): 49–75.
- Whitcher, Brandon, Peter Guttorp, and Donald B Percival.** 2000. "Wavelet Analysis of Covariance with Application to Atmospheric Time Series." *Journal of Geophysical Research: Atmospheres (1984–2012)*, 105(D11): 14941–14962.

- Wu, Chien-Fu Jeff.** 1986. “Jackknife, bootstrap and other resampling methods in regression analysis.” *the Annals of Statistics*, 1261–1295.
- Yamada, Hiroshi.** 2005. “Wavelet-based beta estimation and Japanese industrial stock prices.” *Applied Economics Letters*, 12(2): 85–88.
- Yogo, Motohiro.** 2008. “Measuring business cycles: A wavelet analysis of economic time series.” *Economics Letters*, 100(2): 208–212.

Appendix A

CHAPTER 2 PROOFS: WAVELETS

A.1 Proof of Lemma 1

We seek an expression for $Cov[W_{i,t}, W_{j,u}]$ for any arbitrary t and u , but t and u dictate whether $W_{i,t}$ and $W_{j,u}$ are boundary coefficients – if $t < L_i - 1$ or $u < L_j - 1$ then $W_{i,t}$ and $W_{j,u}$ will be boundary coefficients respectively. Define two indicator variables $b_{i,t}$ and $b_{j,u}$ to denote whether $W_{i,t}$ and $W_{j,u}$, respectively, are boundary coefficients. Then the coefficients can be expressed as:

$$W_{i,t} = \sum_{l=0}^{\min\{L_i-1, t\}} h_{i,l} X_{t-l} + b_{i,t} \sum_{l=t+1}^{L_i-1} h_{i,l} X_{N+t-l}$$

$$W_{j,u} = \sum_{k=0}^{\min\{L_j-1, u\}} h_{j,k} X_{u-k} + b_{j,u} \sum_{k=u+1}^{L_j-1} h_{j,k} X_{N+u-k}$$

As a result, $Cov[W_{i,t}, W_{j,u}]$ can be expressed as:

$$Cov \left[\sum_{l=0}^{\min\{L_i-1, t\}} h_{i,l} X_{t-l} + b_{i,t} \sum_{l=t+1}^{L_i-1} h_{i,l} X_{N+t-l}, \sum_{k=0}^{\min\{L_j-1, u\}} h_{j,k} X_{u-k} + b_{j,u} \sum_{k=u+1}^{L_j-1} h_{j,k} X_{N+u-k} \right]$$

$$= \sum_{l=0}^{\min\{L_i-1, t\}} \sum_{k=0}^{\min\{L_j-1, u\}} h_{i,l} h_{j,k} S_{X, t-u+k-l} \cdots$$

$$+ b_{i,t} \sum_{l=t+1}^{L_i-1} \sum_{k=0}^{\min\{L_j-1, u\}} h_{i,l} h_{j,k} S_{X, N+t-u+k-l} \cdots$$

$$+ b_{j,u} \sum_{l=0}^{\min\{L_j-1, u\}} \sum_{k=u+1}^{L_j-1} h_{i,l} h_{j,k} S_{X, t-N-u+k-l} \cdots$$

$$+ b_{i,t} b_{j,u} \sum_{l=t+1}^{L_i-1} \sum_{k=u+1}^{L_j-1} h_{i,l} h_{j,k} S_{X, t-u+k-l}$$

Since the ACVS for X is square summable this can be represented as:

$$= \sum_{l=0}^{\min\{L_i-1, t\}} \sum_{k=0}^{\min\{L_j-1, u\}} h_{i,l} h_{j,k} \int_{-1/2}^{1/2} S_X(f) e^{i2\pi f(t-u+k-l)} df \dots$$

$$\begin{aligned}
& + b_{i,t} \sum_{l=t+1}^{L_i-1} \sum_{k=0}^{\min\{L_j-1, u\}} h_{i,l} h_{j,k} \int_{-1/2}^{1/2} S_X(f) e^{i2\pi f(N+t-u+k-l)} df \dots \\
& + b_{j,u} \sum_{l=0}^{\min\{L_j-1, u\}} \sum_{k=u+1}^{L_j-1} h_{i,l} h_{j,k} \int_{-1/2}^{1/2} S_X(f) e^{i2\pi f(t-N-u+k-l)} df \dots \\
& + b_{i,t} b_{j,u} \sum_{l=t+1}^{L_i-1} \sum_{k=u+1}^{L_j-1} h_{i,l} h_{j,k} \int_{-1/2}^{1/2} S_X(f) e^{i2\pi f(t-u+k-l)} df
\end{aligned}$$

Rearranging:

$$\begin{aligned}
& = \int_{-1/2}^{1/2} \left(\sum_{l=0}^{\min\{L_i-1, t\}} h_{i,l} e^{-i2\pi fl} \right) \left(\sum_{k=0}^{\min\{L_j-1, u\}} h_{j,k} e^{i2\pi fk} \right) S_X(f) e^{i2\pi f(t-u)} df \dots \\
& + b_{i,t} \int_{-1/2}^{1/2} \left(\sum_{l=t+1}^{L_i-1} h_{i,l} e^{-i2\pi fl} \right) \left(\sum_{k=0}^{\min\{L_j-1, u\}} h_{j,k} e^{i2\pi fk} \right) S_X(f) e^{i2\pi f(N+t-u)} df \dots \\
& + b_{j,u} \int_{-1/2}^{1/2} \left(\sum_{l=0}^{\min\{L_i-1, t\}} h_{i,l} e^{-i2\pi fl} \right) \left(\sum_{k=u+1}^{L_j-1} h_{j,k} e^{i2\pi fk} \right) S_X(f) e^{i2\pi f(t-N-u)} df \dots \\
& + b_{i,t} b_{j,u} \int_{-1/2}^{1/2} \left(\sum_{l=t+1}^{L_i-1} h_{i,l} e^{-i2\pi fl} \right) \left(\sum_{k=u+1}^{L_j-1} h_{j,k} e^{i2\pi fk} \right) S_X(f) e^{i2\pi f(t-u)} df
\end{aligned}$$

To keep notation relatively compact and managable, define four new variables α_0 , α_1 , α_2 , and α_3 by evaluating the above expression for $b_{i,t} = b_{j,u} = 0$, $b_{i,t} = 1, b_{j,u} = 0$, $b_{i,t} = 0, b_{j,u} = 1$, and $b_{i,t} = b_{j,u} = 1$, respectively:

$$\begin{aligned}
\alpha_1 & \equiv \int_{-1/2}^{1/2} H_i(f) H_j(f)^* S_X(f) e^{i2\pi f(t-u)} df \\
\alpha_2 & \equiv \int_{-1/2}^{1/2} \left(e^{i2\pi f(t-u)} \sum_{l=0}^t h_{i,l} e^{-i2\pi fl} + e^{i2\pi f(N+t-u)} \sum_{l=t+1}^{L_i-1} h_{i,l} e^{-i2\pi fl} \right) H_j(f)^* S_X(f) df \\
\alpha_3 & \equiv \int_{-1/2}^{1/2} \left(e^{i2\pi f(t-u)} \sum_{k=0}^u h_{j,k} e^{i2\pi fk} + e^{i2\pi f(t-N-u)} \sum_{k=u+1}^{L_j-1} h_{j,k} e^{i2\pi fk} \right) H_i(f) S_X(f) df
\end{aligned}$$

$$\begin{aligned}
\alpha_4 &\equiv \int_{-1/2}^{1/2} \left(\sum_{l=0}^t h_{i,l} e^{-i2\pi fl} \right) \left(\sum_{k=0}^u h_{j,k} e^{i2\pi fk} \right) S_X(f) e^{i2\pi f(t-u)} df \dots \\
&+ \int_{-1/2}^{1/2} \left(\sum_{l=t+1}^{L_i-1} h_{i,l} e^{-i2\pi fl} \right) \left(\sum_{k=0}^u h_{j,k} e^{i2\pi fk} \right) S_X(f) e^{i2\pi f(N+t-u)} df \dots \\
&+ \int_{-1/2}^{1/2} \left(\sum_{l=0}^t h_{i,l} e^{-i2\pi fl} \right) \left(\sum_{k=u+1}^{L_j-1} h_{j,k} e^{i2\pi fk} \right) S_X(f) e^{i2\pi f(t-N-u)} df \dots \\
&+ \int_{-1/2}^{1/2} \left(\sum_{l=t+1}^{L_i-1} h_{i,l} e^{-i2\pi fl} \right) \left(\sum_{k=u+1}^{L_j-1} h_{j,k} e^{i2\pi fk} \right) S_X(f) e^{i2\pi f(t-u)} df
\end{aligned}$$

I now consider each case separately. If neither coefficient is a boundary coefficient then:

$$|Cov[W_{i,t}, W_{j,u}]| \leq |\alpha_1| = \left| \int_{-1/2}^{1/2} H_i(f) H_j(f)^* S_X(f) e^{i2\pi f(t-u)} df \right| \leq \int_{-1/2}^{1/2} |H_i(f)| |H_j(f)| S_X(f) df$$

Lai (1995) implies that as $L \rightarrow \infty$, $|H_i(f)| |H_j(f)| = 0$ for $i \neq j$. Since $S_X(f)$ is bounded this in turn implies that $|\alpha_1| \rightarrow 0$ as $L, n \rightarrow \infty$ and $L/n \rightarrow 0$, and as a result

$$Cov[W_{i,t}, W_{j,u}] \rightarrow 0$$

If $W_{i,t}$ is a boundary coefficient but $W_{j,u}$ is not, then

$$\begin{aligned}
&|Cov[W_{i,t}, W_{j,u}]| \leq |\alpha_2| \\
&\leq \int_{-1/2}^{1/2} \left| e^{i2\pi f(t-u)} \sum_{l=0}^t h_{i,l} e^{-i2\pi fl} + e^{i2\pi f(N+t-u)} \sum_{l=t+1}^{L_i-1} h_{i,l} e^{-i2\pi fl} \right| |H_j(f)| S_X(f) df
\end{aligned}$$

Note that we have:

$$\left| e^{i2\pi f(t-u)} \sum_{l=0}^t h_{i,l} e^{-i2\pi fl} + e^{i2\pi f(N-1-u)} \sum_{l=t+1}^{L_i-1} h_{i,l} e^{-i2\pi fl} \right| \leq \left| \sum_{l=0}^t h_{i,l} e^{-i2\pi fl} \right| + \left| \sum_{l=t+1}^{L_i-1} h_{i,l} e^{-i2\pi fl} \right|$$

but since

$$\begin{aligned} \left| \sum_{l=0}^t h_{i,l} e^{-i2\pi fl} \right| + \left| \sum_{l=t+1}^{L_i-1} h_{i,l} e^{-i2\pi fl} \right| &= \left| H_i(f) - \sum_{l=t+1}^{L_i-1} h_{i,l} e^{-i2\pi fl} \right| + \left| \sum_{l=t+1}^{L_i-1} h_{i,l} e^{-i2\pi fl} \right| \\ &\leq |H_i(f)| + 2 \left| \sum_{l=t+1}^{L_i-1} h_{i,l} e^{-i2\pi fl} \right| \end{aligned}$$

and since $\sum_{l=0}^{L_i-1} h_{i,l}^2 = 1/2$ this implies that as $L \rightarrow \infty$, $\sum_{l=t+1}^{L_i-1} h_{i,l} e^{-i2\pi fl} \rightarrow H_i(f)$, i.e. the magnitude of the missing part of the impulse response sequence is finite and vanishes in the limit. As a result, as $L \rightarrow \infty$

$$\int_{-1/2}^{1/2} \left(\left| \sum_{l=0}^t h_{i,l} e^{-i2\pi fl} \right| + \left| \sum_{l=t+1}^{L_i-1} h_{i,l} e^{-i2\pi fl} \right| \right) |H_j(f)| S_X(f) df \leq 3 \int_{-1/2}^{1/2} |H_i(f)| |H_j(f)| S_X(f) df$$

Since $S_X(f)$ is bounded and in the limit $|H_i(f)| |H_j(f)| = 0$ for $i \neq j$, then if $W_{i,t}$ is a boundary coefficient but $W_{j,u}$ is not we have:

$$Cov[W_{i,t}, W_{j,u}] \rightarrow 0$$

By symmetry, the same holds if $W_{j,u}$ is a boundary coefficient but $W_{i,t}$ is not. Finally, considering the case when both $W_{i,t}$ and $W_{j,u}$ are boundary coefficients

$$|Cov[W_{i,t}, W_{j,u}]| \leq |\alpha_4|$$

By the same argument as above, the absolute value of each integral in α_4 is bounded above by 0 as $L \rightarrow \infty$ and thus we conclude that when both $W_{i,t}$ and $W_{j,u}$ are boundary coefficients

$$Cov[W_{i,t}, W_{j,u}] \rightarrow 0$$

thereby concluding the proof.

A.2 Proof of Corollary 1

Lemma 1 establishes that the wavelet and scaling coefficients are asymptotically uncorrelated, and hence the details and smooth – as linear functions of the wavelet and scaling coefficients – will be uncorrelated as well.

A.3 Proof of Lemma 2

The energy of the wavelet coefficients is given by:

$$\|W_j\|^2 = \frac{1}{N} \sum_{k=0}^{N-1} \left| H_j \left(\frac{k}{N} \right) \right|^2 |\mathcal{X}_k|^2 = \frac{1}{N} \sum_{k=0}^{N-1} \mathcal{H}_j \left(\frac{k}{N} \right) |\mathcal{X}_k|^2$$

where $\mathcal{H}_j \left(\frac{k}{N} \right)$ is the squared gain function for the wavelet filter. The energy of the details is given by:

$$\|\mathcal{D}_j\|^2 = \frac{1}{N} \sum_{k=0}^{N-1} \mathcal{H}_j^2 \left(\frac{k}{N} \right) |\mathcal{X}_k|^2$$

Percival and Walden (2000) observe that since $\mathcal{H}_j(f) + \mathcal{G}_j(f) = 1$, this implies $\mathcal{H}_j(f) \leq 1$ and $\mathcal{G}_j(f) \leq 1$ and, in general, $\|\mathcal{D}_j\|^2 \leq \|W_j\|^2$. However, Lai (1995) shows that $\mathcal{H}_j(f)$ asymptotically in L approaches the squared gain function of an ideal bandpass filter in which for any frequency f , $\mathcal{H}_j(f) \in \{0, 1\}$ and since $\mathcal{H}_j(f)$ is a positive real-valued function, $\mathcal{H}_j(f)$ is then asymptotically identical to $\mathcal{H}_j^2(f)$ thus proving the lemma. The proof for scaling coefficients and the wavelet smooth follows from an analogous argument.

A.4 Proof of Corollary 2

Since the wavelet and scaling coefficients are asymptotically uncorrelated, the details and smooth – as linear functions of the wavelet and scaling coefficients – will be uncorrelated as well. As a result,

$$Var[X] = \sum_{j=1}^{J_0} Var[\mathcal{D}_j] + Var[\mathcal{S}_j]$$

The lemma above proves that this variance decomposition is asymptotically identical to that given by the wavelet variance decomposition:

$$\text{Var}[X] = \sum_{j=1}^{J_0} \text{Var}[W_j] + \text{Var}[V_j]$$

since $\|W_j\|^2 = \|\mathcal{D}_j\|^2$ implies $\text{Var}[\mathcal{D}_j] = \text{Var}[W_j]$.

Appendix B

**CHAPTER 3 PROOFS: A MULTIREOLUTION REGRESSION
FRAMEWORK**

B.1 Proof of Lemma 3

Without loss of generality let $\mu_X = \mu_U = 0$. The absolute summability of the CCVS implies that $\{\mathcal{D}_{j,t}\mathcal{Q}_t\}$ is an ergodic series and hence by the ergodic theorem:

$$\frac{1}{n} \sum_{t=1}^n \mathcal{D}_{j,t}\mathcal{Q}_t \xrightarrow{p} Cov[\mathcal{D}_{j,t}, \mathcal{Q}_t]$$

The spectral representation of X_t is given by

$$X_t = \int_{-1/2}^{1/2} e^{i2\pi ft} dZ_X(f)$$

and hence the spectral representation of $\mathcal{D}_{j,t}$ is given by

$$\begin{aligned} \mathcal{D}_{j,t} &= \sum_{l=0}^{L_j-1} h_{j,l}^{(\mathcal{D})} X_{t-l} = \sum_{l=0}^{L_j-1} h_{j,l}^{(\mathcal{D})} \int_{-1/2}^{1/2} e^{i2\pi f(t-l)} dZ_X(f) \\ &= \int_{-1/2}^{1/2} e^{i2\pi ft} \left(\sum_{l=0}^{L_j-1} h_{j,l}^{(\mathcal{D})} e^{-i2\pi fl} \right) dZ_X(f) = \int_{-1/2}^{1/2} e^{i2\pi ft} \mathcal{H}_j(f) dZ_X(f) \end{aligned}$$

Therefore $dZ_{\mathcal{D}_j}(f) = \mathcal{H}_j(f) dZ_X(f)$. Similarly, the spectral representation of \mathcal{Q}_t is

$$\mathcal{Q}_t = \int_{-1/2}^{1/2} e^{i2\pi ft} H^{\mathcal{Q}}(f) dZ_U(f)$$

where $H^{\mathcal{Q}}(f)$ is the purely real transfer function (zero phase) of the filter on $\{U_t\}$. These spectral representations imply the CCVS is given by:

$$s_{\mathcal{D}_j\mathcal{Q},\tau} = \int_{-1/2}^{1/2} e^{i2\pi f\tau} \mathcal{H}_j(f) H^{\mathcal{Q}}(f) E[dZ_X^*(f) dZ_U(f)]$$

Because $\{X_t\}$ and $\{U_t\}$ have purely continuous spectra and the CCVS is absolutely summable by assumption, $E[dZ_X^*(f) dZ_U(f)] = S_{XU}(f) df$ and hence:

$$s_{\mathcal{D}_j\mathcal{Q},\tau} = \int_{-1/2}^{1/2} e^{i2\pi f\tau} \mathcal{H}_j(f) H^{\mathcal{Q}}(f) e^{i\phi_{\mathcal{Q}}(f)} S_{XU}(f) df$$

Consequently the cross spectrum is given by

$$S_{\mathcal{D}_j, \mathcal{Q}}(f) = \mathcal{H}_j(f)H^{\mathcal{Q}}(f)S_{XU}(f)$$

Since both $\mathcal{H}_j(f)$ and $H^{\mathcal{Q}}(f)$ are real valued, the cospectrum is simply

$$C_{\mathcal{D}_j, \mathcal{Q}}(f) = \mathcal{H}_j(f)H^{\mathcal{Q}}(f)C_{XU}(f)$$

where $C_{XU}(f)$ is the cospectrum between $\{X_t\}$ and $\{U_t\}$. This defines a frequency domain representation of the covariance between $\{\mathcal{D}_{j,t}\}$ and $\{\mathcal{Q}_t\}$:

$$Cov[\mathcal{D}_{j,t}, \mathcal{Q}_t] = \int_{-1/2}^{1/2} \mathcal{H}_j(f)H^{\mathcal{Q}}(f)C_{XU}(f) df$$

As $L, n \rightarrow \infty$ with $L/n \rightarrow 0$, $\mathcal{H}_j(f)H^{\mathcal{Q}}(f) = 1$ for any $f \in \mathcal{F}_{\mathcal{Q}} \cap (1/2^{j+1}, 1/2^j]$ and 0 otherwise, hence:

$$\frac{1}{n} \sum_{t=1}^n \mathcal{D}_{j,t} \mathcal{Q}_t \xrightarrow{p} 2 \int_{\mathcal{F}_{\mathcal{Q}} \cap (1/2^{j+1}, 1/2^j]} \mathcal{H}_j(f)H^{\mathcal{Q}}(f)C_{XU}(f) df$$

since the cospectrum is an even function of frequency. An analogous argument for the MODWT-based smooth concludes the proof.

B.2 Proof of Lemma 4

Asymptotically the product $\mathcal{H}_j(f)\mathcal{H}_i(f) = 0$ for any frequency f and any $i \neq j$, hence Lemma 3 implies:

$$n^{-1} \mathcal{D}'_i \mathcal{D}_j = \frac{1}{n} \sum_{t=1}^n \mathcal{D}_{i,t} \mathcal{D}_{j,t} \xrightarrow{p} Cov[\mathcal{D}_{i,t}, \mathcal{D}_{j,t}] = 0 \quad i \neq j$$

From Lemma 3, the on-diagonal elements converge to the partitioned auto spectrum:

$$n^{-1} \mathcal{D}'_j \mathcal{D}_j = \frac{1}{n} \sum_{t=1}^n \mathcal{D}_{j,t}^2 \xrightarrow{p} Var[\mathcal{D}_{j,t}] = 2 \int_{1/2^{j+1}}^{1/2^j} S_x(f) df$$

Now consider the matrix $n^{-1}\mathcal{D}'Y$. From the lemma:

$$n^{-1}\mathcal{D}'_jY = \frac{1}{n} \sum_{t=1}^n \mathcal{D}_{j,t}Y_t \xrightarrow{p} Cov[\mathcal{D}_{j,t}, Y_t] = \int_{1/2^{j+1}}^{1/2^j} C_{xy}(f)df$$

where $C_{xy}(f)$ is the cospectrum which partitions the covariance between X and Y . The remainder of the proof follows from analogous arguments for the MRA smooth \mathcal{S}_{J_0} with the lower and upper bounds of integration being 0 and $1/2^{J_0}$ respectively.

B.3 Proof of Theorem 1

The small sample (ie. FIR filter) MRR estimator:

$$\hat{\beta} = (\mathcal{D}'\mathcal{D})^{-1}\mathcal{D}'Y$$

is equal to the following under DGP (3.5) (correct specification):

$$\hat{\beta} = \beta + (\mathcal{D}'\mathcal{D})^{-1}\mathcal{D}'\nu$$

The estimator is unbiased if

$$E[(\mathcal{D}'\mathcal{D})^{-1}\mathcal{D}'\nu] = \mathbf{0}$$

Applying the law of iterated expectations under Assumption 2 yields:

$$E[(\mathcal{D}'\mathcal{D})^{-1}\mathcal{D}'\nu] = E[(\mathcal{D}'\mathcal{D})^{-1}\mathcal{D}'E[\nu|\mathcal{D}]] = \mathbf{0}$$

thereby proving unbiasedness. The proof concludes by observing that the continuous mapping theorem and lemma 4 imply that $n(\mathcal{D}'\mathcal{D})^{-1}$ converges to a diagonal matrix of inverses of the partitioned auto spectrum of X while the ergodic theorem and scale-exogeneity assumption together imply $n^{-1}\mathcal{D}'\nu \rightarrow \mathbf{0}$. Application of Slutsky's theorem then implies that $(\mathcal{D}'\mathcal{D})^{-1}\mathcal{D}'\nu \rightarrow \mathbf{0}$, thereby proving consistency.

B.4 Proof of Theorem 2

Under GDP (3.7), the MRR estimator is

$$\hat{\beta} = (\mathcal{D}'\mathcal{D})^{-1}\mathcal{D}'\mathcal{B}\theta + (\mathcal{D}'\mathcal{D})^{-1}\mathcal{D}'\zeta$$

Together the exogeneity assumption, Lemma 4, the Ergodic theorem, and Slutsky's theorem imply $(\mathcal{D}'\mathcal{D})^{-1}\mathcal{D}'\zeta \rightarrow \mathbf{0}$. As in Lemma 3, let $\mathcal{F}_{jk} \equiv \mathcal{F}_{\mathcal{B}_k} \cap (1/2^{j+1}, 1/2^j]$ where $\mathcal{F}_{\mathcal{B}_k}$ is the (perfect) passband for \mathcal{B}_k . Since \mathcal{B} is composed of scale components of X , Lemma 3 implies that if $\mathcal{F}_{jk} = \mathcal{F}_{\mathcal{D}_j}$ then $n^{-1}\mathcal{D}'_j\mathcal{B}_k \rightarrow 2 \int_{1/2^{j+1}}^{1/2^j} S_X(f) df$ and if $\mathcal{F}_{jk} = \emptyset$ then $n^{-1}\mathcal{D}'_j\mathcal{B}_k \rightarrow 0$. If the frequency band of \mathcal{B}_k covers that of \mathcal{D}_j then that of \mathcal{B}_i cannot for all $i \neq k$, hence the k th element of the j th row of the $(J_0 + 1) \times K$ matrix $n^{-1}\mathcal{D}'\mathcal{B}$ converges to $2 \int_{1/2^{j+1}}^{1/2^j} S_X(f) df$ while the remaining $K - 1$ elements converge to 0. Since Lemma 4 states that $n(\mathcal{D}'\mathcal{D})^{-1}$ converges to a diagonal matrix whose j th diagonal is $\left(2 \int_{1/2^{j+1}}^{1/2^j} S_X(f) df\right)^{-1}$, Slutsky's theorem implies that the j th row of the $(J_0 + 1) \times K$ matrix $(\mathcal{D}'\mathcal{D})^{-1}\mathcal{D}'\mathcal{B}$ converges to a vector of zeros with the k th element equal to unity. Hence $\hat{\beta}_j \rightarrow \theta_k$.

Of note, by a similar argument to that above, it can be shown that if the frequency band of \mathcal{D}_j spans that of \mathcal{B}_i and \mathcal{B}_k , then the MRR estimator converges to a weighted average of θ_i and θ_k :

$$\hat{\beta}_j \rightarrow \theta_i w_i + \theta_k w_k$$

where

$$w_i = \frac{\int_{\mathcal{F}_{ji}} S_X(f) df}{\int_{1/2^{j+1}}^{1/2^j} S_X(f) df}$$

and

$$w_k = \frac{\int_{\mathcal{F}_{jk}} S_X(f) df}{\int_{1/2^{j+1}}^{1/2^j} S_X(f) df}$$

B.5 Proof of Corollary 3

Let $\mathcal{B} = X$. Then $K = 1$ and $\mathcal{F}_{j1} = (1/2^{j+1}, 1/2^j]$ for all $j = 1, 2, \dots, J_0 + 1$ (denoting \mathcal{S}_{J_0} by \mathcal{D}_{J_0+1}). The corollary then follows from Theorem 2.

B.6 Proof of Lemma 5

The asymptotic normality, as well as the ω_{11} and ω_{22} elements of the asymptotic covariance matrix follow from Serroukh and Walden (2000) with slight modification by appealing to the CLT for stationary and ergodic martingale difference sequences rather than the standard Lindberg-Lévy CLT.

It remains to derive the off-diagonal elements of the Ω matrix – the asymptotic covariance in the theorem:

$$E \left[\sum_{t=1}^n X_t Y_{t+m} \sum_{s=1}^n X_s^2 \right] - E \left[\sum_{t=1}^n X_t Y_{t+m} \right] E \left[\sum_{s=1}^n X_s^2 \right]$$

Consider first the quantity $E[X_t Y_{t+m} X_s^2]$:

$$\begin{aligned} E[X_t Y_{t+m} X_s^2] &= E \left[\sum_{i \in \mathbb{Z}} \psi_{t-i} \epsilon_i \sum_{j \in \mathbb{Z}} \phi_{t+m-j} \eta_j \sum_{k \in \mathbb{Z}} \psi_{s-k} \epsilon_k \sum_{l \in \mathbb{Z}} \psi_{s-l} \epsilon_l \right] \\ &= \sum_{i \in \mathbb{Z}} \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \sum_{l \in \mathbb{Z}} \psi_{t-i} \phi_{t+m-j} \psi_{s-k} \psi_{s-l} E[\epsilon_i \eta_j \epsilon_k \epsilon_l] \end{aligned}$$

and note that since i, j, k, l index time, the law of iterated expectations and the MDS assumption about $[\epsilon_t \ \eta_t]'$ lead to:

$$E[\epsilon_i \eta_j \epsilon_k \epsilon_l] = \begin{cases} \sigma_\epsilon^2 \Delta & \text{if } i = j \neq k = l \\ \sigma_\epsilon^2 \Delta & \text{if } i = k \neq j = l \\ \sigma_\epsilon^2 \Delta & \text{if } i = l \neq j = k \\ E[\epsilon_0^3 \eta_0] & \text{if } i = j = k = l \\ 0 & \text{otherwise} \end{cases}$$

Let $B \equiv \sum_{i \in \mathbb{Z}} \psi_{t-i} \phi_{t+m-i} \psi_{s-i} \psi_{s-i}$, then

$$E[X_t Y_{t+m} X_s^2] = E[\epsilon_0^3 \eta_0] B +$$

$$\begin{aligned}
& \sigma_\epsilon^2 \Delta \sum_{i \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \psi_{t-i} \phi_{t+m-i} \psi_{s-k}^2 - \sigma_\epsilon^2 \Delta B + \\
& \sigma_\epsilon^2 \Delta \sum_{i \in \mathbb{Z}} \sum_{j \in \mathbb{Z}} \psi_{t-i} \phi_{t+m-j} \psi_{s-i} \psi_{s-j} - \sigma_\epsilon^2 \Delta B + \\
& \sigma_\epsilon^2 \Delta \sum_{i \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \psi_{t-i} \phi_{t+m-k} \psi_{s-k} \psi_{s-i} - \sigma_\epsilon^2 \Delta B
\end{aligned}$$

hence

$$\begin{aligned}
E[X_t Y_{t+m} X_s^2] &= E[\epsilon_0^3 \eta_0] B - 3\sigma_\epsilon^2 \Delta B + \\
& \sigma_\epsilon^2 \Delta \left(\sum_{i \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \psi_{t-i} \phi_{t+m-i} \psi_{s-k}^2 + 2 \sum_{i \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \psi_{t-i} \phi_{t+m-k} \psi_{s-i} \psi_{s-k} \right)
\end{aligned}$$

and therefore

$$E[X_t Y_{t+m} X_s^2] = s_{XY,m} s_X^2 + 2s_{XY,t-s+m} s_{X,t-s} + (E[\epsilon_0^3 \eta_0] B - 3\sigma_\epsilon^2 \Delta) B$$

As a result

$$E \left[\sum_{t=1}^n X_t Y_{t+m} \sum_{s=1}^n X_s^2 \right] = \sum_{t=1}^n \sum_{s=1}^n E [X_t Y_{t+m} X_s^2]$$

$$\sum_{t=1}^n \sum_{s=1}^n E [X_t Y_{t+m} X_s^2] = \sum_{t=1}^n \sum_{s=1}^n \{s_{XY,m} s_X^2 + 2s_{XY,t-s+m} s_{X,t-s} + (E[\epsilon_0^3 \eta_0] B - 3\sigma_\epsilon^2 \Delta) B\}$$

Note now that

$$E \left[\sum_{t=1}^n X_t Y_{t+m} \right] E \left[\sum_{s=1}^n X_s^2 \right] = \sum_{t=1}^n \sum_{s=1}^n s_{XY,m} s_X^2$$

Hence

$$Cov \left[\sum_{t=1}^n X_t Y_{t+m}, \sum_{s=1}^n X_s^2 \right] = \sum_{t=1}^n \sum_{s=1}^n \{2s_{XY,t-s+m} s_{X,t-s} + (E[\epsilon_0^3 \eta_0] B - 3\sigma_\epsilon^2 \Delta) B\}$$

By analogue arguments to those in Serroukh and Walden (1999), it can be shown that

$$\lim_{n \rightarrow \infty} n^{-1} \sum_{t=1}^n \sum_{s=1}^n B = \frac{1}{\sigma_\epsilon^2 \Delta} s_{XY,m} s_X^2$$

and

$$\lim_{n \rightarrow \infty} n^{-1} \sum_{t=1}^n \sum_{s=1}^n 2s_{XY,t-s+m} s_{X,t-s} = \sum_{k \in \mathbb{Z}} 2s_{XY,k+m} s_{X,k}$$

Therefore

$$\lim_{n \rightarrow \infty} n^{-1} Cov \left[\sum_{t=1}^n X_t Y_{t+m}, \sum_{s=1}^n X_s^2 \right] = \omega_{12}$$

which concludes the proof.

B.7 Proof of Corollary 4

Application of Slutsky's theorem establishes normality while simple application of the delta method to the statistic $\bar{Z}_{1,m}/\bar{Z}_2$ yields the remainder of the corollary.

B.8 Proof of Lemma 6

The proof follows from a slight modification of Brockwell and Davis (1991) pg. 219:

$$nCov[\bar{X}, \bar{Y}] = \frac{1}{n} \sum_{t=1}^n \sum_{s=1}^n E[X_t Y_s] = \sum_{|k| < n} \left(1 - \frac{|k|}{n}\right) s_{XY,k}$$

Note that this is an analogue to Brockwell and Davis (1991) p. 219 who present a similar formulation it in the context of the asymptotic variance of a sample mean using auto covariance functions, however their notation easily adapts to the cross covariance even though, in general $s_{XY,k} \neq s_{XY,-k}$. Now, clearly

$$s_{XY,k} \leq |s_{XY,k}|$$

and multiplying both sides of the inequality by the strictly positive value for $|k| < n$:

$$\left(1 - \frac{|k|}{n}\right) s_{XY,k} \leq \left(1 - \frac{|k|}{n}\right) |s_{XY,k}| \leq |s_{XY,k}|$$

and summing over $|k| < n$:

$$\sum_{|k| < n} \left(1 - \frac{|k|}{n}\right) s_{XY,k} \leq \sum_{|k| < n} |s_{XY,k}|$$

Serroukh and Walden (2000) show that under Assumption 3, the cross covariance sequence $s_{XY,k}$ is absolutely summable. Thus taking the limit as $n \rightarrow \infty$:

$$\lim_{n \rightarrow \infty} nCov[\bar{X}, \bar{Y}] = \sum_{k \in \mathbb{Z}} s_{XY,k}$$

by the dominated convergence theorem, thereby concluding the proof.

B.9 Proof of Lemma 7

$$s_{Z_1 Z_2, k} = E[X_t Y_{t+m} X_{t+k}^2] - E[X_t Y_{t+m}] E[X_{t+k}^2]$$

Using the triangle inequality

$$|s_{Z_1 Z_2, k}| \leq |E[X_t Y_{t+m} X_{t+k}^2]| + |E[X_t Y_{t+m}]| E[X_{t+k}^2]$$

and summing over k :

$$\sum_{k \in \mathbb{Z}} |s_{Z_1 Z_2, k}| \leq \sum_{k \in \mathbb{Z}} |E[X_t Y_{t+m} X_{t+k}^2]| + |E[X_t Y_{t+m}]| \sum_{k \in \mathbb{Z}} E[X_{t+k}^2]$$

Under Assumption 3, $|E[X_t Y_{t+m}]| < \infty$ and $\sum_{k \in \mathbb{Z}} E[X_{t+k}^2] < \infty$, hence $|E[X_t Y_{t+m}]| \sum_{k \in \mathbb{Z}} E[X_{t+k}^2] < \infty$. It remains to show that $\sum_{k \in \mathbb{Z}} |E[X_t Y_{t+m} X_{t+k}^2]| < \infty$. First looking at $|E[X_t Y_{t+m} X_{t+k}^2]|$ it is clear:

$$|E[X_t Y_{t+m} X_{t+k}^2]| \leq E[|X_t Y_{t+m} X_{t+k}^2|]$$

Under Assumption 3, X_t and Y_t can be written as below:

$$E[|X_t Y_{t+m} X_{t+k}^2|] = E \left[\left| \sum_{a,b,c,d \in \mathbb{Z}} \psi_{t-a} \epsilon_a \phi_{t+m-b} \eta_b \psi_{t+k-c} \epsilon_c \psi_{t+k-d} \epsilon_d \right| \right]$$

and by the generalization of the triangle inequality

$$E \left[\left| \sum_{a,b,c,d \in \mathbb{Z}} \psi_{t-a} \epsilon_a \phi_{t+m-b} \eta_b \psi_{t+k-c} \epsilon_c \psi_{t+k-d} \epsilon_d \right| \right] \leq \sum_{a,b,c,d \in \mathbb{Z}} |\psi_{t-a} \phi_{t+m-b} \psi_{t+k-c} \psi_{t+k-d}| E[|\epsilon_a \eta_b \epsilon_c \epsilon_d|]$$

By the generalization of Hölder's inequality and our assumptions on ϵ and η :

$$E [|\epsilon_a \eta_b \epsilon_c \epsilon_d|] \leq (E [|\epsilon_0|^4])^{\frac{3}{4}} (E [|\eta_0|^4])^{\frac{1}{4}} < \infty$$

Let $a_\epsilon \equiv E [\epsilon_0^4]$ and $a_\eta \equiv E [\eta_0^4]$, then

$$\sum_{a,b,c,d \in \mathbb{Z}} |\psi_{t-a} \phi_{t+m-b} \psi_{t+k-c} \psi_{t+k-d}| E [|\epsilon_a \eta_b \epsilon_c \epsilon_d|] \leq a_\epsilon^{3/4} a_\eta^{1/4} \sum_{a,b,c,d \in \mathbb{Z}} |\psi_{t-a} \phi_{t+m-b} \psi_{t+k-c} \psi_{t+k-d}|$$

Summing over k yields:

$$a_\epsilon^{3/4} a_\eta^{1/4} \sum_{a,b,c,d,k \in \mathbb{Z}} |\psi_{t-a} \phi_{t+m-b} \psi_{t+k-c} \psi_{t+k-d}|$$

but since all the summations are over \mathbb{Z} , the above summation can be reindexed to

$$a_\epsilon^{3/4} a_\eta^{1/4} \sum_{a',b',c',d' \in \mathbb{Z}} |\psi_{a'} \phi_{b'} \psi_{c'} \psi_{d'}| < \infty$$

This sum is bounded because, under Assumption 3, $\sum_{a',b',c',d' \in \mathbb{Z}} |\psi_{a'} \phi_{b'} \psi_{c'} \psi_{d'}|$ is a Cauchy product of absolutely convergent sequences and, as such, is absolutely convergent as well.

Hence,

$$\sum_{k \in \mathbb{Z}} |E[X_t Y_{t+m} X_{t+k}^2]| < \infty$$

thereby establishing the absolute summability of the cross covariance sequence:

$$\sum_{k \in \mathbb{Z}} |s_{Z_1 Z_2, k}| < \infty$$

B.10 Proof of Corollary 5

Lemma 7 and Serroukh and Walden (2000) establishes the absolute summability of the auto covariance matrix $s_{Z,k}$ which establishes proof of the corollary.

B.11 Proof of Theorem 3

Proof is established by Corollaries 4 and 5.