Directed attention and nonparametric learning

Ian Dew-Becker and Charles G. Nathanson*

March 2, 2018

Abstract

This paper examines the implications of learning for the effects of ambiguity aversion. The key result is that since agents naturally choose to learn about the sources of uncertainty that reduce utility the most, information acquisition reduces the most severe effects of ambiguity aversion. The specific setting we study is the canonical consumption-savings problem. Agents endogenously learn most about income dynamics at the very lowest frequencies. While ambiguity aversion typically implies in this setting excessive extrapolation of income shocks (which can lead, for example, to a high and volatile equity premium), that effect is eliminated here. Furthermore, deviations of consumption from the full-information benchmark are largest at high frequencies, so consumption responds strongly to predictable changes in income in the short-run but is closer to a random walk in the long-run.

A large recent literature studies model uncertainty, and ambiguity aversion in particular, as a major driver of macroeconomic dynamics and asset prices. Ambiguity aversion has been shown to be able to generate realistic business cycles (Ilut and Schneider (2014) and Bianchi, Ilut, and Schneider (2017)), to help rationalize variation in survey expectations (Bhandari, Borovicka, and Ho (2017)), to generate large and time-varying equity risk premia (Hansen, Sargent, and Tallarini (1999), Ju and Miao (2012), Hansen and Sargent (2015), and Bidder and Dew-Becker (2016)) and to help explain the VIX and the variance risk premium (Drechsler (2013) and Bidder and Smith (2015)).

The central idea behind the ambiguity aversion literature is that people are uncertain about the true model driving the economy and that they choose policies that are designed to be robust against unfavorable models. While the ambiguity literature has taken the model uncertainty to be exogenous, one would naturally expect that if people were highly averse to model uncertainty that they would try to learn and reduce that ambiguity. And that learning should be focused on precisely the parts of the underlying model where errors are most painful. So learning should be expected to reduce the most severe effects of ambiguity aversion and model uncertainty.

*Dew-Becker: Northwestern University and NBER; ian.dewbecker@gmail.com. Nathanson: Northwestern University; nathanson@kellogg.northwestern.edu. We appreciate helpful comments from Ben Hebert, Peter Klibanoff, Konstantin Milbradt, Mikkel Plagborg-Møller, and seminar participants.

In this paper, we study ambiguity and learning in a simple dynamic consumption/savings problem. The basic intuition above could be captured in many settings, but we choose the consumption/savings problem because it is a canonical dynamic optimization with well understood analytic solutions that shares the same basic structure as much richer models used in asset pricing and macroeconomics.[1] Future work could extend the ideas developed here into more general dynamic settings. Contemporaneous work by Epstein and Ji (2017) also examines learning under ambugity, but in a setting in which there is no choice about how to allocate attention and in which the source of ambiguity is not dynamic. They study behavior in a model with an ambiguous urn, which is extremely important in the decision theory literature, but it less directly connected to the standard framework in macroeconomics and asset pricing.

The agents in our model face an exogenous income process with uninsurable risk and unknown dynamics, and they are ambiguity averse over potential models for income. The key innovation compared to past work on ambiguity aversion is that agents acquire information that can reduce the degree of ambiguity. Our goal is to find the optimal information acquisition policy and understand how it changes the effects of ambiguity on an agent's behavior. Most notably, there is a simple benchmark case in which optimal learning completely eliminates the primary effects of ambiguity aversion. More generally, learning always acts to reduce the effects, with the degree depending on the details of the cost specification for information.

The optimization problem that agents face has three components: the consumption choice, nature's choice of a model (embodying ambiguity aversion), and learning. Conditional on a particular model of the world, agents have standard Bayesian expected utility.[2] The agents are unsure of the true model, though, which is where their ambiguity aversion appears: agents act as though nature chooses the process for income, among all sufficiently plausible processes, that will yield the lowest utility from consumption (this is the second phase of the optimization). This selection criterion ensures that consumption decisions are robust to uncertainty about the true model.

The third phase of the optimization represents our contribution. Agents allocate attention to different aspects of the income process, which allows them to endogenously limit the degree of ambiguity they face. When agents pay more attention to a particular aspect of income, such as its low-frequency behavior, they receive information about its true behavior along that dimension and the set of plausible models narrows. A contribution of the paper to the learning literature is in providing a general description of how a person might learn about different aspects of a dynamic process.[3]

[1] See Wang (2004, 2009) and Luo (2008) for analyses of consumption under model uncertainty and information processing constraints and Caballero (1990) for an analysis of the setup with a known model.

[2] During this phase of the optimization, no dynamic learning about the model occurs. For boundedly rational models of dynamic learning, see Abel, Eberly, and Panageas (2007, 2013), Wang (2009), Bansal and Shaliastovich (2010), Hansen and Sargent (2010), Ju and Miao (2012), and Collin-Dufresne, Johannes, and Lochstoer (2015).

[3] There is substantial past work on directed learning (e.g. Van Nieuwerburgh and Veldkamp (2006), Peng and Xiong (2006), Veldkamp (2006), and Barron and Ni (2008)), but we are not aware of work that examines the choice of what part of a dynamic process to learn about. See Sims (2003), Veldkamp (2011), and many citations therein for work on directed attention more generally.

Given that optimal consumption depends on permanent income, it is the low-frequency features of income that are generally most beneficial to learn about. But one must also ask how costly it is to learn about dynamics at different frequencies. Textbook results from the time series econometrics literature say that learning about all frequencies is equally hard (e.g. Brillinger (1981), Priestly (1981), Brockwell and Davis (1991), and Hamilton (1994)). But since intuition suggests that low frequencies might be more difficult to learn about, we also consider a general specification that allows for arbitrary costs across frequencies.

We solve three phases of the optimization analytically and are therefore able to sharply establish our main result: when information is equally costly across frequencies, the agent directs almost all attention to the behavior of income at the lowest frequencies (i.e. at long horizons), which have the largest impact on utility through their ambiguity aversion. The agent's learning in that case perfectly cancels out the most most harmful effects of ambiguity. More specifically, we obtain two key results:

1. In past work, ambiguity aversion causes agents to generally overextrapolate income shocks (Hansen and Sargent (2010, 2015) and Bidder and Dew-Becker (2016)),[4] but with endogenous learning, that result is completely eliminated: agents neither over- nor under-extrapolate shocks when forecasting long-run future income. The lack of bias results from the fact that agents acquire the most information at low frequencies.

2. The agents' focus on low frequencies yields high-frequency mistakes: at short horizons, consumption growth is positively correlated with the *predictable* component of income growth. This comovement violates the permanent income hypothesis (Friedman (1957), Hall (1978)) but matches the extensive empirical evidence on the excess sensitivity of consumption to income (Jappelli and Pistaferri (2010), Kaplan and Violante (2014)). Because agents fail to learn about the high-frequency characteristics of the income process, much of the predictable variation in income is surprising and therefore leads agents to adjust consumption.

What connects the two theoretical results is that high-frequency mistakes have minimal implications for lifetime utility, while low-frequency mistakes can have substantial effects. That idea has been suggested as an explanation for the excess sensitivity puzzle, and our model formalizes it.[5] People cannot achieve perfection, so they choose to make mistakes that are minimally costly. Two aspects of the results are surprising. First, while one might expect that learning would reduce the effects of ambiguity (though that point has not been made previously), the fact that it can perfectly cancel those effects in some cases – even though the learning is incomplete in the sense that not all uncertainty is resolved – has important implications for the interpretation of ambiguity models. Second, while it is understood that the utility cost of high-frequency mistakes is relatively

[4] A bias towards belief in overly persistent processes is present also in the boundedly rational frameworks of Fuster, Hebert, and Laibson (2011) and Bordalo, Gennaioli, and Shleifer (2016).

[5] See Cochrane (1989), Eichenbaum (2011), and Kueng (2016) for discussions of the small utility costs of excess sensitity to transitory income shocks.

small, this is the first paper to obtain such mistakes endogenously, and we show that they can be quantitatively realistic. The paper is thus important both to the literature on ambiguity aversion in dynamic models and also the literature on "mistakes" in household consumption.

The main findings hold in the textbook benchmark where information is equally costly at all frequencies. We also develop a formalization of the idea that low-frequency information should be more expensive to obtain than high frequency information. In that case, the agent continues to focus primarily on low frequencies, but with a less extreme tilt. Weighted by the precision of the signals, the median unit of attention is focused on cycles lasting 250 years in the benchmark case (consistent with results in Dew-Becker and Giglio (2014)) and 47 years in the frequency-dependent cost case. So while attention shifts to much shorter cycles, it is still focused on extremely long-lived shocks. In terms of observable behavior, the results in this case lie between the equal information benchmark and the case of ambiguity with no attention allocation. The bottom line of the analysis, then, is that the effects of ambiguity aversion depend critically on how easily agents can acquire information. In a reasonable benchmark, the main effects can be completely eliminated, but there are also specifications for information costs that generate intermediate outcomes.

In addition to the work on ambiguity aversion above, our work links to the literature on learning and attention allocation more generally (e.g. Sims (2003)). Most past work has focused on learning about hidden states. A potentially useful contribution of the paper is to propose a framework for analyzing information acquisition about specific aspects of a dynamic process and for motivating how the cost might vary across different features.[6]

1 Environment and information

1.1 Preferences

We study agents who solve a consumption-savings problem under ambiguity aversion over model uncertainty. They face a standard budget constraint.

Assumption 1 *Financial wealth, W_t, follows the process*

$$W_t = RW_{t-1} + Y_t - C_t \tag{1}$$

where C_t is consumption, Y_t is income, and R is a fixed gross interest rate.

We denote possible income processes $\hat{f}$. The agent's preferences are represented by the following optimization.

Assumption 2 *Agents choose signal precision τ and a consumption policy C^{policy} according to*

$$\max_{\tau} E\left[G\left(\max_{C^{policy}} \min_{\hat{f}\in F(x;\tau)} E\left[\sum_{t=0}^{\infty} -\alpha^{-1}\beta^t \exp\left(-\alpha C_t\right) \mid \hat{f}\right]\right)\right], \tag{2}$$

[6] See Gabaix (2016) for a recent alternative model of directed attention in a dynamic setting.

where E denotes the expectation operator and G is a strictly increasing function that will be defined in equation (26) below. C^{policy} is a consumption rule mapping current wealth, W_t, and the income history, $Y_0, ..., Y_t$, to consumption, C_t. $F(x;\tau)$ is the set of models the agent considers, and x and τ are defined below. The optimization is subject to the budget constraint (1) and a transversality condition.

The inner max min pair represents ambiguity averse preferences similar to those of Gilboa and Schmeidler (1989). The agent's aim is to maximize discounted utility over consumption, where α is the coefficient of absolute risk aversion and β the time discount factor.

The source of uncertainty that the inner expectation applies to is the future realizations of income. Conditional on a (functional) parameter $\hat{f}$, agents calculate expectations over income realizations, and hence future consumption, using Bayes' rule.

The parameter $\hat{f}$ is unknown. If people were Bayesian expected utility maximizers, they would choose the consumption policy to maximize expected utility under a probability measure for $\hat{f}$. That is, we would have

$$\max_{C^{policy}} \int E\left[\sum_{t=0}^{\infty} -\alpha^{-1}\beta^t \exp\left(-\alpha C_t\right) \mid \hat{f}\right] d\Phi\left(\hat{f}\right) \tag{3}$$

where $d\Phi\left(\hat{f}\right)$ represents a probability density over models. $\hat{f}$ is a potentially infinite dimensional object. We therefore take the position that it is not reasonable to assume that people are able to fully articulate a probability distribution over all possible values of $\hat{f}$ (the work of Hansen and Sargent (2007) on robust control is motivated similarly).

Instead, we model agents as ambiguity averse. They believe that $\hat{f}$ may fall into a set $F(x;\tau)$, where x is a set of signals about the true model that they receive, which have precision τ. The consumption policy is chosen to maximize expected utility with the understanding that nature will then select the least favorable value of $\hat{f} \in F(x;\tau)$.

The outer expectation in (2) is taken over possible realizations of the signals x. The agent chooses the signal precisions, τ, to maximize the expected outcome of the ambiguity-averse consumption/savings problem. An agent that receives high-quality signals about income dynamics will have a smaller set $F(x;\tau)$, thus reducing the effects of ambiguity. The choice of τ and its implications are the centerpiece of our analysis.

The function G is applied to expected utility conditional on the signals for reasons that we discuss below.

Note that if there were no model uncertainty, so that the true model f is known, then the agent is solving a standard consumption-savings problem under CARA preferences: $\max E\left[\sum_{t=0}^{\infty} -\alpha^{-1}\beta^t \exp\left(-\alpha C_t\right)\right]$. Our analysis therefore ignores wealth effects, but it is also more realistic than the assumption of quadratic utility over consumption used in Hansen, Sargent, and Tallarini (1999), among others in that the preferences are monotone in consumption. Appendix F shows that it is possible to obtain similar results to what is described in the main text in a case in which agents have Epstein–

Zin (1991) preferences with a unit elasticity of substitution when there is uncertainty about the dynamics of the return on wealth.

Finally, it is also important to note that this is in certain regards a date-0 problem. Agents receive signals about the income process once. They then choose an optimal consumption policy that is meant to be robust to model uncertainty. We do not model how people update information about the income process ($\hat{f}$) over time. The model will imply that agents choose to learn primarily about the low-frequency features of income, and those are also the features that one would expect to be updated most slowly, making dynamic learning a secondary concern here. Furthermore, note that consumption is chosen fully dynamically in that the policy conditions on the past histories of wealth and income.

1.2 Income

Assumption 3 *Consumers face an exogenous and untradeable stochastic income stream, Y_t, that follows the process*

$$Y_t = a(L) Y_{t-1} + b_0 \varepsilon_t \tag{4}$$

$$\varepsilon_t \sim i.i.d.\ N(0,1) \tag{5}$$

where $a(L)$ is a power series in the lag operator, L. We assume $a(L)$ is such that Y is well behaved (in particular, has a spectrum that is positive and bounded).[7]

While linearity and Gaussianity are certainly restrictive assumptions, they are in line with the past work we build on. Much of our analysis will apply to the Wold representation,

$$Y_t = b(L)\varepsilon_t, \tag{6}$$

$$\text{where } b(L) \equiv \frac{b_0}{1 - La(L)}. \tag{7}$$

The coefficients in the power series $b(L)$ are denoted b_j (i.e. $b(L) = \sum_{j=0}^{\infty} b_j L^j$). Throughout the paper, we refer to models in the time domain in terms of $b(L)$. Since the distribution of ε_t is fixed, $b(L)$ completely characterizes the statistical distribution of income. To be clear, though, the agent forecasts the future using only the past history of income. The ε_t are not directly observable.

Agents do not know the true income process. Alternative possible income processes are denoted $\hat{b}$.[8] Our focus is on uncertainty about the dynamics of income, rather than about the distribution

[7] The assumption that Y_t is a linear Gaussian process is not necessary for most of the results. The critical assumptions about the true income process are that it is second-order stationary and that it has a spectral density that is finite and bounded away from zero. The distribution of the innovations is largely irrelevant (though it is important that it is fixed over time).

[8] Agents forecast with $\hat{b}$ the same way they would with b,

$$E_t\left[Y_{t+n} \mid \hat{b}\right] = \sum_{k=0}^{\infty} \hat{b}_{n+k} \hat{b}(L)^{-1} Y_{t-k} \tag{8}$$

of its innovations. The latter question is also interesting, but our goal is to understand how consumption responds to changes in income, and how well people understand the difference between permanent and transitory dynamics. An interpretation of our analysis is that it derives optimal attention to different aspects of income dynamics conditional on a choice having been made about how much attention to pay overall to dynamics versus the distribution of innovations.

1.3 Signals about the spectrum of income

The key type of uncertainty that agents face is over the model that drives income. The definitions above are in the time domain, but our analysis examines a rotation, using the Fourier transform, into the frequency domain. The Fourier transform is used in time series analysis because it orthogonalizes stationary processes. One way for an agent to model income is to estimate its autocovariances. But estimates of autocovariances are in general correlated across lags, and one must impose complicated restrictions to guarantee positive definiteness, both of which substantially complicate the analysis.

Those issues do not arise in the frequency domain. In what we describe here, agents receive signals about features of the income process that are mutually orthogonal and always generate a positive definite covariance matrix for income. So a primary reason that we model agents as learning in the frequency domain is that such learning represents acquiring information about fundamentally independent aspects of the income process.

The next subsection defines the frequency transform. We then describe precisely how agents acquire information about income dynamics. Last, we discuss the relationship between the information scheme used here and those studied in models of rational inattention.

1.3.1 The Fourier transform and its distribution

The spectral density of the income process is defined as the Fourier transform of the autocovariances and thus also fully represents income dynamics:

$$\exp f(\omega) \equiv \sum_{j=-\infty}^{\infty} \cos(\omega j) \operatorname{cov}(Y_t, Y_{t-j}). \tag{9}$$

The notation f is used for the log spectrum because that is what maps most directly into utility. As f is periodic, we may restrict attention to the domain $\omega \in [0, \pi]$. There is a one-to-one mapping between the spectral density and the autocovariances since they are a Fourier transform pair. The spectrum is a variance decomposition for income in terms of fluctuations at different frequencies:

$$var\left(Y_t\right) = \frac{1}{\pi} \int_0^{\pi} \exp\left(f\left(\omega\right)\right) d\omega. \tag{10}$$

$\exp(f(\omega))$ measures the contribution of fluctuations in income at frequency ω to the total variance of income. The relative magnitude of f across frequencies determines the extent to which variation in income is driven by low- versus high-frequency fluctuations. An AR(1) process with an autocorrelation near 1 has a spectrum whose mass is isolated at low frequencies, whereas a process that features reversals, such as $Y_t = \varepsilon_t - (1/2)\varepsilon_{t-1}$, has a spectrum with mass concentrated at high frequencies (those near π).

As with b and $\hat{b}$, f is the true log spectral density of income, while alternative hypothetical spectra are denoted $\hat{f}$. The agent can construct forecasts of future income based on $\hat{f}$ since there is a unique $\hat{b}$ associated with each $\hat{f}$ (Priestley (1981) section 10.1).

Estimates of the spectral density at a given frequency may be obtained by calculating the Fourier transform of the data itself at that frequency. Specifically, given a sample $\{Y_t\}_{t=1}^{T}$, an estimate of the spectrum at frequency ω, $\tilde{f}(\omega)$, is

$$\tilde{f}(\omega) \equiv \log \left| \sum_{t=1}^{T} Y_t \exp(i\omega t) \right|^2 . \tag{11}$$

Estimating the specturm at a single frequency does not require taking the entire Fourier transform. Instead, it is a single inner product of the time series with a complex exponential. This is the basis for standard nonparametric estimation of the spectrum (e.g., Brillinger (1981), chapter 5), and it is consistent with our desire to give agents unstructured uncertainty about dynamics (as opposed to having them estimate a known $ARMA(p,q)$ specification, for example).

Lemma 1 *As $T \to \infty$,*

$$E\left[\tilde{f}(\omega) - f(\omega)\right] \rightarrow -\varrho \tag{12}$$

$$cov\left(\tilde{f}(\omega_1) - f(\omega_1), \tilde{f}(\omega_2) - f(\omega_2)\right) \rightarrow \frac{\pi^2}{6} 1\{\omega_1 = \omega_2\} \tag{13}$$

for $\omega \in (0,\pi)$, where ϱ is Euler's constant and $1\{\cdot\}$ is the indicator function. These results hold exactly for all T when $f(\omega)$ is constant.

Proof. This result follows directly from Brillinger (1981) theorem 5.2.6 combined with the continuous mapping theorem. ■

Lemma 1 shows why we study income dynamics in the frequency domain: the sample spectrum yields an estimate of the true spectrum with errors that are asymptotically uncorrelated across frequencies. Estimation in the time domain does not have that orthogonality property. If one fits an ARMA model, for example, the coefficients are in general correlated. In what follows, we assume people use the asymptotic approximation and treat the spectral estimates as truly uncorrelated across frequencies.

An important feature of the result in lemma 1 is that it implies that all frequencies are equally difficult to learn about. That result in fact holds exactly under the agent's prior mean that we specify below that the spectrum is flat.[9] While surprising at first, that result is standard in the time series literature (see Brillinger (1981), Priestley (1981) and Hamilton (1994) for textbook treatments; see Shao and Wu (2007) and Hashimzade and Vogelsang (2008) for similar results under weaker conditions).[10] Moreover, it is not specific to the sample spectrum analyzed here: Berk (1974) shows that when the log spectrum is estimated using autoregressive models, estimation error is also uncorrelated and homoskedastic across frequencies (even though, again, the AR coefficients themselves are in general correlated).

Given that the time series literature uniformly argues that the spectral density is equally difficult to estimate at all frequencies, we take that position as our benchmark. However, we share the common intuition is that low-frequency dynamics seem as though they should be more difficult to learn about. We therefore specify in the next section a model of information acquisition that can accomodate both the benchmark view that all frequencies are equally difficult to learn about and an alternative that low frequencies might be more costly to learn about.

1.3.2 Information acquisition method

Information acquisition happens prior to the consumption policy being chosen; this step can be thought of as happening behind a veil of ignorance, before realizations of the agent's own income history have occurred. We assume that people gather information by estimating the log spectrum from observed income histories driven by the same model that will drive their own income. For example, a person who knows they will become an economist might investigate the income histories that older economists have received. These observations and calculations are costly, though, so we assume that people are limited in the number of measurements of spectra that they can obtain.

More specifically, we assume that there is a large dataset available that reports the income histories of many people i, $\{Y_{i,t}\}_{t=1}^{T}$, all of whom have the same parameters determining their income processes (i.e. the same f and hence b), but different realizations (different ε's). That dataset can be thought of as representing the information that people can get from talking to family members, teachers, or other mentors who are old enough to have long income histories.

To obtain information about the log spectrum of income at some frequency ω, the agent calculates the sample spectrum $\tilde{f}(\omega)$ for $\tau(\omega)$ of the income histories from the dataset. Combining the central limit theorem with lemma 1 above, the mean of $\tilde{f}(\omega) + \varrho$ across many income histories is asymptotically normally distributed with mean $f(\omega)$ and variance $\frac{\pi^2}{6}\tau(\omega)^{-1}$ (we ignore the $\frac{\pi^2}{6}$ scaling in what follows).

[9] When the spectrum is not flat, there is essentially bleeding across frequencies, which makes the estimates no longer exactly uncorrelated in finite samples except in the white-noise case. That fact can also potentially cause the variance of the errors to vary across frequencies, but they need not be higher at low frequencies. It is possible, but tedious to obtain an exact finite-sample distribution for $\tilde{f}$.

[10] Technically, estimates of the spectrum at the frequencies 0 and π – i.e. both the lowest and highest frequencies – have variances that are twice as high as points on the interior of the interval. When we allow for costs that vary across frequencies in section 5, we consider a much more exptreme case where the variance are infinitely high at frequency 0.

We assume that agents face a cost for each inner product that they calculate, representing a friction in information acquisition. Each observation of a sample spectrum at a single frequency requires calculating the inner product of an income history with $\{\exp(i\omega t)\}_{t=1}^{T}$. Agents face a cost of calculating each inner product. Allowing the cost to vary across frequencies according to some function $\gamma(\omega)$ makes some frequencies more expensive to learn about than others. In the main results, $\gamma(\omega) = 1$, and we examine the more general case in section 5.

For technical reasons (to avoid infinite information flows, for example), we assume that the agent gains information on the spectrum on the uniform discretization of $[0, \pi]$ given by $\omega_j = \pi j/n$ for $j \in \{1, ..., n\}$ and we take n as large. We scale the variances by $d\omega \equiv \pi/n$ so that they can be interpreted as the information density at each point. This leads to the following assumption on information:

Assumption 4 *The agent receives signals $\{x(\omega_j)\}_{j=1,...,n}$ that are distributed as*

$$x(\omega_j) \sim N\left(f(\omega_j), \tau(\omega_j)^{-1}/d\omega\right) \tag{14}$$

where the errors are uncorrelated across frequencies. The total cost of information is proportional to

$$\sum_{j=1}^{n} \gamma(\omega_j)\,\tau(\omega_j)\,d\omega. \tag{15}$$

If τ differs across frequencies, that means that the agent calculated the sample spectrum for more income histories at some frequencies than others. That is, they have many income histories to examine, but for frequencies that they learn less about, they only estimate the spectrum using a small number of them, and the effort saved is allocated elsewhere.

A natural benchmark in the absence of optimization would be for an agent to allocate equal information costs to all frequencies, so that $\tau(\omega_j) \propto \gamma(\omega_j)^{-1}$. When $\gamma(\omega_j) = 1$, this corresponds to obtaining signals with equal variances at all frequencies, which is the usual time series benchmark. We therefore refer to that as the statistical benchmark allocation.

1.3.3 Relationship with rational inattention

Rational inattention provides an alternative and equally important interpretation of the information structure. It is possible that complete information about the spectrum of income is available, but agents have trouble processing it. Then the noise in the signals represents cognitive errors that people make in interpreting the available information. The frequencies at which τ is larger are the ones the agent pays the most attention to. Such a specification also makes the benchmark with $\gamma = 1$ natural since it is less clear why a mental information processing constraint would bind especially tightly on any particular frequency.

In terms of the literature, the signal structure we analyze is highly similar to that in Kacperczyk, Van Nieuwerburgh, and Veldkamp (2016) in that agents receive signals with normally distributed error and they are constrained by the total precision of the signals. This constraint is most natural when each independent observation of the spectrum is equally costly to obtain. Sims (2003) proposes an alternative constraint based on information flow or entropy. In our setting, the total entropy of the signals is $\sum_{j=1}^{n} \log(\tau(\omega_j)\, d\omega)$, so high-precision signals are relatively less costly under an entropy constraint. Section 5 provides results for that alternative cost function.

That said, our setting is more restricted than the fully general rational inattention specification: the information the agents acquire is always independent across frequencies (which is motivated by the fact that statistical estimates of the spectrum are independent across frequencies) and the errors are Gaussian (motivated by the central limit theorem). In the most general form of the models that Sims (2003) studies, those restrictions need not hold. However, they are commonly imposed elsewhere, as in Kacperczyk, Van Nieuwerburgh, and Veldkamp (2016).

1.4 Priors and model plausibility

Agents measure the plausibility of models, and define the set they worry about, $F(x;\tau)$, based on their signals and a prior. Given that the model space is infinite-dimensional, it is difficult to imagine that a person would have a fully defined prior, though. People likely cannot place a formal probability on every possible model, or even necessarily express a view about the relative likelihood of all possible pairs of models. That fact motivates our use of ambiguity aversion, and it leads us to specify prior beliefs as loosely as possible.

Agents believe that the log spectrum is likely to be smooth in the sense that its differences across frequencies have limited variation. The smoothness prior is a belief in simplicity: agents believe that spectra typically have limited variation across frequencies, rather than fluctuating wildly. Following Shiller (1973), Akaike (1979), and Kitagawa and Gersch (1984, 1996), the prior is represented by a penalty on variability that is appended to the likelihood of the data.[11] Given assumption 4, the penalized log likelihood of the data given a model $\hat{f}$ is

$$PL\left(x \mid \hat{f}, \tau\right) = \underbrace{-\frac{1}{2}\sum_{j=1}^{n}\left(x(\omega_j) - \hat{f}(\omega_j)\right)^2 \tau(\omega_j)\, d\omega}_{\text{Data likelihood}} \underbrace{- \frac{\lambda}{2}\sum_{j=2}^{n}\left(\frac{\hat{f}(\omega_j) - \hat{f}(\omega_{j-1})}{d\omega}\right)^2 d\omega}_{\text{Roughness penalty}}. \tag{16}$$

$PL\left(x \mid \hat{f}, \tau\right)$ depends on two factors: the log likelihood for normally distributed data and a term encoding the belief in smoothness. Models are viewed as less plausible when they are rougher or more complicated in the sense of having a larger average squared derivative. The most plausible models have perfectly flat spectra – white noise – while the least plausible have highly variable

[11] The smoothness prior is often explicitly jutified as a belief in simplicity. In Shiller (1973), which is the first application of such a prior, a justification is that "[i]n most applications...the researcher will feel that...the lag coefficients should trace out a 'smooth' or 'simple' curve." While Shiller's (1973) smoothness prior is stated in the time domain, those in Akaike (1979) and Kitagawa and Gersch (1985, 1989) are specified in the frequency domain in a manner almost identical to ours.

spectra.[12,13]

The parameter λ controls the strength of prior. For any fixed λ, as the signal precision grows large, the smoothness penalty becomes irrelevant. One reason we include the smoothness prior is that without it, $\hat{f} = x$ is the maximum-likelihood estimate, which would imply that $\hat{f}$ has infinite variation and would yield an inconsistent estimate of f, even as $n \to \infty$ (Wahba (1980)). The smoothness prior also implies that when people have weak signals, they use simple and smooth models. Complexity here only arises when people have a wealth of information. When signals are more precise, so that τ is large relative to λ, the roughness penalty is relatively less important and the agent will consider more complex models.

In addition to the roughness penalty, we also assume that agents are able to express a prior mean over possible models. In the absence of any information about the world, they believe the average spectrum is flat at $\bar{f}$. This assumption is introduced so that it is possible for the agent to calculate expectations for $\hat{f}$ prior to observing signals (i.e. the outer expectation operator in assumption 2).

The penalized likelihood leads to the following assumption:

Assumption 5 *Nature chooses the income process from the set*

$$F(x;\tau) = \left\{\hat{f} : PL\left(x \mid \hat{f}, \tau\right) \geq \bar{L}\right\}. \tag{17}$$

Agents assume that nature can choose models that are not too inconsistent with the data in the sense that they have a log likelihood above some lower bound. Another way to state the assumption is that, compared to the maximum likelihood estimator of the spectrum, nature can impose an alternative model that cannot be rejected on the basis of a likelihood ratio test at some confidence level (depending on $\bar{L}$).

Assumption 5 determines the set of models (or "priors", in the language of the ambiguity aversion literature) that the agent believes nature will choose from. This set is typically determined exogenously, whereas here the set is endogenous to the agent's information acquisition decision. That is the key qualitative way in which we build on the past work on multiple priors.

In our specification, the set of models that the agent deems sufficiently plausible that nature might choose them is constrained by a likelihood ratio test. Hansen and Sargent (2007) study an alternative measure of plausibility. They assume that agents have some specific (exogenous) benchmark model and that nature is constrained to choose an alternative that is close to the benchmark in terms of Kullback–Leibler (KL) divergence. The KL divergence is similar to the constraint here, in that it involves a likelihood ratio. The difference is that the KL divergence is the expected log likelihood ratio comparing the alternative and benchmark models when data is

[12] That white noise is treated as the most plausible is also sensible from an information theoretic perspective since Gaussian white noise has the greatest Shannon entropy among all time series processes with a given variance.

[13] An alternative way to penalize complexity in models would use the coefficients of the ARMA representation. We will see below, though, that the smoothness prior we impose here also ends up imposing smoothness on the AR and MA coefficients.

generated under the alternative. Our set of models from which nature can choose is determined by the likelihood of the model conditional on the data the agent has observed.

The difference between our set of possible models and those in the Hansen–Sargent framework is therefore that we replace their benchmark model by the (endogenous) signals x, and that deviations between an alternative model and x are weighted by the chosen precisions τ, whereas the KL divergence in the Hansen–Sargent framework puts equal weight on deviations at all frequencies (see Dahlhaus (1996) and Bidder and Dew-becker (2016)). That is the sense in which the set of plausible models is made endogenous here.

It is also worth noting here that the KL divergence, which is sometimes also called an entropy distance, is separate from entropy as a potential measure of information flow to the agents. The KL divergence used in the Hansen–Sargent framework is the relative entropy between the benchmark model and an alternative that nature might choose. The relative entropy used in the rational inattention framework for measuring information flows is the relative entropy between a prior and a posterior. We discuss in section 5 how our analysis changes if the Shannon/Sims use of entropy to measure information flow is applied in our setting.

At this point, all the basic terms in the preferences in assumption 2 have been defined. The next section examines the three optimizations.

2 Solution

All three optimizations in the preferences – the consumption policy, nature's choice of a model, and the information decision – are analytically solvable. The solution is itself an important contribution of the paper. There is little work that obtains closed-form solutions for optimal consumption under model uncertainty and rational inattention, and the fact that the model can be solved when model uncertainty and attention are themselves endogenous is even more surprising. We analyze the three pieces of the optimization in turn.

2.1 Optimal consumption conditional on a model

The minimax theorem implies that the inner maximization and minimization in the preferences (2) can be reversed. Intuitively, the operations represent a zero-sum game with a pure-strategy Nash equilibrium. We first solve for optimal consumption given a model.

Lemma 2 *The consumption policy that maximizes expected utility conditional on some $\hat{b}$ is*

$$C_t = (R-1)\left(W_{t-1} + \hat{z}(L)\,\hat{\varepsilon}_t\right) - \frac{\alpha}{2}R^{-1}\left(1-R^{-1}\right)\hat{b}\left(R^{-1}\right)^2 - \alpha^{-1}\frac{\log\beta R}{R-1}, \tag{18}$$

where $\hat{z}(L)$ is a lag polynomial with coefficients

$$\hat{z}_j = \sum_{k=j}^{\infty} R^{-(k-j)}\hat{b}_k. \tag{19}$$

Expected utility from consumption is then

$$\begin{aligned}
& \max_{C^{policy}} E\left[-\alpha^{-1}\sum_{t=0}^{\infty}\beta^t \exp\left(-\alpha C_t\right) \mid \hat{b}\right] \\
= & \frac{-\alpha^{-1}}{1-\beta}\exp\left(\frac{\alpha^2}{2}R^{-1}\left(1-R^{-1}\right)\hat{b}\left(R^{-1}\right)^2+\log\frac{(1-\beta)}{1-R}+\frac{\log\beta R}{R-1}\right). \qquad (20)
\end{aligned}$$

Lemma 2 provides two useful results. First, we obtain a standard consumption function: agents consume the annuity value of financial plus human wealth, $(R-1)\left(W_{t-1}+\hat{z}\left(L\right)\hat{\varepsilon}_t\right)$, minus a precautionary saving term $\frac{\alpha}{2}R^{-1}\left(1-R^{-1}\right)\hat{b}\left(R^{-1}\right)^2$. The behavior of consumption thus depends on beliefs about income dynamics through two channels. First, $\hat{b}$ affects the riskiness of the income stream, and hence the amount of precautionary savings agents desire to hold. Second, and more importantly, current consumption depends on beliefs about future income, which are driven by $\hat{b}$. When $\hat{b}$ implies that income shocks are more persistent ($\sum_{k=j}^{\infty}R^{-(k-j)}\hat{b}_k$ is larger) consumption responds more strongly to them. These are all standard results. Deviations of the behavior of the agents in our model from the standard permanent-income predictions are caused by deviations of their model, $\hat{b}$, from the truth.

Lemma 2 also characterizes optimized expected utility from consumption for a given income process $\hat{b}$. The only term that differs across models is $\hat{b}\left(R^{-1}\right)^2$, which measures the variance of innovations to permanent income, and hence the variance of consumption growth. Expected utility is lower when the variance of consumption growth is higher.

The information structure laid out in the previous section refers entirely to the log spectrum, but utility is derived in lemma 2 terms of the lag polynomial $\hat{b}$. The two are linked through the following result.

Lemma 3 *For a log spectrum $\hat{f}$ that is bounded from above and below, where $\hat{b}\left(L\right)$ is the associated Wold representation,*

$$\log\hat{b}\left(R^{-1}\right)^2 = \frac{1}{\pi}\int_0^{\pi}Z\left(\omega\right)\hat{f}\left(\omega\right)d\omega, \qquad (21)$$

$$\textit{where } Z\left(\omega\right) \equiv 1+2\sum_{j=1}^{\infty}\cos\left(\omega j\right)R^{-j}. \qquad (22)$$

Proof. This is known as the Poisson representation in complex analysis. Insert R^{-1} for z in equation 10.2.10 of Szegő (1975) or equation 2.11 of Inoue and Kasahara (2006). ■

Lemma 3 gives us a powerful result: $\log\hat{b}\left(R^{-1}\right)^2$, the statistic that determines expected utility from consumption conditional on a model, is linear in the log spectrum. This is the key result that allows us to solve the model analytically. It does not seem to have been noted previously in the economic theory literature, but it is likely useful in other contexts, since it is a general result for

NPV innovations.[14]

Lemma 3 shows that utility is always decreasing in $\hat{f}$ and hence the overall variance of income growth. The function Z determines the importance of risk at each frequency, illustrating the use of the frequency-domain approach. The left-hand panel of figure 1 plots Z for an annual calibration with $R = 1.025$. $Z(\omega) > 0$ for all ω, it is bounded from above for $R > 1$, reaching its maximum at $\omega = 0$, and it is decreasing on $(0, \pi)$. When $R = 1$, Z is equivalent to the Dirac delta function. The mass of Z primarily lies on extremely low frequencies. So what matters for the agent's utility, through $\hat{b}\left(R^{-1}\right)^2$, is the variance of the most persistent components of income. Transitory fluctuations in income do not meaningfully affect consumption. Rather, permanent income shocks change human wealth, and thus consumption, so permanent volatility reduces utility. These characteristics of the utility function and Z are robust features of the model, as they do not depend on any sort of detailed calibration – the only parameter affecting Z is the gross interest rate.[15]

2.2 Nature's minimization

Since $\hat{b}\left(R^{-1}\right)^2$ is the only term in (20) that differs across models, nature's minimization problem in (2) is equivalent to choosing $\hat{f}$ from the set $F(x;\tau)$ to maximize $\int_0^\pi Z(\kappa)\hat{f}(\kappa)\,d\kappa$. Nature's Lagrangian is

$$\min_{\hat{f}\in F(x;\tau)} -\frac{1}{\pi}\int_0^\pi Z(\omega)\hat{f}(\omega)\,d\omega - \psi PL\left(x \mid \hat{f}, \tau\right), \tag{23}$$

where ψ is a Lagrange multiplier. We refer to the model that achieves the minimum in (23) as $f^w(\omega; x, \tau)$.

It is straightforward to solve for f^w from the first-order conditions for the nature's optimization. Define vectors (in boldface) of the form $\mathbf{f}^w(x;\boldsymbol{\tau}) \equiv [f^w(\omega_1; x, \tau), ..., f^w(\omega_n; x, \tau)]'$ (recall that the frequencies $\omega_j = \pi j/n$ are the uniform discretization of the interval $[0, \pi]$ on which the agent receives signals and that we think of n as large). $diag(\cdot)$ is an operator that creates a matrix with its argument on the main diagonal and zeros elsewhere.

Proposition 1 *The model that solves (23) is*

$$\mathbf{f}^w(x;\boldsymbol{\tau}) = \left(I_{n\times n} - \lambda diag\left(\boldsymbol{\tau}^{-1}\right)D\right)^{-1}\left(\psi diag\left(\boldsymbol{\tau}^{-1}\right)\mathbf{Z} + \mathbf{x}\right) \tag{24}$$

[14] The innovation variance for the NPV of a time series arises naturally in many economic settings, such as the consumption/savings problem here, equilibrium macroeconomic models (Hansen and Sargent (1980, 1981)), models with generalized recursive preferences (Bidder and Dew-Becker (2016); Dew-Becker and Giglio (2016); Dew-Becker (2016)), the q theory of investment, and Calvo-type price setting.

[15] The analysis so far has assumed income is stationary. That assumption has no effects on our results. In the presence of a unit root, the analysis applies to the first difference of income. If $\hat{g}(L)$ is the Wold representation for the first difference of income, then $\hat{b}\left(R^{-1}\right) = \hat{g}\left(R^{-1}\right)/\left(1 - R^{-1}\right)$. The agent then can calculate $\log \hat{b}\left(R^{-1}\right)^2$ by using Lemma 3 applied to the log spectrum of income *growth* and subtracting $\log\left(1 - R^{-1}\right)$. The loading of utility on frequencies for the level of income is the same as for the first difference.

where $I_{n\times n}$ is an $n \times n$ identity matrix and D is a differencing matrix of the form

$$D \equiv \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 1 & -2 & 1 & & \\ 0 & 1 & -2 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & -1 \end{bmatrix} d\omega^{-2}. \tag{25}$$

f^w is a linear function of x and Z. The worst-case spectrum is higher at frequencies where τ is smaller – there is more uncertainty about the spectrum – and where Z is larger – increases in the spectrum are more painful. Similarly, when ψ is larger, so that the agent is more ambiguity-averse (the constraint on nature's choice of a model is looser), the worst-case model tilts more in the direction of Z.

Before analyzing the implications of proposition 1 in detail, we first solve the agent's optimal τ to help frame the effects of information choice on consumption behavior.

2.3 Optimal information choice

The fundamental goal of the paper is to understand how information acquisition interacts with ambiguity aversion. Ambiguity aversion here is used to model agents' aversion to model uncertainty. But even without ambiguity aversion, the specification of utility from consumption can affect attitudes towards model uncertainty, since $E\left[\sum_{t=0}^{\infty} -\alpha^{-1}\beta^t \exp\left(-\alpha C_t\right) \mid \hat{f}\right]$ is not in general a linear function of the model $\hat{f}$. We choose the function G so that in the absence of ambiguity aversion, agents would be risk-neutral over models. That is, G is chosen so that utility is linear in the riskiness of income, $\hat{b}\left(R^{-1}\right)^2$ (which means utility is also linear in $\hat{f}$).

Assumption 6

$$G(x) \equiv -\log\log\left(-\alpha\left(1-R\right)\left(\beta R\right)^{1-R} x\right) \tag{26}$$

The log log specification for G implies that the preferences (2) have the equivalent representation, after optimizing over consumption,

$$\max_{\tau} E\left[\min_{\hat{f}\in F(x;\tau)} -\frac{1}{\pi}\int_0^{\pi} Z(\omega)\,\hat{f}(\omega)\,d\omega\right]. \tag{27}$$

The aversion to model uncertainty is captured purely by the ambiguity aversion in the sense that if the minimization operator were removed, utility would not by affected by uncertainty about the spectrum of income, f.

Proposition 2 *The optimal information policy under the preferences (2) (and (27)) when $\gamma(\omega) = 1$ is*

$$\tau^*(\omega_j) = \underbrace{\theta^{-1/2}}_{\textit{Shadow cost of info.}} \times \underbrace{\psi^{1/2}}_{\textit{Ambiguity aversion}} \times \underbrace{Z(\omega_j)}_{\textit{Utility weights}}. \tag{28}$$

where θ is the Lagrange multiplier on the information constraint, $\sum_j \tau(\omega_j) d\omega \leq \bar{\tau}$.

Recall that the function Z measures how the level of the log spectrum, f, affects utility. Agents optimally gather information exactly in proportion to Z, learning the most about the frequencies that are most important for utility. In terms of the adversarial game with nature, the agent chooses precision to constrain nature most at the frequencies that are potentially most painful.

While Z controls the shape of τ^*, θ and ψ determine its scale. An increase in the available precision $\overline{\tau}$ lowers the Lagrange multiplier θ, leading to more precision at all frequencies. ψ determines the extent to which nature is constrained by the penalized likelihood, i.e. how ambiguity-averse people are. Holding the shadow cost θ of precision constant, a decrease in ambiguity-aversion through ψ lowers the chosen precisions.

To see the implication of proposition 2 for noise in the signals at each frequency, the right-hand panel of figure 1 plots $Z(\omega)^{-1} \propto \tau^*(\omega)^{-1}$. The variance of the signals that the agents receive is a simple function of frequency, rising smoothly as the frequency increases (it is an affine function of $\cos(\omega)$).

Lemma 2 and propositions 1 and 2 give the complete analytic solution to the model in the case $\gamma(\omega) = 1$. The remainder of the paper analyzes the implications of the solution for the types of models that agents optimally use and how those choices affect observable consumption behavior. Section 5 then examines how more general specifications for γ affect the results.

3 Behavior of the model agents use

We have two relevant cases for τ. The utility-optimal information policy, $\tau^*(\omega)$, says that it is proportional to $Z(\omega)$, while the statistical benchmark in the absence of optimization is to set $\tau(\omega)$ to equal a constant. We focus on two key results for f^w under those policies:

1. **Optimal learning eliminates excessive extrapolation:** Without an optimal information policy, the worst-case model displays excessive persistence compared to the truth – people over-extrapolate shocks. But under optimal information (τ^*), that bias disappears.

2. **Agents make mistakes primarily about the transitory component of income:** Under the optimal policy, agents use models that tend to deviate from the truth more at high than at low frequencies.

This section derives those results theoretically and examines them in numerical simulations of the model.

3.1 Optimal learning eliminates excessive extrapolation

Taking an expansion around an infinite level of precision, the appendix derives the following first-order approximation in the continuous limit of the problem ($d\omega \to 0$) for arbitrary τ:

$$E\left[f^{w}(\omega;x,\tau)-f\mid f\right]\approx\psi\tau(\omega)^{-1}Z(\omega)+\lambda\tau(\omega)^{-1}f''(\omega). \tag{29}$$

Equation (29) yields our first important result. In the statistical benchmark case where τ is constant across frequencies, f^{w} is biased in the direction of $Z(\omega)$. Recall from figure 1 that Z is large at low frequencies and close to zero elsewhere. So under the statistical benchmark, the worst-case model has excessively high power at low frequencies, which means that it is more persistent than the truth (f). That result is almost exactly what is obtained in Bidder and Dew-Becker (2016), and is closely related to results in Hansen and Sargent (2010, 2016). Intuitively, since highly persistent models lead to the lowest utility, the agent naturally fears them.

Equation (29) also yields the more important part of the result, though, which is that under the optimal policy, τ^{*}, there is no systematic bias towards either under- or over-extrapolation. Specifically, we have under the optimal policy

$$E\left[f^{w}(\omega;x,\tau^{*})-f\mid f\right]\approx\psi^{1/2}\theta^{1/2}+\lambda\tau^{*}(\omega)^{-1}f''(\omega). \tag{30}$$

Since $\tau^{*}(\omega)\propto Z(\omega)$, the frequencies that are most important for utility are also the ones that the agent learns the most about, thus constraining the worst-case model. The proportionality completely cancels Z out of the bias, leaving just a constant.

When f^{w} deviates from f by only a constant, the two models have identical autocorrelations and differ only in the conditional variances. For example (ignoring the effects of f'' for the moment; i.e. for small λ), if income follows an AR(1) process with persistence ρ, then $E\left[f^{w}\right]$ is the log spectrum for an AR(1) also with persistence ρ, but with innovations that have a greater variance.

Equation (30) is a key result of the paper. It shows that endogenous learning can completely eliminate overextrapolation. Intuitively, ambiguity averse agents tend to focus on models with excessive persistence because they are associated with low utility. But that fact also causes them to obtain the most information about those frequencies, thus entirely canceling out the effect of ambiguity.

This result stands in conflict with recent work that argues that ambiguity aversion and information processing constraints lead to overextrapolation. What we show here is that when people are able to choose what aspects of income to learn about, they naturally focus on the low frequencies, since those are most important for utility. But it is precisely that focus that then eliminates any bias towards excessive extrapolation.

3.1.1 Numerical example

To make the result above more concrete, we consider a simple numerical example. Suppose income is truly i.i.d. over time, $Y_t = \varepsilon_t$, so that the true model has zero persistence. Since $f''(\omega) = 0$, the second term in equations (29) and (30) is equal to zero. The left-hand panel of figure 2 plots the true (flat) log spectrum $f(\omega)$ along with the mean worst-case spectra under the optimal information policy τ^* and for the statistical benchmark in which τ is constant across frequencies (the calibration is set so that they have equal total precision: $\sum_j \tau^*(\omega_j) = \sum_j \tau$), which we denote with $\bar{f}_*^w$ and $\bar{f}_F^w$, respectively. The figure shows that $\bar{f}_*^w$ is shifted up by a constant compared to f, while $\bar{f}_F^w$ actually has a significantly different shape, with a peak at low frequencies indicating persistence in income.

The right-hand panel of figure 2 plots the impulse response functions (the b's) associated with the three models. Since income is truly i.i.d., $b_j = 0$ for $j \geq 1$ under the true model. Under the optimal information policy with model uncertainty, the only thing that changes on average is that b_0 becomes larger – people fear a higher variance, but they do not on average act as though income actually has any persistence. Under the statistical benchmark, though, there is clearly persistence in income: the impulse response is consistently positive after the initial impact. Figure 2 thus illustrates our first basic result. While ambiguity aversion and model uncertainty can often drive agents to act as though income is excessively persistent, that result is delicate: it disappears when people can allocate attention and information acquisition optimally.

3.2 Agents make mistakes about the transitory component of income

The primary mistakes in the agent's worst-case model come from the term involving $f''(\omega)$. That part of the formula is driven by the agent's smoothness prior. In the face of noisy data, agents estimate the spectrum of income by smoothing information across frequencies. Since $f^w(\omega; x, \tau)$ is a convex combination of the data x local to ω, it is biased upward when $f'' > 0$ and downward when $f'' < 0$. Intuitively, if there is a narrow peak in f, a simple model will tend to smooth the peak out, and thus be biased downward.

In that sense, the agents also have a bias towards simplicity: they use models with smaller variations across frequencies when they have less information.[16] When the true spectrum is in fact complex, in the sense that it has local peaks and troughs, the worst-case model will tend to make mistakes in smoothing those peaks out. So the errors appear exactly where f'' is large.

Since the optimal information policy gives the agents noisier signals about the spectrum at high frequencies, that is also where they make the largest smoothing errors. In (30), $f''(\omega)$ is multiplied by $\tau^*(\omega)^{-1}$. So when precision is high, the term is scaled down and the worst-case spectrum tracks the true spectrum closely. But when τ^* is small – at high frequencies – agents do more smoothing across frequencies and make larger mistakes.

[16] That intuition can be formalized. It is possible to show that correlations in the estimated spectrum, $f^w(\omega; x, \tau)$, are higher across frequencies, implying that complexity is lower, in regions where τ is smaller.

3.2.1 Numerical example

To illustrate the errors caused by smoothing, we now consider a richer numerical example with multiple peaks in the spectrum that we view as more realistic. The left-hand and middle panels of figure 3 plot the log spectrum of the data-generating process for income, while the right-hand panel plots the impulse response of income to a shock. The calibration is chosen to have both high-and low-frequency components (see evidence discussed in Kaplan and Violante (2010)). The high-frequency piece – which generates the middle peak in the spectrum – is driven by the fact that a component of the shocks to income reverts: when income rises higher by \$1 today, it is lower on average by 50 cents over the next three periods. That behavior can be caused by forces that shift income over time but have little effect on total lifetime income. For example, many people overpay taxes during the year and then receive refunds (e.g. Souleles (1999)). The low-frequency component of income – the left-hand peak in the spectrum – comes from the fact that the impulse response is persistently positive in the later periods following a shock. This represents a persistent component in income growth, and could come from variation over time in the average growth rate of the economy or the performance of one's employer. The preference parameters are chosen to illustrate the main mechanisms in the model. Quantitatively, θ is chosen so that agents make quantitatively large consumption mistakes (see below).[17]

We examine two specifications for τ: the first is the optimum derived above, τ^*, which is proportional to $Z(\omega)$; the second specification is the statistical benchmark that sets $\tau(\omega)$ to be constant at the mean of τ^*:

$$\tau^F(\omega_j) = \tau^F \equiv n^{-1} \sum_{i=1}^{n} \tau^*(\omega_j). \tag{31}$$

As in the previous example, the choice of the mean for τ^F implies that it has the exact same information cost as τ^*. Note, though, that since precision is the inverse of variance, the average variance of the errors across frequencies is in fact much smaller under τ^F than under τ^*.

Figure 3 plots $\bar{f}^w_*$ and $\bar{f}^w_F$ for the two-peak calibration. The average model under the optimal policy, $\bar{f}^w_*$, matches f very well at the lowest frequencies, but it does a poor job of matching the middle-frequency peak in f and also deviates substantially at higher frequencies. The average model under the statistical benchmark, $\bar{f}^w_F$, has the opposite behavior: it matches the middle-frequency peak and high-frequency behavior well, and in fact matches f well at almost all frequencies, but it fits relatively poorly at low frequencies. That is exactly what the formulas predict: optimal learning

[17] Technically, the impulse response function for income is equal to [1,-0.15,-0.3,-0.15,0,...] plus $0.095\exp(-0.1j)$. It is then scaled so that the standard deviation of consumption growth is 1.56 percent (when initial consumption is equal to 1).

As discussed above, n is intended to be taken as large – it is only used to avoid infinities – so we set it to 4000. $\beta = 0.975$ to represent an annual calibration, and $R = \beta^{-1}$ for simplicity. $\bar{\tau}$, λ, and ψ are chosen in order to ensure that the agents make non-trivial mistakes in modeling consumption and that the behavior is visibly different across the two policies for τ. $\psi = 10^{-4}$; $\lambda = 0.00075$; $\bar{\tau} = 405.83$; $\theta = 49.35$. The parameterization is meant to illustrate the qualitative behavior of the model rather than match specific quantitative data. The degree of ambiguity aversion, ψ, has minimal effects on behavior under the optimal information policy, but it matters much more under the flat information policy.

causes models to be relatively more accurate at low than high frequencies. Overall, though, $\bar{f}_F^w$ has a much better fit than $\bar{f}_*^w$, with a root mean squared error that is 42 percent smaller, due to the fact that $\bar{f}_F^w$ spreads information evenly across frequencies.

The right-hand panel of figure 3 plots the lag polynomials, b, $\bar{b}_*^w$, and $\bar{b}_F^w$, associated with the log spectra f, $\bar{f}_*^w$, and $\bar{f}_F^w$, respectively. $\bar{b}_*^w$ fails to match the short-run mean-reversion in the income process, while the lag polynomial for the suboptimal information policy, $\bar{b}_F^w$, does not, as predicted by the analytic results. The figure shows that the greater smoothness of $\bar{f}_*^w$ also translates into smoothness in the associated lag polynomial, and in particular errors in the transitory behavior of income. But the figure shows that the optimal policy performs better at longer lags, giving a closer fit to the persistent component of the impulse response function. Since it is the long-run part that determines human wealth, and hence optimal consumption, it is optimal from an expected utility perspective for agents to use models that fit the persistent component at the cost of missing the transitory dynamics.

4 Implications for observable consumption behavior

We now explore the implications of the results in the previous section for the observable behavior of consumption.

4.1 Consumption function

The consumption function from (18) implies that consumption growth follows

$$\begin{aligned} \Delta C_t &= \left(1-R^{-1}\right) b^w\left(R^{-1}\right) \varepsilon_{t+1}^w + \frac{\alpha}{2}\left(1-R^{-1}\right)^2 b^w\left(R^{-1}\right)^2 + \alpha^{-1}\log\beta R && (32)\\ \text{where } \varepsilon_{t+1}^w &\equiv b^w(L)^{-1} Y_t, && (33)\end{aligned}$$

$\Delta \equiv 1-L$ is the first-difference operator, and $b^w(L)$ is the Wold representation associated with the worst-case model f^w. In the case where agents use the true model, so that $b^w = b$ (i.e. under complete information), the filtered shocks, ε^w are equal to the true shocks, ε, and consumption follows a random walk with innovations equal to the innovation in the annuity value of the NPV of future income, $\left(1-R^{-1}\right) b\left(R^{-1}\right)\varepsilon_{t+1}$. When the agent uses a model that differs from the truth, though, ε_{t+1}^w is no longer an i.i.d. process and consumption growth is no longer uncorrelated over time. That is, the agent's estimated shocks, ε^w, are in general serially correlated, which leads to (suboptimal) serial correlation in consumption growth.

To better understand the implications of the worst-case model for the behavior of consumption growth, we analyze the log spectrum of consumption growth,

$$f_{\Delta C}^w(\omega; x,\tau) = \log\left(\left(1-R^{-1}\right)^2 b^w\left(R^{-1}; x,\tau\right)^2\right) + f(\omega) - f^w(\omega; x,\tau). \tag{34}$$

When the agent knows the true model, $f_{\Delta C}^w$ is perfectly flat, which means that consumption growth is uncorrelated over time and the level of consumption is a random walk. But in general the agent does not know the true model. For example, if the true spectral density has a peak at some frequency but the worst-case spectrum does not, then $f_{\Delta C}^{w}$ will inherit the same peak through the term $f(\omega) - f^{w}(\omega; x, \tau)$. That is, features of the income spectrum that the agent "ignores" in the sense that they do not appear in f^{w} are passed through to the spectrum of consumption growth.

Using (34), we can immediately map the results in the previous subsections into the spectrum of consumption growth. Specifically, for general information policies and for the optimal policy, we have

$$\begin{aligned} E\left[f_{\Delta C}^{w}(\omega; x, \tau) \mid f\right] &\approx E \log\left(\left(1 - R^{-1}\right)^{2} b^{w}\left(R^{-1}; x, \tau\right)^{2}\right) \\ &\quad - \psi \tau(\omega)^{-1} Z(\omega) - \lambda \tau(\omega)^{-1} f''(\omega), \end{aligned} \tag{35}$$

$$\begin{aligned} E\left[f_{\Delta C}^{w}(\omega; x, \tau^{*}) \mid f\right] &\approx E \log\left(\left(1 - R^{-1}\right)^{2} b^{w}\left(R^{-1}; x, \tau^{*}\right)^{2}\right) \\ &\quad - \psi^{1/2} \theta^{1/2} - \lambda \tau^{*}(\omega)^{-1} f''(\omega). \end{aligned} \tag{36}$$

Again, the information policies differ in two key ways. First, comparing the terms $\psi \tau(\omega)^{-1} Z(\omega)$ and $\psi^{1/2}\theta^{1/2}$, there are no systematic deviations of consumption growth from white noise under the optimal information policy. Under other policies, though, since people overextrapolate income shocks, consumption is actually mean reverting in the long-run – there is a trough in $f_{\Delta C}^{w}$ at frequency zero. Intuitively, overextrapolation causes people to consume more than they can afford (more than human wealth) following positive shocks. Eventually, then, they must reduce consumption, causing long-run mean reversion. So the observable prediction of the model is that we actually *should not* observe long-run mean reversion in consumption growth. By the same token, people should also not underreact to shocks (as under rational inattention), which would lead to long-run persistence in consumption growth.

The second class of mistakes is the smoothing errors due to the term $\lambda \tau(\omega)^{-1} f''(\omega)$. This term says essentially that variation in the spectrum of income that the agent is not aware of passes directly into consumption growth. When f'' is negative, for example, there is a local peak in the spectrum of income, and the spectrum of consumption growth then is also relatively high. Again, these errors are scaled by the precision of signals. The model predicts that consumption should track income relatively more closely – have a similar impulse-response function – at high than low frequencies. Transitory variation in income, such as the shifts in income over time studied by Souleles (1999), is predicted to pass directly into consumption. We illustrate that behavior below in a numerical example.

Compared to the behavior under the standard setup with no model uncertainty, our model generates, through limited information, excess sensitivity of consumption to high-frequency shocks to income. This result is not obtained, though, by appealing to some sort of irrationality; rather, it arises simply from people optimally choosing to focus their attention on low frequencies. Endogenous attention leads to our second difference from the literature, which is that unlike other recent work on model uncertainty (Fuster, Hebert, and Laibson (2012), Bidder and Dew-Becker (2016), and Hansen and Sargent (2016)), the model does *not* predict excessive extrapolation of shocks. The model predicts excess sensitivity to transitory variation in income, but in fact the *correct* sensitivity to the permanent component.

The model also has rather different predictions from rational inattention over state variables (as opposed to rational inattention over model specifications), which suggests that they could be tested against each other empirically. As discussed by Sims (2003), the most prominent prediction of rational inattention is delayed reaction to shocks, due to the fact that people observe the shocks imperfectly. If income rises permanently, Sims (2003) shows that in general people will take a number of periods to fully realize that such a shock has occurred, meaning that consumption responds slowly to permanent shocks to income. Here, on the other hand, agents respond rapidly to permanent shocks because it is precisely the low-frequency part of income that they understand best.

Sims (2003) and Luo (2008) show that rational inattention can also generate excess sensitivity of consumption to income shocks, but the effects are calibration-specific and may be quantitatively small (e.g. the simulations in Sims (2003)). Intuitively, excess sensitivity arises because agents are not able to distinguish permanent from transitory shocks. So to obtain high-frequency mistakes, the rational inattention model must also predict low-frequency mistakes. In our model, though, the prediction of optimal information acquisition is in fact that the same attention choice both induces high-frequency mistakes and eliminates low-frequency mistakes. Furthermore, we see in the next section that the high-frequency mistakes can be quantitatively large and realistic.[18]

4.2 Numerical example

We examine the behavior of consumption under the numerical simulation when income has both transitory and persistent components. Figure 4 plots the log spectra of consumption growth under the various models. $\bar{f}_*^w$ provides a closer fit to the utility optimal consumption spectrum at all frequencies. On the other hand, the statistical information policy produces a spectrum that is flatter – and closer to white noise – across most frequencies, but it has a very large peak at the lowest frequencies. The key question, then, will be which type of deviation – low- or high-frequency – is more relevant for utility.

To see how the fitting errors affect the behavior of consumption growth in the time domain, the right-hand panel of figure 4 plots the impulse response of the level of consumption to a unit shock to ε_t (i.e. a true innovation, not a filtered one) under the three consumption rules along with the cumulative impulse response of income (multiplied by $\left(1 - R^{-1}\right)$). As we would expect, the response of consumption under the full-information rule is flat: the permanent income hypothesis holds, and the response of consumption is approximately equal to the cumulative increase in income. The line for consumption under the optimal information policy shows that it inherits some of the

[18] It is also worth noting that the models in Sims (2003) and Luo (2008) can only be solved under quadratic utility, whereas we are able to accommodate CARA preferences here.

short-run mean-reversion in income, rising and falling in the first few periods. It does not include the persistent component in income, though – consumption immediately jumps to approximately its long-run level, but the fluctuates around that level excessively. So the consumption policy is "right" in the long-run, but it is excessively sensitive to transitory variation in income in the short-run.

The behavior of a person using the model $\bar{f}_*^w$ is again notably different from one using $\bar{f}_F^w$. The latter model does a better job of eliminating high-frequency fluctuations in consumption, but at the cost of inheriting the low-frequency behavior of income. The initial response of consumption under $\bar{f}_F^w$ is too small, and consumption slowly drifts upward over the 80 periods of the IRF plotted here, eventually overshooting. So the τ^F policy, counter to what is observed empirically, eliminates the sensitivity of consumption to transitory fluctuations in income, but causes consumption growth to deviate from white noise at long horizons. This result argues that empirically, τ^* is a better description of consumption behavior than a setting where agents do not choose information optimally, τ^F.

Those results may also be observed in more standard time series regressions for consumption growth. Table 1 below reports the coefficients from simulated regressions of consumption growth on the predictable and unpredictable components of income growth under the two information policies and also under the full-information optimum.[19]

Information policy	Predictable income	Unpredictable income
τ^*	0.50	1.13
τ^F	0.12	0.72
Full-info. optimum	0	1.11

Table 1. Coefficients from regressions of consumption growth on income growth

The coefficient from the regression of consumption growth on the predictable part of income is of the same order of magnitude as the coefficient on the unpredictable part under τ^*. The model can thus replicate the empirical result that consumption responds strongly to predictable income changes. That value is consistent (by calibration) with the results of Parker (1999) and Souleles (1999), who both find that consumption rises by 0.5 percent following a 1-percent anticipated increase in income.

Under the statistical benchmark, τ^F, on the other hand, that relationship is much weaker, with the response to predictable income being, at 0.12, smaller by a factor of 4. It is precisely the fact that agents optimally (under τ^*) fail to learn about high-frequency features of the model that causes them to overreact to predictable parts of income. Furthermore, note that the response of consumption to true income shocks is far closer to the full-information optimum under τ^* than

[19] Here we use the version of the model in which income is difference-stationary. That is, the calibration of the impulse responses and the spectrum are applied to income growth instead of its level. As discussed in footnote 15, the results go through identically in that case. The difference is simply that then consumption and income have volatilities that are of the same order of magnitude, as observed in the data. the calibration is otherwise identical to what is discussed in footnote 17.

under τ^F. This again demonstrates that in many ways, τ^* helps agents get long-run responses right.

An alternative way to examine the behavior of consumption in the time domain is to study its autocorrelations. The left-hand panel of figure 5 plots the autocorrelations of consumption growth under τ^* and τ^F. Obviously under the full-information optimum, the autocorrelations are zero. At short lags, the autocorrelations are higher under τ^*. Subsequently, though, the autocorrelations are substantially lower – by nearly a factor of 10. The right-hand panel plots the first autocorrelation of consumption growth over different spans. For a horizon denoted by n on the x-axis, we plot $corr\left(\sum_{j=0}^{n-1}\Delta C_{t+j},\sum_{j=0}^{n-1}\Delta C_{t-n+j}\right)$. So the figure represents how consumption growth is correlated over neighboring intervals of length n. Consistent with the left-hand panel, for short intervals the correlations are higher under τ^* than τ^F. As we claimed above, though, the figure shows that consumption growth over long periods is substantially less autocorrelated under τ^* than τ^F.

To summarize, this example confirms the analytic results above that the optimal information policy generates consumption growth that is close to white noise in the long-run, but that it causes consumption to be excessively sensitive to variation in income in the short-run. It also shows that the model can generate the empirical result that consumption responds to predictable variation in income.

4.3 Empirical evidence

Since the optimal information policy implies that people learn the most about low-frequency features of the income process, it says that deviations of consumption growth from white noise should be observed primarily at high frequencies. Specifically, if the agent's model of income dynamics, $f^w(\omega; x, \tau^*)$, is flat at high frequencies, then any variation in the shape of the true spectrum passes directly into consumption. The shape of the spectrum of $f^w_{\Delta C}(\omega; x, \tau^*)$ will typically be similar to that of $f(\omega)$ at high frequencies as the model predicts that people use simple (flat) models there.

Another way to build intuition for that prediction of the model is to note that high-frequency shocks also have relatively small effects on the net present value of income compared to more persistent shocks (which is why the function Z is relatively small at high frequencies). So the model essentially predicts that people spend excessively out of relatively small high-frequency increases in income compared to the larger low-frequency shocks.

Those predictions of the model are consistent with recent empirical evidence. Parker (1999) and Souleles (1999) provide classic evidence on the response of consumption to predictable changes in income due to the tax code (the cap on social security taxes and tax refunds, respectively). The shocks studied in those papers essentially shift income over time, exactly as in our numerical example. The results above show that consumption in the model does in fact respond to such variation in income, and that it tracks predictable income variation strongly.

Kaplan and Violante (2014) review extensive evidence on the effectiveness of fiscal stimulus payments, finding that people tend to spend approximately 25 percent of these transitory payments in the quarter that they are received, even though the standard frictionless model would imply that they should spend a fraction near the level of the real interest rate (i.e. less than 1 percent per quarter). Moreover, these responses occur even among people with high incomes, who are less likely to be liquidity constrained (see also Kueng (2016)).

Kaplan and Violante explain the empirical evidence by arguing that when people hold illiquid assets, their consumption is excessively sensitive to transitory shocks because the benefit of smoothing is smaller than the cost of adjusting the stock of illiquid assets (e.g. housing). The intuition behind our results is similar to theirs (and also that of Cochrane (1989)) in that our results are also driven by the relatively small welfare benefit of smoothing transitory shocks. We differ in emphasizing the cost of learning about high-frequency dynamics, as opposed to assuming that saving is costly. Kaplan and Violante (2016) note that their model is consistent with the finding of Hsieh (2003) that consumption seems to respond relatively more to small than to large income shocks. That intuition is consistent with our argument that it is most natural for people to learn about shocks that have large effects on human wealth.

While the key source of variation for Kaplan and Violante (2014) is the size of shocks to income, for us it is their duration. Consumption mistakes should appear in response to short-duration shocks in our setting, and the empirical research finding violations of the permanent income hypothesis typically studies transitory income shocks.

Cochrane and Sbordone (1988) examine the joint relationship between aggregate consumption and output at long horizons and find that consumption helps forecast future output growth, but output does not help forecast consumption (nor do lags of consumption itself), implying that consumption growth is approximately white noise at long horizons. In other words, our model is consistent with the view that consumption growth may deviate from white noise and respond excessively to income in the short-term, but at longer horizons it is well described as white noise.

That implication requires aggregation, though, which is a nontrivial step. Since the consumption function in our model is linear, it will have desirable aggregation properties, but the exact details will depend on how income is driven by aggregate and idiosyncratic shocks at each frequency. Aggregate empirical results are thus not an ideal test of the model. The most direct test would be to measure the extent to which individual consumption growth is close to white noise over long horizons.

An alternative way to test the model, instead of examining consumption, would be to directly ask people what they are willing to pay for information. If they are at the optimum τ^*, then information is equally valuable at all frequencies. On the other hand, under the standard models of ambiguity aversion without endogenous information acquisition, people would value low-frequency information most highly and be willing to pay the most for it. That said, this paper faces the same problem as others in the information acquisition literature that there is no direct data on the type or quantity of information that people have (see Angeletos, Collard, and Hellas (2017) for a related discussion).

4.4 Relationship with the full-information optimal consumption rule

Our information-constrained agent uses a consumption rule that is suboptimal to the extent that $b^w(L)$ differs from $b(L)$. b^w is not chosen to directly generate a path for consumption that necessarily maximizes realized utility; rather, ambiguity aversion causes it to be chosen to maximize utility under a pessimistic probability measure. However, the agent's worst-case optimization problem is closely related to an optimization that approximates the correct consumption rule.

Remark 1 *A second-order expansion of the Kullback–Leibler (KL) divergence between the full-information rational expectations consumption process and that used by an agent with model f^w around the point $f^w = f$ is*

$$KL\left(f_{\Delta C}; f^w_{\Delta C}\right) \approx \frac{1}{4\pi}\int_0^{2\pi}\left(\left(Z(\omega)-1\right)^2 + 2\left(R^{-1}\frac{\alpha}{2}\left(1-R^{-1}\right)\right)^2 Z(\omega)^2\right)\left(f^w(\omega)-f(\omega)\right)^2 d\omega. \tag{37}$$

The KL divergence is a likelihood-based measure of the deviation between the two random processes (one interpretation is that it measures how likely one would be to reject the hypothesis that consumption is driven by one process after observing data generated by the other). Squared errors in the model f^w are weighted by a quadratic function of $Z(\omega)$. As long as R is close enough to 1, this weighting function is strictly maximized at $\omega = 0$, meaning that reducing the distance between f^w and f at low frequencies reduces the KL divergence the most. The optimization problem that our agent solves involves minimizing squared errors in $f^w(\omega)$ weighted by $\tau^*(\omega)$, and proposition 2 shows that $\tau^*(\omega) \propto Z(\omega)$. The estimations of the agent and of someone minimizing KL divergence both involve using the weights given by Z to put more emphasis on the precision of the estimate at low frequencies.

5 Alternative information cost specifications

5.1 Costs varying by frequency

The baseline case has equal information costs across frequencies – $\gamma(\omega) = 1$ – consistent with textbook time series analysis results. There is a common intuition, though, that information about low frequencies should be more costly to obtain. We are not aware of a formalization of that viewpoint, so we now provide a simple model that can generate it.

Recall that our model of information acquisition is that agents have a databse of income histories that they can query. If the people whose incomes are in the database (i.e. the friends, mentors, etc. who the agent learns from) exit the labor force at a constant rate over time, then the length of the histories in the database is naturally geometrically distributed; we would expect the number of histories to decrease as the length increases. Specifically, if people exit the labor force in each period with probability δ, then the fraction of people in the dataset who worked for at least k periods is $(1-\delta)^{k-1}$.

Loosely, in order to get information about cycles that least k periods, the agent needs to look at an income history that is at least k periods long (if we assume that the agent can use no frequency lower than the first periodogram ordinate). So to learn about a frequency ω, the agent must find a history lasting $2\pi/\omega$ periods. On average, that will require looking at $(1-\delta)^{1-2\pi/\omega}$ histories. For lower ω, there are fewer sufficiently long histories, and thus less available information.

We therefore set in this section

$$\gamma(\omega) = (1-\delta)^{1-2\pi/\omega}. \tag{38}$$

The benchmark results so far correspond to the case where $\delta = 0$. We now study how the results are affects by assuming $\delta > 0$. The only direct effect this has on the model is to change the optimal information policy. The general expressions obtained above for f^w ((24) and (29)) and consumption growth ((20), (32), and (34)) as functions of τ continue to hold. As in the base case, we calibrate the parameter θ in this section so that the response of consumption to a one-percent increase in the predictable consumption is 0.5 percent.

It is important to note that the justification for the cost function $\gamma(\omega)$ is only informal. There is not an extant treatment of why learning about low frequencies would be harder than learning about high frequencies. Appendix G provides an alternative and more formal model that generates the result that it is more difficult for agents to learn about low-frequency variation based on the idea that agents believe the spectrum may be particularly variable at the lowest frequencies (e.g. it might have a very narrow peak due to high persistence). In that specification, even though low frequencies are harder to learn about, the information policy remains $\tau^* \propto Z$, as in the baseline. In what follows with $\gamma(\omega) \neq 1$, on the other hand, that will not be true.

5.1.1 Optimal information policy

While we cannot solve the model in closed form when $\delta > 0$, there is a solution in the case with no smoothness prior ($\lambda = 0$):

Proposition 3 *The optimal information policy under the preferences (2) (and (27)) for arbitrary γ and for $\lambda = 0$ is*

$$\tau^{\gamma}(\omega_j) = \underbrace{\gamma(\omega)^{-1/2}}_{1/\textit{Frequency-specific cost}} \times \underbrace{\theta^{-1/2}}_{\textit{Shadow cost of info.}} \times \underbrace{\psi^{1/2}}_{\textit{Ambiguity aversion}} \times \underbrace{Z(\omega_j)}_{\textit{Utility weights}}. \tag{39}$$

The only difference between this result and our main specification is that τ^{γ} now is decreasing in the frequency-specific information costs – agents obtain less information about expensive than about inexpensive frequencies. At the same time, though, they continue to obtain information in proportion to the utility weights. What this result shows, then, is that even with frequency-specific information costs, agents always undo the effects of the evil agent's minimization by setting τ larger where Z is larger, all else equal. At the same time, though, τ^{γ} is obviously smaller when information is more costly, which can cause nature to distort the model at the costly frequencies.

We calibrate $\delta = 0.975$, corresponding to an exit probability of 2 percent, which would imply people have on average a 50-year working life if each period is a year. The top panels of figure 6 then plot the optimal information policies $\tau^* \propto Z$ and

$$\tau^{\gamma}(\omega) = \theta^{-1/2}\psi^{1/2}(1-\delta)^{1-2\pi/\omega} Z(\omega). \tag{40}$$

Both lines again peak at low frequencies, but whereas τ^* peaks at frequency zero, τ^{δ} peaks at a slightly interior frequency (the lines are normalized to have equal information costs). That peak comes at a frequency corresponding to cycles lasting approximately 160 years, though. So while the function γ in this case causes agents to learn less about the very lowest frequencies, the function Z is sufficiently strong that it still causes people to focus their attention on extremely low frequencies.

5.1.2 Behavior of the model agents use

It is more difficult to study the behavior of the model agents use analytically in this case since we are restricted to solving at $\lambda = 0$. The approximate results on the bias from above continue to hold,

$$E\left[f^w(\omega; x, \tau) - f \mid f\right] \approx \psi\tau(\omega)^{-1} Z(\omega) + \lambda\tau(\omega)^{-1} f''(\omega). \tag{41}$$

If we insert the solution for τ^{γ} from the case where $\lambda = 0$ – i.e. ignoring the effects of smoothing on the optimal τ – we have

$$\begin{aligned} E\left[f^w(\omega; x, \tau^{\gamma}) - f \mid f\right] \approx\ & (1-\delta)^{(1-2\pi/\omega)/2}\theta^{1/2}\psi^{1/2} \\ & +\lambda\theta^{1/2}\psi^{-1/2}(1-\delta)^{(1-2\pi/\omega)/2} Z(\omega) f''(\omega). \end{aligned} \tag{42}$$

The first term again represents the average bias. When the cost of information at low frequencies is larger, the model tends to be biased upward at low frequencies, inducing excessive extrapolation. So we now obtain excess extrapolation, but through a different mechanism than in past work on ambiguity in dynamic models (Hansen and Sargent (2010, 2016) and Bidder and Dew-Becker (2016)). Whereas excess extrapolation in those models appears because low-frequency shocks are most painful to agents, in (42) it appears because low frequencies are most difficult to learn about.

As discussed above, we compare the results under τ^* and τ^{γ} to the case where an equal fraction of the total information budget is allocated to each frequency. In this case, then, we set $\tau^{F\gamma} \propto (1-\delta)^{-(1-2\pi/\omega)/2}$.

The middle panels of figure 6 plot the worst-case spectra under various τ policies now also including τ^{γ}.[20] That policy leads to results between the benchmark τ^* and the equal-cost $\left(\tau^{F\gamma}\right)$

[20] For non-zero λ with γ varying across frequencies, $\tau^{\gamma}(\omega) \propto Z(\omega)\gamma(\omega)^{-1/2}$ is not technically the optimal policy – it must be solved for numerically. We focus on the analytic case for the sake of simplicity. Furthermore, the calibration in figure 6 is set up so that the total precision under τ^* is the same as that under τ^{γ} – they differ only in how that precision is allocated across frequencies.

policy. At the very lowest frequencies, the τ^γ model does not match the true spectrum as well as τ^*, but it still does much better than τ^F. At the middle frequency peak and at higher frequencies, on the other hand, the policy τ^γ does a better job of matching the log spectrum than τ^* but still worse than τ^F.

5.1.3 The response of consumption to shocks

The bottom two panels of figure 6 plot the behavior of consumption under the various models. The left-hand panel plots the spectra. The spectral density of consumption growth under the τ^γ policy again lies between those of the τ^* policy and the statistical benchmark. While it has a hump at middle frequencies, it is somewhat smaller than that for the optimal policy. The bottom-right panel plots the impulse responses of consumption. The τ^γ policy is nearly as effective as the τ^* policy at replicating the optimal initial response of consumption, and it shares some of the excess sensitivity to the short-run variation in income.

The table below calculates the coefficients from regressions of consumption growth on the predictable and unpredictable parts of income, as we did for the benchmark results.

Information policy	Predictable income	Unpredictable income
τ^γ	0.50	1.00
$\tau^{F\gamma}$	0.29	0.75
Full-info. optimum	0	1.11

Table 2. Coefficients from regressions of consumption growth on income growth with frequency-specific information costs

The optimal information policy again generates excess sensitivity to income shocks. In this case, compared to the baseline, the response of consumption to unpredictable income shocks is slightly smaller, but it is still economically very close to the full-information optimum. The equal-cost information policy, $\tau^{F\gamma}$, as with τ^F above, again generates much smaller responses to both predictable and unpredictable shocks to income.

Overall, when low frequencies are more costly to learn about, the main results are weakened, but by a quantitatively small amount. Agents continue to allocate the most attention to low frequencies, just not to the *very* lowest – the peak is at an interior frequency, but one corresponding to cycles lasting a century or more. Agents under both the baseline and the case with frequency-specific information costs make relatively larger mistakes in their models at middle than low frequencies, they respond suboptimally strongly to predictable variation income, and they respond nearly optimally to the unpredictable component. The model therefore generates similar overall behavior in beliefs and consumption under both information cost specifications.

5.2 Entropy cost for information

The literature on information acquistion in some cases uses linear costs on precision and in others uses an entropy constraint. We now show that when the linear cost of precision is replaced with an entropy cost, the results become more extreme than in our benchmark case.

In terms of entropy, the total information flow to the agents in the model can be measured based on the difference between the prior variance that they have at each frequency and the posterior. In our case, since we do not write down a fully specified prior, the information flow cannot be calculated exactly. One way to interpret our agents' prior information, though, would be as a limit in which the prior variance becomes infinite. In such a case, the frequency-by-frequency information flow approaches $\sum_j \log\left(\tau\left(\omega_j\right)\right) d\omega$.

As with the $\delta > 0$ case above, we are not able to solve the model in general with entropy costs. However, assuming $\lambda = 0$ again (and with $\gamma = 1$), one may show that the optimal information policy takes the form

$$\tau^{entropy}\left(\omega_j\right) = \theta^{-1}\psi Z\left(\omega_j\right)^2 \tag{43}$$

So whereas in the benchmark we had $\tau^* \propto Z$, with the entropy constraint we find $\tau^{entropy} \propto Z^2$, thus making the focus on the lowest frequencies even stronger and further emphasizing our main results. Intuitively, this result appears because the entropy cost is logarithmic in the precision of the signals, rather than linear, making particularly precise signals relatively less costly in this case than in the benchmark.

6 Conclusion

This paper studies how people can direct their attention to different features of a model. We consider a nonparametric class of income processes and show precisely how agents optimally allocate attention to the behavior of income at different frequencies. The utility maximizing policy is to pay the most attention to the behavior of income at very low frequencies, and use a relatively simple and inaccurate model at high frequencies.

While there is extensive past work on learning, the innovation of this paper is to provide an exactly solvable framework for studying how learning can be applied to different aspects of a model of the world, as opposed to learning about state variables. The theory can be used to describe what people pay attention to, what aspects of the world they try to model accurately and what they use coarser approximations for, and the set of mistakes that people should be expected to make.

We show that optimal learning implies people are most likely to make mistakes at high frequencies, as those are the aspects of the income process least important for utility. Consistent with empirical evidence, the model implies that consumption tends to track transitory fluctuations in income in the short-run, but at lower frequencies consumption growth is close to white noise (which it would be under the full-information optimal policy). In other words, the consumption mistakes that the empirical literature has documented are consistent with optimal learning.

References

Abel, Andrew B., Janice C. Eberly, and Stavros Panageas, "Optimal Inattention to the Stock Market," *The American Economic Review*, 2007, *97* (2), 244–249.

___ , ___ , **and** ___ , "Optimal Inattention to the Stock Market with Information Costs and Transactions Costs," *Econometrica*, 2013, *81* (4), 1455–1481.

Akaike, Hirotugu, "Smoothness Priors and the Distributed Lag Estimator.," Technical Report, DTIC Document 1979.

Angeletos, George-Marios, Fabrice Collard, and Harris Dellas, "Quantifying Confidence," 2017. Working paper.

Bansal, Ravi and Ivan Shaliastovich, "Confidence Risk and Asset Prices," *The American Economic Review*, 2010, *100* (2), 537–541.

Barron, John M. and Jinlan Ni, "Endogenous Asymmetric Information and International Equity Home Bias: The Effects of Portfolio Size and Information Costs," *Journal of International Money and Finance*, 2008, *27* (4), 617–635.

Berk, Kenneth N., "Consistent Autoregressive Spectral Estimates," *The Annals of Statistics*, 1974, *2(3)*, 489–502.

Bhandari, Anmol, Jaroslav Borovicka, and Paul Ho, "Survey data and subjective beliefs in business cycle models," 2017. Working paper.

Bianchi, Francesco, Cosmin Ilut, and Martin Schneider, "Uncertainty shocks, asset supply and pricing over the business cycle," 2017. Working paper.

Bidder, Rhys and Ian Dew-Becker, "Long-Run Risk is the Worst-Case Scenario," *The American Economic Review*, September 2016, *106* (9), 2494–2527.

Bidder, Rhys M and Matthew E Smith, "Robust animal spirits," *Journal of Monetary Economics*, 2012, *59* (8), 738–750.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, "Diagnostic Expectations and Credit Cycles," *Working Paper*, 2016.

Brillinger, David R., *Time Series: Data Analysis and Theory*, McGraw Hill, 1981.

Brockwell, Peter J. and Richard A. Davis, *Time Series: Theory and Methods*, Springer, 1991.

Caballero, Ricardo J., "Consumption puzzles and precautionary savings," *Journal of Monetary Economics*, 1990, *25* (1), 113–136.

Cochrane, John H., "The Sensitivity of Tests of the Intertemporal Allocation of Consumption to Near-Rational Alternatives," *The American Economic Review*, 1989, *79* (3), 319–337.

___ **and Argia Sbordone**, "Multivariate Estimates of the Permanent Components of GNP and Stock Prices," *Journal of Economic Dynamics and Control*, 1988, *12(2–3)*, 255–296.

Collin-Dufresne, Pierre, Michael Johannes, and Lars A Lochstoer, "Parameter Learning in General Equilibrium: The Asset Pricing Implications," *The American Economic Review*, 2016, *106* (3), 664–698.

Dahlhaus, R., "On the Kullback-Leibler information divergence of locally stationary processes," *Stochastic Processes and their Applications*, 1996, *62* (1), 139–168.

Dew-Becker, Ian, "How Risky is Consumption in the Long-run? Benchmark Estimates From a Robust Estimator," *Review of Financial Studies*, 2017, *30* (2), 631–666.

___ **and Stefano Giglio**, "Asset pricing in the frequency domain: theory and empirics," *Review of Financial Studies*, 2016, *29* (8), 2029–2068.

Drechsler, Itamar, "Uncertainty, Time-Varying Fear, and Asset Prices," *The Journal of Finance*, 2013, *68* (5), 1843–1889.

Eichenbaum, Martin, "Comment on "Natural Expectations, Macroeconomic Dynamics, and Asset Pricing"," in "NBER Macroeconomics Annual 2011, Volume 26," University of Chicago Press, 2011, pp. 49–60.

Epstein, Larry G. and Shaolin Ji, "Optimal Learning and EllsbergŠs Urns," 2017. Working paper.

___ **and Stanley E. Zin**, "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: An Empirical Analysis," *The Journal of Political Economy*, 1991, *99(2)*, 263–286.

Friedman, Milton, *A Theory of the Consumption Function*, Princeton University Press, 1957.

Fuster, Andreas, Benjamin Hebert, and David Laibson, "Natural Expectations, Macroeconomic Dynamics, and Asset Pricing," *NBER Macroeconomics Annual*, 2011, *26* (1), 1–48.

Gabaix, Xavier, "Behavioral Macroeconomics Via Sparse Dynamic Programming." Working paper.

Gersch, Will and Genshiro Kitagawa, "Smoothness Priors Transfer Function Estimation," *Automatica*, 1989, *25* (4), 603–608.

Hall, Robert E., "Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence," *Journal of Political Economy*, 1988, *86* (6), 971–987.

Hamilton, James Douglas, *Time series analysis*, Princeton university press Princeton, 1994.

Hansen, Lars Peter and Thomas J. Sargent, "Formulating and Estimating Dynamic Linear Rational Expectations Models," *Journal of Economic Dynamics and Control*, 1980, *2*, 7–46.

___ **and** ___ , "A note on Wiener-Kolmogorov Prediction Formulas for Rational Expectations Models," *Economics Letters*, 1981, *8* (3), 255–260.

___ **and** ___ , "Fragile Beliefs and the Price of Uncertainty," *Quantitative Economics*, 2010, *1(1)*, 129–162.

___ **and** ___ , "Sets of Models and Prices of Uncertainty," 2015. Working paper.

___ **, Thomas J Sargent, and Thomas D Tallarini**, "Robust permanent income and pricing," *Review of Economic Studies*, 1999, *66* (4), 873–907.

Hashimzade, Nigar and Timothy J. Vogelsang, "Fixed-b Asymptotic Approximation of the Sampling Behavior of Nonparametric Spectral Density Estimators," *Journal of Time Series Analysis*, 2008, *29(1)*, 142–162.

Hsieh, Chang-Tai, "Do Consumers React to Anticipated Income Changes? Evidence from the Alaska Permanent Fund," *The American Economic Review*, 2003, *93* (1), 397–405.

Ilut, Cosmin L and Martin Schneider, "Ambiguous business cycles," *The American Economic Review*, 2014, *104* (8), 2368–2399.

Jappelli, Tullio and Luigi Pistaferri, "The Consumption Response to Income Changes," *Annual Review of Economics*, 2010, *2*, 479–506.

Ju, Nengjiu and Jianjun Miao, "Ambiguity, Learning, and Asset Returns," *Econometrica*, 2012, *80(2)*, 559–591.

Kacperczyk, Marcin, Stijn van Nieuwerburgh, and Laura Veldkamp, "A Rational Theory of Mutual Funds' Attention Allocation," *Econometrica*, 2016, *84* (2), 571–626.

Kaplan, Greg and Giovanni L Violante, "A Model of the Consumption Response to Fiscal Stimulus Payments," *Econometrica*, 2014, *82* (4), 1199–1239.

Kitagawa, Genshiro and Will Gersch, "A Smoothness Priors–State Space Modeling of Time Series with Trend and Seasonality," *Journal of the American Statistical Association*, 1984, *79* (386), 378–389.

___ **and** ___ , "A smoothness priors long AR model method for spectral estimation," *IEEE transactions on automatic control*, 1985, *30* (1), 57–65.

—— **and** —— , *Smoothness Priors Analysis of Time Series*, Springer Science & Business Media, 1996.

Kueng, Lorenz, "Explaining Consumption Excess Sensitivity with Near-Rationality: Evidence from Large Predetermined Payments," 2016. Working paper.

Luo, Yulei, "Consumption Dynamics under Information Processing Constraints," *Review of Economic Dynamics*, 2008, *11* (2), 366–385.

Parker, Jonathan A, "The Reaction of Household Consumption to Predictable Changes in Social Security Taxes," *The American Economic Review*, 1999, *89* (4), 959–973.

Peng, Lin and Wei Xiong, "Investor Attention, Overconfidence and Category Learning," *Journal of Financial Economics*, 2006, *80* (3), 563–602.

Priestley, M.B., *Spectral Analysis and Time Series*, Elsevier, 1981.

Shao, Xiaofeng and Wei Biao Wu, "Asymptotic Spectral Theory for Nonlinear Time Series," *The Annals of Statistics*, 2007, *35* (4), 1773–1801.

Shiller, Robert J., "A Distributed Lag Estimator Derived from Smoothness Priors," *Econometrica*, 1973, pp. 775–788.

Sims, Christopher A, "Implications of Rational Inattention," *Journal of Monetary Economics*, 2003, *50* (3), 665–690.

Souleles, Nicholas S., "The Response of Household Consumption to Income Tax Refunds," *The American Economic Review*, 1999, *89* (4), 947–958.

van Nieuwerburgh, Stijn and Laura Veldkamp, "Information acquisition and under-diversification," *The Review of Economic Studies*, 2010, *77* (2), 779–805.

Veldkamp, Laura L., "Information Markets and the Comovement of Asset Prices," *Review of Economic Studies*, 2006, *73* (3), 823–845.

—— , *Information Choice in Macroeconomics and Finance*, Princeton University Press, 2011.

Wahba, Grace, "Automatic Smoothing of the Log Periodogram," *Journal of the American Statistical Association*, 1980, *75* (369), 122–132.

Wang, Neng, "Precautionary Saving and Partially Observed Income," *Journal of Monetary Economics*, 2004, *51* (8), 1645–1681.

—— , "Optimal consumption and asset allocation with unknown income growth," *Journal of Monetary Economics*, 2009, *56* (4), 524–534.

Figure 1: Weighting function $Z(\omega)$ and its multiplicative inverse

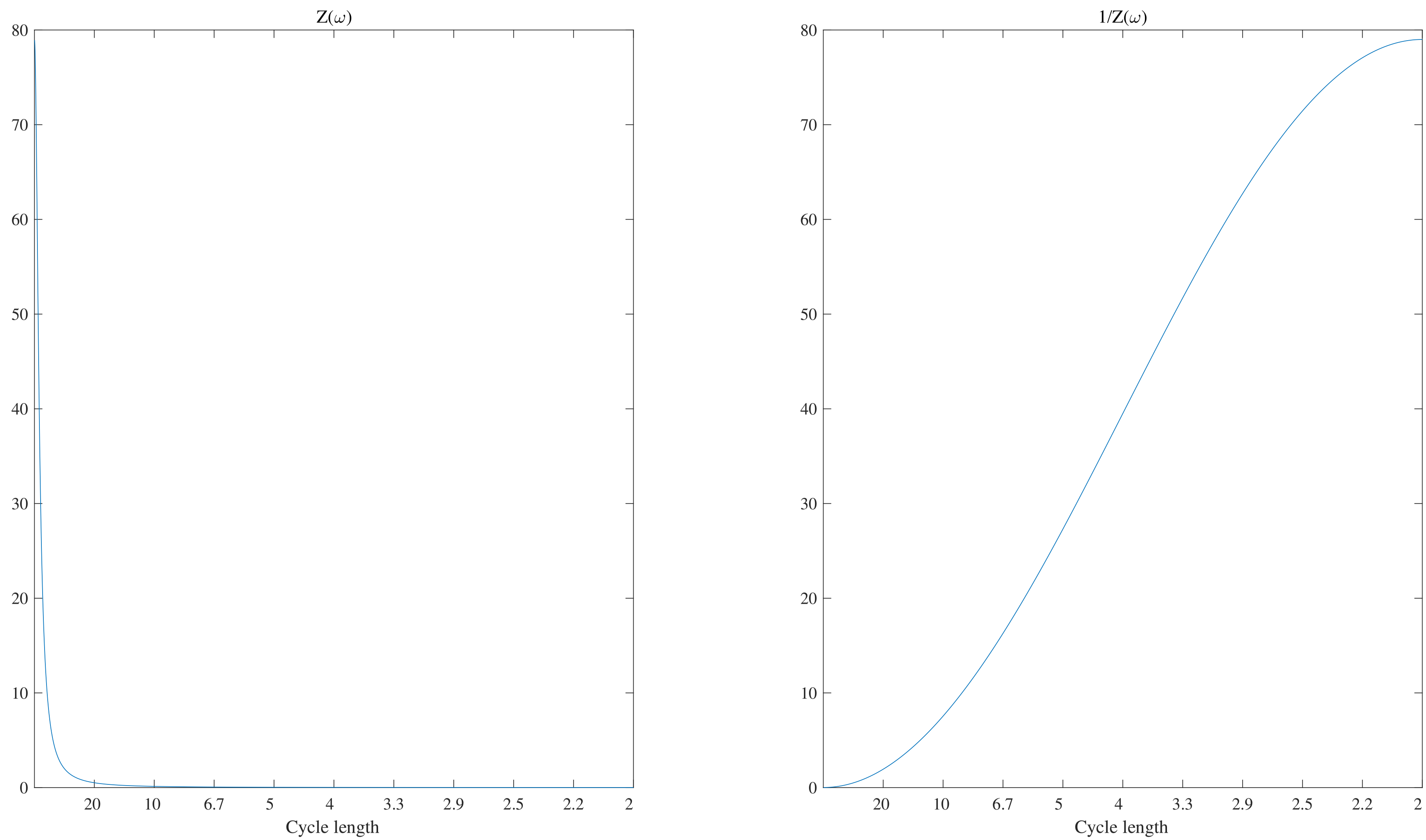
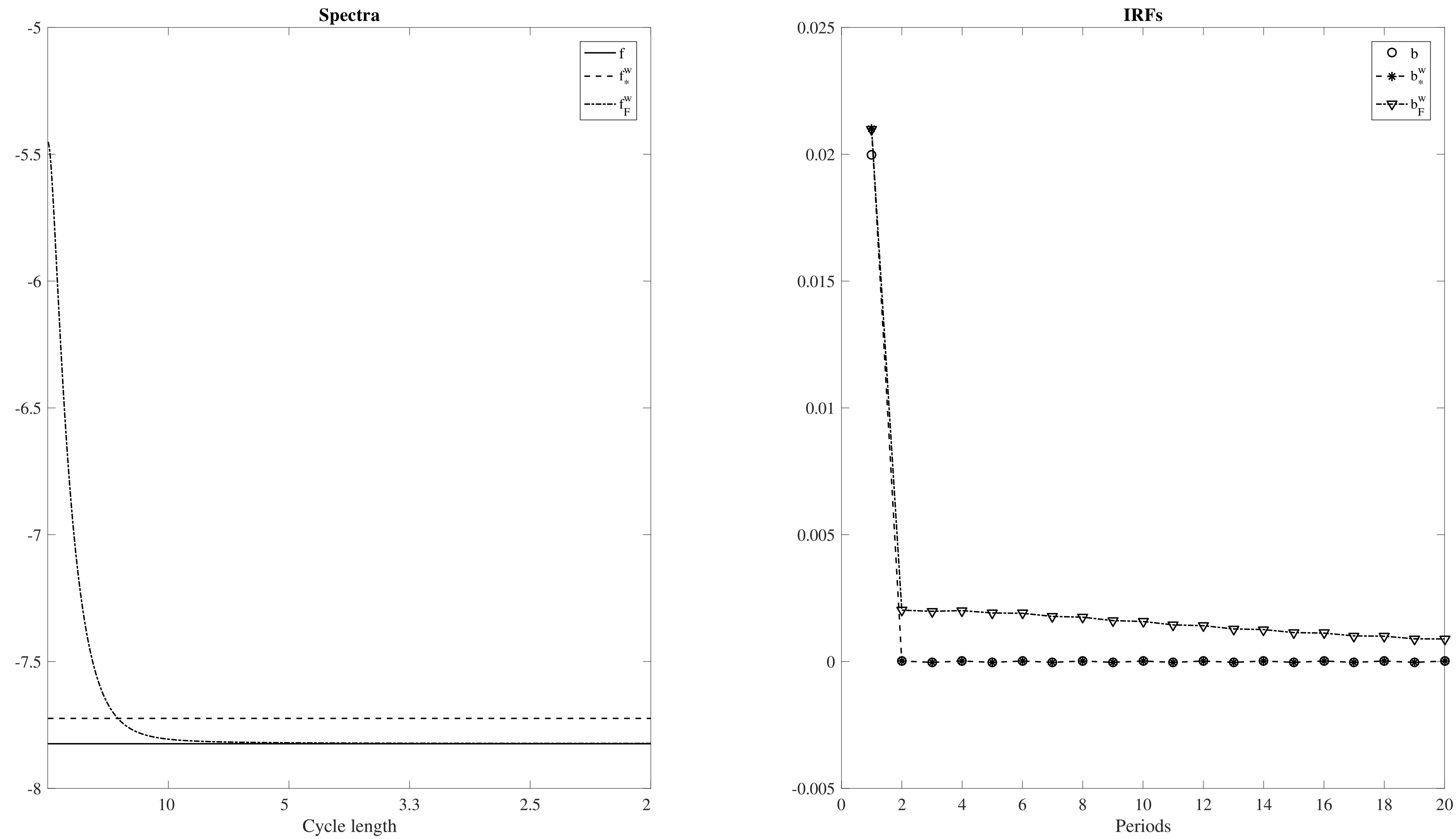
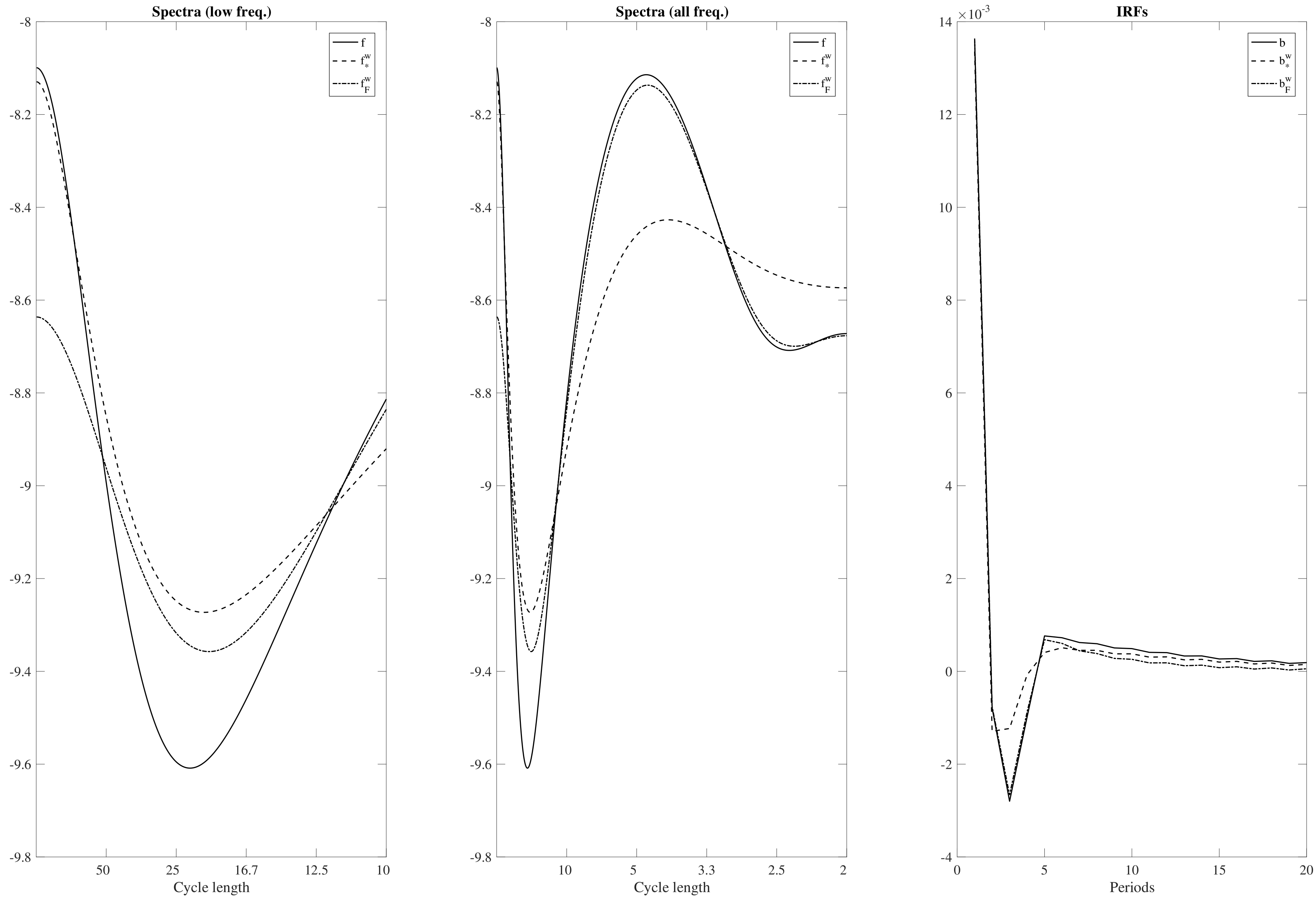


Figure 2: Average estimated log spectra and IRFs for white-noise income



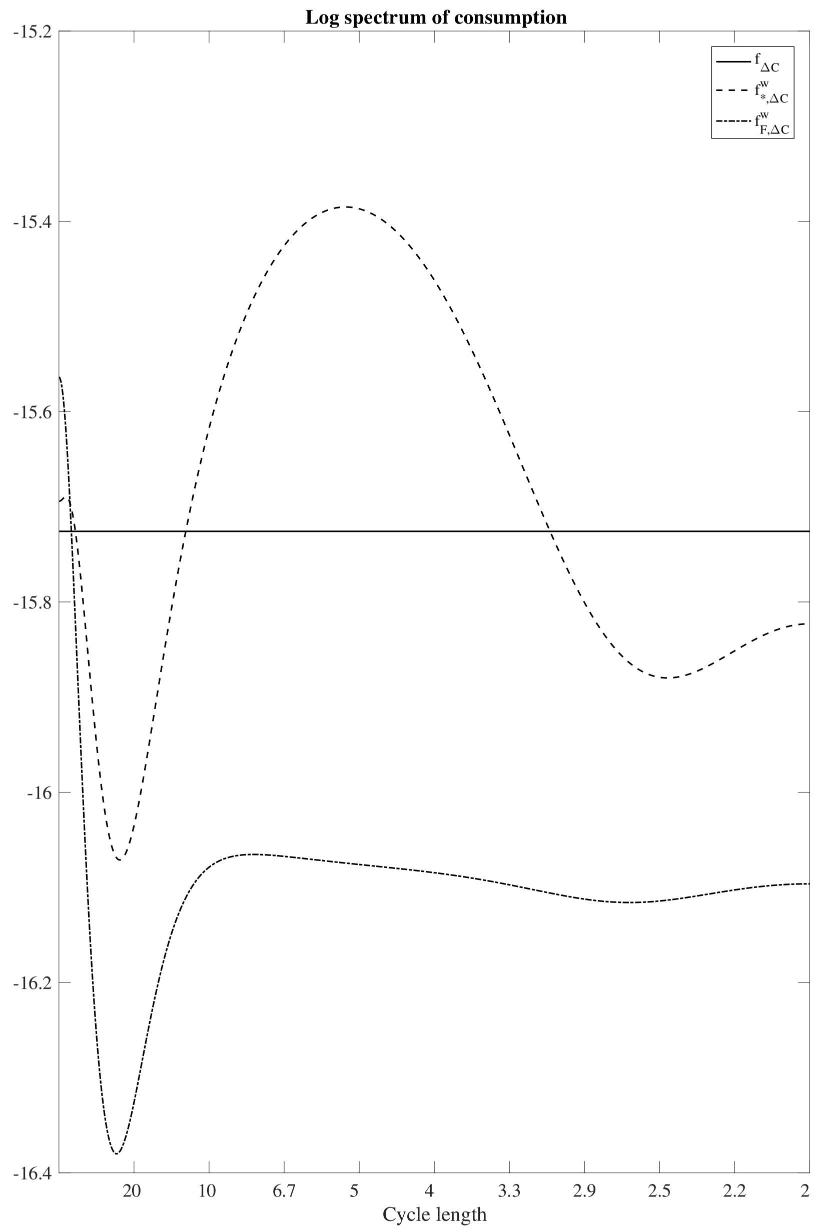
Notes: In the left-hand panel, f is the true log spectrum of income (flat), the line for f_*^w is the average worst-case log spectrum under the optimal information policy, and the line for f_F^w is the average worse-case log spectrum under the statistical benchmark that yields equally precise signals at all frequencies. The right-hand panel plots the impulse response functions (Wold moving average representations) for income corresponding to the three log spectra.

Figure 3: Average estimated log spectra and IRFs with transitory and persistent components in income



Notes: The middle and left-hand panel correspond to the left-hand panel in figure 2, except for a different value for the true spectrum, f. The right-hand panel here corresponds to the right-hand panel in figure 2, but for this alternative example with an income process that has both persistent and transitory components.

Figure 4: Behavior of consumption with permanent and transitory income fluctuations



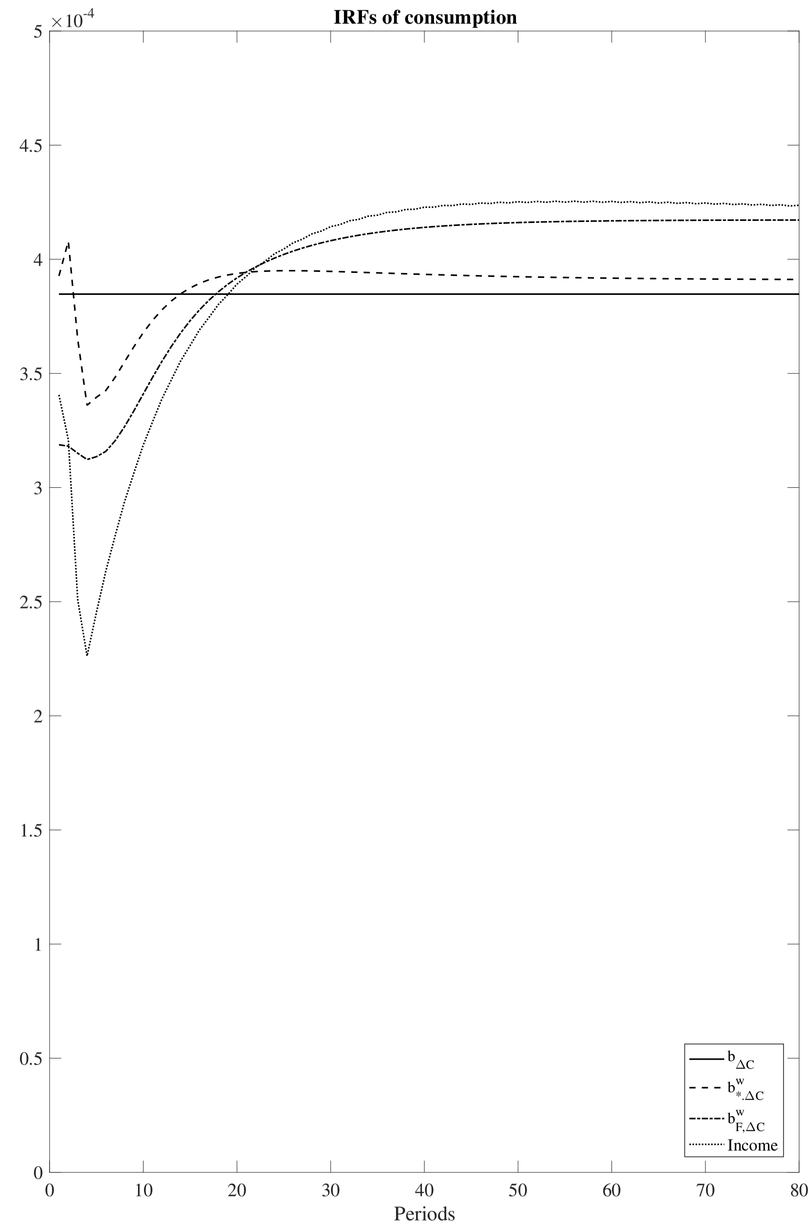
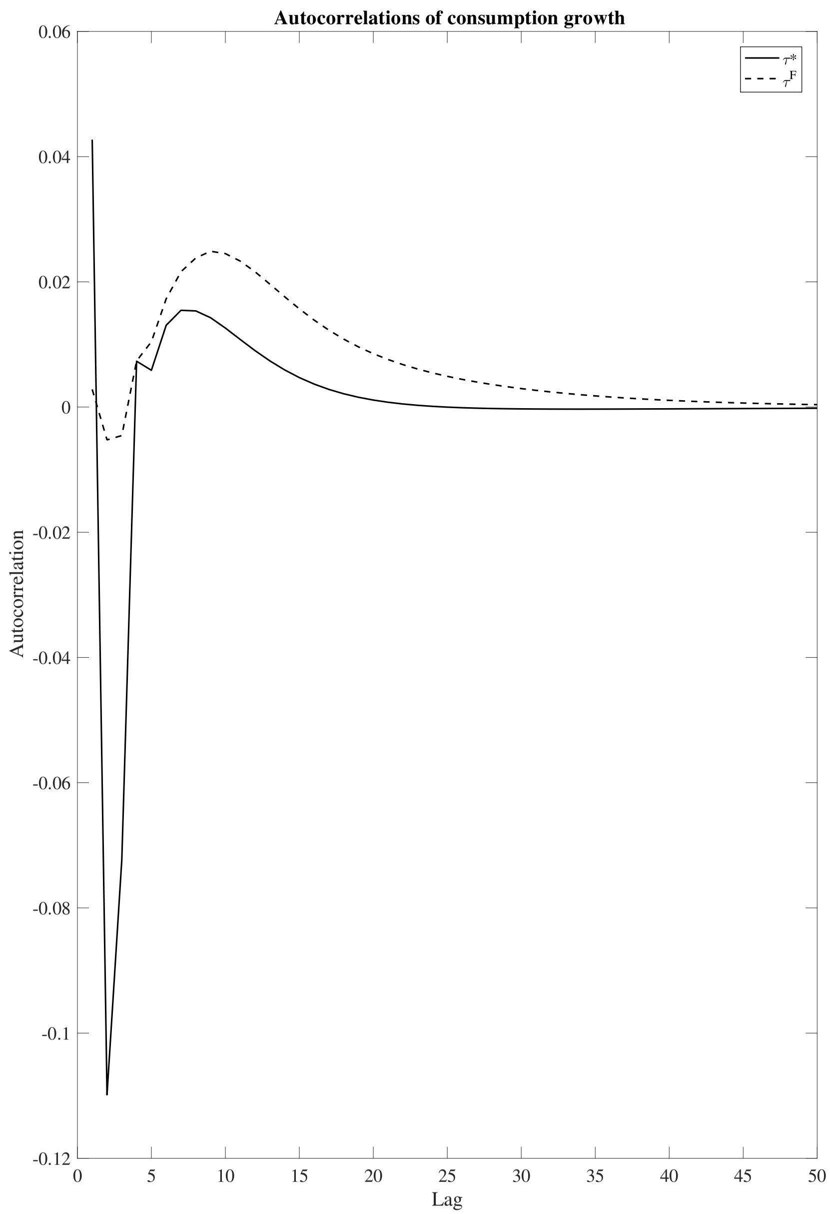
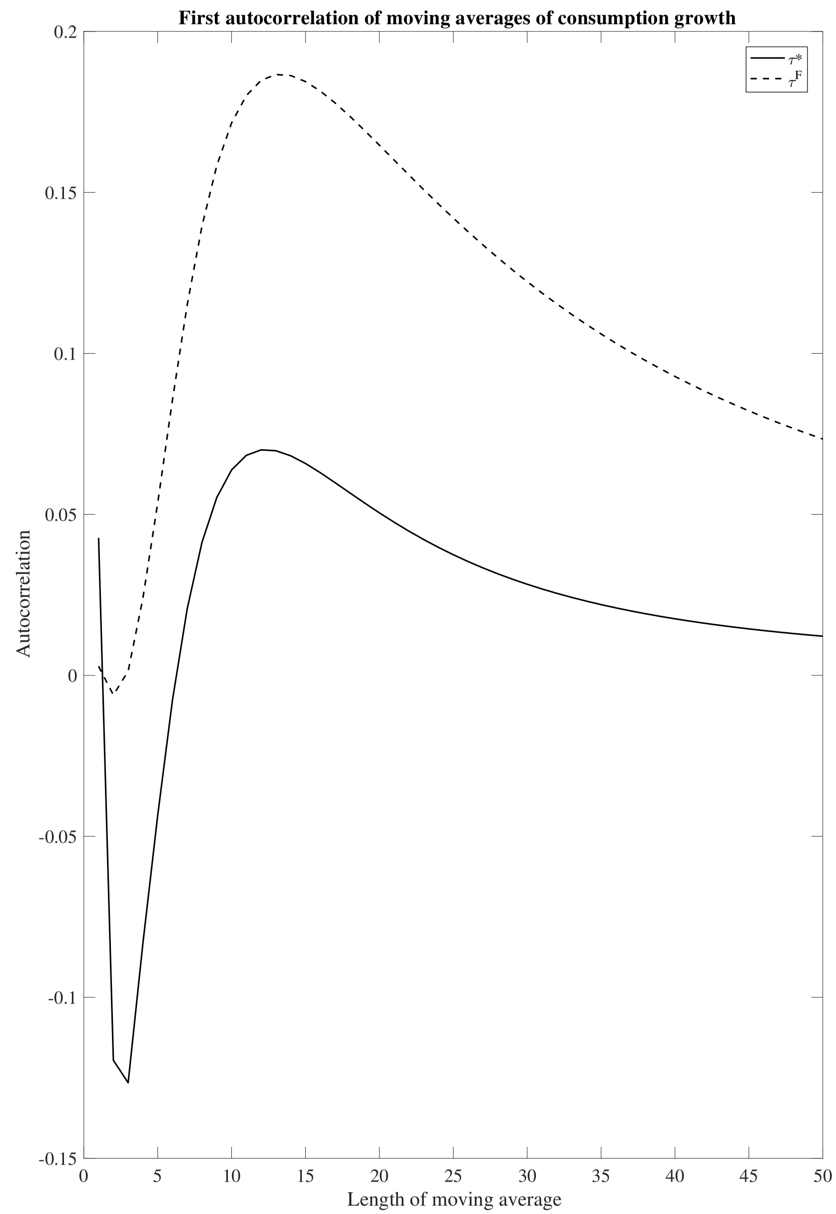


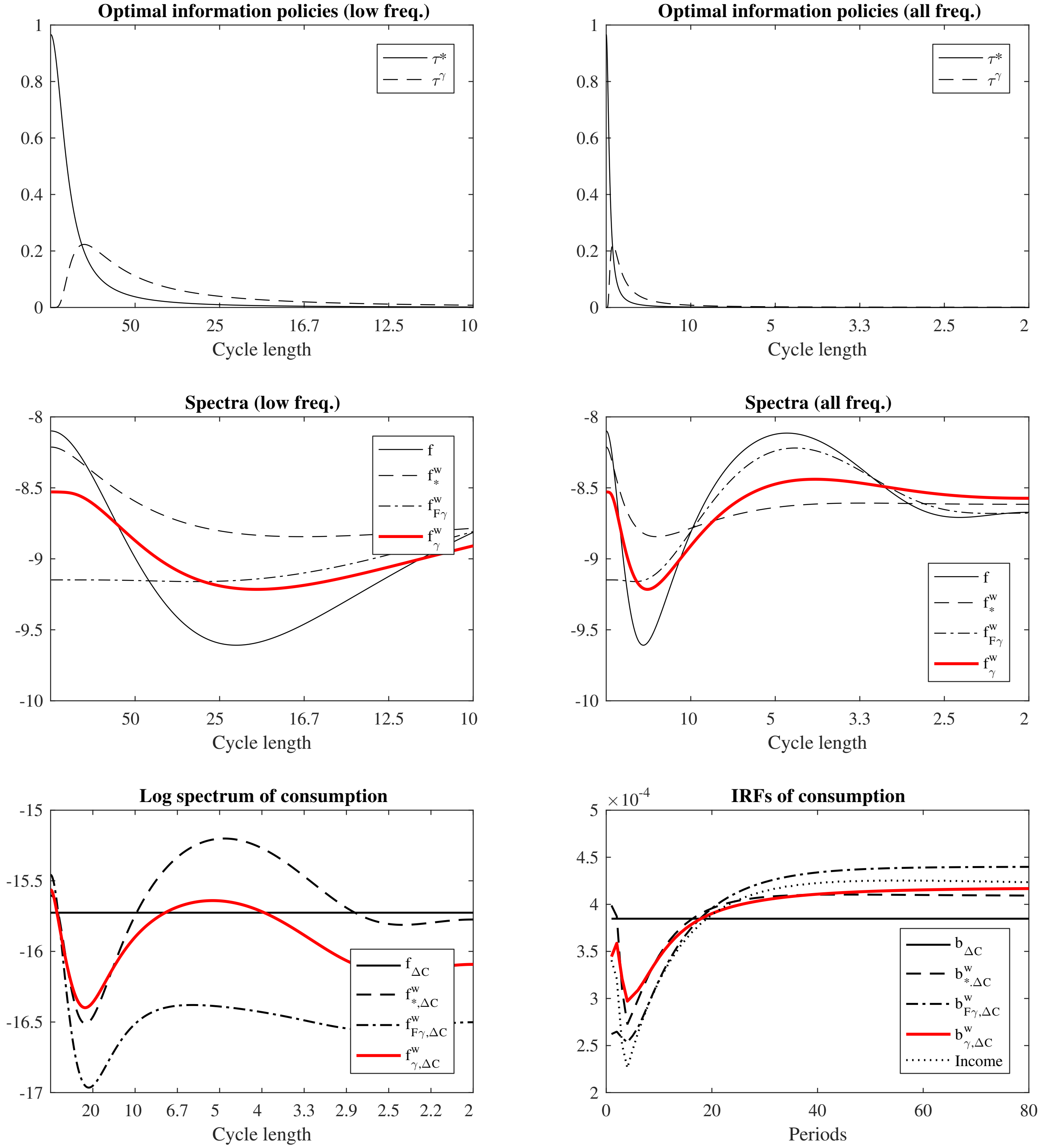
Figure 5: Persistence of consumption growth with transitory and persistent components in income





Notes: The left-hand panel plots the autocorrelations of consumption growth, $corr(\Delta C_t, \Delta C_{t-1})$. The right-hand panel plots the autocorrelation of moving averages, $corr\left(\sum_{j=0}^{n-1} \Delta C_{t+j}, \sum_{j=0}^{n-1} \Delta C_{t-n+j}\right)$, where n varies along the x-axis.

Figure 6: Effects of information cost varying across frequencies



Notes: The middle panels of this figure replicate the left and middle panels in figure 3, but using the optimal information policy when information costs vary across frequencies, denoted τ^{δ}. The bottom panels replicate figure 4. f_{δ}^{w} denotes the average worst-case log spectrum under that information policy, whereas f_{*}^{w} continues to denote the average worst-case spectrum under the optimal policy in the case where information costs are constant across frequencies. $f_{F\delta}^{w}$ is the average worst-case spectrum under the policy that allocates equal attention cost to each frequency. $f_{\delta,\Delta C}^{w}$ and $f_{F\delta,\Delta C}^{w}$ are the spectra for consumption growth associated with those worst-case models.

A Proof of lemma 2

From Dew-Becker (2016), the optimal consumption rule is

$$C_t = (R-1) W_{t-1} + -\frac{\alpha}{2} R^{-1} \left(1 - R^{-1}\right) \hat{b} \left(R^{-1}\right)^2 + z(L) \hat{\varepsilon}_t - \alpha^{-1} \frac{\log \beta R}{R-1} \tag{1}$$

for a lag polynomial $z(L)$. Dew-Becker (2016) also shows that

$$E_t \left[-\alpha^{-1} \sum_{j=0}^{\infty} \beta^j \exp\left(-\alpha C_{t+j}\right) \right] = \frac{-\alpha^{-1}}{1-R} \exp\left(-\alpha C_t\right) \tag{2}$$

(note that the probability measure for the expectation operator here is arbitrary) which implies

$$\begin{aligned} -\alpha^{-1} \log E_t \left[(1-\beta) \sum_{j=0}^{\infty} \beta^j \exp\left(-\alpha C_{t+j}\right) \right] \quad &= \quad -\alpha^{-1} \log \frac{(1-\beta)}{1-R} + (R-1) W_{t-1} \\ & \quad -\frac{\alpha}{2} R^{-1} \left(1 - R^{-1}\right) \hat{b} \left(R^{-1}\right)^2 + z(L) \hat{\varepsilon}_t - \alpha^{-1} \frac{\log \beta R}{R-1}. \end{aligned} \tag{3}$$

The result in the text then immediately follows.

B Finding the worst-case spectrum (proposition 1)

Nature chooses $\left\{ \hat{f}(\omega_j) \right\}$ to solve

$$\{f^w(\omega_j)\} = \arg \max_{\{\hat{f}(\omega_j)\}} \sum_{j=1}^{n} Z(\omega_j) \hat{f}(\omega_j) d\omega - \frac{\psi^{-1}}{2} \sum_{j=1}^{n} \left(x(\omega_j) - \hat{f}(\omega_j) \right)^2 \tau(\omega_j) d\omega - \frac{\psi^{-1}}{2} \lambda \sum_{j=2}^{n} \left(\frac{\hat{f}(\omega_j) - \hat{f}(\omega_{j-1})}{d\omega} \right)^2 d\omega. \tag{4}$$

The first-order conditions for interior points $(1 < j < n)$ are

$$0 = Z(\omega_j) + \psi^{-1} \left(x(\omega_j) - f^w(\omega_j) \right) \tau(\omega) + \frac{\psi^{-1} \lambda}{d\omega} \left(\left(\frac{f^w(\omega_{j+1}) - f^w(\omega_j)}{d\omega} \right) - \left(\frac{f^w(\omega_j) - f^w(\omega_{j-1})}{d\omega} \right) \right).$$

At the boundaries they are

$$0 = Z(\omega_1) + \psi^{-1} \left(x(\omega_1) - f^w(\omega_1) \right) \tau(\omega_1) + \psi^{-1} \lambda \frac{f^w(\omega_2) - f^w(\omega_1)}{d\omega^2} \tag{5}$$

$$0 = Z(\omega_n) + \psi^{-1} \left(x(\omega_n) - f^w(\omega_n) \right) \tau(\omega_n) - \psi^{-1} \lambda \frac{f^w(\omega_n) - f^w(\omega_{n-1})}{d\omega^2}. \tag{6}$$

We define here vectors containing the various objects at the frequencies ω_j using variables with no subscript. For example, $\tau \equiv \left[\tau(\omega_1), \tau(\omega_2), ..., \tau(\omega_n)\right]'$. We can then write the first-order conditions as

$$0 = Z + \psi^{-1} diag(\tau)(x - f^w) + \psi^{-1} \lambda D f^w, \tag{7}$$

where $diag(\tau)$ is a matrix with τ on the diagonal and zero elsewhere and D is a differencing matrix:

$$D \equiv \begin{bmatrix} -1 & 1 & 0 & 0 & \cdots & 0 \\ 1 & -2 & 1 & 0 & & \\ 0 & 1 & -2 & 1 & & \vdots \\ \vdots & & & \ddots & & 0 \\ & & 0 & 1 & -2 & 1 \\ 0 & \cdots & 0 & 0 & 1 & -1 \end{bmatrix} d\omega^{-2}. \tag{8}$$

The second-order condition is that

$$-diag(\tau) + \lambda D \tag{9}$$

is negative definite, i.e. that all of its eigenvalues are negative.

The solution to nature's optimization problem is then obtained by directly solving (7):

$$\begin{aligned} f^w &= (diag(\tau) - \lambda D)^{-1} (\psi Z + diag(\tau) x) & (10)\\ &= \left(I - \lambda diag\left(\tau^{-1}\right) D\right)^{-1} \left(\psi diag\left(\tau^{-1}\right) Z + x\right), & (11) \end{aligned}$$

where τ^{-1} here is an elementwise inverse of the vector τ. Since this is a linear problem, the solution is unique as long as the matrix inverse exists.

C Proposition 2

Consider a total derivative of (7) with respect to τ' at the point $x = \bar{f}$:

$$0 = \psi^{-1} diag\left(\bar{f} - f^w\right) - \psi^{-1} diag(\tau) \frac{dE[f^w]}{d\tau'} + \psi^{-1} \lambda D \frac{dE[f^w]}{d\tau'}. \tag{12}$$

We can then solve for $\frac{dE[f^w]}{d\tau'}$:

$$\frac{dE[f^w]}{d\tau'} = (\lambda D - diag(\tau'))^{-1} diag\left(E[f^w] - \bar{f}\right) \tag{13}$$

Now the objective is to minimize

$$\begin{aligned} \{\tau^*(\omega_j)\} &= \arg \min_{\{\tau(\omega_j)\}} \log E\left[b^w \left(R^{-1}\right)^2\right] + \theta \sum_j \tau(\omega_j) \, d\omega & (14)\\ &= \arg \min_{\{\tau(\omega_j)\}} Z' E[f^w] \, d\omega + \theta \sum_j \tau(\omega_j) \, d\omega. & (15) \end{aligned}$$

The first-order condition for that problem is

$$0 = Z' \frac{dE[f^w]}{d\tau'} + \theta 1_{1\times n}, \tag{16}$$

where $1_{1\times n}$ is a $1 \times n$ vector of ones. Inserting the formula for $\frac{df^w}{d\tau'}$ yields

$$0 = Z' (\lambda D - diag(\tau^{*\prime}))^{-1} diag\left(E[f^w] - \bar{f}\right) + \theta 1_{1\times n} \tag{17}$$

$$Z' = -\theta 1_{1\times n} diag\left(E[f^w] - \bar{f}\right)^{-1} (\lambda D - diag(\tau^{*\prime})). \tag{18}$$

Now we conjecture that $E[f^w] - \bar{f}$ is equal to a constant c multiplied by a column of ones. We then have

$$\begin{aligned} Z' &= -\theta c^{-1} 1_{1\times n} (\lambda D - diag(\tau^{*\prime})) & (19)\\ &= \theta c^{-1} \tau^{*\prime}, & (20) \end{aligned}$$

where the second line uses the fact that $1_{1\times n} D = 0_{1\times n}$ since the columns of D sum to zero.

In order to confirm that result, we must now show that when $Z = \theta c^{-1} \tau^*$, $E[f^w] - \bar{f} = c 1_{n\times 1}$. Inserting $Z = \theta c^{-1} \tau^*$ into (7) yields

$$0 = \theta c^{-1} \tau^* + \psi^{-1} diag(\tau^*) \left(\bar{f} - E[f^w]\right) + \psi^{-1} \lambda D E[f^w]. \tag{21}$$

In order for it to be the case that $E[f^w] - \bar{f} = c 1_{n\times 1}$, we must have

$$\begin{aligned} 0 &= \theta c^{-1} \tau^* - \psi^{-1} diag(\tau^*) 1_{n\times 1} c + \psi^{-1} \lambda D 1_{n\times 1} \left(\bar{f} + c\right) & (22)\\ &= \theta c^{-1} \tau^* - \psi^{-1} \tau^* c. & (23) \end{aligned}$$

where the second line uses the fact that $D1_{n\times 1} = 0_{n\times 1}$. This is solved by

$$\sqrt{\theta\psi} = c \tag{24}$$

$$\begin{aligned} Z &= \theta c^{-1} \tau^* && (25) \\ \tau^* &= (\theta/\psi)^{-1/2} Z. && (26) \end{aligned}$$

We can then plug the value of τ^* into the equation for $E[f^w]$:

$$\begin{aligned} E[f^w] &= \left(I - \lambda diag\left(\tau^{*-1}\right) D\right)^{-1} \left(\psi diag\left((\theta/\psi)^{1/2} Z^{-1}\right) Z + x\right) && (27) \\ &= \left(I - \lambda diag\left(\tau^{*-1}\right) D\right)^{-1} \left(1_{n\times 1} \theta^{1/2} \psi^{1/2} + x\right), && (28) \end{aligned}$$

where, as with τ^{-1}, Z^{-1} is an elementwise inverse of the vector Z. It follows that

$$\begin{aligned} E[f^w] &= \left(I - \lambda diag\left(\tau^{*-1}\right) D\right)^{-1} \left(1_{n\times 1} \theta^{1/2} \psi^{1/2} + \bar{f}\right) && (29) \\ &= 1_{n\times 1} \theta^{1/2} \psi^{1/2} + \bar{f}, && (30) \end{aligned}$$

where the last line follows from the fact that the rows of $\left(I - \lambda diag\left(\tau^{*-1}\right) D\right)^{-1}$ sum to 1. To see why, note that

$$\left(I - \lambda diag\left(\tau^{*-1}\right) D\right)^{-1} = I + \lambda diag\left(\tau^{*-1}\right) D + \left(\lambda diag\left(\tau^{*-1}\right) D\right)^2 + ... \tag{31}$$

The rows of $\lambda diag\left(\tau^{*-1}\right) D$ sum to zero, meaning that $1_{n\times 1}$ is an eigenvector with eigenvalue zero. When a matrix is raised to a power, its eigenvectors are unchanged and its eigenvalues are raised to the same power, meaning that $1_{n\times 1}$ remains an eigenvector with 0 the associated eigenvalue, and the rows sum to zero. Since the rows of I sum to 1, the rows of $\left(I - \lambda diag\left(\tau^{*-1}\right) D\right)^{-1}$ then do also.

C.1 Bias of $f^w(\omega; x, \tau)$

From above, the solution for the vector f^w is

$$\begin{aligned} f^w(x,\tau) &= \left(I - \lambda diag\left(\tau^{-1}\right) D\right)^{-1} \left(\psi diag\left(\tau^{-1}\right) Z + x\right) && (32) \\ f^w(x,\tau) &= \left(I + \sum_{j=1}^{\infty} \left(\lambda diag\left(\tau^{-1}\right) D\right)^j\right) \left(\psi diag\left(\tau^{-1}\right) Z + x\right) && (33) \\ f^w(x,\tau) - \psi diag\left(\tau^{-1}\right) Z - x &= \left(\sum_{j=1}^{\infty} \left(\lambda diag\left(\tau^{-1}\right) D\right)^j\right) \left(\psi diag\left(\tau^{-1}\right) Z + x\right). && (34) \end{aligned}$$

Now scale τ^{-1} by c and divide both sides by c

$$c^{-1} f^w(x, \tau/c) - \psi diag\left(\tau^{-1}\right) Z - c^{-1} x = \begin{pmatrix} \lambda diag\left(\tau^{-1}\right) D \\ + \sum_{j=2}^{\infty} c^{j-1} \left(\lambda diag\left(\tau^{-1}\right) D\right)^j \end{pmatrix} \left(c\psi diag\left(\tau^{-1}\right) Z + x\right). \tag{35}$$

Since both sides are linear in x, we can take take the expectation and then the limit as $c \to 0$ to yield

$$\lim_{c\to 0} \frac{E[f^w(x, \tau/c)] - f}{c} = \psi diag\left(\tau^{-1}\right) Z + \lambda diag\left(\tau^{-1}\right) Df. \tag{36}$$

In the limit as $n \to \infty$, Df becomes f''.

C.2 Alternative information cost specifications

When $\lambda = 0$, the formula for f^w as a function of τ becomes

$$f^w(\omega_j) = \psi\tau(\omega_j)^{-1} Z(\omega_j) + x(\omega_j) \tag{37}$$

and the first-order condition for the optimization of $\tau(\omega_j)$ is in the case with arbitrary $\gamma(\omega_j)$

$$0 = -\psi\tau(\omega_j)^{-2} Z(\omega_j)^2 + \theta\gamma(\omega_j) \tag{38}$$

The result from the text then follows immediately. For the entropy information cost, the first-order condition is

$$0 = -\psi\tau(\omega_j)^{-2} Z(\omega_j)^2 + \theta\tau(\omega_j)^{-1} \tag{39}$$

which again immediately yields the desired result.

D Consumption and income forecasts

D.1 The behavior of consumption

From Dew-Becker (2016), consumption follows

$$\begin{aligned} C_t &= (R-1)W_{t-1} + Z_t - (R-1)^{-1}\alpha^{-1}\log(\beta R) & (40)\\ W_t &= W_{t-1} + Y_t - Z_t + (R-1)^{-1}\alpha^{-1}\log(\beta R), & (41) \end{aligned}$$

where

$$Z_t = \left(1 - R^{-1}\right)Y_t - \frac{1}{\alpha}R^{-1}\log E_t \exp(-\alpha Z_{t+1}). \tag{42}$$

We then have

$$\begin{aligned} C_{t+1} &= (R-1)W_t + Z_{t+1} - (R-1)^{-1}\alpha^{-1}\log(\beta R) & (43)\\ \Delta C_{t+1} &= (R-1)W_t + Z_{t+1} - (R-1)W_{t-1} - Z_t & (44)\\ \Delta C_{t+1} &= (R-1)(RW_{t-1} + Y_t - C_t) + Z_{t+1} - (R-1)W_{t-1} - Z_t & (45) \end{aligned}$$

$$\Delta C_{t+1} = (R-1)Y_t + Z_{t+1} - RZ_t + \alpha^{-1}\log(\beta R). \tag{46}$$

Now define H as follows:

$$\begin{aligned} Z_t &= \left(1 - R^{-1}\right)Y_t - \frac{1}{\alpha}R^{-1}\log E_t \exp(-\alpha Z_{t+1}) & (47)\\ H_t &\equiv Z_t - \left(1 - R^{-1}\right)Y_t & (48)\\ H_t &= -\frac{1}{\alpha}R^{-1}\log E_t \exp\left(-\alpha\left(H_{t+1} + \left(1 - R^{-1}\right)Y_{t+1}\right)\right). & (49) \end{aligned}$$

This definition yields

$$\begin{aligned} \Delta C_{t+1} &= R\left(\left(1 - R^{-1}\right)Y_t - Z_t\right) + Z_{t+1} + \alpha^{-1}\log(\beta R) & (50)\\ &= H_{t+1} - RH_t + \left(1 - R^{-1}\right)Y_{t+1} + \alpha^{-1}\log(\beta R), & (51) \end{aligned}$$

with the recursion

$$\begin{aligned} \bar{h} + h(L)\varepsilon_t &= -\frac{1}{\alpha}R^{-1}\log E_t\left[\exp\left(-\alpha\left(\bar{h} + h(L)\varepsilon_{t+1} + \left(1 - R^{-1}\right)b(L)\varepsilon_{t+1}\right)\right) \mid f^w\right] & (52)\\ &= R^{-1}\left(\bar{h} + \sum_{j=1}^{\infty}\left(h_j + \left(1 - R^{-1}\right)b_j^w\right)\varepsilon_{t+1-j}\right) - R^{-1}\frac{\alpha}{2}\left(h_0 + \left(1 - R^{-1}\right)b_0^w\right)^2 & (53) \end{aligned}$$

and solution

$$h_j = R^{-1}h_{j+1} + R^{-1}\left(1-R^{-1}\right)b^w_{j+1} \tag{54}$$

$$\bar{h} = -\frac{R^{-1}}{1-R^{-1}}\frac{\alpha}{2}\left(h_0 + \left(1-R^{-1}\right)b^w_0\right)^2 \tag{55}$$

$$= -R^{-1}\frac{\alpha}{2}\left(1-R^{-1}\right)b^w\left(R^{-1}\right)^2 \tag{56}$$

$$\Delta C_{t+1} = \left(1-R^{-1}\right)Y_{t+1} + H_{t+1} - RH_t + \alpha^{-1}\log\beta R. \tag{57}$$

Now we can insert the formulas for the various objects:

$$\Delta C_t = \left(1-R^{-1}\right)b(L)\varepsilon_{t+1} + (1-R)\bar{h} + h_0\varepsilon^w_{t+1} + \sum_{j=0}^{\infty}\left(h_{j+1} - Rh_j\right)\varepsilon^w_{t-j} + \alpha^{-1}\log\beta R \tag{58}$$

$$= \left(1-R^{-1}\right)b(L)\varepsilon_{t+1} + (1-R)\bar{h} + \left(1-R^{-1}\right)\left(b^w\left(R^{-1}\right) - b^w_0\right)\varepsilon^w_{t+1} \tag{59}$$

$$-\sum_{j=0}^{\infty}\left(1-R^{-1}\right)b^w_{j+1}\varepsilon^w_{t-j} + \alpha^{-1}\log\beta R \tag{60}$$

$$= \left(1-R^{-1}\right)b(L)\varepsilon_{t+1} + (1-R)\bar{h} + \left(1-R^{-1}\right)\left(b^w\left(R^{-1}\right) - b^w_0\right)\frac{b(L)}{b^w(L)}\varepsilon_{t+1} \tag{61}$$

$$-\sum_{j=0}^{\infty}\left(1-R^{-1}\right)b^w_{j+1}\frac{b(L)}{b^w(L)}\varepsilon_{t-j} + \alpha^{-1}\log\beta R \tag{62}$$

$$= \left(1-R^{-1}\right)b(L)\varepsilon_{t+1} + (1-R)\bar{h} + \left(1-R^{-1}\right)b^w\left(R^{-1}\right)\frac{b(L)}{b^w(L)}\varepsilon_{t+1} \tag{63}$$

$$-\left(1-R^{-1}\right)b^w(L)\frac{b(L)}{b^w(L)}\varepsilon_{t+1} + \alpha^{-1}\log\beta R \tag{64}$$

$$= \left(1-R^{-1}\right)b^w\left(R^{-1}\right)\frac{b(L)}{b^w(L)}\varepsilon_{t+1} + (1-R)\bar{h} + \alpha^{-1}\log\beta R. \tag{65}$$

So consumption growth is equal to a constant plus $\left(1-R^{-1}\right)b^w\left(R^{-1}\right)\frac{b(L)}{b^w(L)}\varepsilon_{t+1}$. The dynamic behavior of consumption growth is therefore determined by $b^w\left(R^{-1}\right)\frac{b(L)}{b^w(L)}$. The spectral density of consumption growth is

$$f^w_{\Delta C}(\omega) = b^w\left(R^{-1}\right)^2\frac{f(\omega)}{f^w(\omega)}. \tag{66}$$

E KL divergence for consumption process

We consider the relative entropy of consumption growth under the worst-case model compared to the true model. If we have two models of consumption growth defined by their spectra and means, $\{f_{\Delta C}(\omega), \mu_{\Delta C}\}$, then the Kullback–Leibler divergence is

$$\int \frac{\exp f^w_{\Delta C}(\omega)}{\exp f_{\Delta C}(\omega)} - \log\frac{\exp f^w_{\Delta C}(\omega)}{\exp f_{\Delta C}(\omega)}d\omega + \frac{\left(\mu^w_{\Delta C} - \mu_{\Delta C}\right)^2}{\exp f_{\Delta C}(0)}. \tag{67}$$

In our case, the ratio of the spectra is

$$\frac{\exp f^w_{\Delta C}(\omega)}{\exp f_{\Delta C}(\omega)} = \frac{b^w\left(R^{-1}\right)^2\frac{\exp f(\omega)}{\exp f^w(\omega)}}{b\left(R^{-1}\right)^2} \tag{68}$$

$$= \exp\left(\int Z(\kappa)f^w(\kappa)\,d\kappa\right)\exp\left(-f^w(\omega)\right)\frac{\exp f(\omega)}{b\left(R^{-1}\right)^2}, \tag{69}$$

and the difference in the means is

$$\mu^w_{\Delta C} - \mu_{\Delta C} = -R^{-1}\frac{\alpha}{2}\left(1 - R^{-1}\right)\left(b^w\left(R^{-1}\right)^2 - b\left(R^{-1}\right)^2\right). \tag{70}$$

So the KL divergence is, ignoring additive constants,

$$\begin{aligned}
KL \quad = \quad & \exp\left(\int Z(\kappa) f^w(\kappa)\, d\kappa\right)\int \exp\left(-f^w(\omega)\right)\frac{\exp f(\omega)}{b\left(R^{-1}\right)^2}d\omega && (71)\\
& + \int f^w(\omega)\, d\omega - \int Z(\kappa) f^w(\kappa)\, d\kappa && (72)\\
& + \frac{\left(R^{-1}\frac{\alpha}{2}\left(1-R^{-1}\right)\right)^2\left(b^w\left(R^{-1}\right)^2 - b\left(R^{-1}\right)^2\right)^2}{\exp f_{\Delta C}(0)}. && (73)
\end{aligned}$$

The derivative with respect to $f^w(m)$ is

$$\begin{aligned}
\frac{dKL}{df^w(m)} \quad = \quad & -\exp\left(\int Z(\kappa) f^w(\kappa)\, d\kappa\right)\exp\left(-f^w(m)\right)\frac{\exp f(m)}{b\left(R^{-1}\right)^2} && (74)\\
& + Z(m)\exp\left(\int Z(\kappa) f^w(\kappa)\, d\kappa\right)\int \exp\left(-f^w(\omega)\right)\frac{f(\omega)}{b\left(R^{-1}\right)^2}d\omega + 1 - Z(m) && (75)\\
& + 2\frac{\left(R^{-1}\frac{\alpha}{2}\left(1-R^{-1}\right)\right)^2\left(b^w\left(R^{-1}\right)^2 - b\left(R^{-1}\right)^2\right)}{\exp f_{\Delta C}(0)}\exp\left(\int Z(\kappa) f^w(\kappa)\, d\kappa\right) Z(m). && (76)
\end{aligned}$$

Evaluating at $f^w = \bar{f}$, we obtain $\frac{dKL}{dlf^w(m)}|_{f^w=\bar{f}} = 0$, as we would expect. The second derivative is

$$\begin{aligned}
\frac{d^2KL}{dlf^w(m)^2} \quad = \quad & \exp\left(\int Z(\kappa) f^w(\kappa)\, d\kappa\right)\exp\left(-f^w(m)\right)\frac{\exp f(m)}{b\left(R^{-1}\right)^2} && (77)\\
& - \exp\left(\int Z(\kappa) f^w(\kappa)\, d\kappa\right)\exp\left(-f^w(m)\right)\frac{\exp f(m)}{b\left(R^{-1}\right)^2}Z(m) && (78)\\
& + Z(m)^2\exp\left(\int Z(\kappa) f^w(\kappa)\, d\kappa\right)\int \exp\left(-f^w(\omega)\right)\frac{\exp f(\omega)}{b\left(R^{-1}\right)^2}d\omega && (79)\\
& - Z(m)\exp\left(\int Z(\kappa) f^w(\kappa)\, d\kappa\right)\exp\left(-f^w(m)\right)\frac{\exp f(m)}{b\left(R^{-1}\right)^2} && (80)\\
& + 2\frac{\left(R^{-1}\frac{\alpha}{2}\left(1-R^{-1}\right)\right)^2\left(b^w\left(R^{-1}\right)^2 - b\left(R^{-1}\right)^2\right)}{\exp f_{\Delta C}(0)}\exp\left(\int Z(\kappa) f^w(\kappa)\, d\kappa\right) Z(m)^2 && (81)\\
& + 2\frac{\left(R^{-1}\frac{\alpha}{2}\left(1-R^{-1}\right)\right)^2}{\exp f_{\Delta C}(0)}\left(\exp\left(\int Z(\kappa) f^w(\kappa)\, d\kappa\right) Z(m)\right)^2. && (82)
\end{aligned}$$

Evaluating now at $f^w = f$,

$$\begin{aligned}
\frac{d^2KL}{dlf^w(m)^2}|_{f^w=\bar{f}} \quad &= \quad (Z(m)-1)^2 + 2\left(R^{-1}\frac{\alpha}{2}\left(1-R^{-1}\right)\right)^2\frac{b\left(R^{-1}\right)^2}{\exp f_{\Delta C}(0)}Z(m)^2 && (83)\\
&= \quad (Z(m)-1)^2 + 2\left(R^{-1}\frac{\alpha}{2}\left(1-R^{-1}\right)\right)^2 Z(m)^2. && (84)
\end{aligned}$$

So the weights across frequencies in the KL divergence are approximately a function of $Z(\omega)^2$.

F Epstein–Zin preferences

Suppose people have preferences of the form

$$v_t = (1-\beta)c_t + \frac{\beta}{1-\alpha}\log E_t\left[\exp\left((1-\alpha)v_{t+1}\right)\right] \tag{85}$$

where $c_t = \log C_t$. They face the budget constraint

$$W_{t+1} = R_{t+1}(W_t - C_t) \tag{86}$$

Returns follow the process

$$r_{t+1} = \log R_{t+1} = \bar{r} + b(L)\varepsilon_{t+1} \tag{87}$$

lower-case letters from here on denote logs.

Since people have a unit elasticity of intertemporal substitution, the consumption-wealth ratio will be constant. We write

$$c_t = \bar{c} + w_t \tag{88}$$

The budget constraint can be rewritten as

$$\Delta w_{t+1} = r_{t+1} + \log(1 - \exp(\bar{c})) \tag{89}$$

where Δ is the first-difference operator.

For consumption growth, we then have

$$\begin{aligned} \Delta c_{t+1} &= \Delta w_{t+1} \\ &= r_{t+1} + \log(1 - \exp(\bar{c})) \end{aligned}$$

and we guess that lifetime utility is

$$v_t = \bar{v} + c_t + v(L)\varepsilon_t \tag{90}$$

We can confirm this guess,

$$\bar{v} + c_t + v(L)\varepsilon_t = (1-\beta)c_t + \frac{\beta}{1-\alpha}\log E_t\left[\exp\left((1-\alpha)(\bar{v} + c_{t+1} + v(L)\varepsilon_{t+1})\right)\right] \tag{91}$$

$$\bar{v} + v(L)\varepsilon_t = \frac{\beta}{1-\alpha}\log E_t\left[\exp\left((1-\alpha)(\bar{v} + \log(1-\exp(\bar{c})) + \bar{r} + b(L)\varepsilon_{t+1} + v(L)\varepsilon_{t+1})\right)\right] \tag{92}$$

$$= \beta\left(\bar{v} + \log(1-\exp(\bar{c})) + \bar{r} + b_+(L)\varepsilon_{t+1} + v_+(L)\varepsilon_{t+1}\right) + \beta\frac{1-\alpha}{2}(b_0 + v_0)^2\sigma^2 \tag{93}$$

where v_+ and b_+ denote the lag polynomials with the constants (b_0 and v_0) removed and the coefficients in the polynomials are denoted b_j and v_j.

Matching coefficients yields

$$\bar{v} = \frac{\beta}{1-\beta}\left(\log(1-\exp(\bar{c})) + \bar{r} + \frac{1-\alpha}{2}(b_0+v_0)^2\sigma^2\right) \tag{94}$$

$$v_j = \beta(v_{j+1} + b_{j+1}) \tag{95}$$

$$\Rightarrow v_0 + b_0 = b(\beta) \tag{96}$$

So we have

$$\bar{v} = \frac{\beta}{1-\beta}\left(E[\Delta c] + \frac{1-\alpha}{2}b(\beta)^2\sigma_\varepsilon^2\right) \tag{97}$$

Next we insert this into the Euler equation along with the expression for consumption growth

$$\begin{aligned} 1 &= E_t\left[\beta\frac{\exp((1-\alpha)v_{t+1} - \Delta c_{t+1} + r_{t+1})}{E_t[\exp((1-\alpha)v_{t+1})]}\right] \\ \bar{c} &= \log(1-\beta) \end{aligned}$$

finally yielding

$$\bar{v} = \frac{\beta}{1-\beta}\left(\log(\beta) + \bar{r} + \frac{1-\alpha}{2}b(\beta)^2\sigma^2\right) \tag{98}$$

The key result here is that lifetime utility depends on $b(\beta)^2\sigma_\varepsilon^2$, which is the same term as in the main text. Note also that log utility is the special case of the above in which $\alpha = 1$. In that case, agents are indifferent to return risk.

G Prior on smoothness in terms of cycle length

The main analysis studies the spectrum in the frequency domain, with a prior on smoothness that is equally strong at all frequencies. This section considers an alternative specification where the analysis is in terms of cycle length and shows that the optimal information acquisition policy remains unchanged, even though lower-frequency fluctuations become more difficult to learn about.

Recall from the main text that for a fluctuation at a frequency ω, the length of the associated cycle is $\zeta = 2\pi/\omega$. It is obviously possible through a simple change of variables to write the entire model in terms of cycles instead of frequencies.

The key difference in this section from the main text is that we assume that agents have a prior on the smoothness of the spectrum in terms of cycles, rather than frequencies. Note that in the frequency domain, the upper half of the range $[0, \pi]$ is associated with cycles lasting four or fewer periods, whereas the lower half of the range is associated with all longer cycles. In economic terms, we might think there is potentially much more interesting variation in the model in the range of cycles lasting longer than four periods. Put another way, any cycle lasting less than, say, four quarters, could be said to be "high", with little meaningful to distinguish them, whereas cycles lasting longer than four quarters could include business cycles, medium-frequency trends, and long-term growth rates. It might be more natural, then, for the agent to have a prior on the smoothness of the spectrum written in terms of cycles than frequencies.

Formally, define

$$\tilde{f}(\zeta) \equiv f\left(\frac{2\pi}{\zeta}\right) \tag{99}$$

We now say that the agent has a prior that restricts the total squared variation in $\tilde{f}$, which would be

$$\begin{aligned} \int_2^\infty \tilde{f}'(\zeta)^2 \, d\zeta &= \int_0^\pi \tilde{f}'\left(\frac{2\pi}{\omega}\right)^2 \frac{2\pi}{\omega^2} d\omega && (100) \\ &= \int_0^\pi f'\left(\frac{2\pi}{\omega}\right)^2 \omega^2 d\omega && (101) \end{aligned}$$

In other words, since the transformation $2\pi/\omega$ stretches the space around the very lowest frequencies, the agent is essentially open to the possibility that the spectrum might be infinitely variable at the very lowest frequencies.

Going back to the discretization used in the main analysis, we write the penalized likelihood in this case as

$$PL\left(x \mid \hat{f}, \tau\right) = -\frac{1}{2} d\omega \sum_{j=1}^n \left(x(\omega_j) - f(\omega_j)\right)^2 \tau(\omega_j) - \frac{\lambda}{2} \sum_{j=2}^n \left(\frac{\hat{f}(\omega_j) - \hat{f}(\omega_{j-1})}{d\omega}\right)^2 \omega_j^2 d\omega \tag{102}$$

with the only difference now being the added ω_j^2 in the second summation, showing that the smoothness prior is tigher at high than low frequencies. This can be written in terms of vectors and matrices as

$$PL\left(x \mid \hat{f}, \tau\right) = -\frac{1}{2} d\omega \left(x - \hat{f}\right)' diag(\tau) \left(x - \hat{f}\right) - \frac{\lambda}{2} \hat{f}' D_\omega \hat{f} d\omega \tag{103}$$

where

$$D_\omega \equiv \begin{bmatrix} -\omega_2^2 & \omega_2^2 & 0 & 0 & \cdots & 0 \\ \omega_2^2 & -\omega_3^2 - \omega_2^2 & \omega_3^2 & 0 & & \\ 0 & \omega_3^2 & -\omega_4^2 - \omega_3^2 & \omega_4^2 & & \vdots \\ \vdots & & & \ddots & & 0 \\ & & 0 & \omega_{n-1}^2 & -\omega_n^2 - \omega_{n-1}^2 & \omega_n^2 \\ 0 & \cdots & 0 & 0 & \omega_n^2 & -\omega_n^2 \end{bmatrix} d\omega^{-2}. \tag{104}$$

G.1 Estimation precision

There is a formal sense in which the change in the smoothness prior makes it more difficult for an agent to learn about low frequencies. Consider the simple estimation problem of choosing $\hat{f}$ to maximize $PL\left(x \mid \hat{f}, \tau\right)$. The point estimate is then

$$f^* \equiv \arg\max_{\hat{f}} PL\left(x \mid \hat{f}, \tau\right) \quad (105)$$

$$= \left(diag\left(\tau\right) - \lambda D_\omega\right)^{-1} diag\left(\tau\right) x \quad (106)$$

If we set $\tau = \bar{\tau} 1_{n\times 1}$ for a scalar $\bar{\tau}$, so that the agent has signals with equal precision at all frequencies, we obtain

$$f^* = \left(I - \lambda \bar{\tau}^{-1} D_\omega\right)^{-1} x \quad (107)$$

The variance matrix of f^* is

$$var\left(f^*\right) = \bar{\tau}^{-1}\left(I - \lambda \bar{\tau}^{-1} D_\omega\right)^{-1}\left(I - \lambda \bar{\tau}^{-1} D_\omega\right)^{-1} \quad (108)$$

The variance is straightforward to analyze numerically. Figure G.1 plots the main diagonal of the variance matrix for various values of $\lambda \bar{\tau}^{-1}$. Each case is rescaled so that they are equal for the lowest frequency, illustraing how the variances differ across frequencies. In all cases, the variance of the estimator $f^*(\omega)$ is lower at high frequencies. So when agents have a weaker prior on smoothness at low than high frequencies (due to the assumption of equal smoothness in terms of cycles), it is more difficult to learn about the spectrum at low frequencies.

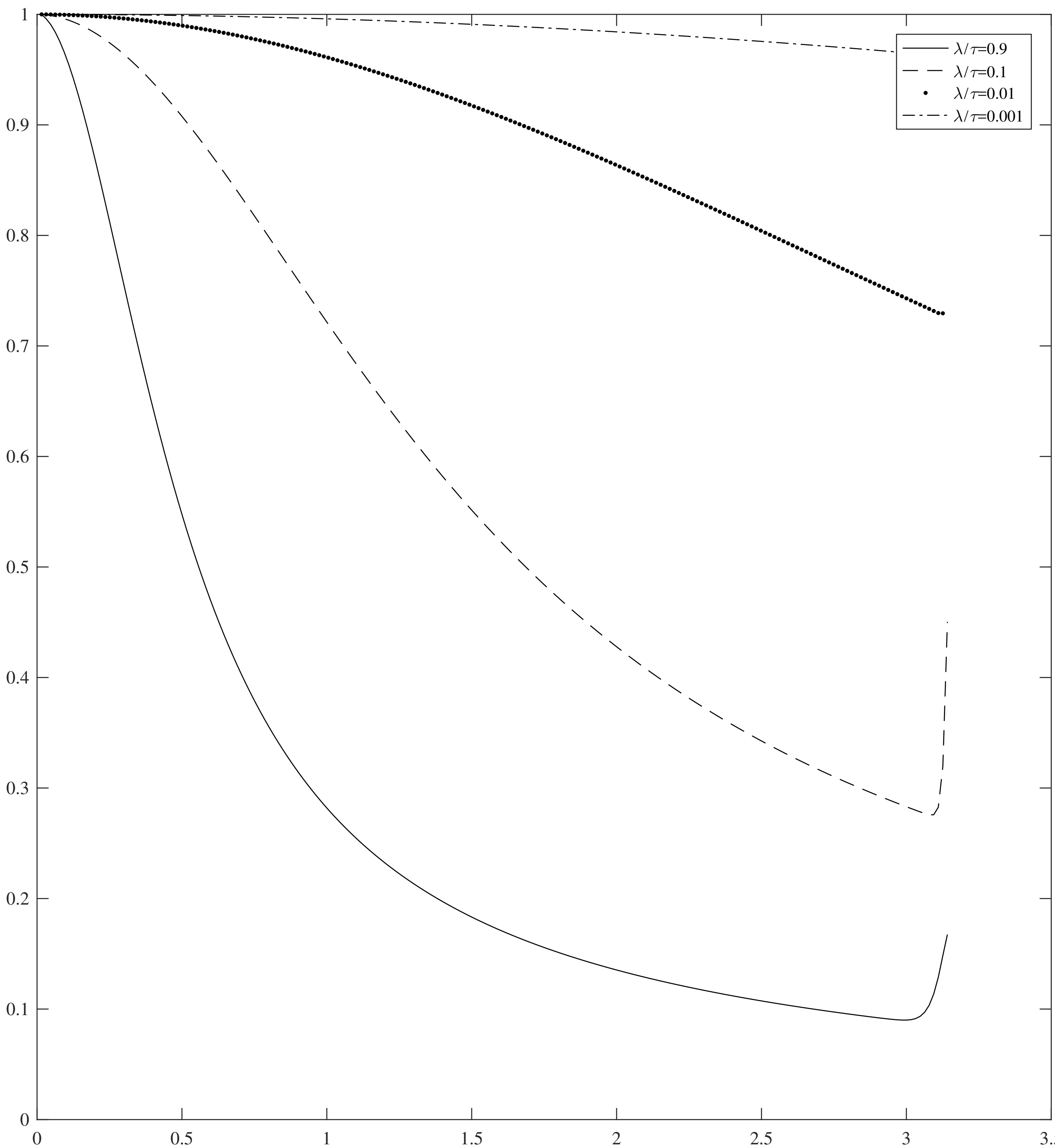
G.2 Optimal information policy

It is straightforward to confirm that the optimal information policy, $\tau \propto Z$, is unchanged in this case. The result follows from the fact that the two chatacteristics of D that are necessary for the main result – that its rows and columns sum to zero – also hold for D_ω.

References

Dew-Becker, Ian, "The pricing of economic risks under time-separable and recursive preferences," 2016. Working paper.

Figure G.1: Variance of spectral estimates across frequencies



Notes: Variance of estimates of the spectrum across frequencies for the smoothness prior in terms of cycles.