Structural Asset Pricing Theory with Wavelets

Elizabeth Ann Housworth* Todd B. Walker† Chen Xu‡

November 2018

Abstract

We apply Fourier and wavelet decompositions to structural asset pricing models with time non-separable utility. Through simulations, we show how Fourier decompositions of the utility function, coupled with isolating certain frequencies of the stochastic consumption process, reveal a preference for temporal allocations. We demonstrate the usefulness of wavelets by highlighting their ability to isolate frequency and time, simultaneously. While much work has been devoted to wavelet applications of financial data, we are unaware of papers that use wavelets to analyze structural aspects of asset pricing models.

Keywords: Asset Pricing, Wavelets
JEL Classification Numbers: G12

*Departments of Mathematics, Biology, and Statistics, Indiana University, ehouswor@indiana.edu
†Department of Economics, Indiana University, walkertb@indiana.edu
‡Department of Mathematics, Indiana University, xu50@imail.iu.edu

1 Introduction

Time non-separable utility functions have seen increasing use in economics and finance. They have been shown to improve empirical fit in macroeconomic models [Christiano, Eichenbaum, and Evans (2005)], explain financial anomalies [Campbell and Cochrane (1999), Bansal and Yaron (2004)] and have solid microfoundations [Epstein and Zin (1989), Szeidl and Chetty (2005)].

The time non-separable nature of the utility function implies that agents care not only about the volatility of consumption but also the temporal composition of that volatility. Is volatility resolved sooner or later? With standard separable utility functions, agents answer this question with indifference. However, when utility is time non-separable it matters a great deal as to when the uncertainty is resolved. Under habit formation utility with standard preferences, agents prefer lower frequency fluctuations in consumption. They are even willing to tolerate more volatility as long as that volatility is distributed toward the lower end of the spectrum. Conversely, agents with Epstein-Zin (EZ) recursive utility have a preference for higher frequency fluctuations. Long-run risk carries a high price with EZ preferences and a much lower price with habit formation.

The Fourier transform is ideally suited to identify these differences in utility specifications. Whiteman (1985) was the first to show how frequency domain decompositions can be useful in understanding the temporal composition of volatility. Fourier decompositions of the utility function, coupled with isolating certain frequencies of the stochastic consumption process, will reveal the extent to which the preference for early / late resolution interacts with the properties of the consumption process. Standard time series techniques known to economists are not particularly useful for isolating these important properties (e.g., impulse response functions).

The main contribution of the paper is a wavelet-based decomposition of the stochastic discounting factor. Specifically, we address the question: What are the theoretical restrictions imposed by preferences on asset pricing models in the time-scale (or frequency) domain?[1]

With the tool of maximal overlap discrete wavelet transform (MODWT) multi-resolution analysis (MRA), we extend the analysis of Fourier transforms to recursive preferences. Much of the work in the asset pricing literature applies to models with habit formation [e.g., Otrok, Ravikumar, and Whiteman (2002)], few papers examine recursive preferences, and fewer still study spectral properties of these models.[2]

We use wavelets to study the properties of time non-separable utility. Unlike Fourier transforms, wavelets are able to isolate frequency and time, simultaneously. We conduct a simulation exercise to demonstrate the usefulness of the Fourier and wavelet transforms in asset pricing settings (Sections

[1] We thank an anonymous referee for highlighting this contribution.

[2] Recent exceptions include Dew-Becker and Giglio (2016), which will be discussed more below; Ortu, Tamoni, and Tebaldi (2013), who perform a persistence-based decomposition of the log consumption growth process across time-scales; Bandi and Tamoni (2016) demonstrate that the spectral component corresponding to business cycle frequencies explains cross-sectional differences in asset prices; Boons and Tamoni (2016) examine the importance of macro volatility components with persistence greater than business cycle frequency in determining asset prices; Xyngis (2016), examines multi-resolution analysis of macroeconomic uncertainty; Xyngis (2017) demonstrates that macroeconomic shocks with frequencies lower than the business-cycle are not robustly priced in the cross-section of expected returns; and Kang, In, and Kim (2017) found that among all the scale components of the of the Fama-French (FF) three factors, the cycle period of 8-16 months explained the most variation in expected cross-sectional returns.

2 and 4). As wavelet analysis is not ubiquitous in economics, we provide a brief introduction in Section 3. This serves to fix notation and lays the foundation for the remainder of the paper. Section 5 contains our empirical application. We follow the decomposition of the stochastic discount factor found in Dew-Becker and Giglio (2016), which assumes the underlying stochastic process can be summarized by a vector autoregression. We then perform a discreet wavelet transform of both the data and utility specification, which is our primary contribution. Comparing the two tells us how the different utility specifications price risk. The low frequency movements are much more costly when preferences are of the recursive type as opposed to habit formation.

2 Time Non-Separable Utility

We examine the two most popular time non-separable utility functions: habit formation, which dates back at least to Pollak (1970), and the recursive preference structure of Epstein-Zin (EZ) utility (1989), which is built on a Kreps-Portus framework. Contemporaneous utility with a habit formation specification is given by

$$u_t := \frac{(C_t - X_t)^{(1-\gamma)}}{1-\gamma}$$

where C_t is the consumption level at time t, and $X_t := \sum_{i=1}^{K} b_i C_{t-i}$ is the influence of past consumption on the current level of utility. The parameters $\gamma \in (0,1)$, $K \in \mathbb{Z}^+$ and $b_i > 0, i = 1, 2..., K$ determine the risk aversion and the extent to which past consumption alters today's utility. For example, when $K = 1$, the more a person consumes at t than at $t-1$, the more he/she has to consume at $t+1$ than at t in order to attain the same level of utility.[3]

EZ preferences do not directly model the temporal utility u_t, but instead, model the life-time utility U_t recursively:

$$U_t = \left\{ (1-\beta) C_t^{1-\rho} + \beta \left(\mathbb{E}_t \left[U_{t+1}^{1-\alpha} \right] \right)^{\frac{1-\rho}{1-\alpha}} \right\}^{\frac{1}{1-\rho}}$$

where $\beta \in (0,1)$ is the subjective discount factor, ρ is the inverse of the intertemporal elasticity of substitution (IES), and α is the risk aversion coefficient. Having a separation between risk aversion and the elasticity of substitution is a key benefit of EZ preferences.

We work within the context of the ubiquitous asset pricing model of Lucas (1978), where the equilibrium equates the price of the asset at time t to its discounted expected return at $t+1$,

$$p_t = \beta \mathbb{E}_t \left[\left(\frac{u'(C_{t+1})}{u'(C_t)} \right) (p_{t+1} + d_{t+1}) \right] \tag{1}$$

p_t is the price, $u'(\cdot)$ is marginal utility, and d is the dividend of the asset (equal to the endowment

[3] Habit formation preferences can take alternative forms. For example, X_t is sometimes defined as the average consumption of all agents, $X_t := \sum_{i=1}^{K} b_i \bar{C}_{t-i}$, giving the contemporaneous utility function the interpretation of "Keeping up with the Joneses." Our focus here is on the temporal nature of these preferences and in that regard, the alternative habit specifications all maintain the same feature.

in the original setup of Lucas (1978)). Following the usual convention, we define the stochastic discount factor (SDF) as $M_{t+1} = \beta \mathbb{E}_t(u'(C_{t+1})/u'(C_t))$.

Whiteman (1985) was the first to demonstrate how frequency domain analysis could be used to provide important insights on the time non-separable behavior of utility. Particularly, he showed that the *timing of uncertainty* would not be constant under these alternative preferences, a point cleanly made in the frequency domain. As such, consider a thought experiment similar to that of Otrok, Ravikumar, and Whiteman (2002).[4]

1. Let c_t be the log consumption at time t, $c_t = \ln(C_t)$, and generate $N = 1,000$ samples of $T = 2,000$ draws from the stochastic i.i.d log consumption series $\{c_t\}_{t=1}^{2000}$,

$$C_t = e^{\text{Norm}\left(-\frac{1}{2}\sigma^2,\, \sigma^2\right)}, \quad t = 1, 2, \cdots, 2000.$$

   Note that the normal distribution is constructed in such a way that no matter how the variance is changed, the mean of the consumption is always one.

2. Decompose each sample series into 32 frequency component series using a band-pass filter $b_j(L)$, where the j^{th} component has a frequency range of $\left[\frac{j-1}{32} \cdot \frac{1}{2}, \frac{j}{32} \cdot \frac{1}{2}\right]$. This delivers $b_j(\omega) = 1$ for $\omega \in [j\pi/32, (j+1)\pi/32]$ for $j = 0, 1, ..., 31$ and zero otherwise.

3. Calculate utility frequency-by-frequency for each consumption simulation[5]
   $U_n(j) = \sum_{t=1}^{T} \beta^t u(b_j(L)c_{n,t})$, taking the average of life-time utility for each frequency range, $\mathbb{E}(U(j)) = (1/N)\sum_{n=1}^{N} U_n(j)$. That is, take each simulation ($T = 2,000$ draws) and calculate lifetime utility, then average this lifetime utility over the $N = 1,000$ samples to get the expected value.

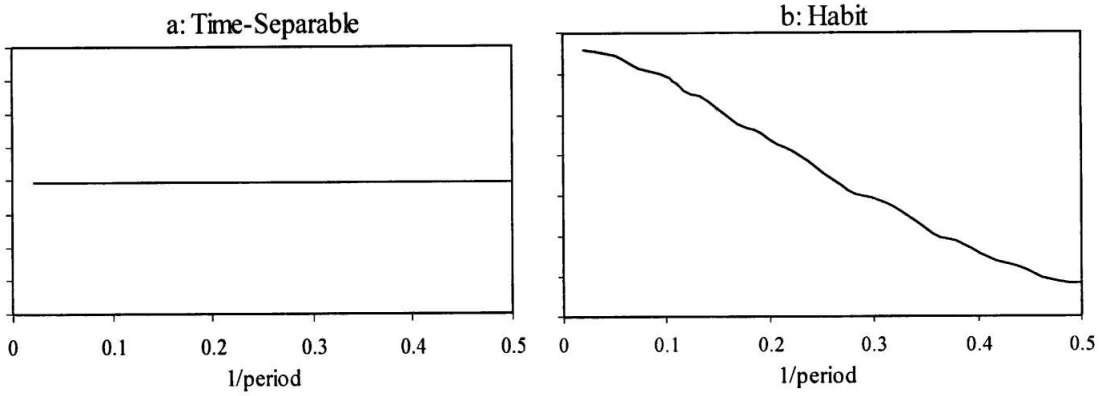


Figure 1: Spectral Utility for Time Separable and Habit Utility

Figure 1 is taken from ORW and plots the results of the computational exercise for habit-formation utility $\frac{(C_t - X_t)^{(1-\gamma)}}{1-\gamma}$ with $X_t = \delta C_{t-1}$, $\delta = -0.615$, $\beta = 0.955$, $\gamma = 0.8$, and standard

[4] We appreciate the comment of an anonymous referee that "raised a slight objection with this approach because it ignores the dynamics of the consumption process across time-scales." Ortu, Tamoni, and Tebaldi (2013) has shown that actual consumption dynamics are richer than our simulation suggests.

[5] Spectral utility is a phrase coined by Whiteman (1985) and refers to frequency decomposition of utility.

time-separable utility ($\delta = 0$). The variance of the consumption process is set to 0.0104.[6] The figure shows utility plotted against different frequencies. With time-separable preferences (Figure 1a), agents are indifferent between fluctuations at different frequencies. As long as volatility remains the same, agents have no preference as to the temporal nature of that volatility. This is not true with habit formation (Figure 1b). Under this preference structure, agents would prefer volatility with a temporal distribution concentrated at lower frequencies. The intuition follows from the idea that last period's consumption enters agents' contemporaneous utility with habit formation. Volatility with higher (lower) frequency makes it more (less) difficult to smooth consumption across periods. Agents would even prefer a higher overall level of volatility as long as the temporal distribution of that volatility was shifted towards the lower end of the spectrum. In other words, low frequency fluctuations in consumption make it easier to support habit formation.

The goal of this paper is to extend this type of analysis through the use of wavelets. As will become clear, wavelets have clear advantages vis-a-vis traditional frequency domain analysis that are particularly beneficial in this context. As wavelets are less ubiquitous, we first provide a brief introduction.

3 Introduction to Wavelet Analysis

In this section, we give a simple, intuitive introduction to discrete-time wavelet analysis. Interested readers should consult Percival and Walden (2006) for a more thorough explanation for both the discrete and continuous cases.

Consider the discrete-time signal in Figure 2, which contains three distinct periods of oscillations. While standard Fourier analysis is able to locate frequency and thus will distinguish amongst the three types of oscillations, it loses information along the time dimension. It is not designed to discriminate *when* these changes in oscillations occur. Wavelets, conversely, can locate these three fluctuations along the *time*-frequency domain.

3.1 Wavelets

A discrete *wavelet* refers to a real bounded sequence $\{w_t\}_{t\in\mathbb{Z}}$ with the following properties

- The term w_t is zero when t is outside a certain range I. That is,

$$w_t = 0, \text{ when } t \notin I = [a, b] \bigcap \mathbb{Z},$$

  where $a < b$ are two real numbers.

- The graph of $\{w_t\}_{t\in I}$ is a wave-like oscillation with a certain frequency (range).

- The sum of all the terms is 0, $\sum_{t\in\mathbb{Z}} w_t = \sum_{t\in I} w_t = 0$.

[6] As stated in Otrok, Ravikumar, and Whiteman (2002), the calibrated parameter values reproduce the average equity premium and risk-free rate over the period 1889-1992.

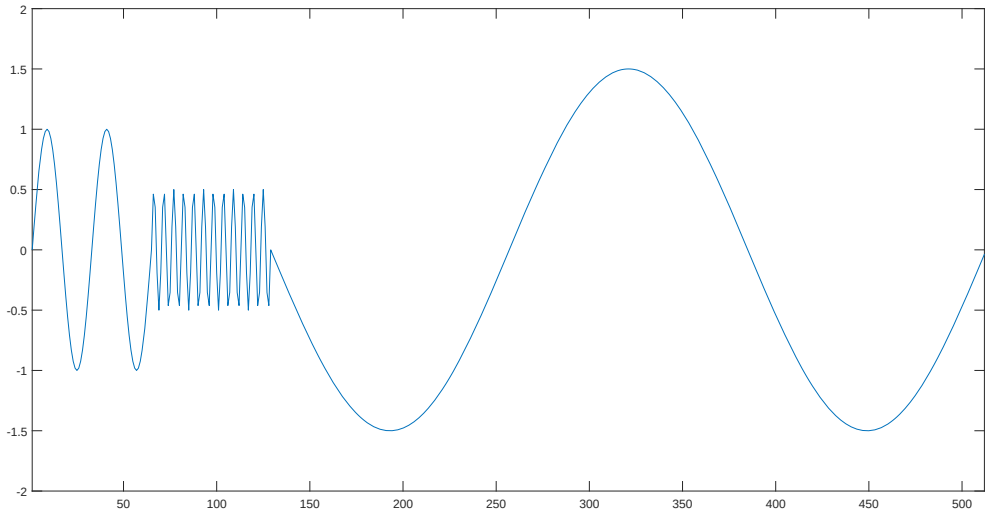


Figure 2: A Hypothetical Signal

Figure 3 is an example of a discrete wavelet $\{w_t\}$ and its Discrete Fourier Transform $F(f) := \sum_{t\in\mathbb{Z}} w_t e^{-i2\pi ft}$. From the graph, we can see that

$$w_t \approx 0, \text{ if } t \notin I = \{18, 19, \cdots, 30\};$$
$$F(f) \approx 0, \text{ if } |f| \in [0, 0.5] \text{ and } |f| \notin [0.25, 0.5].$$

This particular wavelet has a time support of I, where its oscillation frequency approximately lies in the range of $[0.25, 0.5]$.

One important use of wavelets is to measure the power of oscillations in the time-scale domain, where a scale is a frequency range. We use this example wavelet $\{w_t\}$ to illustrate how this can be achieved. First we define the *inner product* of any two real square-summable time series $\{x_t\}_{t\in\mathbb{Z}}$ and $\{y_t\}_{t\in\mathbb{Z}}$ as:

$$\langle x_t, y_t\rangle := \sum_{t\in\mathbb{Z}} x_t \cdot y_t.$$

With this operation, we claim that $\langle w_t, x_t\rangle$ implies the power of the oscillations of $\{x_t\}$ with frequency $f \in [0.25, 0.5]$ in the time range of $t \in I = [18, 19, ..., 30]$. To understand this claim intuitively, look at the two given signals in the Figure 4, where

$$g_t = \begin{cases} \sin(2\pi\cdot(1/16)t) & t \in 1, 2, \cdots, 17; \\ \sin(2\pi\cdot(3/8)t) & t \in 18, 19, \cdots, 30; \\ \sin(2\pi\cdot(1/16)t) & t \in 31, 32, \cdots, 50; \\ 0 & \text{otherwise.} \end{cases} \text{ and } h_t = \begin{cases} \sin(2\pi\cdot(3/8)t) & t \in 1, 2, \cdots, 17; \\ \sin(2\pi\cdot(1/16)t) & t \in 18, 19, \cdots, 30; \\ \sin(2\pi\cdot(3/8)t) & t \in 31, 32, \cdots, 50; \\ 0 & \text{otherwise.} \end{cases}$$

According to the construction of the two given signals, within the time interval of I, the oscil-

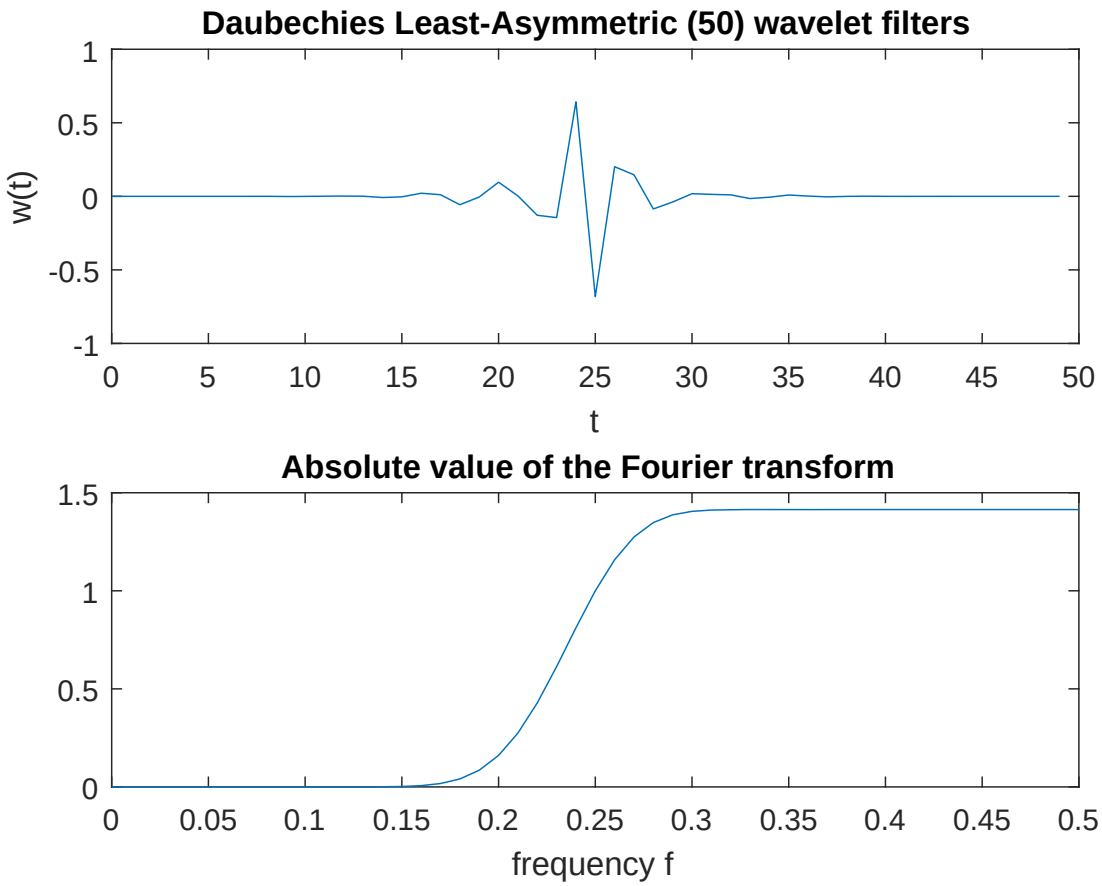


Figure 3: An example of a wavelet

lation frequency in $\{h_t\}$ is $1/16 \notin [0.25, 0.5]$ and the one in $\{g_t\}$ is $3/8 \in [0.25, 0.5]$. Therefore, in I, the oscillations in $\{h_t\}$ with frequency in $[0.25, 0.5]$ has no power, and the oscillations in $\{g_t\}$ with frequency in $[0.25, 0.5]$ has a significant power. The inner products tell us the same thing:

$$\langle w_t, g_t \rangle = \sum_{t \in \mathbb{Z}} w(t) g(t) = 1.2729,$$

$$\langle w_t, h_t \rangle = \sum_{t=-\infty}^{\infty} w_t h_t = 0.022.$$

From this illustration we can see that a wavelet with a oscillation frequency range F and support in the time interval I can be used to determine the power of oscillation of a given signal, in the frequency range F and time interval I. So it is similar to the windowed Fourier transform, but the discrete wavelet transform, which is introduced in the following, has the advantage of adaptive scale-time resolutions.

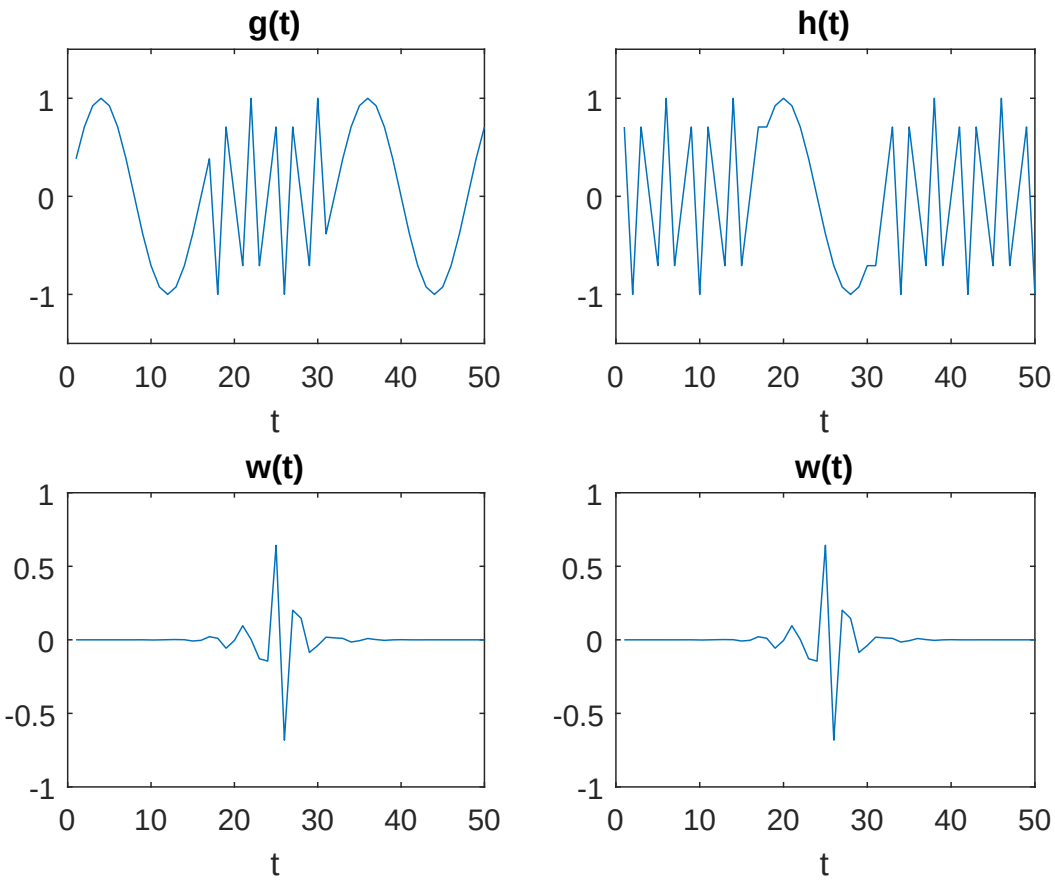


Figure 4: The example wavelet and two given signals

3.2 Discrete Wavelet Transform Given any time series $\{x_t\}$ with real values, the Discrete Wavelet Transform (DWT) describes the powers of its oscillations in the time-scale domain, using a set of wavelets. Each of these wavelets has a unique frequency range and a time support, so the corresponding inner product describes the power of oscillations in $\{x_t\}$ with a distinct frequency range and a time period. For the set of wavelets used in the DWT, we have many options. One option is the popular Daubechies wavelets. In this paper, we will use the set of wavelets constructed with the Daubechies Least-Asymmetry(8) filters. Using these wavelets, the DWT algorithm outputs the inner product of them with any given finite-length signal $\{x_t\}_{t=0}^{2^N}$:

$$\{W_{j,k}\}_{k=0}^{2^{N-j}-1}, j = 1, 2, ..., J \text{ and } \{V_{J,k}\}_{k=0}^{2^{N-j}-1}$$

where

- $W_{j,k}$ is called a wavelet coefficient and it roughly describes the power of the oscillations in $\{x_t\}$ in the time period of $[2^j(k-2), 2^j(k-1)]$ with a frequency range of $[2^{-(j+1)}, 2^{-j}]$. [7]

[7] Note that there are delays in the events in $\{W_{j,k}\}_k$, comparing to the time line of $\{x_t\}$. For a detailed discussion

- The frequency range $[2^{-(j+1)}, 2^{-j}]$ is also called the j^{th} scale.
- $V_{J,k}$ is called the scaling coefficient and it describes the power of the oscillation of $\{x_t\}$ in the time period of $[2^J(k-2), 2^J(k-1)]$ with a frequency range of $[0, 2^{-(J+1)}]$.
- $J \leq N$ is a positive integer determined by us.

If we stack these results in vectors, we get

$$\begin{bmatrix} \vec{\mathbf{w}}_1 \\ \vec{\mathbf{w}}_2 \\ \vdots \\ \vec{\mathbf{w}}_J \\ \vec{\mathbf{v}}_J \end{bmatrix} \quad \text{where } \vec{\mathbf{w}}_j = \begin{bmatrix} W_{j,0} \\ W_{j,1} \\ \vdots \\ W_{j,2^{N-j}} \end{bmatrix} \quad \text{and } \vec{\mathbf{v}}_J = \begin{bmatrix} V_{j,0} \\ V_{j,1} \\ \vdots \\ V_{j,2^{N-J}} \end{bmatrix}$$

Notice that $\vec{\mathbf{w}}_j$ contains the powers of the components in the same scale. Also, the number of wavelet coefficients in $\vec{\mathbf{w}}_j$ decreases as j increases. This is because as j increases, $W_{j,k}$ describes the oscillations with a higher scale and longer time support. To be more specific, the length of $[2^j(k-2), 2^j(k-1)]$ which is the time period that $W_{j,k}$ corresponds to, increases with j. Since the given time series has a fixed time length of 2^N, the number of its scale-j components decreases as j increases.

Next, we give an example of the application of DWT (color plot). The following is an example of applying DWT to a given signal. The results are presented with a color plot.

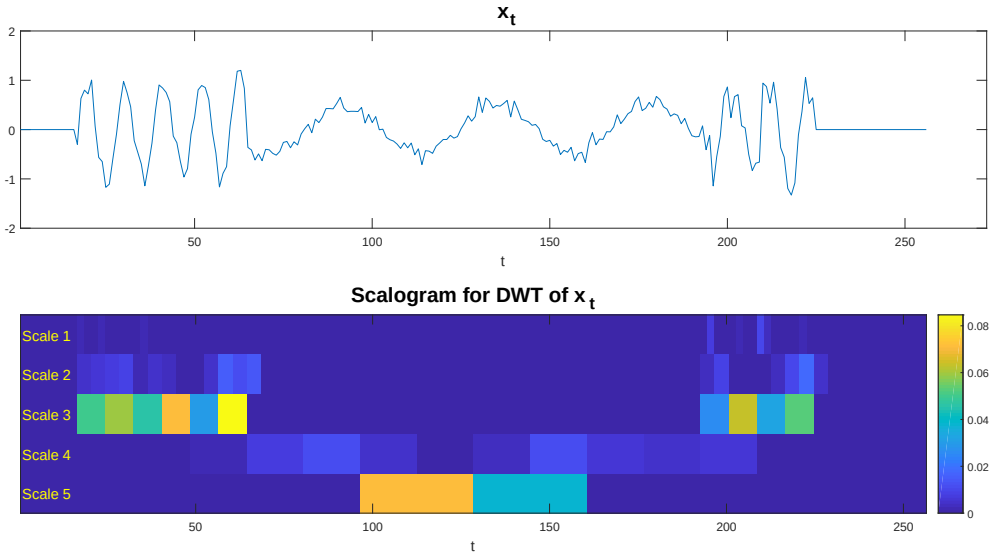


Figure 5: A signal and the color plot of its DWT results

please refer to Section 4.11 in Percival and Walden (2006).

The color plot is on a time-scale plane. Each wavelet coefficient $W_{j,k}$ corresponds to a block. For the corresponding time-scale component in $\{x_t\}$, the width of the block reflects its length in time (so that as scale increases, each block becomes wider), the height of the block reflects the scale, and the brightness of the color reflects the power of the component. The legend on the right of the color plot explains the power of each color. For example, the yellow color indicates a power percentage of 8%, and a green color indicates a power percentage around 6%. It can be seen that the color plot captures in the time-scale domain the powers of different components of the signal above it.

3.3 DWT Algorithm with Matrix Operations

If we stack the scale-j wavelets used by DWT to form a matrix $\mathcal{W}_j$ and write the signal $\{x_t\}_{t=0}^{2^N-1}$ into a vector $X = [x_0\ x_1\ x_2 \cdots x_{2^N-1}]^T$, the scale-$j$ wavelet coefficients $\vec{\mathbf{w}}_j$ can be obtained by:

$$\vec{\mathbf{w}}_j = \mathcal{W}_j X,\ j = 1, 2, ..., J.$$

Similarly, we have

$$\vec{\mathbf{v}}_J = \mathcal{V}_J X.$$

Also, the original signal X can be recovered from the DWT coefficients,

$$X = \sum_{j=1}^{N} \mathcal{W}_j^T \vec{\mathbf{w}}_j + \mathcal{V}_J^T \vec{\mathbf{v}}_J. \tag{2}$$

which will be useful in deriving a key result later.

A technical issue with DWT

There is a practical issue with the DWT when calculating the inner product. In practice, we only have signals with a finite length. Suppose the available signal is $\{x_t\}_{t=0}^{N}$. Then the calculation of certain inner products will involve unavailable x_t's. For example, if we denote the wavelet used to obtain $W_{j,0}$ by $\{h_t\}$, then

$$W_{j,0} = \langle \{h_t\}, \{x_t\} \rangle = \sum_{t \in \mathbb{Z}} h_t x_t.$$

From this formula we can see that the calculation of $W_{j,0}$ involves x_t with negative indices that are not available. To solve this problem, DWT extends $\{x_t\}$ periodically as needed. For example,

$$W_{j,0} = \sum_{t \in \mathbb{Z}} h_{j,0}(t) x_t = \sum_{t \in \mathbb{Z}} h_{j,0}(t) x_{t \mod N}.$$

Therefore, the DWT coefficients at the left boundary of each scale do not reflect the powers of the time-scale components of the original signal unless $\{x_t\}$ is periodic. Our later analysis will exclude such boundary powers.

3.4 Multi-Resolution Analysis with Wavelets

Another important use of wavelets is to construct a set of zero-phase filters, $\{h_t^{(j)}\}_{t\in\mathbb{Z}}, j = 1, 2, ..., J$, so that we can decompose $\{x_t\}$ to a set of time series $\{d_{j,t}\}$, $j = 1, 2, ..., J$ and $\{s_{J,t}\}$ such that:

- the oscillations in $\{d_{j,t}\}$ lies in the scale of $[2^{-(j+1)}, 2^{-j}]$, $j = 1, 2, ..., J$, and the oscillations in $\{s_{J,t}\}$ lies in the scale of $[0, 2^{-(J+1)}]$,
- the oscillations in $\{d_{j,t}\}$ is in the same pace with the scale-j oscillations in $\{x_t\}$, this is due to the fact that $\{h_t^{(j)}\}, j = 1, 2, ..., J$ are constructed in a way that they all have zero phase.
- and the filtering results forms the following scale decomposition of $\{x_t\}$:

$$x_t = d_{1,t} + d_{2,t} + \cdots + d_{J,t} + s_{J,t},\ t \in \mathbb{Z}$$

Again, in practice we only have signals of finite length, say $\{x_t\}_{t=0}^N$ where N is some positive integer. Then we can express the above decomposition using vectors

$$X = D_1 + D_2 + ... + D_J + S_J,$$

where

$$X := [x_0\ x_1\ \cdots x_N]^T, \quad D_j := [d_{j,0}\ d_{j,1}\ \cdots d_{j,N}]^T, \quad \text{and } S_J = [s_{J,0}\ s_{J,1}\ \cdots s_{J,N}]^T.$$

The following is an illustration of decomposition by filtering, where

$$x_t = N(0, 1) + \frac{(t - 512)^{2/3}}{10}, t = 0, 1, 2, ..., 1024$$

is a time series, and the filters are constructed with the modified Daubechies LA-8 filters.

4 Simulation

Recall the simulation exercise of Otrok, Ravikumar, and Whiteman (2002) described in Section 2. We extend this analysis along two dimensions. First, for each simulated series $X_n = \{c_t\}_{t=0}^{2000}$, we use the Daubechies Least-Asymmetric filter described above to decompose the consumption process according to:

$$X_n = D_1 + \cdots + D_J + S_J,$$

where $\{d_{j,t}\}_{t=0}^{2000}$ denote the entries in D_j and $\{s_{J,t}\}_{t=0}^{2000}$ denote the entries in S_J. We set $J = 7$ so that for each $j \in [1, 7]$, the periodicities in each scale are given by $P_{D_1} \in [1/4, 1/2]$, $P_{D_2} \in [1/8, 1/4]$, $P_{D_3} \in [1/16, 1/8]$, $P_{D_4} \in [1/32, 1/16]$, $P_{D_5} \in [1/64, 1/32]$, $P_{D_6} \in [1/128, 1/64]$, and $P_{D_7} \in [1/256, 1/128]$. Therefore, any oscillation in D_1 has a periodicity that ranges from two to four quarters.

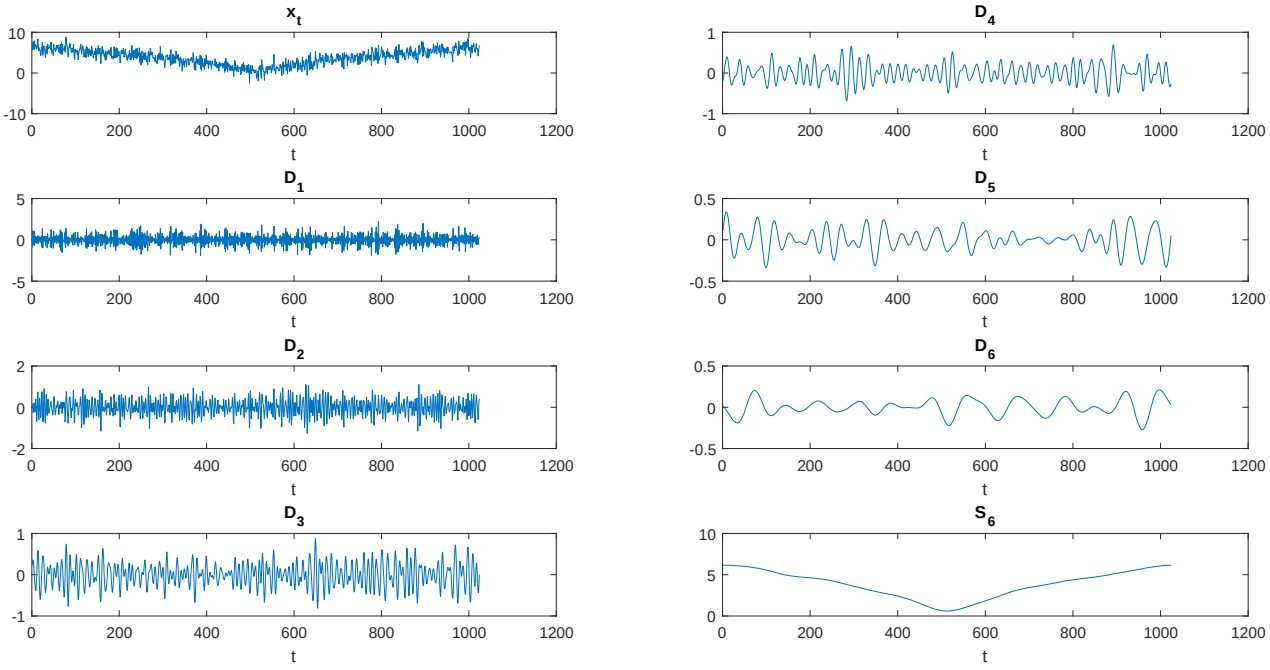


Figure 6: The multi-resolution analysis of a given signal $\{x_t\}$

Second, we examine both habit-formation utility and EZ preferences. The time-t scale-j component of the consumption series is $e^{d_{j,t}}$, and its contribution to the habit-formation, life-time utility is calculated as

$$U_n(d) = \beta^t \frac{\left(e^{d_{j,t}} - \delta e^{d_{j,t-1}}\right)^{1-\gamma}}{1-\gamma}$$

For each pair of fixed (j,t), there are 1000 such contributions, from each simulation. The average will be taken as the expected contribution to the life-time utility. We calculate the expected contribution from each time-scale component to see which one contributes more to lifetime utility, $\mathbb{E}(U_n(d)) = (1/N)\sum_{n=1}^{N} U_n(d)$. As above, we set $\beta = 0.955, \delta = -0.615$, and $\gamma = 0.8$.

Each curve in Figures 7–8 correspond to a particular variance. Recall that scale-j corresponds to the frequency band of $[\frac{1}{2^{j+1}}, \frac{1}{2^j}]$. Scales $1-7$ are the contributions of the D_j coefficients. Figure 7 plots these values for $\delta = -0.615$, while Figure 8 decreases this value to -0.8.

The figures show an increasing pattern in average lifetime utility in the first three scales and a subsequent leveling off thereafter, indicating the agent prefers variation at lower frequencies under the Habit-formation model. This is consistent with the simulation and intuition of Otrok, Ravikumar, and Whiteman (2002).[8] With habit formation preferences, all of the temporal variation in lifetime utility comes from the scales defined over shorter horizons. Scale one (D_1) indicates variation in consumption over a frequency range of two to four quarters. Variation in scales one

[8]Note that Figure 1 plots 1/period on the x-axis, while Figures 7–8 plot scales that are decreasing in frequency on the x-axis.

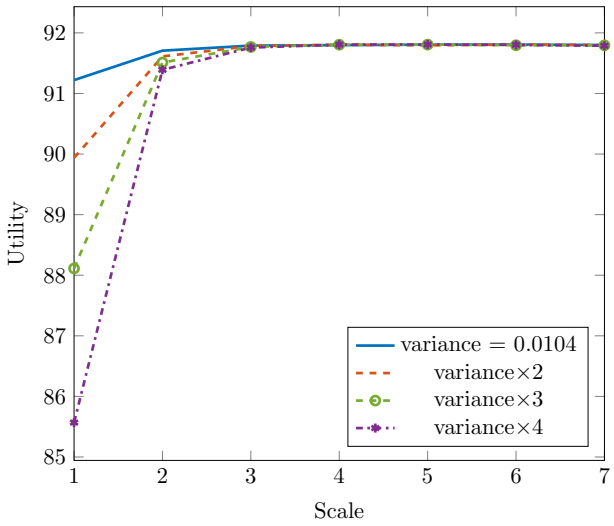


Figure 7: Habit Formation $\delta = -0.615$

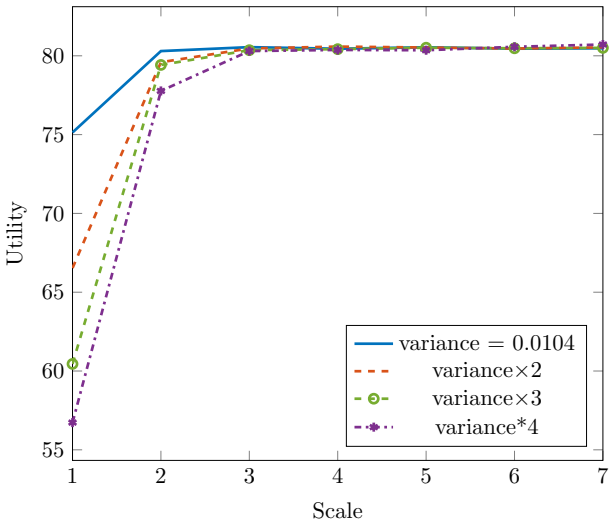


Figure 8: Habit Formation $\delta = -0.8$

through three make it more difficult for agents to smooth consumption across adjacent periods. Agents with habit formation are indifferent to consumption fluctuations occurring at a frequency range above scale three (longer than two to four years). Slightly increasing the degree of habit formation from 0.615 to 0.8, leads to a more significant shift in lifetime utility in scales one through three, but the indifference in scales four through seven remains.

We repeat this exercise for EZ preferences. We calculate lifetime utility $U_{j,T}^{(n)}$ associated with the scale-j consumption component at the last period, where n denotes that it is for the n^{th} simulation. Note that for each simulation we assume $U_{j,t}^{(n)} = 0$ for $t > T$. So

$$U_{j,T}^{(n)} = \left\{(1-\beta)e^{(1-\rho)d_{j,T}^{(n)}}\right\}^{\frac{1}{1-\rho}}, \qquad \mathbb{E}_{T-1}\left(U_{j,T}^{1-\gamma}\right) = \frac{\sum_{n=1}^{1000}\left(U_{j,T}^{(n)}\right)^{1-\gamma}}{1000}$$

For $t = T-1, T-2, \cdots, 0$: $\mathbb{E}_t\left(U_{j,t+1}^{1-\gamma}\right) = \frac{\sum_{n=1}^{1000}\left(U_{j,t+1}^{(n)}\right)^{1-\gamma}}{1000}$, and therefore

$$U_{j,t}^{(n)} = \left\{(1-\beta)e^{(1-\rho)d_{j,t}^{(n)}} + \beta\left(\mathbb{E}_t\left[U_{j,t+1}^{1-\gamma}\right]\right)^{\frac{1-\rho}{1-\gamma}}\right\}^{\frac{1}{1-\rho}} \tag{3}$$

The average expected life-time utility is given by $\mathbb{E}(U_{j,t}) = \sum_{n=1}^{1000}\left(U_{j,t}^{(n)}\right)/1000$.

Figures 9–10 plot the wavelet coefficients and lifetime utility for $\beta = 0.955$, $\gamma = 0.8$, and for various values of ρ. In Figure 9, ρ is set to 0.5 and the variance is increased. The change in lifetime utility is nearly negligible for this specification; utility ranges from 0.8515 to 0.8549 as the variance increases in magnitude by a factor of four (recall that the mean of the consumption process is constant). Figure 10 provides a key insight into the sensitivity of EZ preferences with respect to temporal changes in the consumption process. This figure shows that the parameter ρ governs the quantitative and qualitative response of utility to temporal changes in consumption volatility.

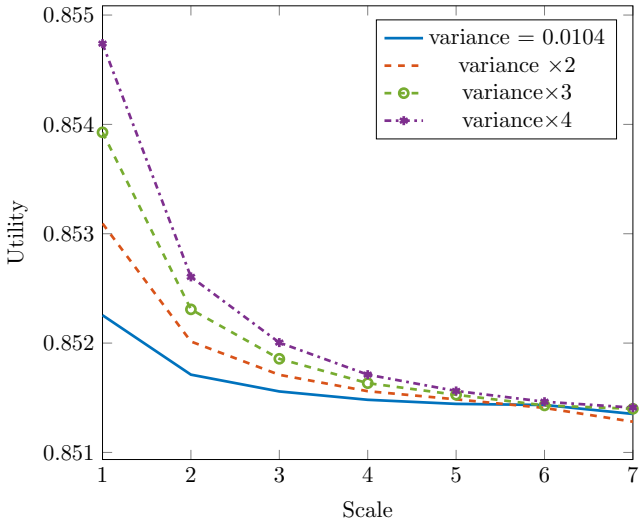


Figure 9: EZ Preferences $\rho = 0.5$

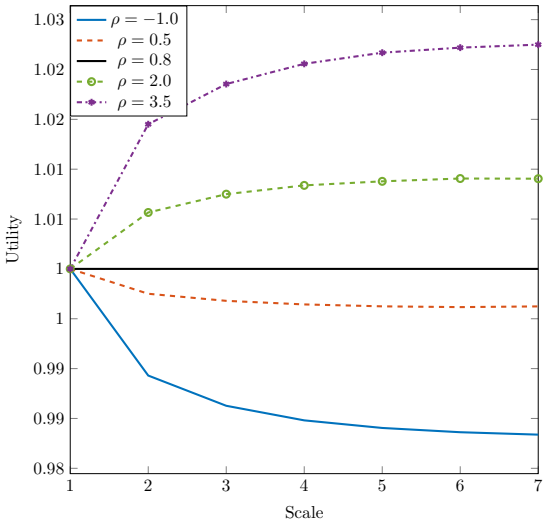


Figure 10: EZ Preferences various ρ

With EZ preferences, we can separate the effects of intertemporal substitution $1/(1-\rho)$ from risk aversion $1-\gamma$. The extent to which EZ preferences have temporal distortions depends upon the relative values of these parameters. For values of ρ equal to γ, (3) reduces to standard time-separable preferences. This is denoted by the flat line at $\rho = 0.8$. Under this scenario, the effects of risk aversion completely offset the degree of intertemporal substitution. For values of ρ less (greater) than α, utility is a monotonically decreasing (increasing) function in scale. When ρ is substantially smaller than α, agents prefer early resolution of risk. That is, they would be indifferent to a slight increase in overall volatility as long as the additional variation was concentrated at higher frequencies. The intertemporal substitution "wins out" in this scenario and consumption smoothing over the long run dominates the risk associated with an increase in overall uncertainty. This preference ordering is the opposite of habit formation. However, when ρ is relatively large, agents prefer variation at lower frequencies, which is identical to habit formation.

5 Empirical Application

The intuition from the previous section carries over to empirical asset pricing applications. We derive a wavelet-based decomposition of the stochastic discount factor of Lucas (1978). In order to do so, we impose Assumptions 1 and 2 of Dew-Becker and Giglio (2016) (D-BG, henceforth).

Assumption 1: The log of the stochastic discount factor $m_{t+1} = \log(M_{t+1})$ can be represented as

$$m_{t+1} = f(\mathcal{I}_t) - \Delta\mathbb{E}_{t+1} \sum_{k=0}^{\infty} z_k x_{t+1+k}, \tag{4}$$

where $f(\mathcal{I}_t)$ is an unspecified function of the time-t information set $\mathcal{I}_t$, $\Delta\mathbb{E}_{t+1} := \mathbb{E}_{t+1} - \mathbb{E}_t$ is the innovation in expectations based on time-t information, x_{t+1+k} are pricing factors with weights, z_k.

The assumption states that the stochastic discount factor can be represented by an unknown function of time-t information $f(\cdot)$ and a linear combination of future values of variables x that contain pricing information at horizon k, where z_k determines the weight of the variables at each horizon. As discussed in D-BG, this assumption provides a log-linearized approximation of the SDF under power utility, habit formation, Epstein-Zin preferences, the CAPM and the ICAPM.

Assumption 2: The pricing factor x_t is driven by an n-dimensional vector moving average process

$$x_t = \mathbf{b}_1 \mathbf{x}_t \tag{5}$$

$$\mathbf{x}_t = \sum_{k=0}^{\infty} \mathbf{\Gamma}_k \mathbf{L}^k \varepsilon_t, \tag{6}$$

where [i.] $\mathbf{x}_t$ is a vector of factors that indicate the status of economy at time t; [ii.] $\varepsilon_t = [\varepsilon_{1,t} \ \varepsilon_{2,t} \ \cdots \ \varepsilon_{n,t}]^T$ is a $n \times 1$ random vector that denotes the fundamental shocks hitting the economy at t. The only assumption placed on the shocks is $\mathbb{E}_{t_1}(\varepsilon_{t_2}) = \mathbf{0}_n$ for any $t_1 < t_2$. The lag operator $\mathbf{L}$ is such that $\mathbf{L}^k \varepsilon_t = \varepsilon_{t-k}$; and the $n \times n$ deterministic matrix $\mathbf{\Gamma}_k$ delivers the effect of the shocks ε_{t-k} to the economy.

Assumption 2 states that the factors driving the economy can we written as a moving-average representation. As stated in D-BG, no additional assumptions (e.g., normality) are imposed on the system. The Wold representation theorem justifies such a representation.

Together these assumptions imply

$$\Delta \mathbb{E}_{t+1}(m_{t+1}) = -\Delta \mathbb{E}_{t+1} \sum_{k=0}^{\infty} z_k x_{t+1+k} \tag{7}$$

$$= -\sum_{i=1}^{n} \left(\sum_{k=0}^{\infty} z_k g_{i,k} \right) \varepsilon_{i,t+1}, \tag{8}$$

where

$$g_{i,k} := \begin{cases} (1,i)\text{th entry of } \mathbf{\Gamma}_k, & \text{if } k \geq 0; \\ 0, & \text{otherwise.} \end{cases}$$

Any particular shock $\varepsilon_{i,t}$ will only affect $\{\mathbf{x}_t, \mathbf{x}_{t+1}, \mathbf{x}_{t+2}, \cdots\}$. Its contributions to $\{x_t, x_{t+1}, x_{t+2}, \cdots\}$ are $\{g_{i,0}\varepsilon_{i,t},\ g_{i,1}\varepsilon_{i,t},\ g_{i,2}\varepsilon_{i,t},\ \cdots\}$, respectively. Therefore, $\{g_{i,k}\}_k$ indicates the impulse response of the i^{th} shock of the variable $\{x_t\}$. The z_k parameters are weights assigned to the impulse responses. The particular values of the weights (z_k) will differ for different specifications of the utility function. Under habit formation, values of z_k are given by

$$z_k = B_{k+1} + D_{k+1} \text{ for } k = 0, 1, 2, \cdots, L, \tag{9}$$

where [9]

$$B_k = \sum_{n=k}^{L} \frac{\alpha h_n}{\beta}, k = 1, 2, ..., L+1; \tag{10}$$

$$D_k = \begin{cases} \sum_{m=1}^{L} \sum_{i=m}^{L} \frac{\alpha b_i h_m}{\beta(1-\sum_{q=1}^{L} b_q)}, & \text{if } k = 1; \\ \sum_{m=0}^{L+1-k} \sum_{i=1+m}^{L} \frac{\alpha b_i h_{m+k}}{\beta(1-\sum_{q=1}^{L} b_q)}, & \text{if } k = 2, 3, ..., L+1. \end{cases} \tag{11}$$

where

$$h_n = \begin{cases} \beta, & \text{if } n = 1; \\ -b_{n-1}\beta^n, & \text{if } n = 2, 3, ..., L+1; \end{cases}$$

Under EZ Preferences, the log-linearization of the stochastic discount factor is

$$\Delta \mathbb{E}_{t+1} m_{t+1} \approx -\left(\rho \Delta \mathbb{E}_{t+1} \Delta c_{t+1} + (\alpha - \rho) \Delta \mathbb{E}_{t+1} \sum_{j=0}^{\infty} \theta^j \Delta c_{t+1+j} \right).$$

Therefore, the values of the weights z_k are given by

$$z_0 = \alpha, \; z_k = (\alpha - \rho) \cdot \theta^k, \text{ for } k \in \mathbb{Z}^+ \tag{12}$$

5.1 Estimation of the Pricing Factor The pricing factor, (5)–(6), is assumed to be driven by an n-dimensional moving average process. The moving average parameters can be estimated by a vector autoregression (VAR) where lag-4 is chosen according to the Akaike information criterion,

$$\bar{\mathbf{x}}_t = \Phi_1 \bar{\mathbf{x}}_{t-1} + \Phi_2 \bar{\mathbf{x}}_{t-2} + \Phi_3 \bar{\mathbf{x}}_{t-3} + \Phi_4 \bar{\mathbf{x}}_{t-4} + \bar{\varepsilon}_t \tag{13}$$

We follow D-BG in using three time series for the estimation, $\bar{\mathbf{x}}_t = (x_{1,t}\ x_{2,t}\ x_{3,t})^T$, where x_{1t} is the natural log of consumption growth; x_{2t} and x_{3t} are the first two principal components of a set of nine financial variables: the aggregate price/earnings and price/dividend ratios; the 10 year/3 month term spread; the Aaa-Baa corporate yield spread (default spread); the small-stock value spread; the unemployment rate minus its 8-year moving average; RREL, the detrended version of the short-term interest rate that Campbell (1991) finds forecasts market returns; the three-month Treasury yield rate; and Lettau and Ludvigson's (2001) CAY. The two principal components are scaled to have the same variance as consumption growth. We employ quarterly data from $1952.1 - 2015.4$.[10] Parameter estimates of the VAR equation (13) are in line with the literature and therefore relegated to Appendix A.

[9] See Appendix A for additional details.

[10] Unlike D-BG, we estimate a four-lag VAR with lag information chosen according to the Akaike information criterion.

5.2 Fourier Transform Letting

$$\mathbf{x}_t := \begin{pmatrix} \bar{\mathbf{x}}_t \\ \bar{\mathbf{x}}_{t-1} \\ \bar{\mathbf{x}}_{t-2} \\ \bar{\mathbf{x}}_{t-3} \end{pmatrix} \text{ and } \Phi := \begin{pmatrix} \Phi_1 & \Phi_2 & \Phi_3 & \Phi_4 \\ I_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} \\ \mathbf{0}_{3\times 3} & I_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} \\ \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & I_{3\times 3} & \mathbf{0}_{3\times 3} \end{pmatrix}$$

then we have $\mathbf{x}_t = \Phi \mathbf{x}_{t-1} + \varepsilon_t$, and we can derive the moving average parameters

$$\mathbf{x}_t = \sum_{k=0}^{\infty} \Phi^k \varepsilon_{(t-k)}$$

as $\Phi^k = \mathbf{\Gamma}_k,$ for $k = 0, 1, 2, \ldots$. We can write (7) as

$$\Delta \mathbb{E}_{t+1}(m_{t+1}) = -\sum_{i=1}^{n} \left(\sum_{k=0}^{\infty} z_k g_{i,k} \right) \varepsilon_{i,t+1},$$

where

$$g_{i,k} := \begin{cases} (1,i)\text{th entry of } \mathbf{\Gamma}_k, & \text{if } k \geq 0; \\ 0, & \text{otherwise.} \end{cases}$$

The sum $\sum_{k=0}^{\infty} z_k g_{i,k}$ measures the contribution of the shock $\varepsilon_{i,t+1}$ to the innovation of the expected log stochastic discount factor $\Delta \mathbb{E}_{t+1}(m_{t+1})$, and is referred as the *price of risk* for shock i. D-BG performed a Fourier transform on the contribution of $\{g_{i,k}\}_k$ and $\{z_k\}_k$ to the risk price. Specifically, they chose three indicators for the economy (so $n = 3$) and for each $i = 1, 2, 3$,

$$\sum_{k=0}^{\infty} z_k g_{i,k} = \frac{1}{2\pi} \int_{-\omega}^{\omega} Z(\omega) G_i(\omega) d\omega,$$

where

$$Z(\omega) := z_0 + 2\sum_{k=1}^{\infty} z_k \cos(\omega k) \quad \text{and} \quad G_i(\omega) := \sum_{k=-\infty}^{\infty} \cos(\omega k) g_{i,k}.$$

Here, $G_i(\omega)$ is the real part of the Fourier transform of $\{g_{i,k}\}_{k\in\mathbb{Z}}$, and $\sum_{k=1}^{\infty} z_k \cos(\omega k)$ is the real part of the Fourier transform of a sequence $\{Z_k\}_{k\in\mathbb{Z}}$ such that

$$Z_k = \begin{cases} 0, & \text{if } k < 0; \\ z_k, & \text{if } k \geq 0. \end{cases}$$

Hence, for any $\omega \in [0, \pi]$, the value of $|G_i(\omega)|$ reflects the contribution of the component with that angular frequency in $\{g_{i,k}\}_k$ to the risk price $\sum_{k=0}^{\infty} z_k g_{i,k}$. This is also true for $Z(\omega)$ and $\{z_k\}_k$.

DG-B's empirical results based on the estimated VAR and standard calibration of preferences show that $G_i(\omega)$ has disproportional mass weight at lower frequencies (e.g., see DG-B Figure 2).

Thus, EZ preferences, which are calibrated to emphasize low-frequency risk, are more sensitive to changes in the price of risk vis-a-vis the habit formation specification.

5.3 Wavelet Transform Instead of the Fourier transform, we apply Discrete Wavelet Transform to the sum $\sum_{k=0}^{\infty} z_k g_{i,k}$. Since $g_{i,k} = 0$ for all $k \in \mathbb{Z}^-$, we can rewrite the sum as

$$\sum_{k=0}^{\infty} z_k g_{i,k} = \sum_{k=-\infty}^{\infty} z_k g_{i,k}.$$

We write it in this form because later we are going to express the sum in terms of the DWT coefficients of $\{g_{i,k}\}_k$ and $\{z_k\}$. When applying the DWT algorithm to $\{g_{i,k}\}_{k=0}^{\infty}$, the "Boundary Condition" (discussed above) is problematic, since the power of $\{g_{i,k}\}$ mainly lies near $k = 0$. We approximate the infinite sum with $\sum_{k=-2^N}^{2^N-1} z_k g_{i,k}$ where 2^N is a large number such that

$$\sum_{k=-2^N}^{2^N-1} z_k g_{i,k} \approx \sum_{k=-\infty}^{\infty} z_k g_{i,k}$$

It is not difficult to find such a number since z_k and $g_{i,k}$ are approximately zero as k increases. We will set $N = 16$, which is appropriate given our data.

To facilitate the wavelet analysis of the approximated risk price $\sum_{k=-2^N}^{2^N-1} z_k g_{i,k}$, let

$$\mathbf{g}_i := (g_{i,-2^N}, g_{i,-2^N+1}, \cdots, g_{i,2^N-1})^T,$$

and

$$\mathbf{z} := (z_{-2^N}, z_{-2^N+1}, \cdots, z_{2^N-1})^T.$$

Let $\vec{\mathbf{w}}_{j,\mathbf{g}_i}$ denote the vector of scale-j wavelet coefficients and $\vec{\mathbf{v}}_{J,\mathbf{g}_i}$ denote the vector of scale-J scaling coefficients, according to equation (2). We then have

$$\mathbf{g}_i = \mathcal{W}_1^T \vec{\mathbf{w}}_{1,\mathbf{g}_i} + \mathcal{W}_2^T \vec{\mathbf{w}}_{2,\mathbf{g}_i} + \cdots + \mathcal{W}_J^T \vec{\mathbf{w}}_{J,\mathbf{g}_i} + \mathcal{V}_J^T \vec{\mathbf{v}}_{J,\mathbf{g}_i}.$$

Applying the above formula to the sum gives

$$\begin{aligned}\sum_{k=-2^N}^{2^N-1} z_k g_{i,k} =& \mathbf{z}^T \cdot \mathbf{g}_i = \mathbf{z}^T \cdot (\mathcal{W}_1^T \vec{\mathbf{w}}_{1,\mathbf{g}_i} + \mathcal{W}_2^T \vec{\mathbf{w}}_{2,\mathbf{g}_i} + \cdots + \mathcal{W}_J^T \vec{\mathbf{w}}_{J,\mathbf{g}_i} + \mathcal{V}_J^T \vec{\mathbf{v}}_{J,\mathbf{g}_i}) \\ =& \vec{\mathbf{w}}_{1,\mathbf{z}}^T \cdot \vec{\mathbf{w}}_{1,\mathbf{g}_i} + \vec{\mathbf{w}}_{2,\mathbf{z}}^T \cdot \vec{\mathbf{w}}_{2,\mathbf{g}_i} + \cdots + \vec{\mathbf{w}}_{J,\mathbf{z}}^T \cdot \vec{\mathbf{w}}_{J,\mathbf{g}_i} + \vec{\mathbf{v}}_{J,\mathbf{z}}^T \cdot \vec{\mathbf{v}}_{J,\mathbf{g}_i}\end{aligned}$$

The last expression above is a time-scale decomposition of $\sum_k z_k g_{i,k}$. Each vector that appears in the expression is a frequency component, with its entries to be the time components of that frequency. In the following sections, we will compare the magnitudes of the time-frequency components of $\mathbf{g}_i$, and see which ones contribute the most to $\sum_k z_k g_{i,k}$ under both the EZ preferences and habit formation utility. To facilitate this comparison, define the power of a term to be its square,

so the total power [11] is $\sum_{l=1}^{J} \|\mathbf{w}\|_{l,g_i}^2 + \|\mathbf{v}\|_{J,g_i}^2$ and the power percentage of $W_{j,k}$ is

$$\text{power percentage of } W_{j,k} := \frac{W_{j,k}^2}{\sum_{l=1}^{J} \|\mathbf{w}\|_{l,g_i}^2 + \|\mathbf{v}\|_{J,g_i}^2}.$$

Specifically, we use the Least Asymmetric Daubechies filter with length 8 (Daubechies LA(8) filter), discussed in Section 3 (see also Percival and Walden (2006)). We select the Least Asymmetric Daubechies wavelet filter because it enables us to line up the events in the transformation coefficients with those in the original signal by shifting. The 'length 8' means the length implies

$$g_l = h_l = 0, \text{ if } l \neq 0, 1, \cdots, 7.$$

5.4 Results Recall that Assumptions 1 and 2 imply a pricing equation given by $\Delta E_{t+1}(m_{t+1}) = -\sum_{i=1}^{n} \left(\sum_{k=0}^{\infty} z_k g_{i,k}\right) \varepsilon_{i,t+1}$, where the z_k coefficients correspond to utility specifications and the $g_{i,k}$ are identified through the wavelet transformation of the VAR. We set $N = 16$, so $\mathbf{g}_i$ cover a period of 2^{32} quarters, which is 2^{30} years.

Figure 11 shows the scalogram for the DWT of $\{g_{1,k}\}$. In this scalogram, each block corresponds to a wavelet coefficient's power (contribution percentage). The larger the scale is, the wider the blocks are, since the width of the block equals the length of the time period the corresponding wavelet coefficient describes. For each wavelet coefficient, the magnitude of the power contribution percentage is reflected by the color of the corresponding block, and the color-magnitude code bar is on the right axis. On the top of this bar, we can see a yellow color represents the percentage around 12%, while the dark blue at the bottom represents 0. The middle of the horizontal axis corresponds to $k = 0$. We can see that the power comes from scale 6 to scale 12, at or right after $k = 0$.

This table shows the wavelet coefficients with significant power percentages. We can see that the wavelet coefficients with the most significant contribution are $W_{6,0}, W_{7,0}, W_{8,0}$ and $W_{9,1}$. The sum of the power contribution percentages from scale-1 to scale-14 wavelet coefficients unaffected by periodic extension is 98.96%, implying the power contribution from higher scale components (contained in unaffected $\{v_{J,k}\}_{k=0}^{2^{N-J}-1}$) is no greater than 1.04%; thus, all significant time-scale components have been captured.

We present the scalograms for the DWT of $\{g_{2,k}\}$ and $\{g_{3,k}\}$ in Appendix A, as similar conclusions can be drawn from them as that of $\{g_{1,k}\}$. The sum of the power contribution percentages of the unaffected wavelet coefficients from scale-1 to scale-14 is 99.61% for $\{g_{2,k}\}$ and 97.10% for $\{g_{3,k}\}$ with most of the contribution coming from coefficients 5 through 12. The implication is that the effects of any given shock $\varepsilon_{i,t}$ on the economy is dominated by the scale-5 to scale-12 components of $\mathbf{g}_i$. This also means that these components dominate the contribution from $\{g_{i,k}\}_k$ for the sum, $\sum_k z_k g_{i,k}$.

[11] Note that $\sum_{l=1}^{J} \|\mathbf{w}\|_{l,g_i}^2 + \|\mathbf{v}\|_{J,g_i}^2$ is constant for all $J = \{1, 2, ..., N\}$, where 2^N is the length of the input signal of DWT.

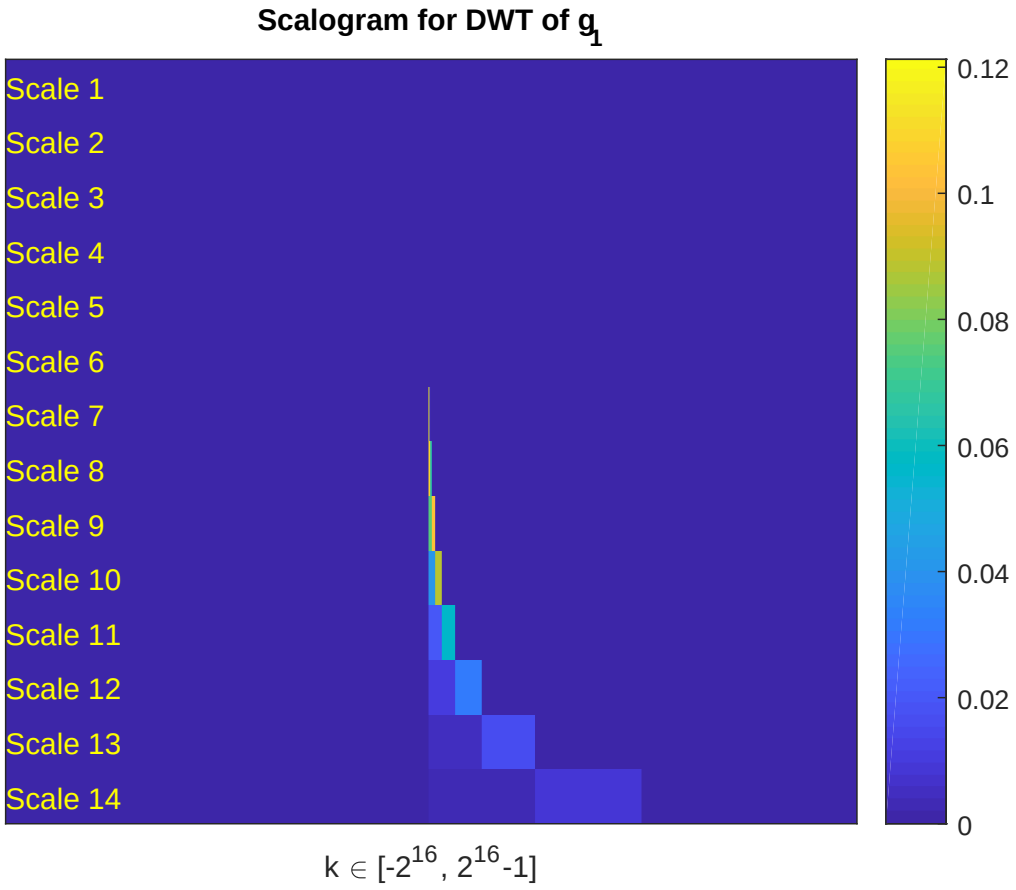


Figure 11: Scalogram for the DWT of $\{g_{1,k}\}$.

How do the wavelet coefficients interact with the two time non-separable utility functions? The $g_{i,k}$ coefficients give the weight at different frequencies and time scales of the DWT, while the $\{z_k\}$ coefficients, determined by the utility functions, provide the weight of the utility at each time scale. To see this point, recall,

$$\sum_k z_k g_{i,k} = \vec{\mathbf{w}}_{1,\mathbf{z}}^T \cdot \vec{\mathbf{w}}_{1,\mathbf{g}_i} + \vec{\mathbf{w}}_{2,\mathbf{z}}^T \cdot \vec{\mathbf{w}}_{2,\mathbf{g}_i} + \cdots + \vec{\mathbf{w}}_{J,\mathbf{z}}^T \cdot \vec{\mathbf{w}}_{J,\mathbf{g}_i} + \vec{\mathbf{v}}_{J,\mathbf{z}}^T \cdot \vec{\mathbf{v}}_{J,\mathbf{g}_i}.$$

In this equation, each $W_{j,t}^{(g)}$ can be viewed as a coefficient of $W_{j,t}^{(z)}$, where $W_{j,t}^{(g)}$ and $W_{j,t}^{(z)}$ denotes the wavelet coefficients of $\{g_{i,k}\}_k$ and $\{z_k\}$, respectively. So for example, when $W_{6,0}^{(g)}$ is large, then the contribution of $W_{6,0}^{(z)}$ to the sum will be magnified through $W_{6,0}^{(g)} \cdot W_{6,0}^{(z)}$.

Habit Formation. For the habit formation specification, we set $N = 10$, $\alpha = 0.5$, $L = 1024$, $\beta = 0.975$, and $b_k = 0.65^k$ for $k = 1, 2, \cdots, L$. The scalogram for the DWT of $\{z_k\}$ under habit formation is in Figure 12 [12]. The values of the most relevant coefficients are in Table 2. From the

[12] To better present the result, we only plot the range where $k \in [-50, 50]$ as the time-scale component beyond this

Coefficient (w_g)	Power $g_{1,k}$	Period Length
$W_{4,0}$	3.49%	4-8 yrs
$W_{5,0}$	6.63%	8-16 yrs
$W_{6,0}$	10.03%	16-32 yrs
$W_{7,0}$	12.12%	32-64 yrs
$W_{8,0}$	10.94%	64-128 yrs
$W_{8,1}$	6.46%	64-128 yrs
$W_{9,0}$	7.43%	128-256 yrs
$W_{9,1}$	10.40%	128-256 yrs
$W_{10,0}$	4.15%	256-512 yrs
$W_{10,1}$	8.85%	256-512 yrs
$W_{11,0}$	2.08%	512-1024 yrs
$W_{11,1}$	5.62%	512-1024 yrs
$W_{12,1}$	3.06%	1024-2048 yrs

Table 1: Power Contribution Percentage for DWT of $g_{1,k}$.

Coefficient	Power Percentage	Period Length
$W_{1,-1}$	6.45%	0.5-1 years
$W_{2,0}$	27.99%	1-2 years
$W_{3,1}$	9.92%	2-4 years
$W_{4,1}$	15.38%	4-8 years
$W_{5,1}$	12.47%	8-16 years
$W_{6,0}$	1.9%	16-32 years

Table 2: Power Contribution Percentage for DWT of z_k under Habit Formation.

table and figure, it is clear that most of the power contribution is coming from coefficients that have scale one to eight. The sum of the power contribution percentages of these coefficients is 97.2%. This implies that the short-term (16 years or less) volatility contributes significantly to the pricing risk.

Comparing Figure 12 / Table 2 to Figure 11 / Table 1, we see that variation in the data is primarily at scales seven and higher, while habit formation (as calibrated) prices risk more significantly at higher frequencies (scales four and lower). Translating to time periods: the habit-formation model is sensitive to fluctuations at periods less than 16 years, while the majority of data variation is for periods greater than 16 years. This result is robust to alternative values for the degree of habit formation b_k, as long as this value is positive.

Epstein-Zin Preferences The calibration for EZ preferences is given by $N = 16, \theta = 0.975, \alpha = 5, \rho = 0.5$. As with habit formation, we plot the scalogram in Figure 13 and provide a table that summarizes the most significant DWT coefficients (z_k), Table 3. In contrast with habit formation, EZ preferences are sensitive to low frequency risk. The sum of the power contribution percentages

range has negligible power percentages

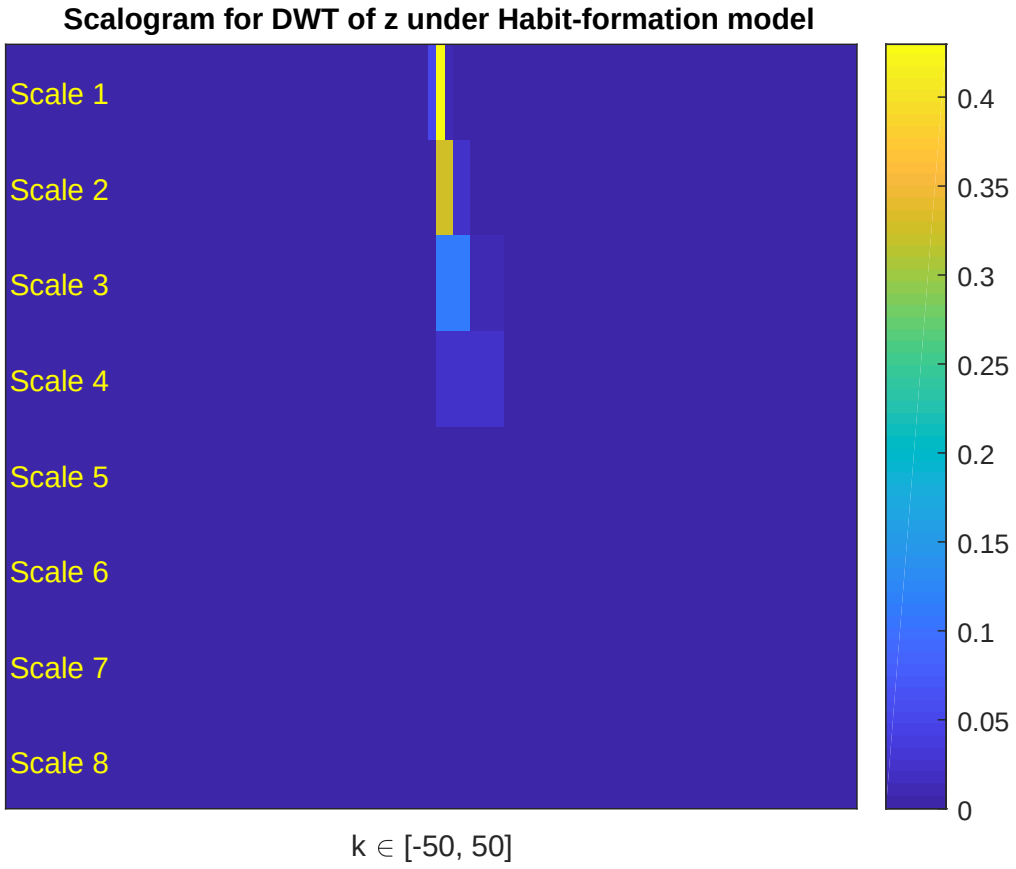


Figure 12: Scalogram for the DWT of $\{z_k\}$ under the Habit-formation model.

from scale-6 to scale-14 is 98.29%. This corresponds to a period of 16 years and higher. Comparing Figure 13 to the DWT of the data, 11, we see that the two are nearly identical. The variation in the data corresponds to where the EZ preferences are most sensitive.

One difference between our results using wavelets and those found using the more standard Fourier transform is the importance of ultra-low frequency variations. Dew-Becker and Giglio (2016) show that consumption cycles lasting 100 years and longer are most important for pricing assets under the standard calibration for EZ preferences. They found very little weight on frequencies shorter than 100 years and no weight on business cycle frequencies. While we find no role at business cycle frequency, we find significant (greater than 25%) contributions coming from horizons less than 100 years using the same data, empirical methodology and calibration.

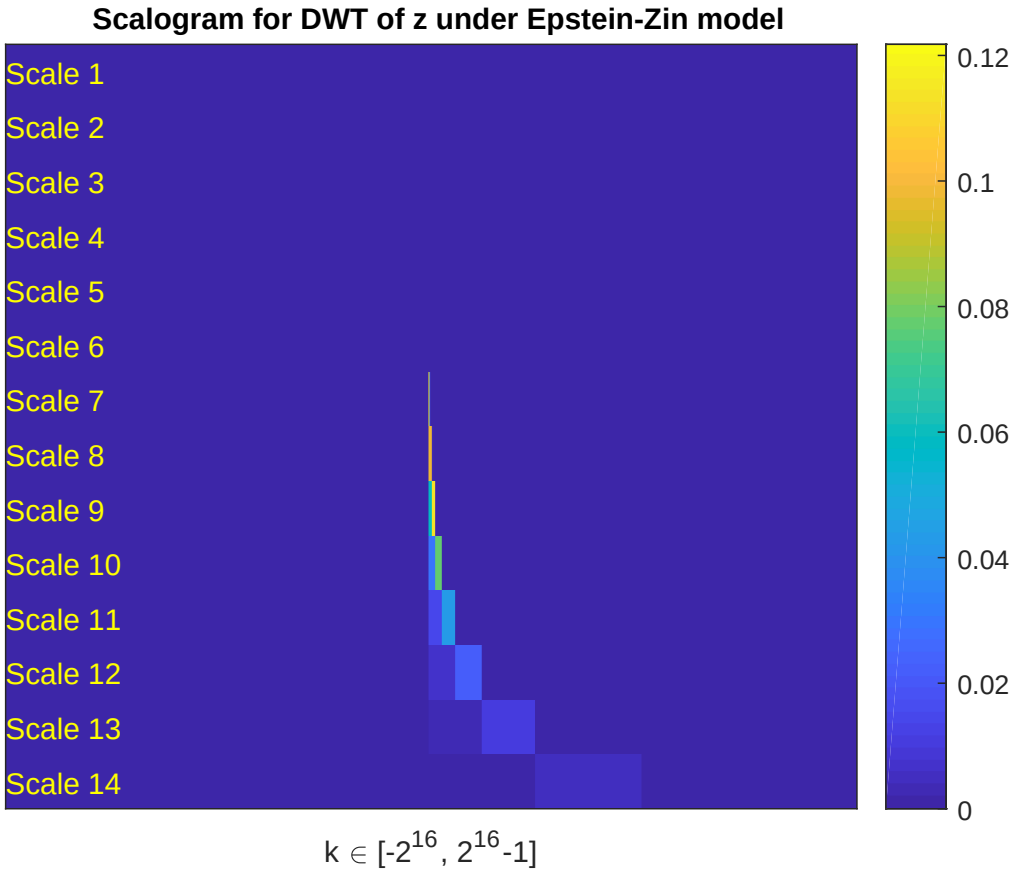


Figure 13: Scalogram for the DWT of $\{z_k\}$ under Epstein-Zin preferences.

5.5 Discrete Wavelet Transformation Based on Haar Filter In this subsection we will repeat the above discrete wavelet transform using the Haar filter instead of the Daubechies LA(8) filter. Although the Haar filter suffers from an important leakage issue[13] and hence is not a good approximation to the ideal band-pass filter, it has a short length of two and therefore can locate events in the time series better in time. Because of these properties, the Haar filter is also used in this literature, e.g. Ortu, Tamoni, and Tebaldi (2013), Bandi and Tamoni (2016), and Boons and Tamoni (2016).

Due to space limitations, we report full results in Online Appendix E, including the scalograms of the DWTs based on the Haar filter. Here we merely note that the analysis using Haar wavelets does not change our underlying message. By comparing the scalograms in Appendix E to Figures 11-13, we see that the dominating components for power contribution are closer to $k = 0$, indicating the effects of the shocks decay more quickly than the previous results (Figure 11-13). This does not change our primary message since the scalograms in Appendix E are very similar to their counterparts in Section 5.4.

[13]Please refer to Percival and Walden (2006) for more details.

Coefficient	Power	Period Length
$W_{6,0}$	11.02%	16-32 yrs
$W_{7,0}$	12.18%	32-64 yrs
$W_{7,1}$	4.15%	32-64 years
$W_{8,0}$	9.72%	64-128 yrs
$W_{8,1}$	9.73%	64-128 yrs
$W_{9,0}$	5.81%	128-256 yrs
$W_{9,1}$	11.39%	128-256 yrs
$W_{10,0}$	2.22%	256-512 yrs
$W_{10,1}$	7.63%	256 years

Table 3: Power Contribution Percentage for DWT of z_k under EZ preferences.

5.6 Empirical Tests In this section, we perform an empirical test to see which utility model can price cross-sectional properties of asset prices. We adopt a method similar to that used in D-BG Dew-Becker and Giglio (2016). They derived the following moment conditions[14] for $\{z_k\}_k$ and $\{g_{i,k}\}_k, i = 1, 2,$ or 3.

$$\mathbb{E}\left((\bar{\mathbf{x}}_{t+1} - \mathbf{\Phi}\mathbf{y}_t) \otimes \mathbf{y}_t,\ \exp(\mathbf{r}_{t+1}) - \exp(r^f_{t+1}) - \mathbf{r}_{t+1}(\bar{\mathbf{x}}_{t+1} - \mathbf{\Phi}^{(1)}\mathbf{y}_t)^T\mathbf{p}\right) = \mathbf{0}, \tag{14}$$

where

- $\bar{\mathbf{x}}_{t+1} = [x_{1,t+1}, x_{2,t+1}, x_{3,t+1}]^T$, which is defined as in (13).
- $\mathbf{y}_t = \begin{bmatrix} \bar{\mathbf{x}}_t \\ \bar{\mathbf{x}}_{t-1} \\ \bar{\mathbf{x}}_{t-2} \\ \bar{\mathbf{x}}_{t-3} \end{bmatrix}$, which is a vector of length 12.
- $\mathbf{r}_{t+1}$ denotes the vector of portfolio returns that are representative of the economy, e.g. the Fama French 25 portfolio returns.
- r^f_{t+1} denotes the risk-free return.
- $\mathbf{\Phi}^{(1)} = [\Phi_1\ \Phi_2\ \Phi_3\ \Phi_4]$ where $\Phi_i, i = 1, 2, 3, 4$ are the coefficient matrices in (13). So $\mathbf{\Phi}^{(1)}$ is a matrix of dimention 3 by 12.
- $\mathbf{p} = 2 \cdot [p_1\ p_2\ p_3]^T$ is two times the vector of risk prices, i.e. for $i = 1, 2,$ or 3,

$$p_i = \sum_{k=1}^{N} z_k \cdot g_{i,k}.$$

[14] Please see Chapter 4 and Appendix F in Dew-Becker and Giglio (2016) for derivation details. We rewrite these conditions using their result 1 and equation (59) for our convenience.

$\mathbf{r_t}$: stacked returns of Fama-French (FF) 25 portfolios sorted on size and BM			
Coefficient	estimate	z-score	p-value
q_1	46.14	3.64	0.0003
q_2	516.82	0.54	0.5893
q_3	−442.16	-0.27	0.7882
$\mathbf{r_t}$: stacked returns of the FF 25 portfolios sorted on size and BM, and the 49 industrial portfolios			
Coefficient	estimate	z-score	p-value
q_1	−11.73	-11.45	0.0000
q_2	−456.05	-4.76	0.0000
q_3	−105.67	-0.75	0.4517

Table 4: Empirical test results

To test which model better prices cross-sectional dynamics, we model z_k in the following way,

$$z_k = q_1 \cdot z_k^{\text{EZ}} + q_2 \cdot z_k^{\text{Power}} + q_3 \cdot z_k^{\text{Habit}},$$

where $\{z_k^{\text{EZ}}\}$ are the weights under the Epstein-Zin preference stated in (12), $z_k^{\text{Power}} = \begin{cases} 1, \text{if } k = 0; \\ 0, \text{otherwise} \end{cases}$ are the weights under the power utility model[15], and $\{z_k^{\text{Habit}}\}$ are the weights under the habit-formation model stated in (9).

We follow the steps taken in Appendix F.3 of Dew-Becker and Giglio (2016) to test which of $[q_1, q_2, q_3]$ is significantly different from zero, using the standard sequential GMM. The estimation/test results are given in Table 4. For more details of the estimation procedures, please refer to our Online Appendix D.

From Table 4, we can see that q_3 is statistically insignificant under both of the return universes, q_2 is significant only when the 49 industrial portfolios are included in the return universe, and q_1 is always significant with the most extreme z-scores. This is prima facie evidence that EZ preferences are relatively more robust at pricing FF portfolios. Note that this does not imply macroeconomic shocks with low frequencies are robustly priced. This is merely a horse race amongst our three models. As shown in Xyngis (2016), preferences that take business cycle dynamics into account are needed to accurately price the risk in the cross-section of returns.

6 Concluding Thoughts

We have shown how to use wavelet analysis to isolate the time variation in non-separable utility. Our primary result is the wavelet decomposition of the stochastic discount factor. We apply spectral analysis to this object and our results are largely consistent with the most recent literature; economic data used to approximate the stochastic discount factor has a strong low frequency component. However, as noted by Xyngis (2016), this does not imply that these low frequency

[15] We set the weights under the power utility model according to equation (12) in Dew-Becker and Giglio (2016)

dynamics are strong predictors of asset prices.

We believe that wavelets can be used to isolate frequency fluctuations in higher-order approximations of the utility function. This would permit one to examine time-varying volatility, an important aspect of financial data. We leave this for future research.

References

Bandi, F. M., and A. Tamoni (2016): "The Horizon of Systematic Risk: A new Beta Representation," https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2337973, Working Paper.

Bansal, R., and A. Yaron (2004): "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 59(4), 1481–1509.

Boons, M., and A. Tamoni (2016): "Horizon-specific macroeconomic risks and the cross section of expected returns," https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2516251, Working Paper.

Campbell, J. Y. (1991): "A Variance Decomposition for Stock Returns," *The Economic Journal*, 101(405), 157–179.

Campbell, J. Y., and J. H. Cochrane (1999): "By Force of Habit: A ConsumptionBased Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, 107(2), 205–251.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (2005): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 113(1), 1–45.

Dew-Becker, I., and S. Giglio (2016): "Asset Pricing in the Frequency Domain: Theory and Empirics," *Review of Financial Studies*, 29(8), 2029–2068.

Epstein, L. G., and S. E. Zin (1989): "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework," *Econometrica*, 57(4), 937–969.

Gençay, R., F. Selçuk, and B. J. Whitcher (2001): *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*. Academic Press, San Diego, CA, 1st edn.

Kang, B. U., F. In, and T. S. Kim (2017): "Timescale betas and the cross section of equity returns: Framework, application, and implications for interpreting the FamaFrench factors," *The Review of Financial Studies*, 42, 15–39.

Lettau, M., and S. Ludvigson (2001): "Consumption, Aggregate Wealth, and Expected Stock Returns," *The Journal of Finance*, 56(3), 815–849.

Lucas, J. R. E. (1978): "Asset Prices in an Exchange Economy," *Econometrica*, 46(6), 1429–1445.

Ortu, F., A. Tamoni, and C. Tebaldi (2013): "Long-Run Risk and the Persistence of Consumption Shocks," *The Review of Financial Studies*, 26(11), 2876–2915.

Otrok, C., B. Ravikumar, and C. H. Whiteman (2002): "Habit Formation: A Resolution of the Equity Premium Puzzle?," *Journal of Monetary Economics*, 49(6), 1261–1288.

Percival, D., and A. Walden (2006): *Wavelet Methods for Time Series Analysis*, Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Pollak, R. (1970): "Habit Formation and Dynamic Demand Functions," *Journal of Political Economy*, 78(4), 745–763.

Szeidl, A., and R. Chetty (2005): "Consumption Commitments: Neoclassical Foundations for Habit Formation," 2005 Meeting Papers 122, Society for Economic Dynamics.

Whiteman, C. H. (1985): "Spectral Utility, Wiener-Hopf Techniques, and Rational Expectations," *Journal of Economic Dynamics and Control*, 9(2), 225–240.

Xyngis, G. (2016): "Are Low-Frequency Macroeconomic Risks Priced in Asset Prices? A Critical Appraisal of Epstein-Zin Preferences," Manuscript, Cardiff Business School.

——— (2017): "Business-Cycle Variation in Macroeconomic Uncertainty and the Cross-Section of Expected Returns: Evidence for Scale-Dependent Risks," *Journal of Empirical Finance*, 44, 43–65.

Online Appendix (Not for Publication)

Appendix A. Weights z_k under the Habit-formation model

This section derives the theoretical values of z_k in equation (4) under the Habit-formation model. It also includes the derivation of equation (10) and (11). We will use the log-linearization method.

The life time utility function The utility at time t in the Habit-formation model is defined as

$$u(C_t, X_t) = \frac{(C_t - X_t)^{(1-\alpha)}}{1-\alpha}, \tag{A.1}$$

where C_t is the consumption level at time t, $\alpha \in (0,1)$ is a parameter, $X_t = \sum_{i=1}^{L} b_i C_{t-i}$, and $b_i \in (0,1)$ are the parameters. The lifetime utility at time t is defined as

$$U_t := \mathbb{E}_t \sum_{n=0}^{\infty} \beta^n u(C_{t+n}, X_{t+n}). \tag{A.2}$$

The asset pricing model In order to define the stochastic discount factor, we first describe an asset pricing model. At the beginning of each period t, assume the representative agent

- receives a constant endowment $\bar{e}$;
- receives the pay-off of his investment at the beginning of period $t-1$, $s_{t-1}(p_t + d_t)$, where p_t is the current price of the asset, d_t is the dividend of the asset, and s_{t-1} is the number of shares he bought at the beginning of $t-1$;
- and invests $s_t \cdot p_t$ on the same asset.

Then the budget constraint condition of the agent at period t is

$$\bar{e} + s_{t-1}(d_t + p_t) = C_t + s_t p_t. \tag{A.3}$$

Next we will derive an expression for the price p_t. At the beginning of t, suppose the agent wants to maximize his life-time utility when deciding how much he invests and consumes, which will lead to the first-order condition of s_t:

$$\begin{aligned} 0 = \mathbb{E}_t \frac{\partial U_t}{\partial s_t} &= \mathbb{E}_t \sum_{n=0}^{\infty} \beta^n \frac{\partial u(C_{t+n}, X_{t+n})}{\partial s_t} \\ = \mathbb{E}_t \sum_{n=0}^{\infty} \beta^n &\left(\frac{\partial u(C_{t+n}, X_{t+n})}{\partial C_{t+n}} \cdot \frac{\partial C_{t+n}}{\partial s_t} + \frac{\partial u(C_{t+n}, X_{t+n})}{\partial X_{t+n}} \cdot \frac{\partial X_{t+n}}{\partial s_t} \right). \end{aligned} \tag{A.4}$$

From (A.3), we can derive that

$$\frac{\partial C_{t+n}}{\partial s_t} = \begin{cases} -p_t, & \text{if } n = 0; \\ p_{t+1} + d_{t+1}, & \text{if } n = 1; \\ 0, & \text{otherwise.} \end{cases} \tag{A.5}$$

Since $X_t = \sum_{i=1}^{L} b_i C_{t-i}$, $X_{t+n} = \sum_{i=1}^{L} b_i C_{t+n-i}$. So

$$\frac{\partial X_{t+n}}{\partial s_t} = \begin{cases} -b_1 p_t, & \text{if } n = 1; \\ -b_n p_t + b_{n-1}(d_{t+1} + p_{t+1}), & \text{if } n = 2, 3, ..., L; \\ b_L(p_{t+1} + d_{t+1}), & \text{if } n = L+1; \\ 0, & \text{otherwise.} \end{cases} \tag{A.6}$$

Plugging (A.5) and (A.6) to (A.4), we get

$$\begin{aligned} 0 =& \mathbb{E}_t \sum_{n=0}^{L+1} \beta^n \left(\frac{\partial u(C_{t+n}, X_{t+n})}{\partial C_{t+n}} \cdot \frac{\partial C_{t+n}}{\partial s_t} + \frac{\partial u(C_{t+n}, X_{t+n})}{\partial X_{t+n}} \cdot \frac{\partial X_{t+n}}{\partial s_t} \right) \\ =& (C_t - X_t)^{-\alpha} \cdot (-p_t) + \mathbb{E}_t \left(\beta (C_{t+1} - X_{t+1})^{-\alpha} (d_{t+1} + p_{t+1} + b_1 p_t) \right) \\ & - \mathbb{E}_t \left(\sum_{n=2}^{L} \beta^n (C_{t+n} - X_{t+n})^{-\alpha} [b_n(-p_t) + b_{n-1}(d_{t+1} + p_{t+1})] \right) \\ & - \mathbb{E}_t \left[\beta^{L+1} (C_{t+L+1} - X_{t+L+1})^{-\alpha} b_L (d_{t+1} + p_{t+1}) \right]. \end{aligned}$$

So if we gather all the terms containing p_t,

$$\begin{aligned} & (C_t - X_t)^{-\alpha} \cdot p_t - \mathbb{E}_t \left(\beta b_1 (C_{t+1} - X_{t+1})^{-\alpha} \right) \cdot p_t \\ & \qquad - \mathbb{E}_t \left(\sum_{n=2}^{L} \beta^n b_n (C_{t+n} - X_{t+n})^{-\alpha} \right) \cdot p_t \\ =& \mathbb{E}_t \left(\beta (C_{t+1} - X_{t+1})^{-\alpha} (d_{t+1} + p_{t+1}) \right) \\ & \qquad - \mathbb{E}_t \left(\sum_{n=2}^{L+1} \beta^n b_{n-1} (C_{t+n} - X_{t+n})^{-\alpha} (d_{t+1} + p_{t+1}) \right) \end{aligned}$$

Therefore, we have

$$\begin{aligned} p_t &= \frac{\mathbb{E}_t \left(\sum_{n=1}^{L+1} h_n (C_{t+n} - X_{t+n})^{-\alpha} (p_{t+1} + d_{t+1}) \right)}{\mathbb{E}_t \left(\sum_{n=0}^{L} g_n (C_{t+n} - X_{t+n})^{-\alpha} \right)} \\ &= \mathbb{E}_t \left(\frac{\left(\sum_{n=1}^{L+1} h_n \cdot (C_{t+n} - X_{t+n})^{-\alpha} \right)}{\mathbb{E}_t \left(\sum_{n=0}^{L} g_n \cdot (C_{t+n} - X_{t+n})^{-\alpha} \right)} \cdot (p_{t+1} + d_{t+1}) \right), \end{aligned} \tag{A.7}$$

where

$$h_n = \begin{cases} \beta, & \text{if } n = 1; \\ -b_{n-1}\beta^n, & \text{if } n = 2, 3, ..., L+1; \end{cases} \quad g_n = \begin{cases} 1, & \text{if } n = 0; \\ -b_n\beta^n, & \text{if } n = 1, 2, 3, ..., L; \end{cases}$$

Stochastic discount factor (SDF) M_{t+1}. By (A.7), we define the stochastic discount factor (SDF) as

$$M_{t+1} := \frac{\left(\sum_{n=1}^{L+1} h_n \cdot (C_{t+n} - X_{t+n})^{-\alpha}\right)}{\mathbb{E}_t\left(\sum_{n=0}^{L} g_n \cdot (C_{t+n} - X_{t+n})^{-\alpha}\right)}, \tag{A.8}$$

so we have

$$p_t = \mathbb{E}_t\left(M_{t+1} \cdot (p_{t+1} + d_{t+1})\right).$$

Let $m_{t+1} = \log M_{t+1}$. Our goal is to express $\Delta\mathbb{E}_{t+1}m_{t+1}$ in the following form and get the values for z_k.

$$\Delta\mathbb{E}_{t+1}m_{t+1} = -\Delta\mathbb{E}_{t+1}\sum_{k=0}^{L} z_k \Delta c_{t+1+k}, \tag{A.9}$$

where Δc_{t+1} denotes the log consumption growth, that is,

$$e^{c_t} = C_t, \text{ and } \Delta c_{t+1} := c_{t+1} - c_t = \log(C_{t+1}/C_t).$$

Notice that

$$\begin{aligned} \Delta\mathbb{E}_{t+1}m_{t+1} =& \Delta\mathbb{E}_{t+1}\log(M_{t+1}) \\ =& \Delta\mathbb{E}_{t+1}\log\left(\sum_{n=1}^{L+1} h_n \cdot (C_{t+n} - X_{t+n})^{-\alpha}\right) \\ & - \Delta\mathbb{E}_{t+1}\log\left[\mathbb{E}_t\left(\sum_{n=0}^{L} g_n \cdot (C_{t+n} - X_{t+n})^{-\alpha}\right)\right] \\ =& \Delta\mathbb{E}_{t+1}\log\left(\sum_{n=1}^{L+1} h_n \cdot (C_{t+n} - X_{t+n})^{-\alpha}\right), \end{aligned} \tag{A.10}$$

where the last equality is because $\Delta\mathbb{E}_t(\text{constant}) = 0$. In the later sections of this appendix, we will rewrite the left-hand side of (A.10) using log-linearization so that it will have the same format as in (A.9).

Expressing $\Delta\mathbb{E}_{t+1}m_{t+1}$ in the format of (A.9)

Claim A.1:

$$(C_t - X_t)^{-\alpha} = \left[e^{c_t} - \sum_{i=1}^{L} b_i e^{c_{t-i}}\right]^{-\alpha} = e^{-\alpha c_t} \cdot B \cdot A_t,$$

where

$$B = [1 - \sum_{i=1}^{L} b_i]^{-\alpha}, \quad \text{and } A_t = 1 + \sum_{i=1}^{L} \frac{-\alpha b_i}{1 - \sum_{q=1}^{L} b_q} \sum_{m=0}^{i-1} \Delta c_{t-m}.$$

Proof. First, we have

$$c_t - c_{t-k} = \Delta c_t + \Delta c_{t-1} + ... + \Delta c_{t-k+1}. \tag{A.11}$$

$$\begin{aligned}
&\left[e^{c_t} - \sum_{i=1}^{L} b_i e^{c_{t-i}}\right]^{-\alpha} \\
=&e^{-\alpha c_t}\left[1 - \sum_{i=1}^{L} b_i e^{c_{t-i}-c_t}\right]^{-\alpha} \\
&\text{using (A.11), we get} \\
=&e^{-\alpha c_t}\left[1 - \sum_{i=1}^{L} b_i e^{-(\Delta c_t + ... + \Delta c_{t-i+1})}\right]^{-\alpha} \\
&\text{use the linear approximation } e^x \approx 1 + x \\
\approx&e^{-\alpha c_t}\left[1 - \sum_{i=1}^{L} b_i [1 - (\Delta c_t + ... + \Delta c_{t-i+1})]\right]^{-\alpha} \\
=&e^{-\alpha c_t}\left[1 - \sum_{i=1}^{L} b_i + \sum_{i=1}^{L} b_i (\Delta c_t + ... + \Delta c_{t-i+1})\right]^{-\alpha} \\
=&e^{-\alpha c_t}[1 - \sum_{i=1}^{L} b_i]^{-\alpha}\left[1 + \sum_{i=1}^{L} \frac{b_i}{1 - \sum_{q=1}^{L} b_q}(\Delta c_t + ... + \Delta c_{t-i+1})\right]^{-\alpha} \\
&\text{use the linear approximation } (1+x)^\alpha = 1 + \alpha x \\
=&e^{-\alpha c_t}[1 - \sum_{i=1}^{L} b_i]^{-\alpha}\left[1 + \sum_{i=1}^{L} \frac{-\alpha b_i}{1 - \sum_{q=1}^{L} b_q}(\Delta c_t + ... + \Delta c_{t-i+1})\right] \\
=&e^{-\alpha c_t}[1 - \sum_{i=1}^{L} b_i]^{-\alpha}\left[1 + \sum_{i=1}^{L} \frac{-\alpha b_i}{1 - \sum_{q=1}^{L} b_q} \sum_{m=0}^{i-1} \Delta c_{t-m}\right] \\
=&e^{-\alpha c_t} \cdot B \cdot A_t,
\end{aligned}$$

This completes the proof of Claim A.1.

If we apply this Claim to the log term in (A.10), we have

$$\begin{aligned}
&\log\left(\sum_{n=1}^{L+1} h_n \cdot (C_{t+n} - X_{t+n})^{-\alpha}\right) \\
=&\log\left(\sum_{n=1}^{L+1} h_n \cdot e^{-\alpha c_{t+n}} \cdot A_{t+n} \cdot B\right) \\
=&\log(\beta e^{-\alpha c_t} B) + \log\left(\sum_{n=1}^{L} \frac{h_n}{\beta} e^{-\alpha(c_{t+n}-c_t)} \cdot A_{t+n}\right) \\
&\text{by (A.11)} \\
=&\log(\beta e^{-\alpha c_t} B) + \log\left(\sum_{n=1}^{L} \frac{h_n}{\beta} e^{-\alpha \sum_{k=1}^{n} \Delta c_{t+k}} \cdot A_{t+n}\right) \\
&\text{use the linear approximation } e^x \approx 1 + x \\
\approx&\log(\beta e^{-\alpha c_t} B) + \log\left(\sum_{n=1}^{L} \frac{h_n}{\beta} (1 - \sum_{k=1}^{n} \alpha \Delta c_{t+k}) \cdot A_{t+n}\right)
\end{aligned}$$

Plugging this result to (A.10), we have

$$\begin{aligned}
\Delta \mathbb{E}_{t+1} m_{t+1} =&\Delta \mathbb{E}_{t+1} \log\left(\sum_{n=1}^{L+1} h_n \cdot (C_{t+n} - X_{t+n})^{-\alpha}\right) \\
=&\Delta \mathbb{E}_{t+1}\left[\log(\beta e^{-\alpha c_t} B) + \log\left(\sum_{n=1}^{L+1} \frac{h_n}{\beta} (1 - \sum_{k=1}^{n} \alpha \Delta c_{t+k}) \cdot A_{t+n}\right)\right] \\
=&\Delta \mathbb{E}_{t+1}\left[\log\left(\sum_{n=1}^{L+1} \frac{h_n}{\beta} (1 - \sum_{k=1}^{n} \alpha \Delta c_{t+k}) \cdot A_{t+n}\right)\right]
\end{aligned} \tag{A.12}$$

Now we rewrite part of (A.12). Using the expression of A_t in Claim A.1, we have

$$\begin{aligned}
&(1 - \sum_{k=1}^{n} \alpha \Delta c_{t+k}) \cdot A_{t+n} \\
=&(1 - \sum_{k=1}^{n} \alpha \Delta c_{t+k})\left[1 + \sum_{i=1}^{L} \frac{-\alpha b_i}{1 - \sum_{q=1}^{L} b_q} (\sum_{m=0}^{i-1} \Delta c_{t+n-m})\right] \\
&\text{neglecting the quadratic terms, we get} \\
\approx&1 - \sum_{k=1}^{n} \alpha \Delta c_{t+k} - \sum_{i=1}^{L} \frac{\alpha b_i}{1 - \sum_{q=1}^{L} b_q} (\sum_{m=0}^{i-1} \Delta c_{t+n-m}) \\
=&1 - \sum_{k=1}^{n} \alpha \Delta c_{t+k} - \sum_{i=1}^{L} \sum_{m=0}^{i-1} \frac{\alpha b_i}{1 - \sum_{q=1}^{L} b_q} (\Delta c_{t+n-m})
\end{aligned} \tag{A.13}$$

Plugging (A.13) to the log-term in (A.12), we have

$$
\begin{aligned}
&\log\left(\sum_{n=1}^{L+1}\frac{h_n}{\beta}(1-\sum_{k=1}^{n}\alpha\Delta c_{t+k})\cdot A_{t+n}\right)\\
=&\log\left[\sum_{n=1}^{L+1}\frac{h_n}{\beta}\left(1-\sum_{k=1}^{n}\alpha\Delta c_{t+k}-\sum_{i=1}^{L}\sum_{m=0}^{i-1}\frac{\alpha b_i}{1-\sum_{q=1}^{L}b_q}(\Delta c_{t+n-m})\right)\right]\\
=&\log\left[\sum_{n=1}^{L+1}\frac{h_n}{\beta}-\sum_{n=1}^{L+1}\sum_{k=1}^{n}\frac{\alpha h_n}{\beta}\Delta c_{t+k}-\sum_{n=1}^{L+1}\sum_{i=1}^{L}\sum_{m=0}^{i-1}\frac{h_n}{\beta}\frac{\alpha b_i}{1-\sum_{q=1}^{L}b_q}(\Delta c_{t+n-m})\right]\\
=&\log\left[\sum_{n=1}^{L+1}\frac{h_n}{\beta}-\sum_{k=1}^{L+1}B_k\Delta c_{t+k}-\left(\sum_{k=2-L}^{0}H_k\Delta c_{t+k}+\sum_{k=1}^{L+1}D_k\Delta c_{t+k}\right)\right]\ ...(*)\\
=&\log\left(\sum_{n=1}^{L+1}\frac{h_n}{\beta}-\sum_{k=1}^{L+1}(B_k+D_k)\Delta c_{t+k}-\sum_{k=2-L}^{0}H_k\Delta c_{t+k}\right)\\
=&\log\left(1+\sum_{n=2}^{L+1}\frac{h_n}{\beta}-\sum_{k=1}^{L+1}(B_k+D_k)\Delta c_{t+k}-\sum_{k=2-L}^{0}H_k\Delta c_{t+k}\right)\\
&\text{Since }\ \log(1+x)\approx x\\
\approx&\sum_{n=2}^{L+1}\frac{h_n}{\beta}-\sum_{k=1}^{L+1}(B_k+D_k)\Delta c_{t+k}-\sum_{k=2-L}^{0}H_k\Delta c_{t+k}
\end{aligned}
\tag{A.14}
$$

where the derivation of $(*)$ will be explained at the end and

$$
\begin{aligned}
B_k&=\sum_{n=k}^{L}\frac{\alpha h_n}{\beta},k=1,2,...,L+1;\\
D_k&=\begin{cases}\sum_{m=1}^{L}\sum_{i=m}^{L}\frac{\alpha b_i h_m}{\beta(1-\sum_{q=1}^{L}b_q)}, & \text{if } k=1;\\ \sum_{m=0}^{L+1-k}\sum_{i=1+m}^{L}\frac{\alpha b_i h_{m+k}}{\beta(1-\sum_{q=1}^{L}b_q)}, & \text{if } k=2,3,...,L+1.\end{cases}
\end{aligned}
$$

The form of H_k is not given as it will not be necessary.

Substituting (A.14) back to (A.12), we get

$$
\begin{aligned}
\Delta\mathbb{E}_{t+1}m_{t+1}=&\Delta\mathbb{E}_{t+1}\left[\log\left(\sum_{n=1}^{L+1}\frac{h_n}{\beta}(1-\sum_{k=1}^{n}\alpha\Delta c_{t+k})\cdot A_{t+n}\right)\right]\\
\approx&\Delta\mathbb{E}_{t+1}\left[\sum_{n=2}^{L+1}\frac{h_n}{\beta}-\sum_{k=1}^{L+1}(B_k+D_k)\Delta c_{t+k}-\sum_{k=2-L}^{0}H_k\Delta c_{t+k}\right]\\
=&-\Delta\mathbb{E}_{t+1}\left(\sum_{k=1}^{L+1}(B_k+D_k)\Delta c_{t+k}\right)
\end{aligned}
$$

This is of the same format as (A.9). So we have

$$z_k = B_{k+1} + D_{k+1}, \text{ for } k = 0, 1, 2, \cdots, L.$$

DERIVATION OF THE EQUATION OF (*) Equation $(*)$ says

$$\begin{aligned}
&\log\left[\sum_{n=1}^{L+1}\frac{h_n}{\beta} - \sum_{n=1}^{L+1}\sum_{k=1}^{n}\frac{\alpha h_n}{\beta}\Delta c_{t+k} - \sum_{n=1}^{L+1}\sum_{i=1}^{L}\sum_{m=0}^{i-1}\frac{h_n}{\beta}\frac{\alpha b_i}{1-\sum_{q=1}^{L} b_q}(\Delta c_{t+n-m})\right] \\
=&\log\left[\sum_{n=1}^{L+1}\frac{h_n}{\beta} - \sum_{k=1}^{L+1} B_k\Delta c_{t+k} - \left(\sum_{k=2-L}^{0} H_k\Delta c_{t+k} + \sum_{k=1}^{L+1} D_k\Delta c_{t+k}\right)\right] ...(*)
\end{aligned}$$

We are going to prove it by looking at the last two sums on the left respectively.

First, we show that

$$\sum_{n=1}^{L+1}\sum_{k=1}^{n}\frac{\alpha h_n}{\beta}\Delta c_{t+k} = \sum_{k=1}^{L+1} B_k\Delta c_{t+k}, \tag{A.15}$$

where

$$B_k = \sum_{n=k}^{L}\frac{\alpha h_n}{\beta}, k = 1, 2, ..., L+1.$$

The following table shows the coefficients for each Δc_{t+k} in the left-hand side of (A.15).

Value of k	Possible values of n	Corresponding subsum
$k=1$	$n \in [1, L+1]$	$\sum_{n=1}^{L+1}\frac{\alpha h_n}{\beta}\Delta c_{t+1}$
$k=2$	$n \in [2, L+1]$	$\sum_{n=2}^{L+1}\frac{\alpha h_n}{\beta}\Delta c_{t+2}$
$k=3$	$n \in [3, L+1]$	$\sum_{n=3}^{L+1}\frac{\alpha h_n}{\beta}\Delta c_{t+3}$
$\vdots$	$\vdots$	$\vdots$
$k=L+1$	$n \in [L+1, L+1]$	$\sum_{n=L+1}^{L+1}\frac{\alpha h_n}{\beta}\Delta c_{t+L+1}$

Through observation, we can see that the coefficient of Δc_{t+k} is $\sum_{n=k}^{L+1}\frac{\alpha h_n}{\beta}$. Hence, (A.15) is true.

Next, we show that

$$\sum_{n=1}^{L+1}\sum_{i=1}^{L}\sum_{m=0}^{i-1}\frac{h_n}{\beta}\frac{\alpha b_i}{1-\sum_{q=1}^{L} b_q}(\Delta c_{t+n-m}) = \sum_{k=2-L}^{0} H_k\Delta c_{t+k} + \sum_{k=1}^{L+1} D_k\Delta c_{t+k},$$

where D_k is as defined in (A.14). To prove this equality, again, we look at the coefficients for each Δc_{t+k}.

COEFFICIENTS OF $\Delta c_{t+k},\ k = 2, \cdots, L+1$. If $n - m = k$, the corresponding possible values for n, m, and i are

Value of n	Value of m	Possible values of i	Corresponding subsum
$n = L+1$	$m = L+1-k$	$i \in [L-k+2, L]$	$\sum_{i=(L-k)+2}^{L} \frac{\alpha b_i h_{L+1}}{\beta(1-\sum_{q=1}^{L} b_q)}(\Delta c_{t+k})$
$n = L$	$m = L-k$	$i \in [(L-k)+1, L]$	$\sum_{i=(L-k)+1}^{L} \frac{\alpha b_i h_{L}}{\beta(1-\sum_{q=1}^{L} b_q)}(\Delta c_{t+k})$
$n = L-1$	$m = L-k-1$	$i \in [(L-k), L]$	$\sum_{i=(L-k)}^{L} \frac{\alpha b_i h_{L-1}}{\beta(1-\sum_{q=1}^{L} b_q)}(\Delta c_{t+k})$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$n = k$	$m = 0$	$i \in [1, L]$	$\sum_{i=1}^{L} \frac{\alpha b_i h_{k}}{\beta(1-\sum_{q=1}^{L} b_q)}(\Delta c_{t+k})$

The sum of the above corresponding subsum is

$$\sum_{m=0}^{L-k+1} \sum_{i=1+m}^{L} \frac{\alpha b_i h_{m+k}}{\beta(1-\sum_{q=1}^{L} b_q)}(\Delta c_{t+k}).$$

So for $k = 2, ..., L+1$, we set

$$D_k = \sum_{m=0}^{L-k+1} \sum_{i=1+m}^{L} \frac{\alpha b_i h_{m+k}}{\beta(1-\sum_{q=1}^{L} b_q)}.$$

COEFFICIENTS OF Δc_{t+1}. If $n - m = k = 1$, the corresponding possible values for n, m, and i are

Value of n	value of m	Associated values of i	corresponding subsum
$n = L$	$m = L-1$	$i \in [L, L]$	$\sum_{i=L}^{L} \frac{\alpha b_i h_{L}}{\beta(1-\sum_{q=1}^{L} b_q)}\Delta c_{t+1}$
$n = L-1$	$m = L-2$	$i \in [L-1, L]$	$\sum_{i=L-1}^{L} \frac{\alpha b_i h_{L-1}}{\beta(1-\sum_{q=1}^{L} b_q)}\Delta c_{t+1}$
...	...	...	...
$n = 1$	$m = 0$	$i \in [1, L]$	$\sum_{i=1}^{L} \frac{\alpha b_i h_{1}}{\beta(1-\sum_{q=1}^{L} b_q)}\Delta c_{t+1}$

Taking the sum of the above corresponding subsums, we have

$$\sum_{m=1}^{L} \sum_{i=m}^{L} \frac{\alpha b_i h_{m}}{\beta(1-\sum_{q=1}^{L} b_q)}\Delta c_{t+1}.$$

So,

$$D_1 = \sum_{m=1}^{L} \sum_{i=m}^{L} \frac{\alpha b_i h_{m}}{\beta(1-\sum_{q=1}^{L} b_q)}.$$

This finishes the proof.

Appendix B. VAR Estimates

The parameter estimates for the vector autoregression are given by

$$\phi_1 = \begin{pmatrix} 0.30343 & 0.89964 & -0.47492 \\ 0.01221 & 1.30330 & -0.15682 \\ 0.08092 & 0.53625 & 0.52959 \end{pmatrix}$$

$$\phi_2 = \begin{pmatrix} 0.30917 & -0.56619 & 0.18495 \\ -0.00683 & -0.39164 & 0.06958 \\ 0.00687 & -0.29411 & 0.05774 \end{pmatrix}$$

$$\phi_3 = \begin{pmatrix} 0.21782 & -0.80314 & 0.12451 \\ -0.02208 & 0.20593 & 0.04314 \\ -0.02098 & 0.10640 & 0.35794 \end{pmatrix}$$

$$\phi_4 = \begin{pmatrix} 0.09434 & 0.43410 & 0.04153 \\ -0.00148 & -0.16712 & -0.02475 \\ -0.08231 & -0.37490 & -0.01229 \end{pmatrix}$$

The corresponding standard errors are,

$$\mathrm{SE}(\Phi_1) = \begin{pmatrix} 0.06446 & 0.24027 & 0.08674 \\ 0.01939 & 0.07228 & 0.02609 \\ 0.05224 & 0.19471 & 0.07029 \end{pmatrix}$$

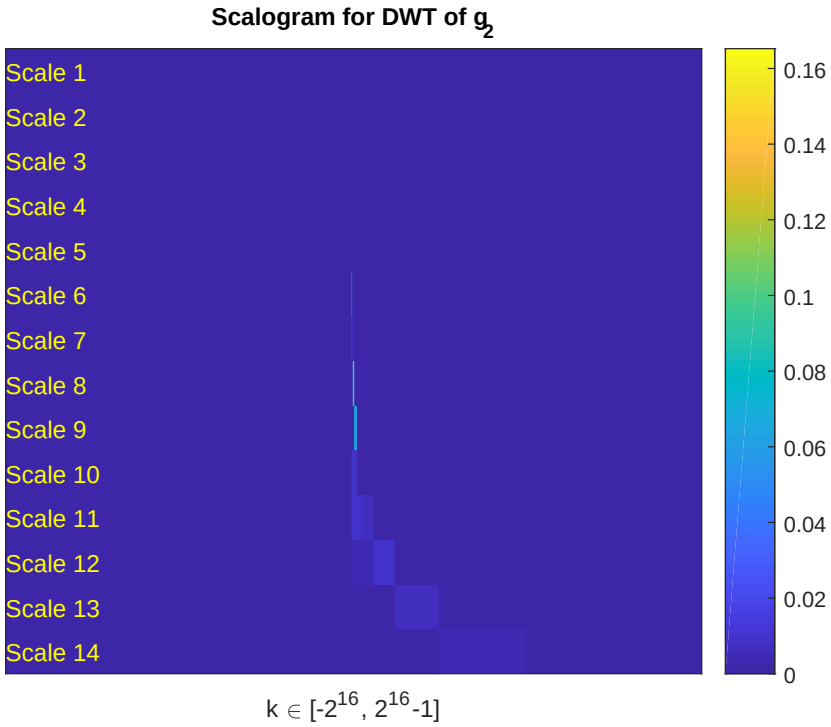
$$\mathrm{SE}(\Phi_2) = \begin{pmatrix} 0.06646 & 0.37375 & 0.11217 \\ 0.01999 & 0.11243 & 0.03374 \\ 0.05386 & 0.30288 & 0.0909 \end{pmatrix}$$

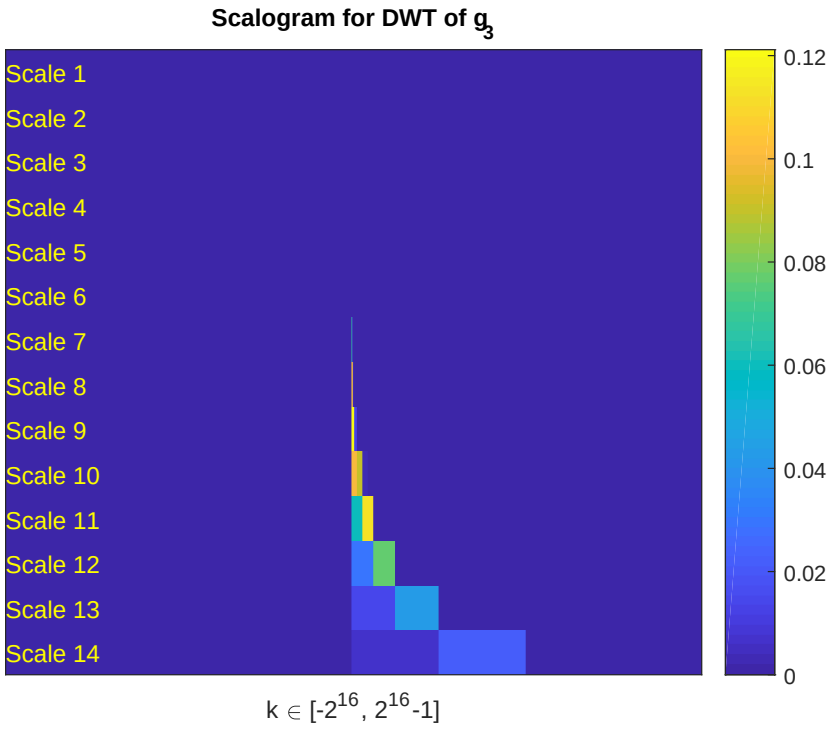
$$\mathrm{SE}(\Phi_3) = \begin{pmatrix} 0.06654 & 0.37548 & 0.11325 \\ 0.02001 & 0.11295 & 0.03407 \\ 0.05392 & 0.30428 & 0.09178 \end{pmatrix}$$

$$\mathrm{SE}(\Phi_4) = \begin{pmatrix} 0.06215 & 0.24218 & 0.09811 \\ 0.0187 & 0.07285 & 0.02951 \\ 0.05037 & 0.19626 & 0.0795 \end{pmatrix}$$

Appendix C. Scalograms for the DWT of $\{g_{2,k}\}$ and $\{g_{3,k}\}$

The scalograms for the DWT of $\{g_{2,k}\}$ and $\{g_{3,k}\}$ are as follows.





Parameter values for the empirical test	
z_k^{EZ}	z_k^{Habit}
$\alpha = 0.5$	$\alpha = 5$
$\alpha = 0.5$	$\rho = 0.5$
$\alpha = 0.5$	$\theta = 0.975$

Table D.1

Appendix D. Details of the Empirical Test

The parameters' values we choose for computing $\{z_k\}_{k=1}^{2^9}$ are listed in Table D.1. The monthly returns (we will convert them to quarterly returns) of the Fama-French 25 portfolio sorted on size and BM, and the 49 industrial portfolios are downloaded from Kenneth R. French's website, with the data period from 1952.01 to 2015.09.

1. We obtain the values of $\Phi^{(1)}$ using the estimates obtained from the lag-4 VAR in (13).
2. Calculate $D_{11} = \frac{\partial G_1}{\partial \Phi^{(1)}}$, where

$$G_1 = \sum_{t=1}^{N} \vec{g}_1(t)/N, \vec{g}_1(t) = (\bar{\mathbf{x}}_{t+1} - \mathbf{\Phi^{(1)}}\mathbf{y}_t) \otimes \mathbf{y}_t, \text{ and } N = 2^9.$$

3. Use the estimated $\Phi^{(1)}$ to estimate $\mathbf{q}$, by solving

$$\text{argmin}_{\mathbf{q}} G_2^T \cdot G_2,$$

where $G_2 = \sum_{t=1}^{N} \vec{g}_2(t)/N$, T denotes matrix transpose, and

$$\vec{g}_2(t) = \exp(\mathbf{r}_{t+1}) - \exp(r_{t+1}^f) - \mathbf{r}_{t+1} \cdot (\bar{\mathbf{x}}_{t+1} - \mathbf{\Phi}^{(1)}\mathbf{y}_t)^T \mathbf{q}.$$

4. Calculate $D_{21} = \frac{\partial G_2}{\partial \Phi^{(1)}}$.
5. Calculate $D_{22} = \frac{\partial G_2}{\partial \mathbf{q}}$.
6. To calculate an efficient GMM estimator of $\mathbf{q}$, we need to estimate $V = \sum_{j=-\infty}^{\infty} \mathbb{E}\left(g(t) \cdot \vec{g}^T(t)\right)$, where $\vec{g}(t) = \begin{bmatrix} \vec{g}_1(t) \\ \vec{g}_2(t) \end{bmatrix}$. The formula we use is stated in section 11.7 Estimating the Spectral Density Matrix of Cochrane's asset-pricing book (we will choose $K = 20$)

$$\hat{V} = \frac{1}{K(N-K)} \sum_{t=k+1}^{N} (\vec{v}(t) - \bar{\mathbf{v}})(\vec{v}(t) - \bar{\mathbf{v}})^T,$$

where

$$\vec{v}(t) = \sum_{j=1}^{K} \vec{g}_{t-j}, \text{ and } \bar{\mathbf{v}} = \sum_{t=K+1}^{N} \vec{v}(t)/(N-K).$$

7. We use the following formula in Hansen (2008) to calculate the selection matrix for the moment conditions $G_2 = 0$.

$$A_{22} = D_{22}^T \cdot W_2,$$

where

- $W_2 = \left\{ C \cdot \hat{V} \cdot C^T \right\}^{-1}$ is the weighting matrix for the moment conditions $G_2 = 0$,
- $C = \left[-D_{21} \cdot (A_{11} \cdot D_{11})^{-1} \cdot A_{11} \quad I \right]$, $A_{11} = D_{11}^T \cdot W_1$,
- and W_1 is the weighting matrix for the VAR estimates of $\Phi^{(1)}$ and hence is the identity matrix.

8. Use W_2 to obtain an efficient GMM estimate $\hat{\mathbf{q}}^{\text{eff}}$ by solving

$$\text{argmin}_{\mathbf{q}} G_2^T \cdot W_2 \cdot G_2.$$

9. Recalculate D_{21} using $\hat{\mathbf{q}}^{\text{eff}}$.

10. Recalculate $\hat{V}$.

11. Recalculate A_{22} using D_{21} and $\hat{V}$.

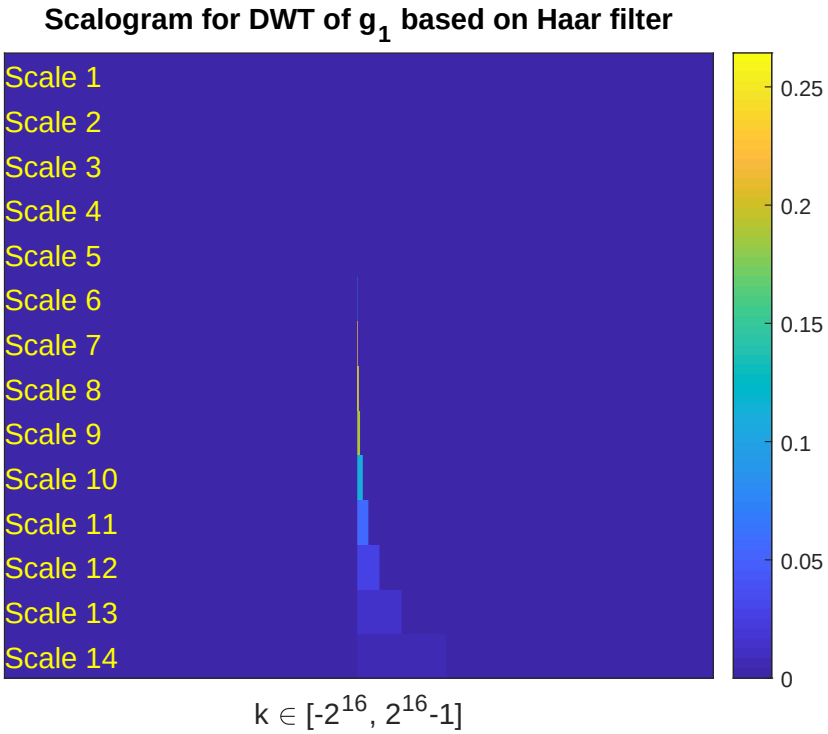
12. Calculate $\text{cov}(\hat{\mathbf{q}}^{\text{eff}} - \mathbf{q})$ where $\mathbf{q}$ denotes the true values, using the following formula derived from Hansen (2008).

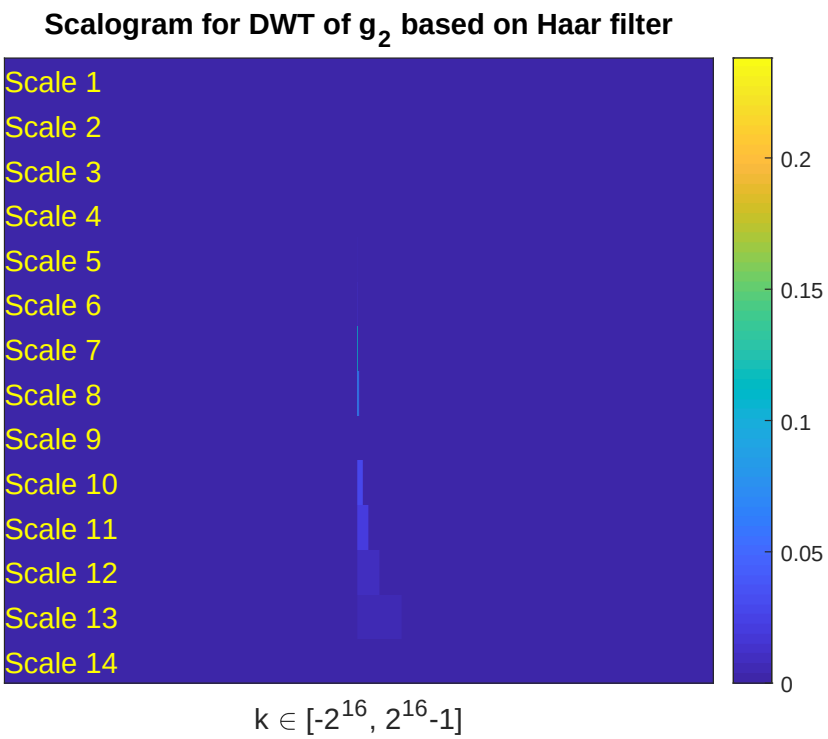
$$\text{cov}(\hat{\mathbf{q}}^{\text{eff}} - \mathbf{q}) = C_1^T \cdot \hat{V} \cdot C_1,$$

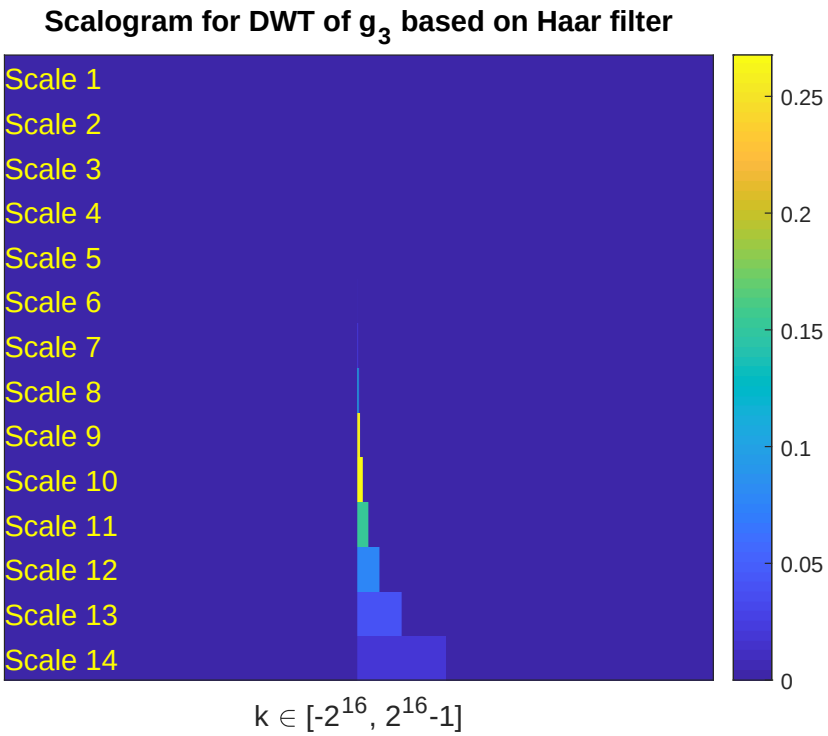
where

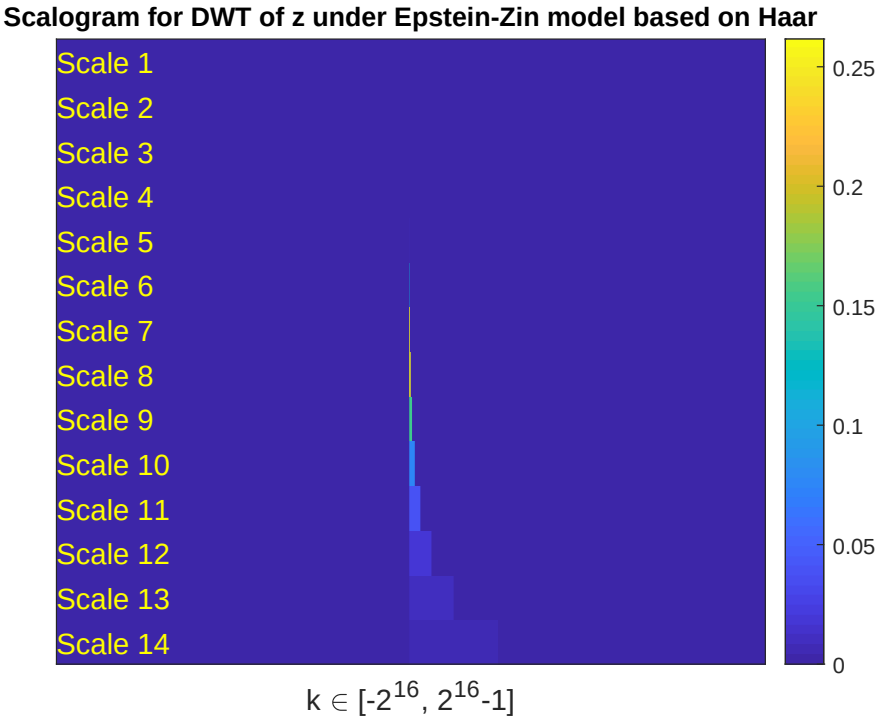
$$C_1 = (A_{22} D_{22})^{-1} \left[-D_{21} \cdot (A_{11} \cdot D_{11})^{-1} \cdot A_{11} \quad I \right]$$

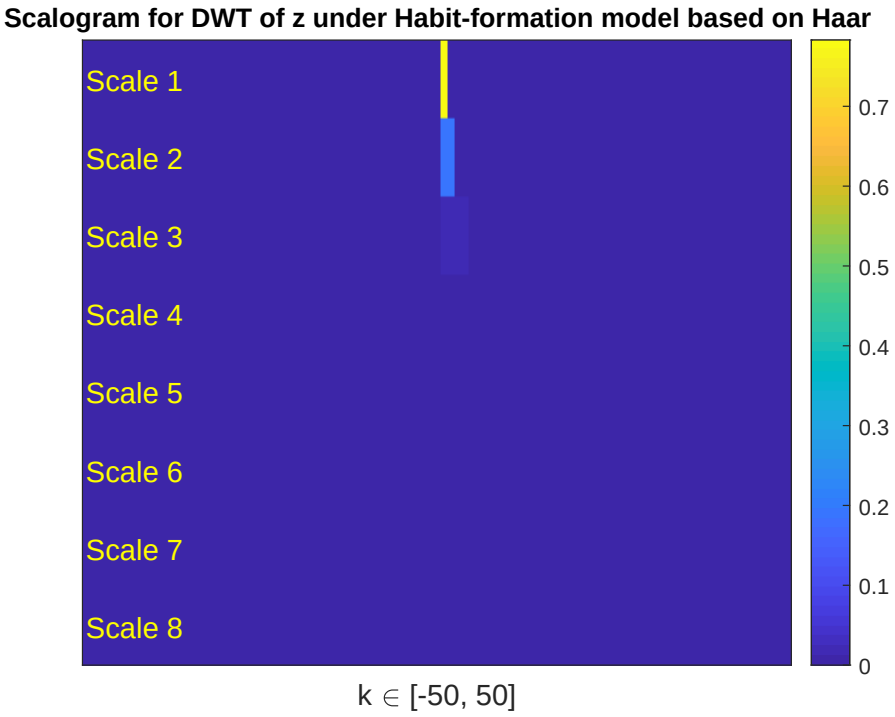
Appendix E. Scalograms for the DWT based on Haar filter


Scalogram for DWT of g1 based on Haar filter
Scale 1
Scale 2
Scale 3
Scale 4
Scale 5
Scale 6
Scale 7
Scale 8
Scale 9
Scale 10
Scale 11
Scale 12
Scale 13
Scale 14
0.25
0.2
0.15
0.1
0.05
0
k ∈ [-2^16, 2^16-1]


Scalogram for DWT of g2 based on Haar filter
Scale 1
Scale 2
Scale 3
Scale 4
Scale 5
Scale 6
Scale 7
Scale 8
Scale 9
Scale 10
Scale 11
Scale 12
Scale 13
Scale 14
0.2
0.15
0.1
0.05
0
k ∈ [-2^16, 2^16-1]


Scalogram for DWT of g_3 based on Haar filter
Scale 1
Scale 2
Scale 3
Scale 4
Scale 5
Scale 6
Scale 7
Scale 8
Scale 9
Scale 10
Scale 11
Scale 12
Scale 13
Scale 14
0.25
0.2
0.15
0.1
0.05
0
$k \in [-2^{16}, 2^{16}-1]$


Scalogram for DWT of z under Epstein-Zin model based on Haar
Scale 1
Scale 2
Scale 3
Scale 4
Scale 5
Scale 6
Scale 7
Scale 8
Scale 9
Scale 10
Scale 11
Scale 12
Scale 13
Scale 14
0.25
0.2
0.15
0.1
0.05
0
$k \in [-2^{16}, 2^{16}-1]$


Scalogram for DWT of z under Habit-formation model based on Haar
Scale 1
Scale 2
Scale 3
Scale 4
Scale 5
Scale 6
Scale 7
Scale 8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
k ∈ [-50, 50]