The Dynamics of Expected Returns: Evidence from Multi-Scale Time Series Modeling*

Daniele Bianchi† Andrea Tamoni‡

First draft: October 2015. This draft: December 18, 2015

Abstract

Conventional wisdom posits that all the relevant investors' information lies at the highest possible frequency of observation, so that long-run expected returns can be mechanically inferred by a forward aggregation of short-run estimates. We reverse such logic and propose a novel framework to model and extract the dynamics of latent short-term expected returns by coherently combining the lower-frequency information embedded in multiple predictors. We show that the information cascade from low- to high-frequency levels allows to identify long-lasting effects on expected returns that cannot be captured by standard persistent ARMA processes. The empirical analysis demonstrates that the ability of the model to capture simultaneously medium- and long-term fluctuations in the dynamics of expected returns, has first order implications for forecasting and investment decisions.

Keywords: Expected Returns, Long-Horizon Predictability, Multi-Scale, Bayesian Methods

JEL codes: G17, G11, C53, C58

*We thank Arie Gozluklu, Michael Moore, and Pietro Veronesi for their helpful comments and suggestions. We also thank seminar participants at the Brown Bag Seminar in Asset Pricing at Nova School of Business and Economics, and Finance Brown Bag Seminar at the Warwick Business School.

†Warwick Business School, University of Warwick, Coventry, UK. Daniele.Bianchi@wbs.ac.uk

‡Department of Finance, London School of Economics, London, UK. A.G.Tamoni@lse.ac.uk

1 Introduction

The predictive power of forecasting variables studied in the literature varies with the time horizon. For example, while the dividend-price ratio is known to reveal long-term tendencies in asset returns, the consumption-wealth ratio shows its highest forecasting power at the intermediate horizon. As such, simple predictive regressions support the view that different predictors should have an horizon-specific effect on the investors' expectations formation process. Nonetheless, to date, the standard approach to extract expected returns relies on the assumption that all the relevant investors' information lies at the highest possible frequency of observation; given this assumption, it is then typical to specify transitional dynamics for state variables over a short horizon, and then use these dynamics to infer the medium- and long-run behavior of expected returns.[1]

In this paper, we propose a novel econometric framework to coherently combine in a unifying model the scale-specific information from multiple predictors, and we investigate whether the dominant view that long-term investors' expectations are the results of an aggregation of short-term expectations is backed up by the empirical evidence. Equivalently, we ask whether the time series behavior of long-run expected returns extracted from low-to-mid frequency predictors is consistent with the long-horizon aggregation of short-run expected returns. Our main results suggest this might not be the case.

Empirically, we show that our model-implied expected return series exhibits aggregation properties which differs wildly from those obtained within the class of standard ARMA processes and, in particular, from those of a standard autoregressive process of order one (henceforth AR(1)). Figure 8 shows this case in point. In Panel A we compare the autocorrelation function (ACF henceforth) for the series of monthly expected returns extracted from the joint information provided by the annual consumption-wealth and a four-year moving average dividend-price ratios, with the autocorrelation of short-run returns with persistent AR(1) dynamics. Although a simple AR(1) process for the monthly series seems to be appropriate to describe the first four lags, Panel A shows that by conditioning on the joint effect of different horizon-specific predictors our expected return series exhibits strong dependence at long lags. Panels B and C reinforce the

[1] Notice that throughout the paper we use the terms *scale*, *frequency*, *level of resolution* and *time horizon* as synonyms.

argument by showing the ACF of the extracted expected return series aggregated accordingly to the temporal scale of the predictors used as conditioning information, i.e. one and four years for the consumption-wealth and dividend-price ratios, respectively. We compare these ACFs with that of an AR(1) process aggregated over the same time window. The evidence suggests that the dynamics of expected returns implied by mid-to-low frequency predictors cannot be reconciled with the one captured by the aggregation of a simple AR(1) process. As a matter of fact, the ACF of the AR(1) collapses at the 4-years horizon while the model that combines information at multiple frequencies still manifests interesting dynamics.[2] This result sheds some new light on the importance of considering the scale-specific information content of different predictors in estimating the dynamics of expected returns (see, e.g. also Bandi et al. 2015).

A natural implication of this first result is that combining information at multiple horizons to construct short-term expected returns can potentially have first order effects for forecasting and investment decisions. For instance, one can think of an agent with, e.g. , a monthly investment horizon, that has to choose her allocation based on predictors available at the same time scale. By doing so, however, she would disregard the fact that the forecasting power of macro-financial predictors typically shows up at longer horizons, as discussed above. This leads to our second research question; we investigate whether extracting short-run expected returns from predictors at specific horizons shows any implication in terms of portfolio choices and forecasting future expected returns. To answer this question, we first investigate the forecasting accuracy of our model compared to standard AR(1) specifications, and show that by extracting coherently information from horizon-specific predictors allows to better capture both average and unusual developments in future expected returns. Then, we compare our framework with the standard VAR-based approach within an optimal asset allocation problem faced by a power utility investor (see, e.g., Kandel and Stambaugh 1996 and Barberis 2000). We show that the optimal buy-and-hold allocation inherits the horizon heterogeneity implied by the extracted future expected returns.

In methodological terms we make two key choices. First, to combine information coherently across horizons as captured, e.g., by the large R^2s typical of long-run forecasting regressions, we adopt a framework that relies on a parsimonious multi-scale model for time series (see Fer-

[2]It is worth pointing out that our approach is based on a very parsimonious parametrization (three parameters effectively) and, yet, it exhibits a variety of autocorrelation structures at the fine level, as shown in Figure 8. See Section 3 for an in-depth discussion.

reira et al. 2006 and Ferreira and Lee 2007).[3] More specifically, our multi-scale model couples standard linear ARMA models at different levels of resolution via a link equation. In order to achieve consistency between, e.g., monthly expected returns and lower frequency predictors we exploit the Jeffrey's rule of conditioning (see Jeffrey 1957, Jeffrey 1965 and Diaconis and Zabell 1982 for more details). Thus, our modeling framework allows to condition the estimates of univariate expected returns on the information obtained from multiple horizon-specific predictors. This is important for two reasons. First, our framework has the ability to use the lower-frequency observations of predictors whilst maintaining the evolution of expected returns at the highest possible frequency, something not provided by earlier work relying on OLS regressions. Second, by relying directly on the lower-frequency dynamics of predictive variables, our framework exploits in a more efficient way the scale-specific information inherent in each predictor. This fact is reflected in the highly stationary dynamics of otherwise close-to-unit-root predictors typically used in the literature, such as the monthly dividend-price ratio (see Panel A of Table 2 for posterior estimates on the dynamics of the observable predictors).

Our second choice in terms of methodology concerns the estimation algorithm. In particular we propose a Markov Chain Monte Carlo (MCMC) algorithm to provide posterior estimates of quantity of interests of the model. This allows to make inference on the dynamics of expected returns and the corresponding parameters in a single step, developing a robust finite-sample framework that allows one to test, in a unifying way, hypothesis on the persistence and dynamics of both scale-specific predictors and expected returns. Also, unlike standard approaches, our MCMC estimation setting allows to fully characterize the marginal predictive distribution of latent future expected returns over time.[4]

Our work takes important steps from existing research. First, the key and novel aspect of our methodology is that we infer the properties of short-run expected returns from the lower-frequency information content of predictors. In fact, this is the opposite of what has been done in the previous literature, which typically postulates a dynamics for short-run expectations and

[3]The econometric framework parallels that in Ferreira et al. (2006). Nonetheless, our setting has four key differences: first, our estimation framework allows for the fine-level time series to be latent; second, we generalize the framework by combining multiple scale-specific sources of information in a unifying setting; third, we investigate the temporal aggregation properties of such multi-scale modeling within an asset pricing context; finally, we formally investigate the forecasting properties of the model on the basis of two forecasting accuracy measures, namely the mean squared error and the log predictive score.

[4]Notice that our approach effectively uses few data points (only fifteen for the log dividend-price ratio), and involves lots of non-linearities. Standard approaches such as OLS and Maximum Likelihood unlikely can deliver unbiased and efficient estimates in our setting.

then infer mechanically the dynamics of long-run returns through forward aggregation (see, e.g., Barberis 2000 and Campbell and Viceira 2005). While this approach leads to consistent modeling across time horizons, it does assume that the only relevant information is localized in the short-run and no valuable gain is obtained by instead directly incorporating longer-horizon information. We reverse such logic by directly condition the short-run dynamics of expected returns to the information conveyed by predictors at the medium- to long-horizons.[5] Second, our econometric framework allows to effectively separate the time frame of forecasting expected returns and investment decision making from the horizon of the predictors. In other words, while existing research posits that a long-horizon investor who rebalances her portfolio frequently, say at monthly frequency, will base her choice on parameters derived from a one-month predictive regression, for our methodology to work one does not need to arbitrarily assume that the forecasting horizon in predictive regressions coincides with the decision interval of the agent.

This paper contributes to the literature on asset returns predictability.[6] Our work is related to Pastor and Stambaugh (2009) and Van Binsbergen and Koijen (2010). Following Pastor and Stambaugh (2009) we think of forecasting variables, such as the dividend-price ratio or the consumption wealth-ratio, as "imperfect predictors". Differently from them, we think of these predictors as scale-specific and, thus, we infer the properties of monthly expected returns using only the information at the relevant scale rather than using monthly observations of raw predictors themselves. Similar to Van Binsbergen and Koijen (2010) we use a latent-variables approach to estimate the expected returns of the aggregate stock market. However, their framework still requires the prediction horizon to be the same as the frequency at which expected returns evolve: in particular Van Binsbergen and Koijen (2010) use the annual information contained in the price-dividend ratio and dividend growth rates to obtain the annual dynamics of expected returns. Instead, our framework combines multiple predictors according to their scale-specific information content, and allows one to use, e.g., annual information to make inference

[5]Similarly, our work is also related to a long econometric literature that studied the impact of the analysis of observations at a scale coarser than the scale of the definition of the process of interest (see, e.g., Amemiya and Wu 1972 and Bollerslev and Wright 2000). This literature usually defines processes at the coarse scales as obtained by sub-sampling or by non-overlapping averages. While this approach may lead to the coherent combination of multi-scale information, it does assume that the only relevant dynamics are at the fine scale of resolution. Differently from this literature, we build our models in a cascade way from coarse to fine levels of resolution, while maintaining the capacity of consistent modeling of time series at different levels of resolution.

[6]See, e.g., Barberis 2000, Lettau and Ludvigson 2001, Lanne 2002, Avramov 2004, Ang and Bekaert 2007, Calvet and Fisher 2007, Campbell and Thompson 2008, Cohen and Frazzini 2008, Lettau and Van Nieuwerburgh 2008, Pastor and Stambaugh 2009, and Dangl and Halling 2012, just to cite a few.

about expected returns at higher, say monthly, frequency.

Finally, our work is related to the Markov-switching multi-fractal framework of Calvet and Fisher (2007). Calvet and Fisher (2007) describe dividends dynamics as driven by shocks with heterogeneous persistence, and then investigate the impact of these shocks with different persistence levels on asset prices. Our framework, although similar in spirit, differs from theirs as it allows to deal with the uncertainty that is naturally embedded in the dynamics of expected returns, i.e. expectations are latent and parameters are uncertain. Moreover our proposed framework is flexible, meaning that it accommodates both standard AR(1) dynamics as well as more complicated long-memory-type of dynamics in a unifying framework.

2 Motivation

In this section we revisit the scale-specific information content of two standard predictors, namely the consumption-wealth ratio, cay_t, and the log dividend-price ratio, dp_t, and then we show the tension inherent in the standard AR(1)-framework to capture the multiple layers of information conveyed by these forecasting variables. Our choice of stock market predictors is motivated by the present value logic, see Campbell and Shiller (1988), and a linearization of the accumulation equation for aggregate wealth in a representative agent economy, see Campbell and Mankiw (1989) and Lettau and Ludvigson (2001). Arguably, cay_t and dp_t are also widely used as benchmark in the forecasting literature.[7]

We start with simple linear predictive regressions of log real stock market returns, r_t, on the predictor z_t, i.e.[8]

$$r_{t+1,t+h} = \alpha + \beta z_{t-h,t} + u_{t,t+h} \ ,$$

where $r_{t+1,t+h}$ represents the total return compounded (forward) from time $t+1$ to time

[7]It is worth mentioning that our framework is general and can accommodate a larger set of alternative predictors. It is also important to stress that our framework exploits the scale-specific information conveyed by a predictor. Thus, our framework rewards especially those setting where each predictor brings information about a range of frequencies not already covered by the other predictors, i.e. those setting where the predictive horizons are mutually exclusive. Our choice of cay_t and dp_t satisfies this condition.

[8]We use the expression "simple" regression to denote regressions that use just one predictor variable at a time. We reserve the word "multiple" regressions for univariate regressions that use information from $p > 1$ predictor variables simultaneously.

$t+h$, $z_{t-h,t}$ denote the h-period backward moving averages of the predictor, z_t is either the consumption-wealth ratio, cay_t, or log dividend-price ratio, dp_t, and we let $h = 1, \ldots, 10$ years. These regressions take the name of forward-backward regressions, and recent empirical evidence shows they are successful at detecting equilibrium relations occurring at specific frequencies or, in our jargon, *scales* (see Bandi et al. 2015 for an in-depth discussion of this approach).[9]

Figure 1 show the resulting R^2 from the predictive regressions. The figure highlights that stock return predictability varies with the horizon. Importantly, the figure also shows that our predictors do contain differential information: this can be readily inferred by the fact that the maximum attainable R^2 depends on whether the predictor carries information about medium- or long-term returns. Indeed, by looking at the log dividend-price ratio, we observe that the dependence between returns and the predictor increases with the time horizon, being stronger in the long run, with R^2-values between 8 and 10 years larger than 20%.[10] Instead, when looking at the consumption-wealth ratio, we find that its predictive power has a tent-shaped pattern with a peak around 1 and 2 years, the R^2-values being in the ballpark of 30%.

[Insert Figure 1 about here]

Given the above evidence, ideally one would like to model expected returns as containing both medium- and long-term fluctuations. Next, we show that the standard practice of modeling expected returns as an autoregressive process of order one (AR(1), henceforth) fails to capture the rich dynamics structure that has emerged from the predictive regressions. To do so we first select the horizon m at which the signal contained in the predictor operates, and then we fit an AR(1) to the predictor temporally aggregated over this horizon.[11]Relying on Figure 1, we initially select m = 1-year for cay_t and m = 4-years for dp_t. The choice of m strikes a balance between the number of non-overlapping observation available for the AR(1) estimation

[9]Alternatively one could have used band spectral regression (Engle 1974) to detect high-, medium- and low-frequency relations. The forward-backward regressions have the advantage of maintaining time adaptation, and localization in both time and frequency. Band spectral methods only guarantee the latter.

[10]This is consistent, e.g., with the evidence in Goyal and Welch (2003). The authors show that it is only for periods of longer than 5 years that the dividend-price stops predicting itself, and instead begins predicting stock market returns and dividend growth. In fact, had we used the longer sample 1930-2013, the R^2 between 8 and 10 years would be larger, at about 40%.

[11]For a time-series z_t, let the temporally aggregated series Z_l be the series consisting of sums of m non-overlapping points of z_t. For example, for $m = 3$, we can aggregate monthly series to obtain a quarterly series. Mathematically, we can define Z_l as $Z_l = \sum_{t=m(l-1)+1}^{ml} z_t = \left(1 + L + \ldots + L^{m-1}\right) z_{ml}$, where L is the lag operator.

and attaining a high (forecasting variable's) signal-to-noise ratio.[12] Panel A of Table 2 shows the results from fitting an AR(1) process to the cay_t and dp_t series, temporally aggregated over one- and four-years respectively. Suppose one wants to describe the annual dynamics of the expected return series x_t. It would be tempting to set the persistence parameter describing expected returns equal to the persistence of cay_t, i.e. $\phi = 0.51$. However, we know that the series obtained by temporally aggregating over m periods an AR(1) behaves approximately as an AR(1) with root equal to ϕ^m (see Amemiya and Wu 1972).[13] Therefore, the dynamics of expected returns over 4-years, as it is implied by the annual-information content of cay_t, would have an ACF decaying as $\phi^4 = 0.07$, a much less persistent process than that implied by our AR(1)-estimate of the four-year moving-average of dp_t, namely $\phi_{dp} \approx 0.8$. Of course, one could think of increasing the persistence of expected returns at annual frequency, e.g. by setting the annual root of expected returns at $\phi = 0.8^{0.25} = 0.94$ so to match the persistence of dp_t at the four-year horizon. By doing so one would capture the low-frequency component of expected returns proxied by dp_t, at the cost of entirely missing the higher-frequency component captured by cay_t.

[Insert Table 2 about here]

The above results show that cay_t and dp_t do contain differential information about future returns. Importantly, we have shown that an AR(1) dynamics for expected returns can hardly capture simultaneously these fluctuations operating at different horizons. In the next section we introduce a method to extract the high-frequency dynamics of expected returns by coherently exploiting the joint information obtained from various predictors at alternative, usually higher, scales (yearly and longer). By combining scale-specific information, our framework exploits the explanatory power of the predictors in a more efficient way than, say, standard OLS regression methods. This gain in efficiency will be reflected in the highly stationary dynamics of otherwise close-to-unit-root predictors, such as the monthly log dividend-price ratio.

[12] In Section 5 we carefully test the optimal value of the coarsening window m on the basis of Marginal likelihood estimates. The choice of m has a deeper meaning. The sums of m non-overlapping points of y_t can be interpreted as the Haar-scaling component operating at scale $j = log_2(m) + 1$ of the raw series z_t. Bandi et al. (2015) showed that if two raw series are related via their components at scale j, then maximum predictability is attained at horizon $h = 2^j$. Therefore, our choice of $m = 4$-years for the dividend-price ratio imply maximum predictability at horizon $h = 8$-years, consistent with the evidence in Figure 1.

[13] To be precise the temporal aggregation of an $AR(p)$ model, is represented by an ARMA(p, r) where $r = \left\lfloor p + 1\frac{-p-1}{m} \right\rfloor$, i.e. the aggregate variable follows an AR model of order p with MA residuals structure. Therefore the ACF does not coincide with that of an $AR(p)$ with rescaled roots; however, we show in the simulation and estimation exercises that the MA effect is quantitatively negligible.

One last remark is in order. Although in this paper we take a reduced form approach to justify our multi-scale framework, it is interesting to observe that the need for multi-frequency dynamics of expected returns can also be justified from a structural point of view. For example, Lochstoer (2009) proposes model where two risk factors, luxury good consumption and relative price growth, operate at a business cycle and a lower frequency, "generational" cycle, respectively. Since the risk factors drive the price of risk, then the risk premium dynamics in the model is predicted to operate at both of these frequencies.

3 Multi-scale time series modeling

We introduce a multi-scale time series approach that allows to coherently combine information across different time horizons (e.g., monthly and yearly financial data) in a unifying setting. From a general viewpoint, our modeling framework couples standard linear models specified at different time scales with a linear stochastic linking equation between scales. For the ease of exposition, in the following we consider the case with only two series: the latent expected return series, $x_{1:n_x}$, and an observable series, $z_{1:n_z}$, which contains information about the low-frequency dynamics of expected returns. In the empirical analysis we extend the framework to multiple scale-specific predictors.

The two series are specified at different temporal scales: in particular, we will assume that $x_{1:n_x}$ evolves at a fine scale (i.e. high frequency), while $z_{1:n_z}$ evolves over a coarser grid (i.e. lower frequency). We thus sometimes refer to $x_{1:n_x}$ as the fine-level process, and $z_{1:n_z}$ as the coarse-level process. The difference between the temporal scales is captured by the relation in the number of observations of the two series: $m \times n_z = n_x$, with $m > 1$ (e.g. if x is at monthly frequency, and $m = 12$ then z consists of yearly observations). To re-iterate, the sequence of x and z are defined as;

$$x_1, x_2, \ldots, x_m, z_1, x_{m+1}, \ldots, x_{2m}, z_2, \ldots, x_{n_x-m+1}, x_{n_x-m+2}, \ldots, x_{n_x}, z_{n_z}$$

Our multiple time-scale framework allows for the dynamics of $x_{1:n_x}$ and $z_{1:n_z}$ to be described independently by different time-series models. Consistent with existing literature on returns predictability, we start by assuming that the latent expected returns, $x_{1:n_x}$, is described by a

stationary linear AR(1) process:

$$x_t = \phi_x x_{t-1} + \epsilon_{x,t}, \qquad \text{with} \qquad \epsilon_{x,t} \sim N\left(0, \sigma_x^2\right), \tag{1}$$

such that the marginal distribution of $x_{1:n_x} = \{x_1, ..., x_{n_x}\}$ has a known form $p\left(x_{1:n_x}\right) = N\left(0, V_x\right)$ with $\{V_x\}_{ij} = \sigma_x^2 \phi_x^{|i-j|} / \left(1 - \phi_x^2\right)$.

We also assume that the coarse level $z_{1:n_z}$ has an n_z-dimensional stationary distribution given by $q\left(z_{1:n_z}\right)$. Although not required, it is useful to think of $z_{1:n_z}$ as another AR(1) evolving over a coarse grid, i.e.

$$z_s = \phi_z z_{s-1} + \eta_s, \qquad \text{with} \qquad \eta_s \sim N\left(0, \sigma_z^2\right), \tag{2}$$

so that, in this case, the implied n_z-dimensional stationary distribution of z is given by $q\left(z_{1:n_z}\right) = N\left(0, Q_z\right)$ with $\{Q_z\}_{ij} = \sigma_z^2 \phi_z^{|i-j|} / \left(1 - \phi_z^2\right)$.[14]

In our framework, the information conveyed by the dynamics of the fine-level process and by the lower frequency predictors is not separated a priori; instead such information coexists on the basis of a stochastic linkage that characterize the relationship between scales. The link between levels of resolution is described by the z values being the averages of non-overlapping groups of m consecutive x values. To this end, let us define the $n_z \times n_x$ matrix A such that $E[z_{1:n_z}] = A x_{1:n_x}$. Thus, A is a sparse matrix whose non-zero elements are all $1/m$; in row i, the non-zero elements are those in columns $(i-1)m + 1$ to im. As a result, the link equation takes the following form;[15]

$$p\left(z_{1:n_z} | x_{1:n_x}\right) = \prod_{s=1}^{n_z} N\left(m^{-1} \sum_{i=1}^{m} x_{(s-1)m+i}, \tau\right) = N\left(A \cdot x_{1:n_x}, \tau\right), \tag{3}$$

where $\tau = \lambda \left(A' V_x A\right)_{11} I$ with I as the n_z-square identity matrix and λ measuring the *between* scales uncertainty. Observe that from the marginal distribution of expected returns $p\left(x_{1:n_x}\right)$ and the conditional distribution implied by (3), we can compute the marginal distribution of

[14] We will show in the empirical section that the AR(1) dynamics provides a good description of the scale-specific predictor (e.g. the four-year component of dp_t) although it might not well capture the properties of the raw predictor series.

[15] More generally one can have $z_s = f(x_1, x_2, \ldots, x_m) + u_s$ where f can take many different forms such as maximizing and averaging.

the observable scale-specific predictor as

$$p\left(z_{1:n_z}\right) = \int p\left(z_{1:n_z}|x_{1:n_x}\right) p\left(x_{1:n_x}\right) dx_{1:n_x} = N\left(0, A'V_xA + \tau\right) \ . \tag{4}$$

Now the problem lies in the fact that the marginals $q(z_{1:n_z})$, $p\left(z_{1:n_z}\right)$, and the stochastic linkage $p\left(z_{1:n_z}|x_{1:n_x}\right)$ are generally incompatible with each other. For example, if $x_{1:n_x}$ is an AR(1) then the conditional distribution $p\left(z_{1:n_z}|x_{1:n_x}\right)$ implies a dynamics which is inconsistent with $q\left(z_{1:n_z}\right)$ being an AR(1). The idea of the multi-scale approach is to revise the fine-scale model, $p(x_{1:n_x})$, by super-imposing the information at the coarse-level (see Eq.(2)) to restore consistency across the different levels of resolution. In particular, consistency across marginal distributions can be restored by using the Jeffrey's rule of conditioning for revising the probability $p\left(x_{1:n_x}\right)$ to a new probability $q\left(x_{1:n_x}\right)$. This ensures consistency across time scales (see Jeffrey 1957, Jeffrey 1965 and Diaconis and Zabell 1982 for more details on Jeffrey's rule of conditioning).[16] In our setting, the Jeffrey's rule implies that:

$$q\left(x_{1:n_x}\right) = \int p\left(x_{1:n_x}|z_{1:n_z}\right) q\left(z_{1:n_z}\right) dz_{1:n_z} \tag{5}$$

with $p\left(x_{1:n_x}|z_{1:n_z}\right) \propto p\left(x_{1:n_x}\right) p\left(z_{1:n_z}|x_{1:n_x}\right)$ and, by standard linear projection,

$$x_{1:n_x}|z_{1:n_z} \sim N\left(Bz_{1:n_z}, V_x - BWB'\right) \tag{6}$$

where $B = V_xA'W^{-1}$ and $W = AV_xA' + \tau$.

Various comments are in order. First, the multi-scale time series framework does not require $x_{1:n_x}$ and $z_{1:n_z}$ to follow AR(1) processes. The framework put forward in this section can be used with any stationary and invertible ARMA processes as building blocks for $x_{1:n_x}$ and $z_{1:n_z}$. We assume an AR(1) for $x_{1:n_x}$ as we want to test whether such simple dynamics is well suited to capture fluctuations at different frequencies, being the AR(1) the standard dynamics used in most, if not all, existing literature (see, e.g. Stambaugh 1999, Barberis 2000, Pastor and Stambaugh 2009, Van Binsbergen and Koijen 2010 and Pástor and Stambaugh 2012 among

[16] More formally, the application of Jeffrey's rule assumes that the revised conditional distribution of $x_{1:n_x}$ given $z_{1:n_z}$, denoted by $q\left(x_{1:n_x}|z_{1:n_z}\right)$, is equal to the conditional distribution of $x_{1:n_x}$ given $z_{1:n_z}$ implied by Equations (1) and (3), denoted by $p\left(x_{1:n_x}|z_{1:n_z}\right)$. This assumption means that given $z_{1:n_z}$, $x_{1:n_x}$ is independent of the new information that led to the revision of beliefs about $z_{1:n_z}$.

others). Second, the multi-scale model is tightly parametrized. Indeed, in Section 3.1 we show that with only three parameters (namely, $\phi_x, \sigma_x, \lambda$) the model is able to generate a variety of autocorrelation structures in the extracted latent expected returns.[17]. For forecasting purposes, the number of parameters grows linearly by a factor of two for each additional scale-specific predictors (i.e. ϕ_z, σ_z for each z). Finally, it is key to observe that the model is not built using a bottom-up aggregation from the fine to the coarser resolutions, i.e. from (1) to (4) via (3), but rather, the information cascade goes from the low frequency predictor to the latent higher frequency expected returns, i.e. from (2) to (5). This allows to use information of the predictors at the "right" scale. The following sections show that this approach based on low-to-high frequency information cascade has important implications, especially for forecasting and investment decisions.

3.1 Properties of multi-scale time series models: Persistence

The parameter λ plays a key role in how information across time scales is combined and how much the distribution of $x_{1:n_x}$ is revised. To illustrate this point, we study the behavior of the covariance structure of the revised process $x_{1:n_x} \sim q\left(x_{1:n_x}\right)$ in the limit. It is possible to show that the revised distribution of $x_{1:n_x}$ is zero-mean normal, $q\left(x_{1:n_x}\right) = N\left(0, Q_x\right)$ with covariance matrix $Q_x = V_x - B\left(W - Q_z\right)B'$. Then, the following results hold in the limit:[18]

$$\lim_{\lambda \to 0} AV_xA' = Q_z \ , \tag{7}$$

$$\lim_{\lambda \to \infty} Q_x = V_x \ . \tag{8}$$

In words, when $\lambda \to 0$, the covariance structure of the fine-level series $x_{1:n_x}$ aggregated to coarser resolution converges to the covariance structure of the coarse-level process $z_{1:n_z}$, i.e. maximal agreement is achieved across fine and coarse scales. On the other hand, when $\lambda \to \infty$ the covariance structure of the fine-level process converges to the original covariance structure, i.e. no revision occurs and $q\left(x_{1:n_x}\right) = p\left(x_{1:n_x}\right)$. Equations (7)-(8) make clear that the parameter λ has a natural interpretation in terms of how much the dynamics of the coarse-level process

[17] See also Sections 4.2 and Section 5.2

[18] See Theorem 11.1 and Theorem 11.2 in Ferreira and Lee (2007). Also, observe that these theorems do not rely on the fine- and coarse level being AR(1) processes, and they are valid for any multi-scale time series model constructed with the link equation (3) and stationary and invertible ARMA processes as building blocks.

influences the behavior of the higher-frequency latent state. The next figure exemplifies the role of λ in determining the correlation structure of the revised fine-level process.

The top panels in Figure 2 show the autocorrelation of the fine-level process obtained from a multi-scale time series model with parameters $\phi_x = \phi_z = 0.9$, $\sigma_x^2 = \sigma_z^2 = 1$, and for different combinations of λ.

[Insert Figure 2 about here]

The top left panel shows the autocorrelation of expected returns obtained by fixing the level of between scale uncertainty at $\lambda = 0.01$. The red and blue lines report the theoretical ACF (i.e. Q_x as implied in Eq. (5)) and the ACF of the series $x_{1:n_x}$ extracted by using the estimation procedure explained in Section 4, respectively. The figure makes clear that, ceteris paribus, for low values of λ the model generates long-lasting effects on the dynamics of the fine-level process which departs from the one that can be obtained by, e.g., an AR(1) model with autoregressive parameter equal to ϕ_x (light blue line). Consistent with (8), the top right panel shows that if the coarse-level does not bring sensible information to revise $x_{1:n_x}$, namely when $\lambda \to \infty$, then the persistence of the latent expected returns collapses to that of a standard AR(1). In this sense an increasing λ has a natural interpretation in terms of the relative increase in uncertainty at the coarse level due to the lack of agreement with the latent fine level.

The structure of the covariance matrix for the revised latent fine-level process $Q_x = V_x - B\,(W - Q_z)\,B'$ implies that the persistence of $x_{1:n_x}$ conditioning on $z_{1:n_z}$ is also influenced by the ratio σ_x^2/σ_z^2. The bottom panels in Figure 2 explore this effect by fixing $\lambda = 0.01$ and decreasing (left panel) or increasing (right panel) the ratio σ_x^2/σ_z^2. The bottom left panel shows that for a low value $\sigma_x^2/\sigma_z^2 = 1/3$ the decay of the autocorrelation function is slower than in the case of higher value $\sigma_x^2/\sigma_z^2 = 3$ (bottom right panel). Interestingly, the bottom right panel shows that the decay of the autocorrelation function of the multiscale process $x_{1:n_x}$ can be faster than the decay of an AR(1) process.

In all, Figure 2 conveys the message that the decay of the autocorrelation function of the multiscale process $x_{1:n_x}$ *at the fine level* is controlled not only by λ but also by the ratio σ_x^2/σ_z^2. However, it is only λ that controls the decay of the autocorrelation function of the process $x_{1:n_x}$ *aggregated at the coarse scale*, i.e. $Ax_{1:n_x}$. This is clearly shown in Figure 3 where each panel

displays the ACF of the process in Figure 2, this time aggregated over $m = 48$ periods.

[Insert Figure 3 about here]

Putting together the evidence provided by the bottom right panels in Figure 2 and Figure 3, it is apparent that the "signal-to-noise" ratio σ_x^2/σ_z^2 mainly governs the short-term time series dependence. When σ_x^2/σ_z^2 is large (bottom right panel in Figure 2) the fine level $x_{1:n_x}$ decays faster than the AR(1) but its non-overlapping averages are more persistent than those implied by the standard aggregation of an autoregressive process (bottom right panel in Figure 3) as dictated by the low value of $\lambda = 0.01$. Indeed, once aggregated over a four-year horizon, the persistence of the expected returns implied by the multi-scale time series model is anywhere higher than a four-year theoretical AR(1) dynamics, as far as λ is small and regardless the value of σ_x^2/σ_z^2. Only when the horizon-specific predictor does not convey information, i.e. λ is high, the long-lasting effects on expected returns disappear, and the ACF of the aggregated process collapses to that implied by the classic AR(1) case (see the top right panel in Figure 3).

To sum up, we have shown in a controlled simulation setting that our multi-scale time series model is able to generate richer autocorrelation structures than those obtained by standard ARMA models. We also studied the aggregation properties of our model, and compare them with the standard forward aggregation of AR(1) dynamics. Finally, as a by product of this analysis, we were able to give a suitable interpretation to the parameters of the model.

3.2 Properties of multi-scale time series models: Forecasting

The revised distribution of the expected returns characterized by equations (5)-(6) makes clear that the multi-scale time series model is built in a cascade way from coarse to fine levels of resolution. This has deep implication for forecasting, especially in the mid-to-long horizon. Indeed, to compute the conditional distribution of future expected returns, not only one needs to generate a sample of future observations at the coarse and fine levels (for the coarse-scale predictors and latent expected returns, respectively) but also these two levels of resolutions interact with one another. Consistent with this intuition, we use a two-stage procedure to compute jointly the forecasts of the $h-$step ahead coarse and fine levels. First, we generate a future realization of the coarse level (i.e. z_{n_z+1}) conditional on the past. After that, a realization

of the (latent) fine level (i.e. $x_{n_x+1} : x_{n_x+m}$) is simulated conditional on the past *and* on the realization of the series evolving at low frequency.

When the multi-scale time series model is built using AR(1) blocks as in Eqs. (1) and (2) the forecasting at the coarse and fine level takes a particular insightful expression. Some additional notation is useful. Let $r = x_{n_x}(\phi_x, \ldots, \phi_x^m)$ and $R_{ij} = \sigma_x^2 \phi_x^{|i-j|} \frac{1-\phi_x^{2\min(i,j)}}{1-\phi_x^2}$ respectively be the m-step ahead predictive mean vector and covariance matrix for a standard (i.e. not revised) AR(1) process at the fine level. Let also denote the collection of parameters by $\Theta = \left(\phi_x, \sigma_x^2, \lambda, \phi_z, \sigma_z^2\right)$. With this notation at hand, the one-step ahead predictive distribution at the coarse level is given by $p\left(z_{n_z+1} \mid x_{1:n_x}, z_{1:n_z}, \Theta\right) \sim N(f_z, F_z)$ where

$$f_z = F_z \left[\sigma_z^{-2} \phi_z z_{n_z} + m^{-1} \left(m^{-2} 1' R 1 + \tau\right)^{-1} 1' r - P_{n_z}^{-1} p_{n_z}\right] , \tag{9}$$

and $F_z = \left[\sigma_z^{-2} + \left(m^{-2} 1' R 1 + \tau\right)^{-1} - P_{n_z}^{-1}\right]^{-1}$. Notice that p_{n_z} and P_{n_z} are the predictive mean and covariance implied by the conditional distribution $p\left(z_{n_z+1} \mid z_{1:n_z}, \Theta\right)$. Conditional on the prediction for the coarse-level process, the predictive distribution for the fine level is given by $p\left(x_{n_x+1} : x_{n_x+m} \mid x_{1:n_x}, z_{n_z+1}, \Theta\right) \sim N(f_x, F_x)$ where[19]

$$f_x = r + m^{-1} R 1 \left(m^{-2} 1' R 1 + \tau\right)^{-1} \left(z_{n_z+1} - m^{-1} 1' r\right) . \tag{10}$$

Equation (10) highlights the importance of using information in low-frequency predictors to revise the distribution of the latent high-frequency process, and, consequently, to revise the future expected path of the fine-level process. Indeed the (mean) forecast at the fine level is the outcome of both the compounding effect of the forecast from a standard AR(1), i.e. r, *and* the revision due to the lower-frequency information in predictors, see the term proportional to $\left(z_{n_z+1} - m^{-1} 1' r\right)$.

The contribution of the multi-scale model in terms of forecasting can be better understood considering two limiting cases. First, when the lower-frequency predictors do not provide sensible information to extract expected returns, i.e. $\lambda \to \infty$ and, as a consequence, τ diverges too, the future path of investors' expectations collapse to the standard iterated AR(1) case, $f_x = r$. This is consistent with the limiting behavior described by . (8) where the low-frequency

[19] See Theorem 6.2, 6.3 and 6.4 in Ferreira et al. (2006).

information does not imply any revision of the fine-level behavior. Second, when σ_z^2 and P_{n_z} are much larger than $\left(m^{-2}1'R1+\tau\right)$, the forecast for the coarse level is given by $f_z = m^{-1}1'r$, i.e. all the information is already contained at the highest frequency and the long-term forecast is determined using models defined at a fine resolution (i.e. $r = x_{n_x}\left(\phi_x, \ldots, \phi_x^m\right)$) and simply aggregated to coarser resolutions, i.e. $m^{-1}1'r$. These are two extreme cases when the multi-scale model has no bite. In all other scenarios, our multi-scale time series model implies that forecasts at low frequency, i.e. z_{n_z+1}, are going to affect future expectations above and beyond the forward iteration of current expectations r.

To quantify this effect, we test the forecasting ability of the model, assuming the true data generating process to be a multi-scale time series with parameters for $x_{1:n_x}$ given by $\phi_x = 0.9$ and $\sigma_x^2 = 0.5$. To resemble the empirical exercise, we set the length of the series to $n_x = 720$, and we fix the coarse-level parameter to $m = 48$. Thus, if x evolves at monthly frequency, then z is observed over a four-year time interval. We also let the observable process $z_{1:n_z}$ follow an AR(1) process as in Eq. (2), with parameters $\phi_z = 0.9$ and $\sigma_z^2 = 1$. In the following, we investigate the role of the between-scales uncertainty parameter by simulating the data for different values of λ.

[Insert Figure 4 about here]

Figure 4 shows the m-step ahead forecast of future expected returns at different point in time. We compare the prediction that one would obtain using the expression in Eq.(10) (magenta line with diamonds), with the m-step ahead forecast obtained from a standard persistent AR(1) process, i.e. $r = x_{n_x}\left(\phi_x, \ldots, \phi_x^m\right)$ (red line with circles). Consistent with empirical evidence on returns predictability based on the dividend-price ratio, we fix the OLS estimates of the AR(1) to $\phi_x = 0.98$.[20] We consider AR(1) dynamics both with and without a constant: the latter is consistent with our data generating process that sets the unconditional mean of the process to zero, while the former accounts for the small sample properties of our simulation setting. In order to better understand the forecasting properties of our model, forecasts are super-imposed to the time series of simulated expected returns $x_{1:n_x}$ (light blue line) and to the low-frequency predictor $z_{1:n_z}$ (dashed, blue with circles).

[20]The value we use is close to the one in, e.g., Table II in Barberis 2000. Alternatively we also fit an AR(1) with an intercept on the average extracted expected returns, i.e. the light blue line on Figure 4, obtaining a value very close to 0.98.

From the top left panel, it is apparent the out-performance of the multi-scale model in capturing m-step ahead expected returns. The forecast from a standard AR(1) quickly converges to its unconditional mean $E\left[x_{1:n_x}|AR(1) \text{ with no intercept}\right] = 0$ (horizontal red line), and it largely misses the increasing expected returns, as well as the decreasing path in the last evaluation point. The top right panel compares our multi-scale forecast with that from an AR(1) with a constant. In this case the AR(1) forecasts (red line with circles) tend to the unconditional mean of the autoregressive process with the intercept $E\left[x_{1:n_x}|AR(1) \text{ with intercept}\right] \approx 0.025$. Also in this case, our multi-scale model better captures the dynamics of future expected returns on each evaluation point.

The bottom panel of Figure 4 investigates how the forecasting properties of our multi-scale model change as we increase the between-scales uncertainty parameter. Consistently with (10), as λ increases, the forecast of the multi-scale model collapses to the m-step ahead prediction from the AR(1) fitted on the extracted expected returns. All in all, when compared with a standard AR(1)-based forward aggregation prediction, our multi-scale model shows a pseudo out-of-sample forecasting performance which is more consistent with future realizations of the latent expected returns. In fact, a separate calculation shows that the mean squared errors of the prediction from the multi-scale model is about 10.6% lower than from an AR(1).

4 Estimation strategy

The dynamics specified in (1)-(3) entails lots of non-linearities and complexities in the conditional likelihood that cannot be easily addressed by using standard estimation methods. For instance, the stochastic link between scales limit the possibility of using standard maximum likelihood methods to filter out the dynamics of latent expected returns. Therefore, for parameter inference, we propose an hybrid Markov Chain Monte Carlo (MCMC) algorithm, which allows to naturally characterize virtually any function of the posterior estimates of latent expected returns and the corresponding structural parameters. Also, the possibility to define priors over the parameter space allows to test assumptions, such as the persistence of expected returns and the information content of scale-specific predictors.

4.1 Prior specification and posterior simulation

A major aspect of our model is that conditional on the latent expected returns, $x_{1:n_x}$, the dynamics of scale-specific predictors are independent to each other, greatly simplifying the sampling from the joint posterior distribution of the parameters. For scale-specific predictors, we specify a standard prior structure assuming independence a priori of the parameters and the following normal-inverse-gamma marginal priors: $\phi_z \sim N(m_{\phi_z}, M_{\phi_z})$, $\sigma_z^2 \sim IG(\nu_{\sigma_z}/2, \nu_{\sigma_z} s_{\sigma_z}/2)$ with m_{ϕ_z}, M_{ϕ_z} the location and scale hyper-parameters of the AR(1) coefficient, and $\nu_{\sigma_z}, \nu_{\sigma_z} s_{\sigma_z}$ the degrees of freedom and scale hyper-parameters for the conditional variance σ_z^2.[21] Similarly, for the fine level process we start from a classic independent normal-inverse-gamma prior on the persistence parameter and the conditional variance of the unrevised process (see Eq. (1)): $\phi_x \sim N(m_{\phi_x}, M_{\phi_x})$, $\sigma_x^2 \sim IG(\nu_{\sigma_x}/2, \nu_{\sigma_x} s_{\sigma_x}/2)$ with m_{ϕ_x}, M_{ϕ_x} the location and scale hyper-parameters of the AR(1) coefficient, and $\nu_{\sigma_x}, \nu_{\sigma_x} s_{\sigma_x}$ the degrees of freedom and scale hyper-parameters for the conditional variance σ_x^2.

Hyper-parameters are set to be uninformative about the shape of the corresponding target posterior distribution. Uninformative priors for the ϕ_x parameter are centered at 0.9 and flat on most of the $(-1, 1)$ range, with prior mass tailing off near ± 1 to ensure stationarity. We also considered alternative priors elicitation by assuming: (i) a zero-mean flat prior; (ii) an informative prior in which 90% of the mass is concentrated around zero; (iii) an informative prior in which 90% of the mass is concentrated around 0.9. For the conditional variances of both observable predictor and latent states we fix the initial degrees of freedom equal to two-percent of the sample sizes. Coupled with a small prior scale hyper-parameter of the inverse gamma, our choice ensures that, a priori, shocks to observable predictors and expected returns are of moderate size. Alternative, diffuse, priors are set such that posterior estimates are consistent with maximum likelihood. More informative priors are set to have a higher probability mass on bigger shocks on both expected returns and observable predictors.

Finally, as far as the between scales uncertainty is concerned, we specify a standard inverse-gamma prior $\lambda \sim IG(\nu_\lambda/2, \nu_\lambda s_\lambda/2)$, with $\nu_{\sigma_\lambda}, \nu_{\sigma_\lambda} s_{\sigma_\lambda}$ denoting the degrees of freedom and scale hyper-parameters, respectively. We start with a benchmark elicitation by assuming a priori a

[21] The prior structure does not change by including more than one scale-specific predictor as they are orthogonal conditional on the latent states. As such, we can specify the same marginal prior structure for each predictor independently.

moderate level of discrepancy between information across different scales and expected returns. Consistent with the indication provided by (7)-(8), we also consider three alternative settings. First, we assume that observable predictors do not bring any information on extracting the latent path of expectations although we keep such prior belief vague. Second, similar to before we assume that predictors do not contain additional information to understand expected returns, but now we also increase the precision of the prior. Third, we assume instead that predictors almost perfectly reveal the underlying investors' expectations formation process.

Although we use a conjugate prior setting, the joint posterior distribution of structural parameters and the latent expected returns is not available in closed form. In order to draw from the posterior distribution we follow a data augmentation principle which relies on the complete likelihood function (e.g. Tanner and Wong 1987). Let the collection of parameters defined as $\Theta = \left(\phi_x, \sigma_x^2, \lambda, \phi_z, \sigma_z^2\right)$.[22] Given the AR(1) dynamics imposed at different scales, and the corresponding linking equation, the complete likelihood can now be defined as

$$p\left(\mathbf{x}_{1:n_x}, \mathbf{z}_{1:n_z} | \phi_x, \lambda, \sigma_x^2\right) = p\left(\mathbf{x}_{1:n_x} | \mathbf{z}_{1:n_z}, \phi_x, \lambda, \sigma_x^2\right) q\left(\mathbf{z}_{1:n_z} | \phi_z, \sigma_z^2\right),$$

with $q\left(\mathbf{z}_{1:n_z} | \boldsymbol{\phi}_z, \boldsymbol{\sigma}_z^2\right)$ describing the information super-imposed at coarse scale by the AR(1) dynamics of expected returns. Consistency across distributions is ensured by the Jeffrey's rule on conditioning information (see Jeffrey 1957, Jeffrey 1965 and Diaconis and Zabell 1982 for more details). Given independence, the full conditional distribution for the parameters of the fine scale and stochastic link equation can be written as

$$\begin{aligned} p\left(\phi_x, \sigma_x^2, \lambda | \mathbf{z}_{1:n_z}, \mathbf{x}_{1:n_x}\right) &\propto p\left(\mathbf{x}_{1:n_x} | \mathbf{z}_{1:n_z}, \phi_x, \lambda, \sigma_x^2\right) p\left(\phi_x, \sigma_x^2, \lambda\right), \\ &= \frac{p\left(\mathbf{z}_{1:n_z} | \mathbf{x}_{1:n_x}, \phi_x, \sigma_x^2, \lambda\right) p\left(\mathbf{x}_{1:n_x} | \phi_x, \sigma_x^2\right)}{p\left(\mathbf{z}_{1:n_z} | \phi_x, \sigma_x^2, \lambda\right)} p\left(\phi_x, \sigma_x^2, \lambda\right), \end{aligned} \tag{11}$$

similarly, the parameters of the observable coarse level can be sampled from the following full conditional distribution

$$p\left(\phi_z, \sigma_z^2 | \mathbf{z}_{1:n_z}\right) \propto q\left(\mathbf{z}_{1:n_z} | \phi_z, \sigma_z^2\right) p\left(\phi_z, \sigma_z^2\right),$$

[22] In the Appendix we generalize the estimation algorithm for $k = 1, \cdots, K$ scale-specific predictors. For the ease of exposition in the main text we represents the set of parameters for the case of $K = 1$.

Posterior simulation is implemented through a Gibbs sampler algorithm (Geman and Geman 1984). We propose a collapsed multi-move Gibbs sampling algorithm (see e.g. Roberts and Sahu 1997, Casella and Robert 2004, Ferreira et al. 2006 and Ferreira and Lee 2007), where expected returns and the parameters are sampled in blocks. More specifically we combine forward filtering backward sampling (see Frühwirth-Schnatter 1994 and Carter and Kohn 1994) to extract the expected returns from scale-specific forecasting variables, a Metropolis-Hasting step for direct inference on a set of parameters for which conjugacy is not obtainable and standard Bayesian updating for the parameters of the predictive regression and the observable predictors. A more detailed description of the MCMC algorithm and its convergence diagnostics are provided in Appendix A and B, respectively. In the following we also provide posterior estimates for each of the alternative prior specification detailed above.

4.2 A simulation-based discussion of the extracted expected returns

In this section we discuss, through simulation examples, the properties of the extracted expected returns obtained from our MCMC sampler scheme for the multi-scale time series model. For each simulation we run our MCMC estimation procedure so that we take into account both uncertainty on the structural parameters and the latent nature of expected returns.[23] In the interest of space, we focus mainly on the ability of the procedure to recover the latent series $x_{1:n_x}$ from some given coarse-level process $z_{1:n_z}$.

For the sake of completeness, we apply the procedure to two simulated data sets. The two simulation exercises differ in the way the dynamics of the observable scale-specific predictor $z_{1:n_z}$ is specified. For both exercise we assume that the initial, unrevised, process $x_{1:n_x}$ for expected returns follows an autoregressive processes of order one, with $\phi_x = 0.5$ and $\sigma_x^2 = 0.5$. We let $n_x = 720$ and $m = 12$, so that the fine and coarse grids can be interpreted as monthly and annual observations, respectively.

In the first exercise, we let the observable process $z_{1:n_z}$ obey Eq. (4), i.e. $z_{1:n_z}$ is truly obtained by temporally aggregating the latent AR(1) process.[24] Therefore, the revised distribution is

[23] Each simulation exercise is repeated fifty times, taking 20 hours on a WorkStation Intel(R) Xeon(R) with 3.20GHz, in total.

[24] Hence, in each simulation, we generate $x_{1:n_x}$ by simulating n_x observations from an AR(1) process; to obtain the series $z_{1:n_z}$, we aggregate the AR(1) process using the matrix A. We then use the observable $z_{1:n_z}$ to estimate the parameters and to extract the latent $x_{1:n_x}$.

consistent across scales, i.e. $q(z_{1:n_z}) = p(z_{1:n_z})$ with $p(z_{1:n_z})$ being the distribution implied by the latent state $p(x_{1:n_x})$ at fine level. In this case the multi-scale model does not affect the dynamic properties of the latent expected returns as all the information is contained at the finest grid of observation and no additional information is gained by considering lower frequency predictors. Thus, if the algorithm is correct, we should recover a standard AR(1) dynamics for $x_{1:n_x}$.

[Insert Figure 5 about here]

Panel A of Figure 5 shows that this is indeed the case. The autocorrelation structure of the extracted latent expected returns closely match that of the underlying standard AR(1) process. In fact, the time series dependence of the filtered process is even weaker than that of the true original dynamics. This is due to the short sample considered and the parameter uncertainty which is embedded in the autocorrelation estimates.[25]

In the second simulation exercise, we assume instead the low-frequency predictor $z_{1:n_z}$ follows an AR(1) process as in Eq. (2), with parameters $\phi_z = 0.9$ and $\sigma_z^2 = 1$. This implies that $p(x_{1:n_x}), q(z_{1:n_z})$, and $p\left(z_{1:n_z}|x_{1:n_x}\right)$ are inconsistent. Thus, unlike for the first simulation example, now the coarse-scale dynamics brings meaningful information about latent expected returns, $x_{1:n_x}$, and consistency across scales can be restored by exploiting the Jeffrey's rule of conditioning. In order to make the information at the coarse-scale highly relevant for the Jeffrey's revision, we fix the parameter $\lambda = 0.1$. This low value implies that we should revise substantially the distributions of $x_{1:n_x}$ toward higher level of persistence, as captured by a slowly decaying ACF.[26] Panel B of figure 5 shows that our estimated ACF obtained from the filtered $x_{1:n_x}$ lines up remarkably well with the true theoretical ACF of the revised process, i.e. $x_{1:n_x} \sim q\left(x_{1:n_x}\right)$ from Eq. (6). The latter falls within the 95% credibility region of posterior estimates, while the ACF of standard stationary linear AR(1) model with root ϕ_x falls apart completely.

To summarize, the evidence provided so far shows that the our modeling framework and the proposed estimation strategy are successful in recovering the time dependence structure of the latent expected returns $x_{1:n_x}$ regardless of the initial assumption about the mapping

[25] In an unreported simulation exercise with a longer time series the estimated autocorrelation are practical indistinguishable to each other in a statistical meaningful sense. Further results are available upon request.

[26] Recall Eq. (7): as $\lambda \to 0$ the covariance structure of the aggregated version of $x_{1:n_x}$ converges to that of the coarse-level process $z_{1:n_z}$. See also Figure 2 and Figure 3.

between fine- and coarse-scale processes. We conclude by observing that, in our second simulation exercise, the ACF of the latent process $x_{1:n_x}$, after revising its distribution using the low frequency information, decays in a geometric fashion rather than in an exponential way typical of autoregressive processes. More specifically, $ACF(k) \approx ck^{2d-1}$ and $d = 0.4$, ie. the process manifest long memory. As a by product of both simulation examples, we show that, despite the parsimonious structure of the model (we have only five parameters), we have been able to generate a much richer autocorrelation structure with respect to a standard AR(1) forward aggregated process. Also, results in Panel B make clear that a small number of resolution levels might be enough to induce an approximate long memory process at the finest time-resolution level.

5 Empirical analysis

The multi-scale time series model (1)-(3) is general and can accommodate a large set of alternative horizon-specific predictors with corresponding sizes m of the coarsening window. In the empirical analysis we focus on two well-known scale-specific predictors to extract the time series of latent expected returns, namely the log dividend-yield, i.e. dp_t, and the consumption-wealth ratio, i.e. cay_t, proposed by Lettau and Ludvigson (2001). We do not take a stand a priori on which predictor gives the highest information. Rather, our choice is motivated by the fact that cay_t and dp_t brings information about a range of frequencies which the empirical evidence shows not being overlapping (see Section 2). This allows a more clear identification of frequency-specific effects on the dynamics of expected returns. Also, arguably, cay_t and dp_t are widely used as benchmark in the forecasting literature.

Although our general specification does not rule out a priori any size of the coarsening window m across predictors, it is worth to carefully investigate ex-ante the "right" scale at which predictors operates in order to set m accordingly in the aggregation (3). A first indication is given by the forward-backward regression results, see Figure 1. The solid blue line with circles indicates that the predictive power of the consumption-wealth ratio sensibly increases for the one-to-two years horizon. This signals that a correct coarsening window would be either $m = 12$ or $m = 24$ for cay_t. Differently, the log dividend-price ratio shows its predictive power being almost constant until year four and steadily increasing afterward, justifying a coarsening window

in the range $m = 12, 48$ months. These results are in line with previous literature (see, e.g. Bandi et al. 2015 and the references therein).

We now couple the evidence from these forward-backward predictive regressions with a formal comparison of the marginal likelihoods of the multi-scale model for each of the two predictors and different sizes of m. In our setting, an analytical evaluation of the marginal likelihood is not possible. Gelfand and Dey (1994a) and Newton and Raftery (1994) showed that a simulation consistent estimate of the marginal likelihood for a model $\mathcal{M}_i$ is obtained by the harmonic mean of the likelihood values, evaluated at each draw of the parameters sampled from the corresponding full conditional distributions.[27] Table 1 reports the (log) marginal likelihoods for both predictors. Panel A shows the results for the log dividend-price ratio. The empirical evidence is in favor of a four-year coarsening window, against a two-year aggregation period, being the marginal likelihood for the case with $m = 48$ around 40% higher than for $m = 24$.

[Insert Table 1 about here]

Panel B shows that for the consumption-wealth ratio a one-year aggregation, i.e. $m = 12$, is preferred to a two-year one, i.e. $m = 24$, as indicated by a 15% higher marginal likelihood. This is consistent with the empirical evidence provided in Figure 1.

Combining the evidence provided by the in-sample R^2 and the marginal likelihood estimates, we assume throughout the rest of the empirical analysis that the mid-frequency fluctuations in monthly extracted expected returns are proxied by the annual consumption wealth-ratio cay_t, i.e. $m = 12$, while the lower-frequency fluctuations are captured by the four-year log dividend-price ratio dp_t, i.e. $m = 48$. The predictor at the annual frequency is constructed as the annual mean of the quarterly cay_t series. The dividend-price ratio is constructed as the total dividends paid by all stocks divided by the total stock market capitalization at the end of the year. The lower-frequency log dividend-price is defined as its mean over a four-year window.[28] The latent

[27] A potential issue in using the harmonic mean of posterior-implied conditional likelihoods is that the inverse likelihood does not have finite variance. This can creates issues, especially when the marginal likelihood cannot be computed in an efficient way (see, e.g. Chib 1995 and Chib and Jeliazkov 2001 for a detailed discussion on alternative efficient procedures). However, in our setting, the marginal likelihood $p\left(z_{1:n_z}|\phi_x, \sigma_x^2, \lambda\right)$ can be efficiently computed through the Kalman filter recursions (see Appendix A), mitigating potential concerns in using an harmonic mean approximation. In that respect, in computing the marginal likelihood we checked the stability of the conditional likelihood for each draw from the posterior distributions. Results of the draw specific likelihood evaluation are available upon request.

[28] We have also estimated the model under an alternative aggregation scheme. In particular, rather than taking the average of quarterly cay_t observation in a year, we use only the last observation at the end of the year.

expected returns are extracted at the monthly frequency to mimic a sensible time horizon for forecasting and investment decision making. The sample period is 1952:01-2013:12. Inference about the latent state and the parameters is made by implementing the MCMC algorithm described in Section 4 and detailed in Appendix A.[29]

5.1 The heterogeneous effect of scale-specific predictors

In this section we investigate the differential information content of our two scale-specific returns predictors, as characterized by posterior distributions of λ_{dp} and λ_{cay}. We super-impose an AR(1) dynamics to the time series of one-year consumption-wealth and four-year log dividend-price ratios. Panel A of Table 2 reports posterior summaries for the persistence ϕ_z and the conditional variance σ_z^2 for both scale-specific predictors, $z = \{cay, dp\}$.

[Insert Table 2 about here]

The magnitude of the posterior mean of ϕ_z and σ_z^2, for $z_t = \{cay_t, dp_t\}$ shows that predictors exhibit different degrees of persistence and volatility. The fact that $\hat{\phi}_{dp} \neq \hat{\phi}_{cay}^4$ highlights that there is no consistency between the mid- and low-frequency components captured by these two predictors. Although the dynamics of z_t does not guarantee coherence across scales, consistency will be restored within our multi-scale model by the Jeffrey's rule of conditioning. Similarly, posterior estimates of conditional variances $\hat{\sigma}_{dp}^2, \hat{\sigma}_{cay}^2$ show that shocks to the one-year consumption-wealth ratio have an impact four times bigger than shocks to the lower-frequency log dividend-price.

As shown in Section 3.1, although the magnitude of the posterior mean of $\hat{\phi}_z$ is large and significant, the degree of informativeness that observable lower-frequency predictors bring in estimating and revising the dynamics of expected returns largely depends on the between-levels uncertainty parameters $\lambda_{dp}, \lambda_{cay}$. In particular, in our model, a predictor associated with low λ

Analogously, rather than taking the average of dp_t over a 4-year window, we use only one observation every four years, starting with December 1955. In both cases, the posterior estimates of expected returns and the corresponding parameters do not change sensibly. Results are available upon request.

[29] To approximate the posterior distribution of the parameters and states we simulate 50,000 draws from the full conditional distributions detailed in Appendix A, and discard 20,000 initial draws as burn-in sample, and keep every other draw in order to increase the efficiency of the algorithm in approximating posterior summaries, such as mean, median and variances. The tuning parameters are set to $\delta_{\phi_x} = 0.01, \delta_{\sigma_x^2} = 1.5$ and $\delta_\lambda = 0.01$. Under this tuning the acceptance rates are 59%, 20%, 29% and 49% for $\phi_x, \sigma_x^2, \lambda_{dp}$ and λ_{cay}, respectively.

carries horizon specific information that helps understanding the dynamics of expected returns. The top panel in Figure 6 reports the posterior estimates of λ_{dp} and λ_{cay} obtained from different priors elicitation.

[Insert Figure 6 about here]

The benchmark prior assumes that a priori the four-year log dividend-price is informative, although such prior is rather flat (blue line with diamond markers). Alternative priors elicitation entail a complete flat distribution centered on the low informativeness of the four-year predictor (dashed green line), a tight prior centered on high informativeness (light blue line with squared markers), or tight prior centered on low informativeness (red line with circles). As we would expect, posterior estimates of λ_{dp} are somehow more sensitive to prior specifications. Let us recall that λ_{dp} is estimated using $n_x/m = 15$ data point only. Despite such minimum amount of information, posterior distributions are largely overlapping, with the magnitude of the differential in posterior means that is in the order of few decimal points, e.g. $\hat{\lambda}_{dp} = 0.013$ for the benchmark and $\hat{\lambda}_{dp} = 0.030$ for the flat prior. Overall, the left panel confirms the high informativeness of the long-term predictor in determining the dynamics of expected returns.

The top right panel in Figure 6 shows that, at least at the monthly frequency, the weight of the yearly average of consumption-wealth ratio, has a lower impact on expected returns. The average between-level uncertainty jumps to $\hat{\lambda}_{cay} = 0.42$ under our benchmark uninformative prior elicitation. Interestingly, the figure makes clear that posterior estimates are practically equivalent regardless the initial prior assumptions about expected informativeness and dispersion of the consumption-wealth ratio. In fact, posterior estimates of $\hat{\lambda}_{cay}$ obtained from the benchmark prior show that the annual consumption-wealth ratio bring around 40 times less information with respect to the four-year log dividend-price.

Consistent with the discussion in Section 3, the posterior estimates of Figure 6 suggest that by jointly considering the differential informational content of the consumption-wealth ratio and the log dividend-price may have first order implications in characterizing the high-frequency dynamics of latent expected returns. Figure 7 confirms this intuition.

[Insert Figure 7 about here]

The blue line with diamonds represents the time series of expected returns at monthly frequency extracted by jointly using the annual information contained in cay_t and the lower-frequency information carried by dp_t. We compare the extracted expected returns from the joint scale-specific predictors with two alternative processes for expected returns, extracted by using only one coarse scale at a time: the expected return series one would obtain using only cay_t at yearly frequency (black line with circles), and the series obtained using only dp_t at the four-years frequency (red dashed line). Figure 7 makes clear that expected returns series filtered from the joint observation of the consumption-wealth and price-dividend ratios accommodate multiple frequency-specific features at once. In the first part of the sample, until the mid 70's, our expected returns series follows closely the (expected returns extracted from just using the) consumption wealth-ratio; it then inherits the trend behavior of the dividend-price ratio until the late 90's but without loosing interesting variations at higher frequency; eventually, our series tracks again the consumption-wealth ratio in the last part of the sample. In other words, by conditioning on the joint scale-specific predictors we allow to model features of expected returns that inherit information coming from different frequency-specific shocks all at once. In the following, we show that this result has first order implications for forecasting and optimal portfolio allocations in the long-run.

5.2 Persistence of expected returns

In this section we discuss the posterior estimates of the expected returns and the corresponding parameters. Importantly, since our estimates for λ_{dp} and λ_{cay} imply a high degree of agreement between coarse and fine levels, the parameter ϕ_x cannot be interpreted as the estimated autoregressive parameter and, thus, we cannot conclude the degree of persistence for expected returns just by inspecting it. We show below that it is the interaction of ϕ_x with the parameters driving *between* resolutions uncertainty, λs, to determine the persistence of discount rates. Bottom panels of Figure 6 show the posteriors of ϕ_x and σ_x^2 computed from alternative priors distributions.

As a whole, posterior estimates are robust to alternative prior specifications. By assuming either a flat or informative prior on either low or high location, the posterior mean $\hat{\phi}_x$ is practically equivalent. This formally implies that the null hypothesis that posterior means

are statistically equivalent under the alternative priors cannot be rejected at a meaningful significance level. Interestingly, priors elicitation play a more significant role in the estimation of the conditional variance of expected returns σ_x^2. Panel B of Table 2 shows that the posterior mean obtained from our benchmark prior is around $\hat{\sigma}_x = 2.4$. Such prior is uninformative and centered around the assumption that shocks to expected returns are typically low in magnitude. Increasing prior informativeness for the low variance case shrinks the posterior mean by a half. However, such prior takes the extreme view that $\sigma_x^2 \approx 0$ a priori, namely, expected returns are constant over time. Finally, imposing a large variance σ_x^2 a priori, generates posterior estimates which are fairly overlapping with respect to the benchmark case. We show below that for these cases the autocorrelation function of the expected returns is equivalent to our benchmark.

At the outset of the paper we argue that a low λ implies low uncertainty at the coarse level due to agreement with the latent fine level, and induces the dynamics of the extracted latent expected returns to be (after revision using the Jeffrey's rule of conditioning) farther away with respect to those of a standard AR(1) process. Figure 8 displays the ACF of our extracted series at monthly horizons (Panel A), as well as the one from the series aggregated over longer horizons, namely 1-year (Panel B) and 4-years (Panel C). The plot compares these ACFs with that of a standard AR(1) process with autoregressive coefficient equal to the posterior mean value $\hat{\phi}_x$ as shown in Table 2.

[Insert Figures 8 about here]

Although a simple AR(1) process for the monthly series seems to be appropriate to describe the first 5 lags, Panel A shows that the series of expected returns extracted from multiple coarse levels exhibits strong dependence at long lags. Panels B and C further highlight this fact by aggregating our extracted expected returns and comparing its ACF to that implied by the aggregation of an AR(1) process. The Figure suggests that the dependence of expected returns to shocks distant in the past is typically higher than that implied by assuming a simple AR(1) dynamics. It is worth to mention that such long-memory-type of property is generated with a highly parsimonious structure that involves the estimation of only three parameters.[30] We also remark that this long-memory feature is the result of modeling the time series from lower

[30]The parameters ϕ_z and σ_z affect only the forecasting. They do not directly enter the ACF of the latent fine-level process, which is only determined by $\phi_x, \sigma_x, \lambda_z$.

to higher frequencies, and of an estimation algorithm that incorporates information at multiple time scales in a coherent way.

One may argue that alternative ARMA specifications might be able to capture low frequency shocks in expected returns as well as our multi-scale model. Therefore, to further challenge our results, we compare the ACF of the expected returns extracted by the joint scale-specific process of the consumption-wealth and the log dividend-price ratio with the one implied by an ARMA(2,1) process. The ARMA(2,1) is an interesting example for two reasons. First, it has the same number of parameters as our model. Second, an ARMA(2,1) can be thought of as the process resulting from summing up two AR(1) with different persistence roots. Thus, one expects the ARMA(2,1) process to be able to capture multiple, low frequency fluctuations in time series. Figure 9 reports the results.

[Insert Figures 9 about here]

Panel A shows that, although the ARMA(2,1) is comparable now up to the first 8 lags, the long-memory-type of persistence of our extracted expected returns remains unmatched as we go further in investigating the impact of past shocks. As before, Panels B and C confirm this evidence by aggregating the different specifications of $\hat{x}_{1:n_x}$ over longer horizons.

Although posterior estimates are comparable across different prior specifications, the results in terms of autocorrelations can be more sensitive. To address this concern we also compute the entire set of autocorrelations for the alternative priors as shown in Figure 6. Figure 10 reports the results.

[Insert Figures 10 about here]

The top left panel shows that the autocorrelation of the extracted expected returns at the monthly frequency is not sensitive to the prior specification of the parameter ϕ_x. In particular the dependence at long lags is present independently of the prior specification. This is becomes apparent once expected returns are aggregated on a 4-year horizon, see top right panel: over this aggregation horizon, the departure of the ACF of expected returns from that of a standard persistent AR(1) is present across different priors. The bottom left panel shows the ACFs for alternative priors on the conditional variance of expected returns σ_x^2. The pattern is consistent

for different prior specifications except for the case with strong beliefs on $\sigma_x^2 \approx 0$. Such prior embeds however the rather extreme view that expect returns are constant over time, a priori. Despite this extreme view, posterior estimates are still consistent with the logic of our multi-scale model as shown on the bottom right panel for the 4-year expected returns. Indeed, although with a higher rate of decay, the persistence of shocks to expected returns is sensibly higher than the one implied by a classic AR(1) specification computed with the same ϕ_x.

The analysis so far suggests a long range dependence behavior for the monthly series of expected returns revised by the scale-specific information brought by two distinct predictors. However, although our expected return series emulates a long memory process, it is different in essence. In our framework, the long memory type of behavior is explicitly modeled as a result of an information cascade from low- to -higher frequency predictors. Thus, the main advantage of our multi-scale framework over actual long memory models is increasing interpretability in the light of economic theory.

5.2.1 Comparison with OLS predictive regressions

In this section we address the concern that our extracted expected returns may simply approximate the fitted value from a standard predictive regression. We first run a simple regression of quarterly returns on the predictor z_t:

$$r_{t+1} = \alpha + \beta z_t + \varepsilon_{t+1} \ ,$$

where z_t is either the consumption-wealth ratio, cay_t, or the log dividend-price ratio, dp_t. In this case the expected returns from OLS regressions, $\widehat{\mu}_t = \widehat{\alpha} + \widehat{\beta} z_t$ inherits the correlation structure of the chosen predictor. We also run a multiple regression of quarterly returns on both cay_t and dp_t,

$$r_{t+1} = \alpha + \beta_{cay} cay_t + \beta_{dp} dp_t + \varepsilon_{t+1} \ , \tag{12}$$

so that in this case the ACF of expected returns, $\widehat{\mu}_t = \widehat{\alpha} + \widehat{\beta}_{cay} cay_t + \widehat{\beta}_{dp} dp_t$, will depend on the relative importance of the predictors at the chosen quarterly forecasting horizon. Figure 11 compares the ACF of our expected returns series with the fitted value from the dividend-price regression (black line with square markers), and with the fitted value from the consumption-

wealth regression (red dashed line with circles), as well as with the fitted value from multiple regression on consumption-wealth and dividend-price ratios (green line with triangle markers).

[Insert Figures 11 about here]

In all cases we observe that the ACF of our multi-scale filtered expected returns is different from the ones implied by OLS predictive regressions. First, the multi-scale ACF has lower first-order autocorrelation compared with OLS fitted expected returns. Second, and more importantly, the rate of decay of the ACF for our multi-scale process lies in between those of the close-to-unit root dp_t series and the fast mean-reverting cay_t. This fact is due to the parameters $\lambda_{dp}, \lambda_{cay}$ which control how much of the coarse-level scale-specific information from predictors influence shorter-term expected returns. Finally, the long range dependence in our series at lags of about 8-years coincide with that implied by the fitted value of an OLS regression exploiting both predictors.

Figure 12 portrays the extracted expected returns implied by our multi-scale mode against the fitted values obtained from a simple OLS regression with the log dividend-price ratio (Panel A, black line with squares), or with the consumption-wealth ratio (Panel B, red dashed line with circles) as the sole predictor variable, as well as against the OLS fitted values computed conditioning on both predictors (Panel B, green line with triangles). Panel A shows that the expected return series from a standard predictive regression based on the log dividend-price ratio partly tracks (the mean of) our extracted expected returns. However, the former breaks the 95% credibility intervals across late 70s and early 80s, as well as throughout early 2000. This is due to the fact that, unlike our multi-scale model, a standard OLS regression with a low-frequency predictor as the log dividend-price ratio cannot capture the higher-frequency fluctuations from the consumption-wealth ratio, a fact that makes our series more transient in the short term as documented from the first few lags of the ACFs shown in Figure 11.

[Insert Figures 12 about here]

Interestingly, Panel B makes clear that by choosing a quarterly horizon for the OLS multivariate predictive regression (12), the fitted value favors consumption-wealth as driving factor of

expected returns, as opposed to the log dividend-price ratio. As a matter of fact, the fitted value from the regression that uses both cay_t and dp_t overlaps almost completely with that from the regression using cay_t as the sole predictor. This is important because it shows that a standard OLS framework is unable to resolve the tension between the horizon-specific information embedded in the predictors. This in turns affects optimal portfolio decisions which are, by construction, influenced by the predictive power of the forecasting variables used. More specifically, panel B of Figure 12 implies that a long-term investor that rebalances her portfolio at quarterly frequency, will likely choose the parameters derived from a quarterly return forecasting regression, thus favoring those fluctuations induced by cay_t only. This is true despite the fact that the dividend-price ratio arguably contains dominating information about longer-run fluctuations in returns. In that respect, by jointly combining different horizon-specific layers of information in extracting expected returns, our multi-scale time series model allows to reconcile the tension between the heterogeneous information content of predictors and the time frame typically assumed for investment decision making.

5.3 Forecasting future expected returns

The simulation exercise in Section 3.2 shows that our multi-scale model helps to better capture future fluctuations in expected returns. We now check the forecasting ability of the model in a realistic setting. More specifically, we compare the m-step ahead predictive performance of our multi-scale model against the performance of an iterated forecasts from an AR(1) benchmark. Forecasts are computed from the marginal predictive distribution

$$p\left(x_{t+1} : x_{t+ml} \mid x_{1:t}, z_{s+l}\right) = \int p\left(x_{t+1} : x_{t+ml} \mid x_{1:t}, z_{s+l}, \Theta\right) p\left(\Theta | x_{1:n_x}, z_{1:n_z}\right) d\Theta \ , \qquad (13)$$

for $s = 1, \cdots, n_z$ and $t = 1, \cdots, n_x$. The conditional predictive $p\left(x_{t+1} : x_{t+ml} \mid x_{1:t}, z_{s+l}, \Theta\right)$ is obtained by iterating $l = h/m$ times the m-step ahead predictive distribution, see Eq. (10), which in turn is obtained by conditioning on the l-step ahead forecast from the coarse-level predictive, see Eq. (9). The marginal $p\left(x_{t+1} : x_{t+ml} \mid x_{1:t}, z_{s+l}\right)$ is then approximated by using an importance sampling (IS) estimator (see, e.g. Geweke 2005, and Geweke and Amisano 2012), in which sampling draws from the posterior distribution of the parameters are obtained using the

entire sample of observable predictors and latent expected returns.[31] In order to compare our results with existing research, we calibrate the persistence parameter of the competing AR(1) to a high value of 0.98, as commonly used in the literature (see, e.g., Table II in Barberis 2000). Figure 13 shows the model-implied future expected returns $E_t\left[x_{t+1:t+h}\right] = x_{t+1:t+h|t}$ for the years 1954-1959, 1978-1983, 1988-1991, and 2007-2011 which represents the period across the great financial crisis.

[Insert Figures 13 about here]

More specifically, at $t = 1953, 1977, 1987$ and $t = 2006$ we compute the $h = 48$ step ahead forecasts, which coincides with the coarsening window in (3).[32] Those forecasts are compared with the series of extracted latent expected returns for 1-to-48 months ahead (solid blue line). It is noteworthy that the forecasts of the multi-scale model (magenta line with diamond markers) are quite different from the one produced by the autoregressive model (red line with circles). While the forecasts of the AR(1) always have a monotone exponential decay shape towards the unconditional mean of the series, the forecasts of our multi-scale expected returns have more interesting non-linear shapes which better anticipate future patterns. Also, the figure highlights how the standard AR(1) model give poor medium term, 1-to-48 months, ahead forecasts.

Figure 13 suggests that by incorporating information at multiple horizons has implications for forecasting future expected returns. We now formally investigate this result on the basis of two complementary measures of forecasting accuracy. We first compute a standard Mean Squared Error (MSE) which allows to investigate the performance of alternative methodologies in terms of point forecasts. Gneiting (2011) showed that MSE is a consistent evaluation measure when the point forecast equals the mean of the predictive distribution. As such we compute point forecasts for each model as the average of the corresponding distribution of future expected returns predictions. Panel A of Table 3 compares the MSE obtained from our multi-scale model, against the one implied from an AR(1) fitted on the series of latent expected returns, and an AR(1) in which the persistent parameter is restricted to imply high persistence of current

[31] The predictive density (13) can be alternatively approximated with an harmonic mean estimator; see Gelfand and Dey (1994b), or the truncated elliptical mean estimator in Sims et al. (2008). However, conditional on using the same set of parameters across forecasting horizons, namely the parameters of the model do not change as h increases, the predictive likelihood can be consistently estimated directly by using draws from the posterior. In that respect the importance-sampling estimator is unbiased, unlike, for instance, the mean harmonic estimator (see, e.g. Geweke 2005 for more details).

[32] In the portfolio allocation exercise below we extend the forecasting horizon to $l > 1$ coherently with the idea of modeling investment decision at longer horizons.

expectations $\phi_x = 0.98$;

[Insert Table 3 about here]

The MSE obtained from the multi-scale model is lower throughout the forecasting horizons $h = 12, 24, 36, 48$. When compared with the persistent AR(1) model, the use of multiple time scales reduces the mean squared prediction error for $h = 48$ from 5.27 to 4.28, a reduction of about 23%. Similarly, the 48-step ahead forecast obtained from OLS estimates of an AR(1) generates a prediction error which is about 15% higher than the one obtained from the multi-scale time series model for the same horizon. This result is due to the fact that forecasts at the coarse level for the predictors guide the forecasts at the fine level for expected returns; see Eq. (10), reducing the prediction error.

Although point forecasts of future expected returns reveal interesting aspects of the predictive ability of our multi-scale model, such conditional means cannot provide insight into the uncertainty that is associated with producing these forecasts. In that respect, a direct evaluation of the marginal predictive likelihood (13) is a natural tool to assess the ability of the multi-scale time series model to predict unusual developments in future expected returns, such as the likelihood of large drops or jumps in future expectations given current information. We therefore also evaluate the forecasting accuracy of our model, compared to the AR(1) specifications as above, on the basis of a log predictive score measure. The log predictive score for the model $\mathcal{M}_i$, for the $h = 48$ steps ahead prediction of expected returns can be expressed as

$$S_{\mathcal{M}_i}(h, n_z) = \sum_{s=1}^{n_z-1} \log p\left(x_{sm+1} : x_{sm+h} \mid x_{1:sm}, z_{s+1}\right) \ , \tag{14}$$

and is computed by evaluating the marginal predictive distribution (13) with the extracted latent expected returns over the prediction sample. Panel B of Table 3 shows the results. The log-predictive score highlights another interesting feature of our multi-scale time series model. The forecasting performance of the model at shorter horizons is sensibly higher for the multi-scale model. Indeed, the log-predictive score is around ten times higher at $h = 12$, more than 40% higher at $h = 24$ months ahead, about 60% higher for $h = 36$, and around 6% higher at $h = 48$. This is consistent with the prediction pattern showed in figure 13. Although, to some extent, the prediction at $h = 48$ is comparable, the fact that short-term forecasts are obtained

by directly including lower frequency information allows to increase the ability of the model to anticipate shorter-term future expected returns. As a whole, Table 3 makes clear that our multi-scale time series model specification allows to better capture both average and unusual developments in future expected returns.

One last remark is in order. In traditional time series analysis, the economic agent defines on which time scale to analyze the time series depending on the decision interval and/or investment horizon. She then finds a reasonable model and performs forecasts at that scale; however, in general, that model will not be reliable for performing forecasts at other time scales. In contrast, our multi-scale model is able to forecast reliably at all levels of time resolution that are included in the model. We next move to study the consequences of multi-scale modeling and forecasting for portfolio allocation.

5.4 Portfolio allocation for a buy-and-hold investor

One of the key advantage of the multi-scale time series mode is to jointly combine the heterogeneous information content from different horizon-specific predictors to construct higher-frequency expected returns. Intuitively, this allows to reconcile the tension between the time scale at which forecasting variables show their highest predictive power and the time frame that is typically assumed for investment decision making. In this section we investigate further this implication on the basis of a standard asset allocation framework for buy-and-hold investor (see, e.g. Barberis 2000 and Avramov 2004).

We assume the investor has a power utility with risk aversion γ and at time $T = n_x$ she has to decide the optimal allocation of wealth in the risky asset with an horizon of h months. There are two assets: the continuously compounded real return on the value-weighted market portfolio in month t, and the real monthly return on Treasury bills r_f. For simplicity, we suppose that the return on T-Bills is a constant (this is similar to, e.g., Barberis 2000). Future, i.e. at $T + h$, realized returns in excess of the risk-free rate are related to expected returns $E_T\left[r_{T+1:T+h}\right] = x_{T+1:T+h|T}$ as follows

$$r_{T+1:T+h} = \alpha_h + \beta_h x_{T+1:T+h|T} + \varepsilon_{T+1:T+h}, \qquad \text{with} \qquad \varepsilon_{T+1:T+h} \sim N\left(0, \sigma_h^2\right), \tag{15}$$

with $x_{T+1:T+h|T}$ denoting the cumulated forecast obtained from the marginal predictive distribution (13). In that respect, our portfolio allocation is purely out-of-sample as it is derived conditioning on the information available at time T.[33]

It is worth to notice that the predictive regression (15) allows to directly compare the optimal allocation obtained from our multi-scale time series model with the one an investor would obtain assuming all the information lies at the highest possible, in this case monthly, frequency. The latter implies that future expected returns are obtained by iterating short-run forecasts. If initial wealth is $W_T = 1$ and $w \in (0, 1)$ is the allocation in the market portfolio, then the end-of-horizon wealth is given by

$$W_{T+h} = (1 - w) \exp(r_f h) + w \exp(r_f h + r_{T+1} + \cdots + r_{T+h}),$$

We assume that investor's preferences over terminal wealth are described by a constant relative risk aversion utility function of the form

$$U(W_{T+h}) = \frac{W_{T+h}^{1-\gamma}}{1-\gamma},$$

with $\gamma > 1$ the coefficient of relative risk aversion. Let $r_{T+1:T+h}$ be the cumulated excess stock return over the investment horizon h, the buy-and-hold optimal allocation is then represented by the solution of the problem

$$w^\star = \max_w E_T \left(\frac{((1 - w) \exp(r_f h) + w \exp(r_f h + r_{T+1:T+h}))^{1-\gamma}}{1-\gamma} \right),$$

E_T denotes the fact that investors calculate their expectations conditional on the information available at time T. The portfolio choice is obtained by averaging the cumulative predicted returns over the corresponding marginal posterior draws obtained from the MCMC algorithm outlined in Section 4 and detailed in Appendix A.

[33] Notice that posterior estimates of the regression parameters are obtained by estimating a forward-backward regression for horizons $h = 1, ..., 120$ for each draw of the in-sample extracted expected returns $x_{1:n_x}$. As far as posterior estimates are concerned, we specify a conjugate independent normal-inverse-gamma prior structure;

$$\alpha, \beta \sim N(m_\beta, M_\beta), \qquad \sigma^2 \sim IG(\nu_\beta/2, \nu_\beta s_\beta/2),$$

with m_β, M_β the location and scale hyper-parameters of the regression coefficients, and $\nu_\beta, \nu_\beta s_\beta$ the degrees of freedom and scale hyper-parameters for the conditional variance if the idiosyncratic error σ^2. Marginal priors do not depend on the investment horizon as the functional form of the parameters does not change for any $h > 1$ (see Appendix A for more details on the estimation procedure).

At the heart of this paper is the issue of which information the investor uses to compute this expectation. As discussed in Section 3.2 and 5.3, expectations on future returns are very different depending on whether we acknowledge the information content of multiple scale-specific predictors. In order to disentangle the impact of horizon-specific predictors we show the outcome of the buy-and-hold portfolio allocation the conditions on the mid-term consumption-wealth ratio only, or on the long-run log dividend-price ratio only, or both. For the sake of comparison with existing literature we also compute the optimal allocation implied by a monthly compounded dividend-price ratio (see, e.g Barberis 2000 and Pástor and Stambaugh 2012). Figure 14 show the optimal portfolio allocation for an investor with a relative risk aversion equal to $\gamma = 5$ and investment horizon equal to $h = 120$ months. For the sake of completeness we investigate the optimal allocations under different settings.

[Insert Figures 14 about here]

Top panels report the exposure to the risky asset computed by fixing current expected returns on the last sample observation and the value of the lower frequency predictor is either at their sample minimum (left panel) and maximum (right panel).[34] The figure makes clear that by conditioning on both the four-year log dividend-price ratio and the one-year consumption wealth ratio the optimal allocation (blue line with diamonds) takes an out-of-sample path which inherits the dynamics of both dp_t (light blue line with squares) and cay_t (red line with circles). Consistent with our initial intuition depicted in Figure 1, in the short term the information predominant comes from cay_t while dp_t dominates in the longer run. Interestingly, by directly using dp_t as perfectly characterizing monthly expected returns, the investor would prefer a rather flat exposure to risky assets across the investment horizon. Such behavior is consistent with the close-to-unit-root evidence shown in Figure 11.

As shown in section 3.2 the multi-scale prediction implies that future expected returns strongly depends on forecasts for the coarse level information. As such, we would expected that the current value of the lower frequency information is indeed crucial in driving investor's future expectations. The bottom panels of Figure 14 confirm this intuition by now fixing the last extracted expected returns to their minimum (bottom left) or maximum (bottom right)

[34] In the standard predictive regression case the last value of the predictor also corresponds to the last value of expected returns.

values. When compared with the respective top panel, we observe that the optimal allocation does not change sensibly, for the same given last observations of economic predictors. In this sense, the frequency-mixing feature of the optimal allocation is preserved.

6 Concluding Remarks

We define expected returns as the outcome of a multiple time-scale information updating process, and investigate whether the dominant view that long-run investors' expectations are the results of an aggregation of short-term expected returns is backed up by empirical evidence. Methodologically, we propose a parsimonious multi-scale time series model that allows to combine information coherently across horizons (or levels of resolution) to extract the dynamics of univariate expected returns. By doing so, we allow the researcher to reconcile the tension between the heterogeneous information content of different scale-specific predictors, with the time frame that is typically considered in empirical finance applications such as forecasting and portfolio allocations.

Empirically, we show that simultaneously considering the information cascade that comes from low- to high-frequency predictors, allows to identify long-lasting effects in the dynamics of expected returns. Importantly these effects cannot be captured by standard persistent ARMA processes. Finally we show that combining information at multiple horizons to extract the series of expected returns has dramatic implications for forecasting and, thus, it is of first order importance for investment decision purposes. We quantify these effects along several dimensions.

Although we investigate the properties of the multi-scale time series model within an asset pricing and portfolio decisions context, our framework is relevant in any setting where it is important to incorporate information available at different time scales in the inference about the process of interest. For instance, our framework with its novel forecasting dynamics might be of interest to government and budget office that need to forecast productivity and GDP over long-horizons, and for central banks that need to incorporate information about the medium- and long-term dynamics of output and inflation to set interest rates. We view these as promising avenues for future research.

Figure 1. Predictive Regressions

Horizon-specific predictive regressions. This figure shows the in-sample R^2 of simple linear regressions (with an intercept) of h-period continuously compounded market returns on the CRSP value-weighted index on h-period past consumption-wealth ratio (solid line with circles) and on h-period past log price-dividend ratio (solid line with asterisks). For the consumption-wealth ratio regressions, the sample is quarterly and spans the period 1952Q1-2013Q4. For the dividend-price ratio regressions, the sample is annual and spans the period 1952-2013.

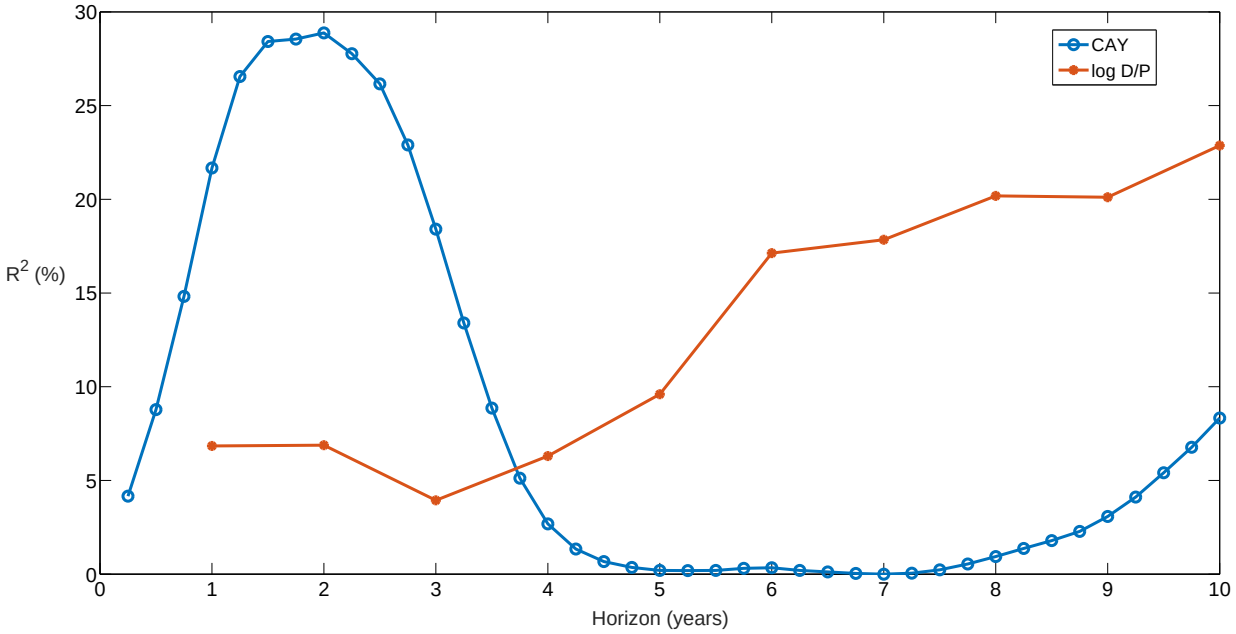
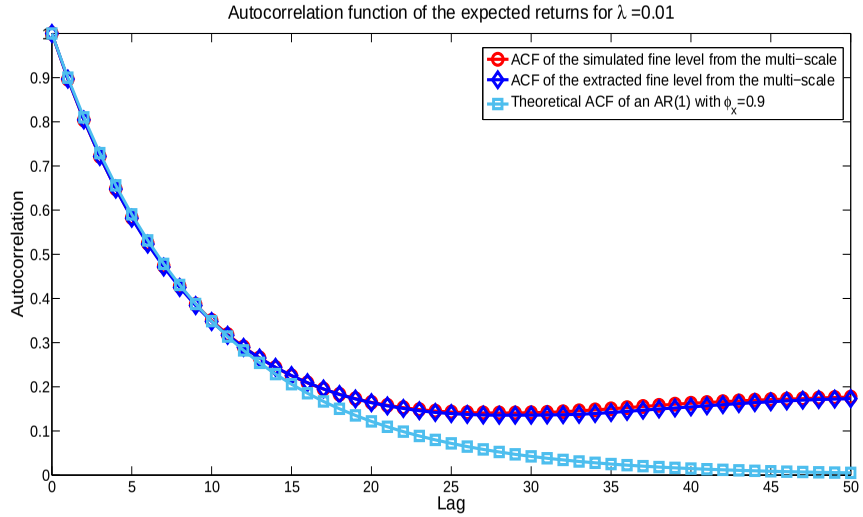
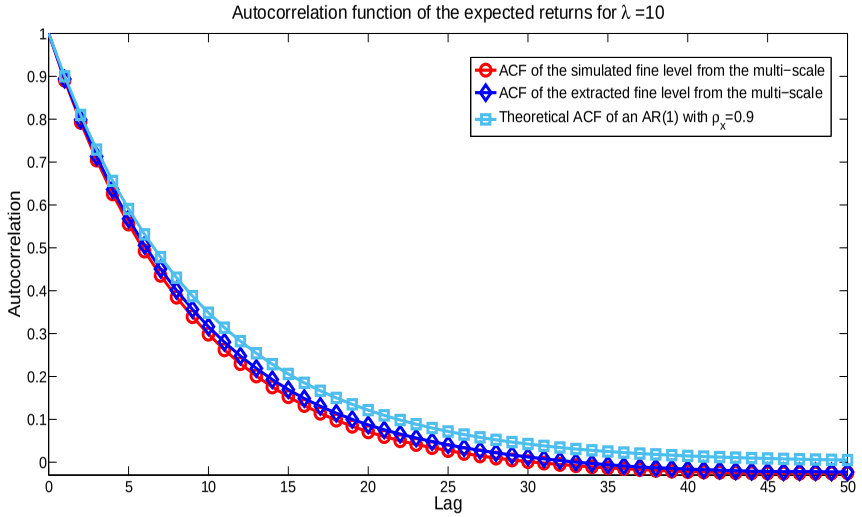
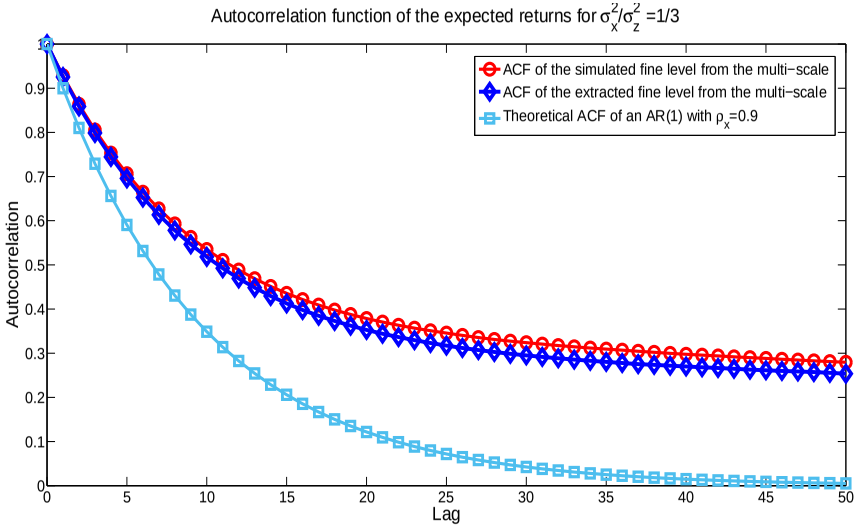


Figure 2. Autocorrelation Functions of Multi-scale time series Vs AR(1)

Multi-scale time series Vs AR(1): autocorrelation functions. This figure shows the theoretical autocorrelation function for a simulated multi-scale time series $x_{1:n_x}$ (red line with circles), the one extracted by conditioning on the lower-frequency series z (blue line with diamonds), and the theoretical autocorrelation of an AR(1) with the same autoregressive parameter as the one used to simulate the multi-scale process (cyan with squares). The parameters $\phi_x = \phi_z = 0.9$ are fixed across panels. The top left panel shows the results assuming $\lambda = 0.01$ and $\sigma_x^2 = \sigma_z^2 = 1$. The top right panel shows the results for a simulation with $\lambda = 10$ and $\sigma_x^2 = \sigma_z^2 = 1$. The bottom right panel shows the results for a simulation with $\lambda = 0.01$ and $\sigma_x^2/\sigma_z^2 = 3$. The bottom left panel shows the results for a simulation with $\lambda = 0.01$ and $\sigma_x^2/\sigma_z^2 = 1/3$. We simulate $n_x = 720$ with $m = 48$ in all cases.







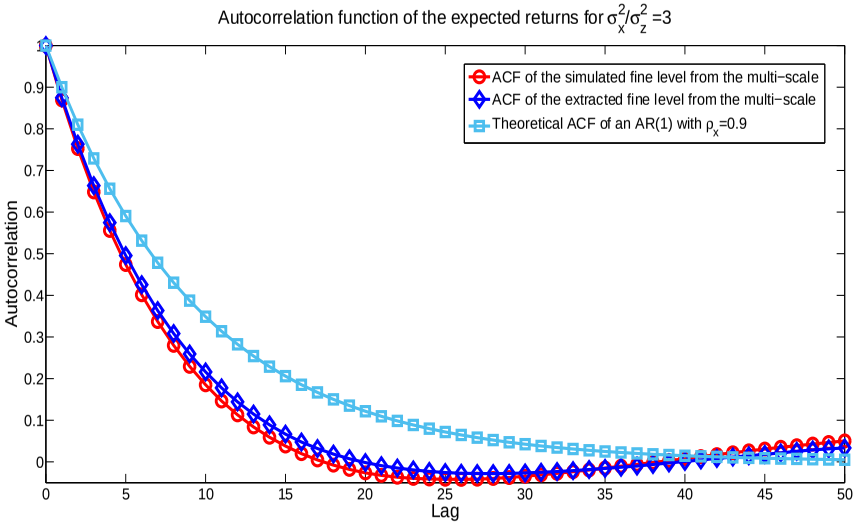


Figure 3. Autocorrelation Functions of Multi-scale time series Vs AR(1): Four-year aggregation

Multi-scale time series Vs AR(1): autocorrelation functions of four-year aggregations. This figure shows the theoretical autocorrelation function for a simulated multi-scale time series (red line with circles) aggregated over four years, the four-year extracted expected returns by conditioning on the lower-frequency series z (blue line with diamonds), and the theoretical four-year autocorrelation of an AR(1) with the same autoregressive parameter as the one used to simulate the multi-scale process (cyan with squares). The parameters $\phi_x = \phi_z = 0.9$ are fixed across panels. The top left panel shows the results assuming $\lambda = 0.01$ and $\sigma_x^2 = \sigma_z^2 = 1$. The top right panel shows the results from the simulation expected returns with $\lambda = 10$ and $\sigma_x^2 = \sigma_z^2 = 1$. The bottom right panel shows the results for a simulation with $\lambda = 0.01$ and $\sigma_x^2/\sigma_z^2 = 3$. The bottom left panel shows the results for a simulation with $\lambda = 0.01$ and $\sigma_x^2/\sigma_z^2 = 1/3$. We simulate $n_x = 720$ with $m = 48$ in all cases.

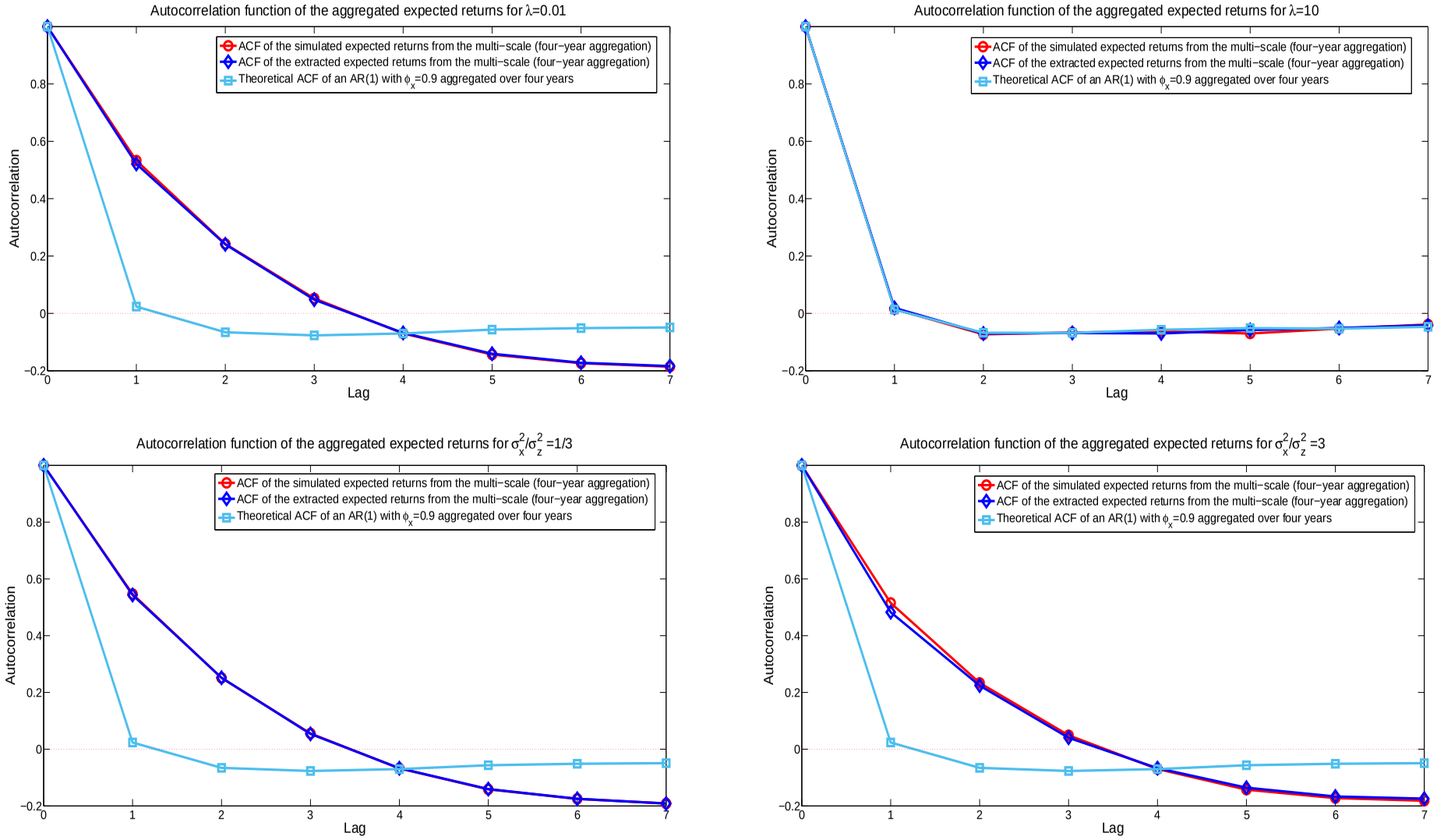
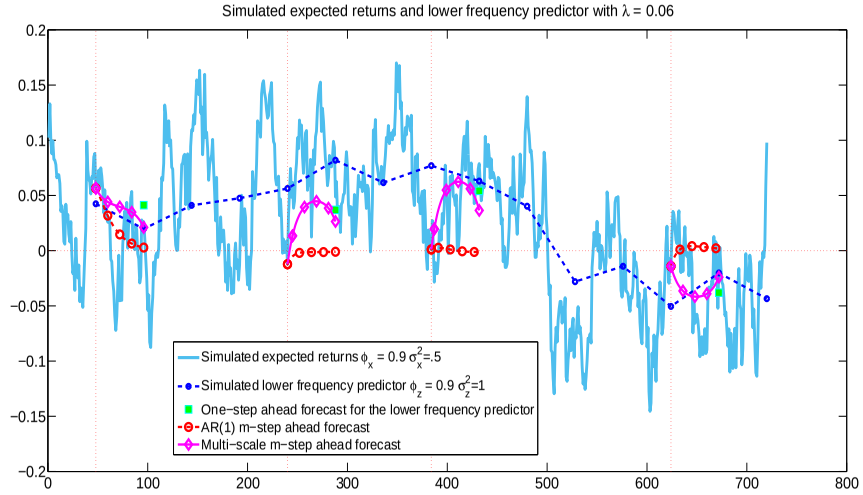
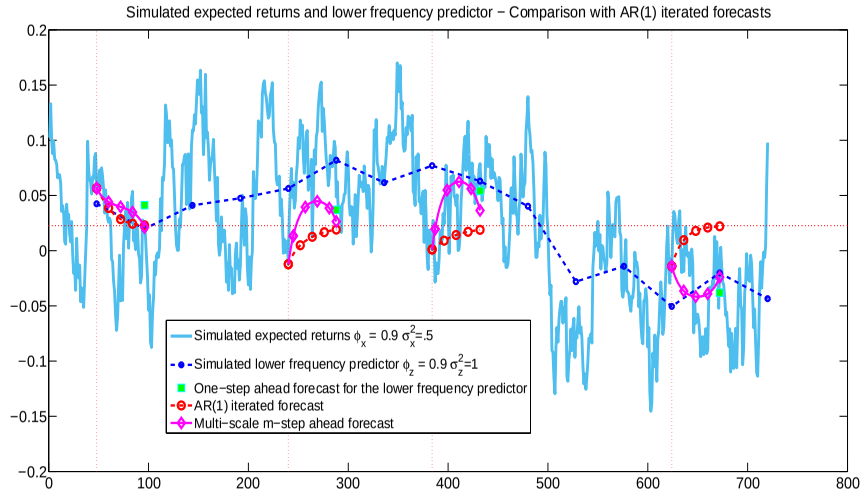
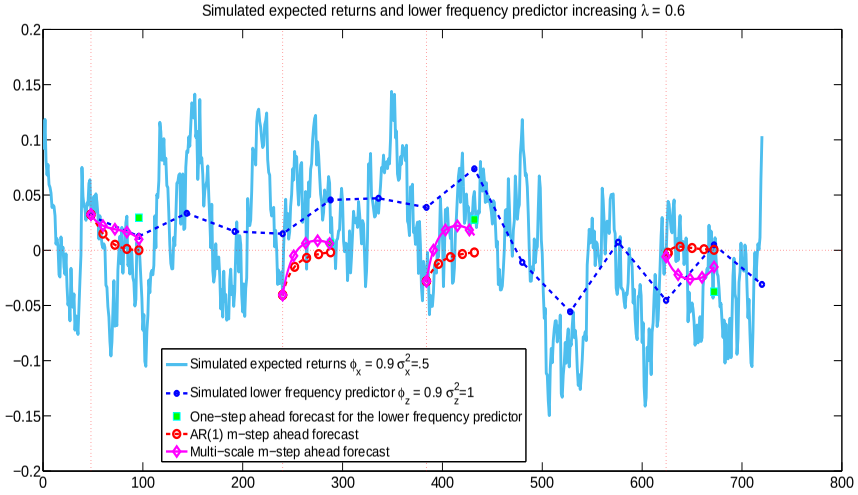


Figure 4. Multi-scale Vs AR(1) Forecasts: Simulation Exercise

Comparing forecasts in simulation. This figure shows the forecasts obtained by the multi-scale time series model in comparison with those obtained by a highly persistent AR(1) model. Except for the between-levels uncertainty all other parameters are fixed, with $\phi_x = \phi_z = 0.9$, $\sigma_x^2 = .5$ and $\sigma_z^2 = 1$. Top left (right) panel shows the m-step ahead forecast obtained by fixing $\lambda = 0.06$ compared with the m-step ahead (iterated) prediction obtained from an AR(1) with $\phi_x = 0.98$. Bottom left (right) panel shows the m-step ahead forecast obtained by fixing $\lambda = 0.6$ ($\lambda = 6$) in comparison to the m-step ahead forecast obtained from an AR(1) with $\phi_x = 0.98$. The results are obtained by averaging forecasts on 20,000 patterns for predictors and expected returns.







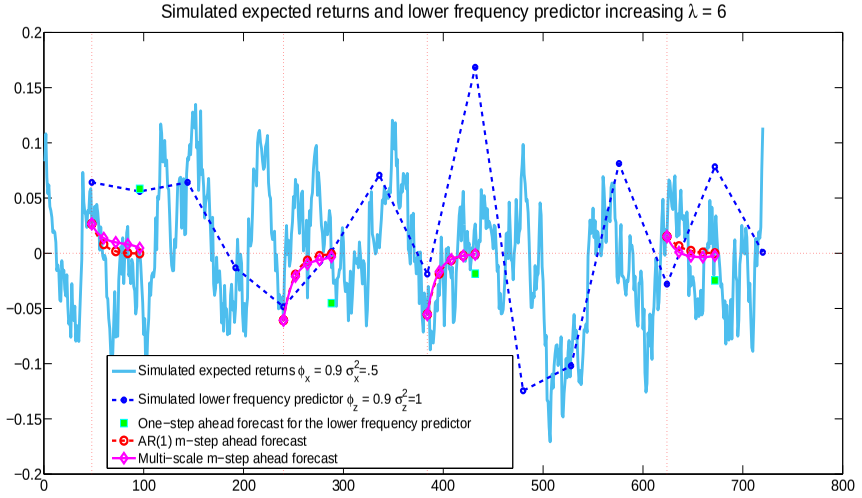
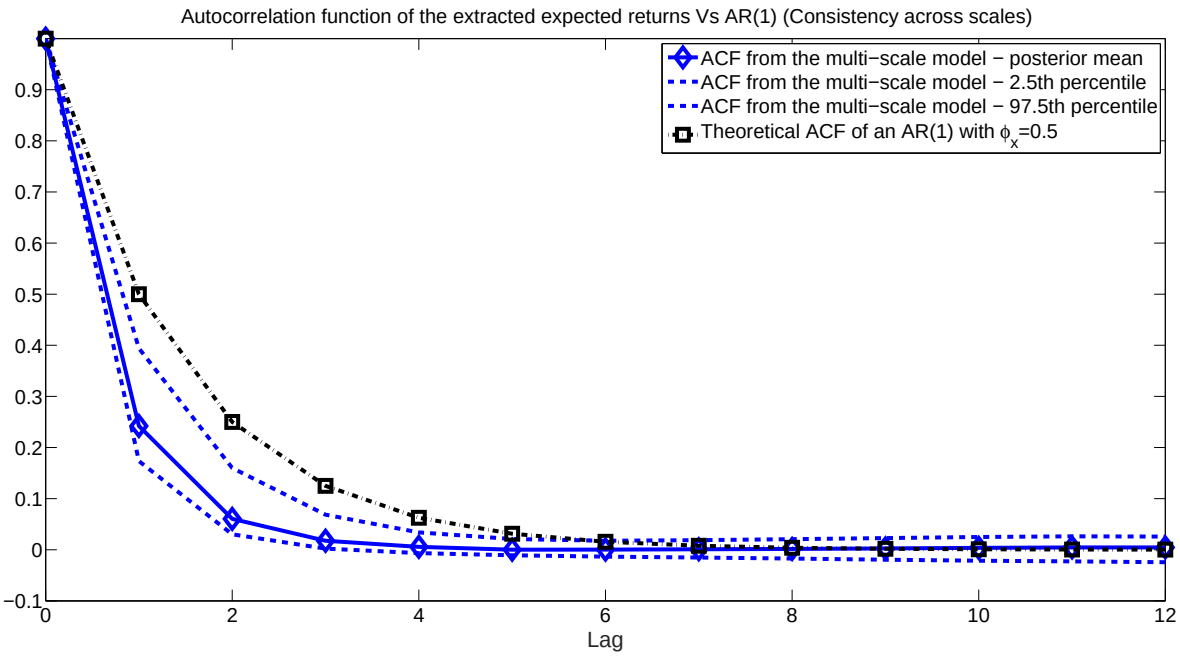


Figure 5. Persistence of the Latent Expected Returns from Alternative Simulation Examples

We study the performance of our MCMC sampler scheme for the multiscale time series model through simulation examples. Panel A displays the posterior mean ACF of the filtered expected returns at monthly frequency (blue line with diamond markers), together with the corresponding credibility intervals (dashed blue lines). The true data generating process consists of latent AR(1) such that the revised distribution is consistent across scales, i.e. $q(z_{1:n_z}) = p(z_{1:n_z})$. Panel B displays instead the ACF of the filtered expected returns aggregated at monthly frequency (blue line with diamond markers), when the true data generating process consists of a multi-scale process built from coarse to fine levels of resolution, namely where $p(x_{1:n_x}), q(z_{1:n_z})$, and $p\left(z_{1:n_z}|x_{1:n_x}\right)$ are inconsistent to each other.



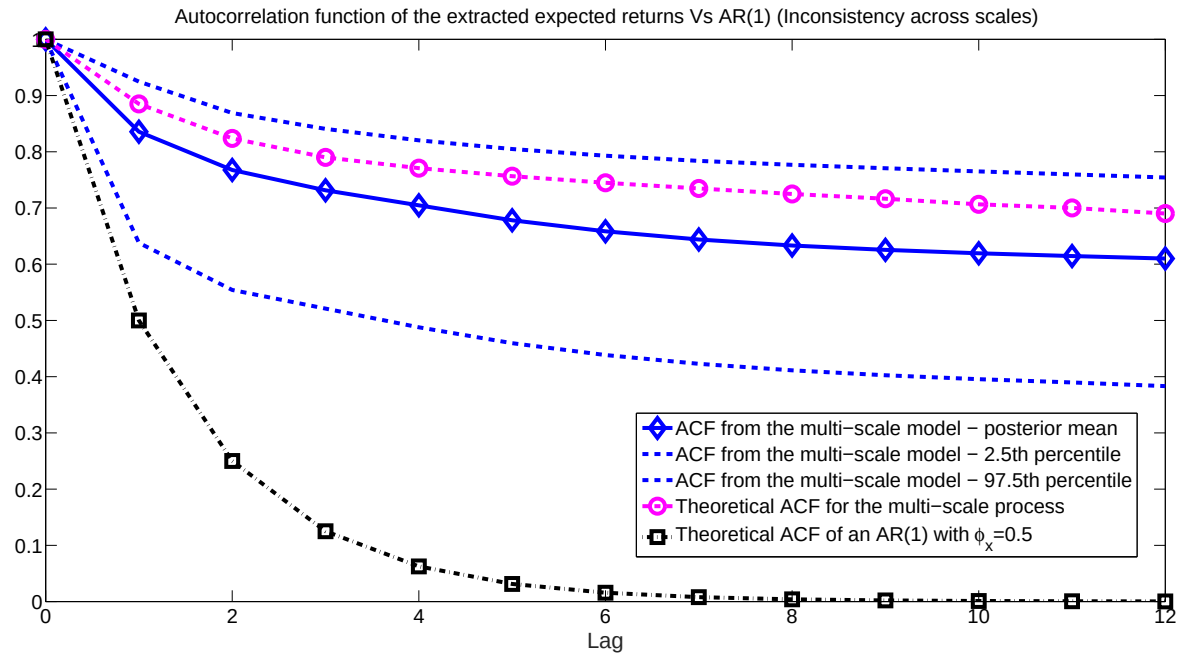
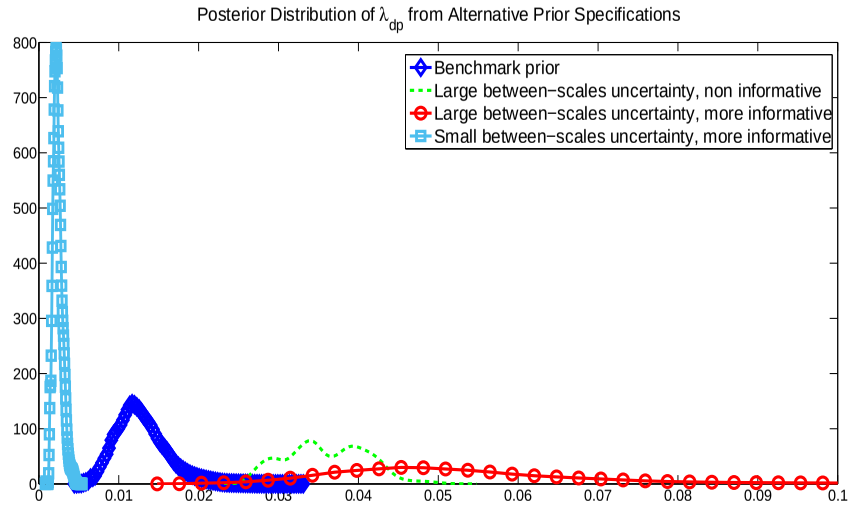
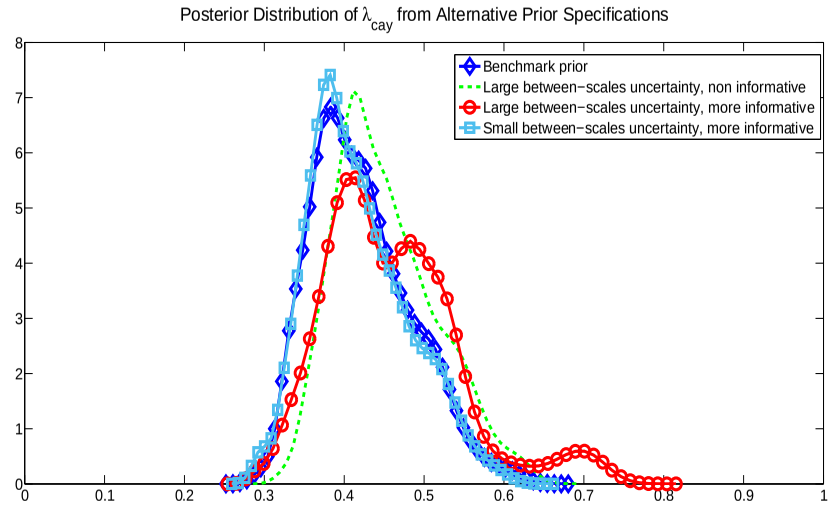
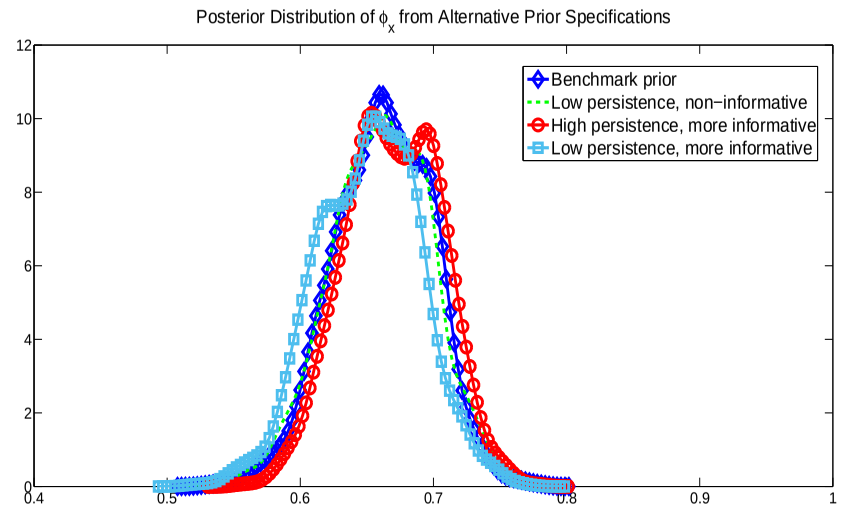


Figure 6. Posterior Distributions of the Parameters for Expected Returns

Posterior summaries for expected returns. This figure reports the posterior distributions of the parameters for the latent expected returns extracted by using jointly the annual consumption-wealth and the four-year log dividend-price ratios. Top left (right) panel shows the posterior distribution of the between-levels uncertainty measure for the log dividend-yield (cay). Bottom left panel shows the posterior distribution of the auto-regressive parameters ϕ_x. Bottom right panel shows the posterior distribution of the conditional variance σ_x^2. The blue line with diamond markers shows the posterior distribution obtained from our benchmark prior specification. The light-blue squared and red-circled lines represent the posterior distribution obtained under tighter prior on either high or low persistence. Finally, the green-dashed line reports the results obtained for an uninformative prior on low persistence. Estimates are obtained from a sample of 30,000 draws out of 50,000 simulations, storing every other draw. The sample period is 1952:01-2013:12.







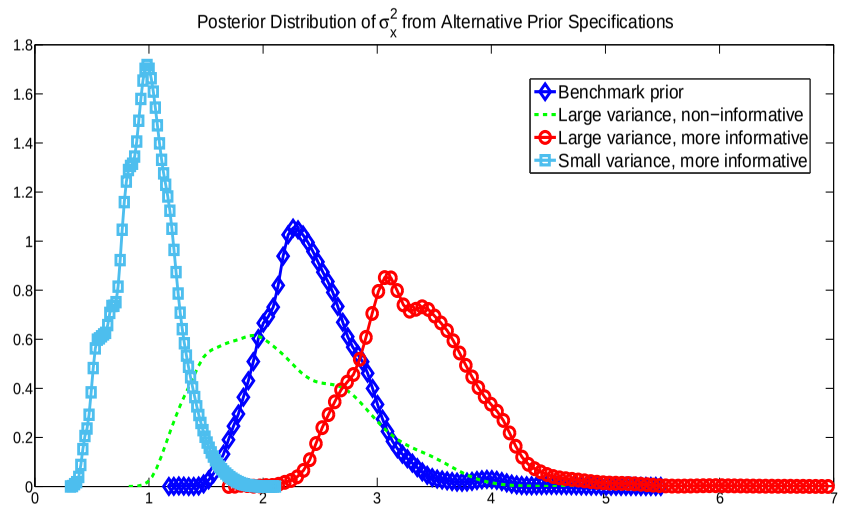


Figure 7. Extracted Expected Returns

Estimated expected returns. This figure shows the expected returns extracted by using the four-year log dividend-price ratio and the one-year consumption-wealth ratio, both singularly and jointly. The blue line with diamonds shows the posterior means of expected returns conditional on both cay_t aggregated annually ($m = 12$) and dp_t aggregated over 4-years ($m = 48$). The solid line with circles shows the time series extracted by using only the consumption-wealth ratio cay_t aggregated annually ($m = 12$). The dashed line shows the series of expected returns filtered from dp_t only, and aggregated over 4-years ($m = 48$) only. Shaded areas mark the NBER recessions. The series has been standardized. The sample period for the expected returns is 1952:01-2013:12.

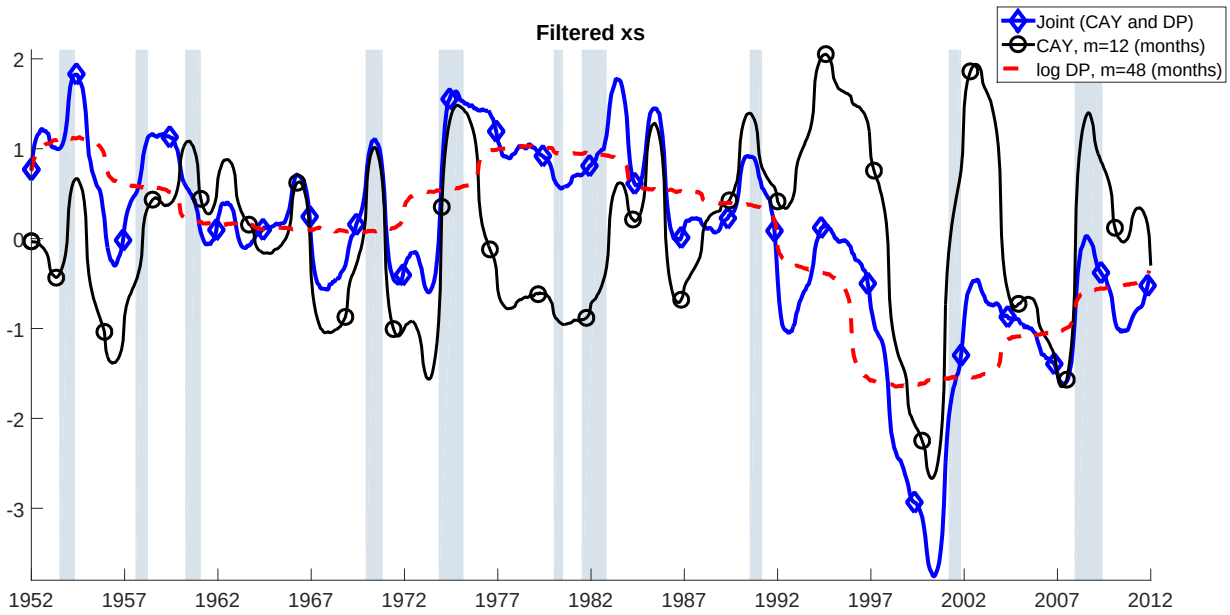


Figure 8. Persistence of Expected Returns Against the AR(1) Aggregation

Persistence of expected returns. This figure shows the autocorrelation function of expected returns estimated by jointly considering both the annual consumption-wealth ratio ($m = 12$) and the 4-year log dividend-yield ($m = 48$) as predictors. Panel A displays the autocorrelation (ACF) function of expected returns at monthly frequency (solid line with diamonds). Panel B displays the ACF function of expected returns aggregated over one year (solid line with diamonds). Panel C displays the ACF function of expected returns aggregated over four years (solid line with diamonds). Each Panel compares the ACF for the expected returns extracted using information at multiple time-scales with the one of an AR(1) process with root equal to ϕ_x and suitably aggregated in Panel B and C.

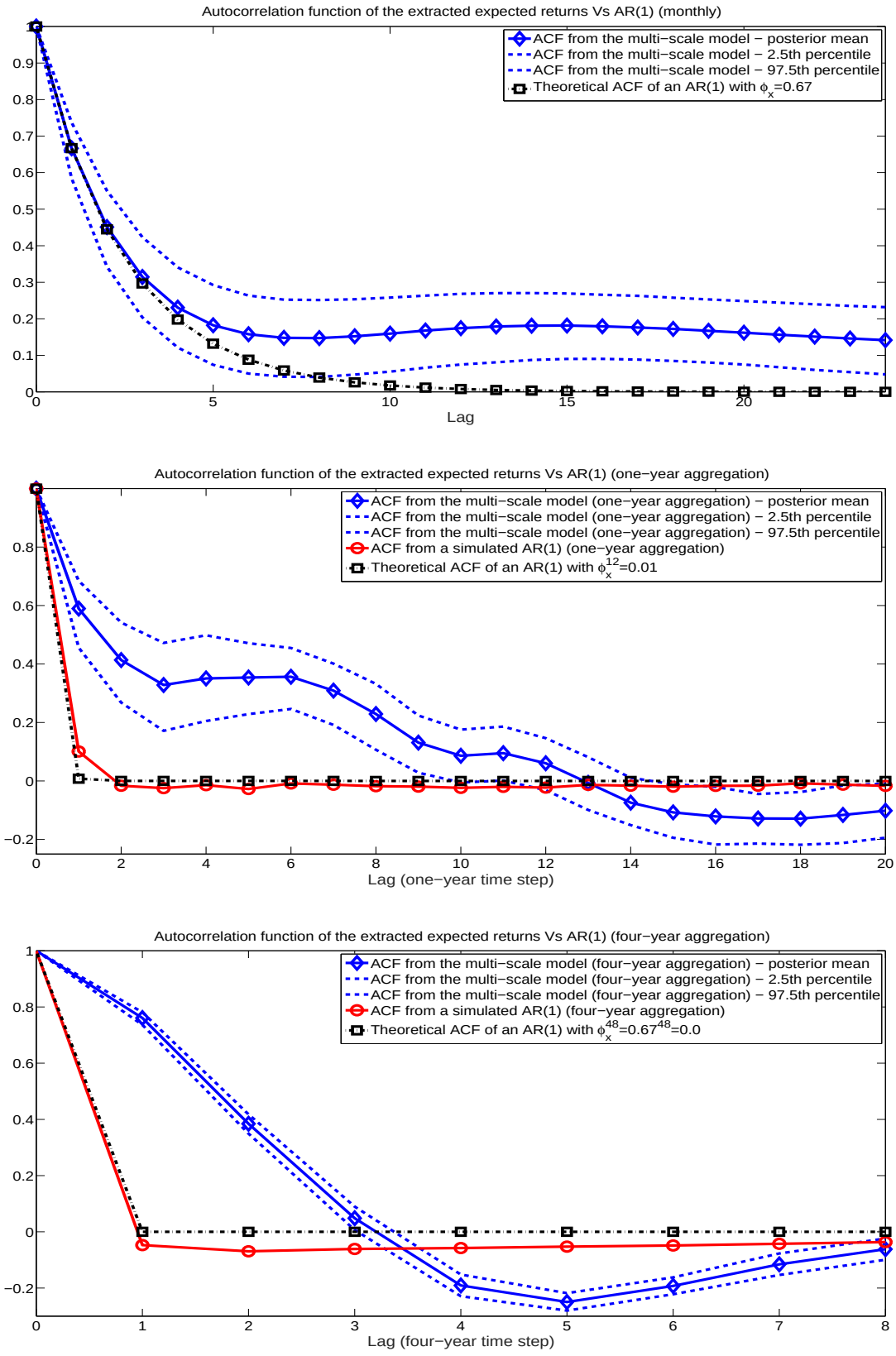
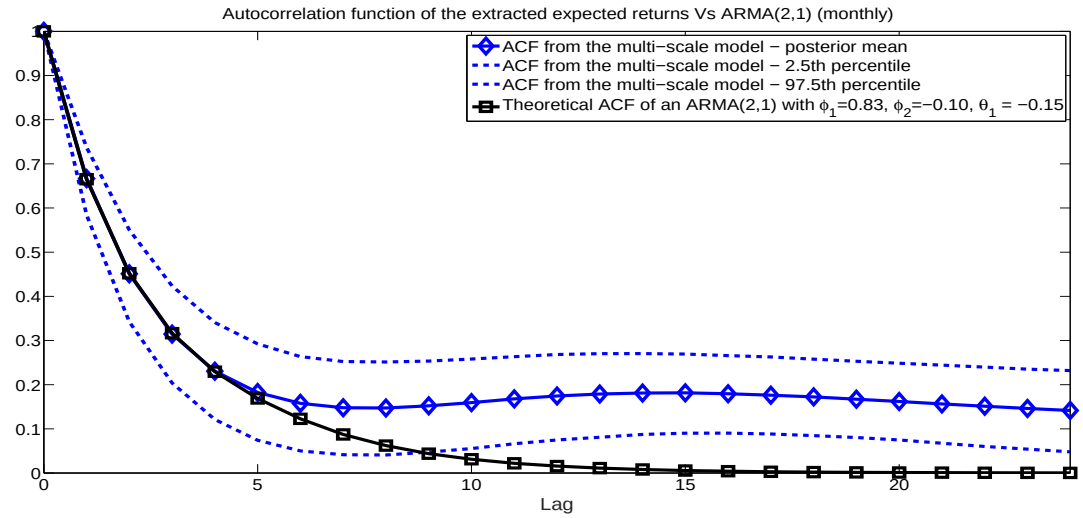
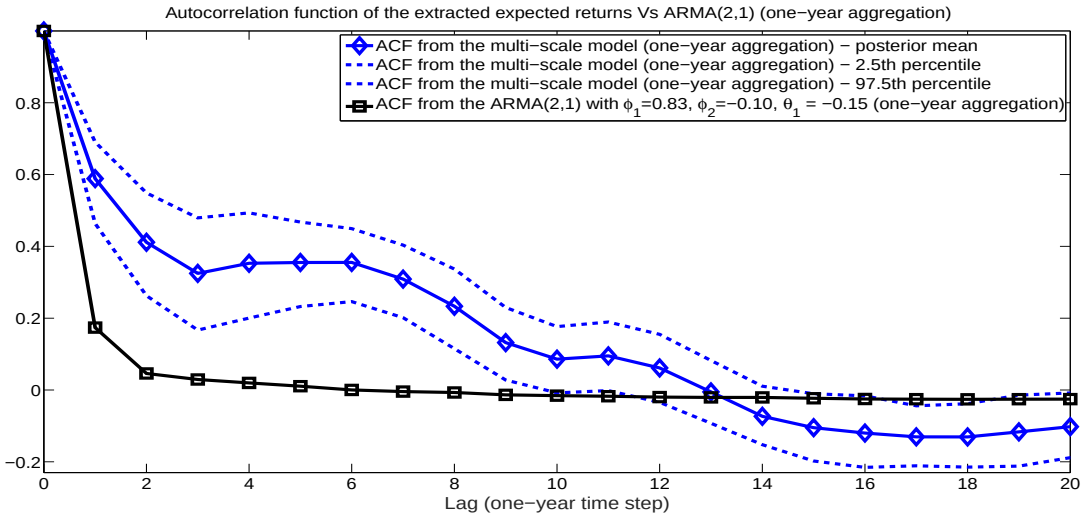


Figure 9. Persistence of Expected Returns Against the ARMA(2,1) Aggregation

Persistence of expected returns. This figure shows the autocorrelation function of expected returns estimated by jointly considering both the annual consumption-wealth ratio ($m = 12$) and the 4-year log dividend-yield ($m = 48$) as predictors. Panel A displays the autocorrelation (ACF) function of expected returns at monthly frequency (solid line with diamonds). Panel B displays the ACF function of expected returns aggregated over one year (solid line with diamonds). Panel C displays the ACF function of expected returns aggregated over four years (solid line with diamonds). Each Panel compares the ACF for the expected returns extracted using information at multiple time-scales with the one of an ARMA(2,1) process with root equal to ϕ_x and suitably aggregated in Panel B and C.





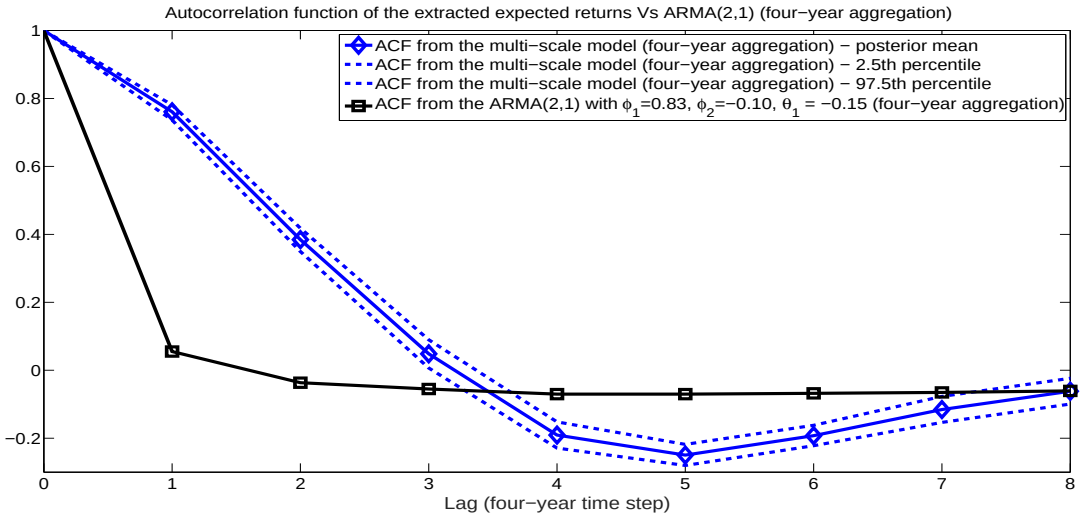
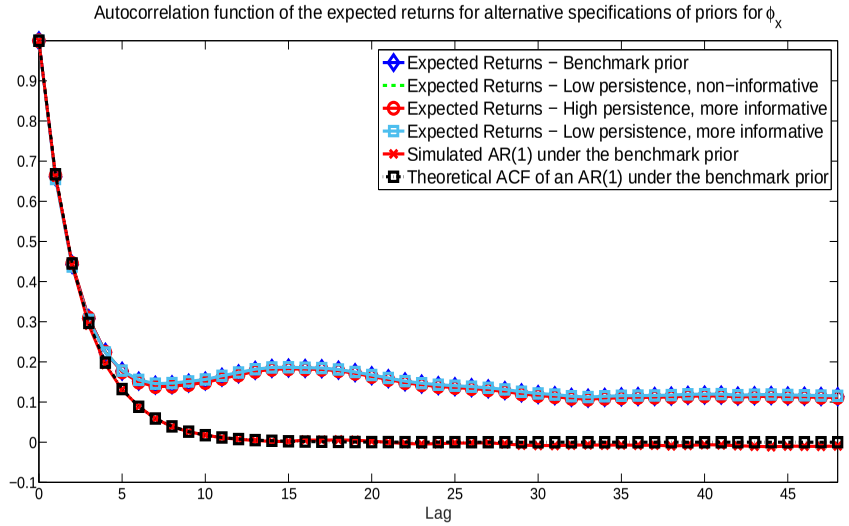
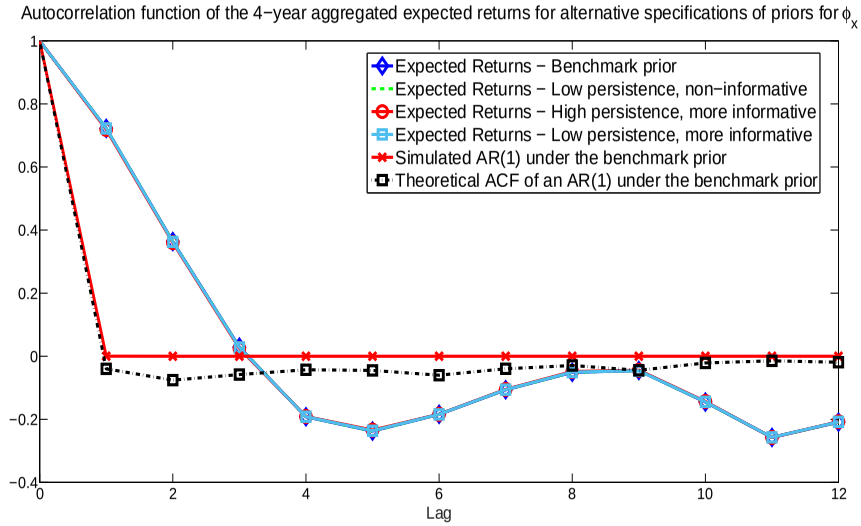
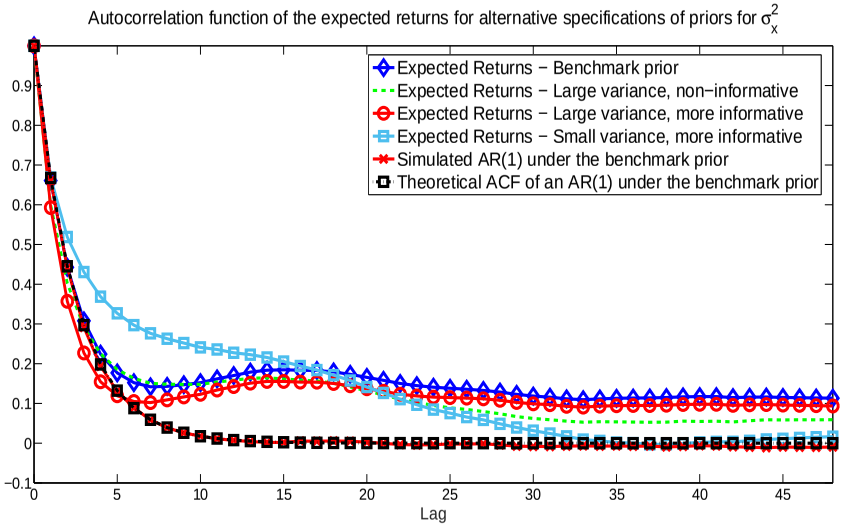


Figure 10. Persistence of Expected Returns Obtained from Alternative Prior Specifications

Autocorrelation functions under alternative priors. This figure reports the autocorrelation functions for alternative prior specifications of the autoregressive parameter ϕ_x (top panel) and the conditional variance σ_x^2 (bottom panel) of the expected returns. Left panels show the ACFs computed from the monthly expected returns and right panels show the ACF for their four-year aggregation. Expected returns are extracted by using jointly the annual consumption-wealth and the four-year log dividend-price ratios. The blue line with diamond marks shows the posterior distribution obtained from our benchmark prior specification. Estimates are obtained from a sample of 30,000 draws out of 50,000 simulations, storing every other draw (see Appendix B for more details). The sample period for the expected returns is 1952:01-2013:12.







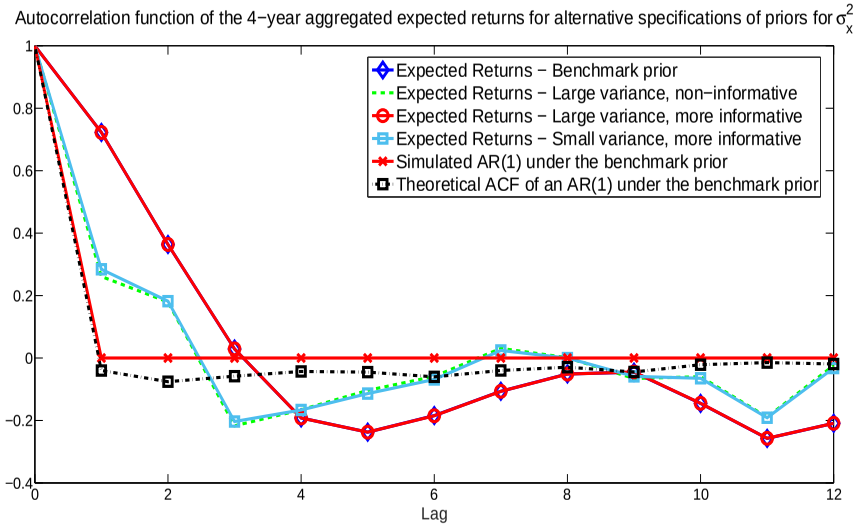


Figure 11. Persistence of expected returns: Comparison with OLS predictive regressions.

Comparison with OLS expected returns: autocorrelation functions. This figure shows the autocorrelation functions of the expected returns extracted by using jointly both the annual consumption-wealth ratio ($m = 12$) and the four-year log dividend-price ($m = 48$), in comparison to the ones implied by standard OLS predictive regressions. The Figure displays the posterior average autocorrelation function (solid line with diamonds) along with the 95% confidence intervals (dashed lines). We compare the autocorrelation from the multi-scale against the one implied by the fitted values of a standard predictive regression with the log dividend-price ratio (black line with squares), with the consumption-wealth ratio (red line with circles), and with both the log dividend-price and the consumption-wealth ratio (green line with triangles).

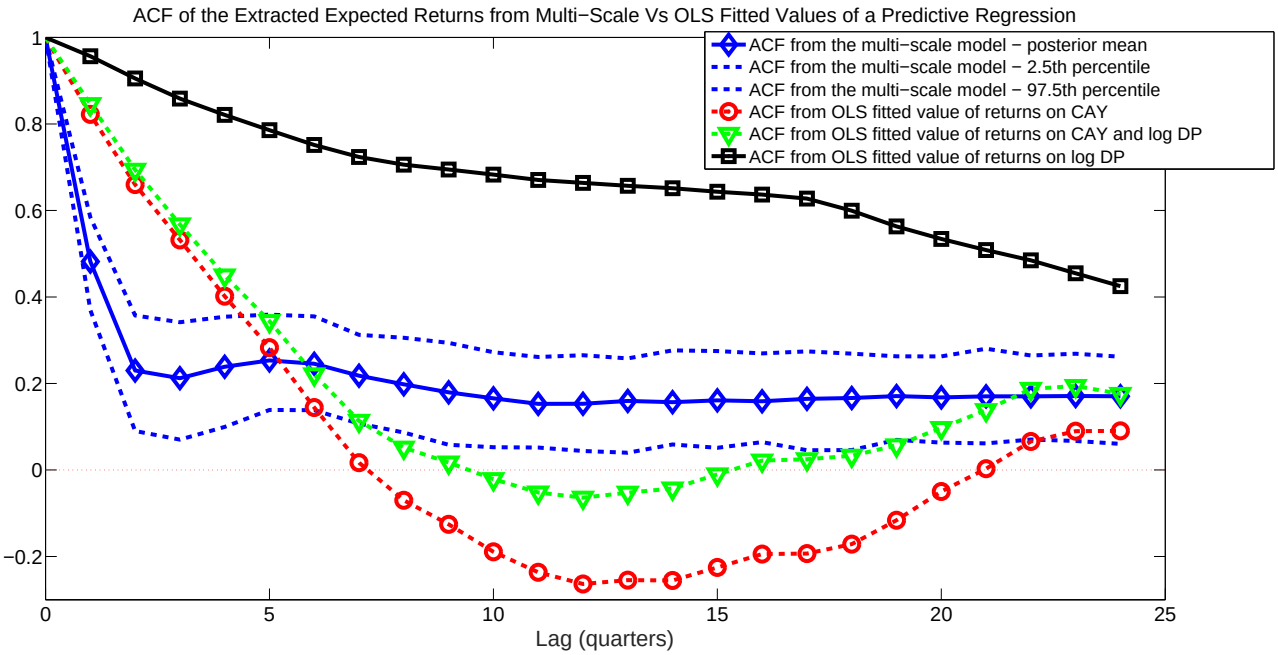
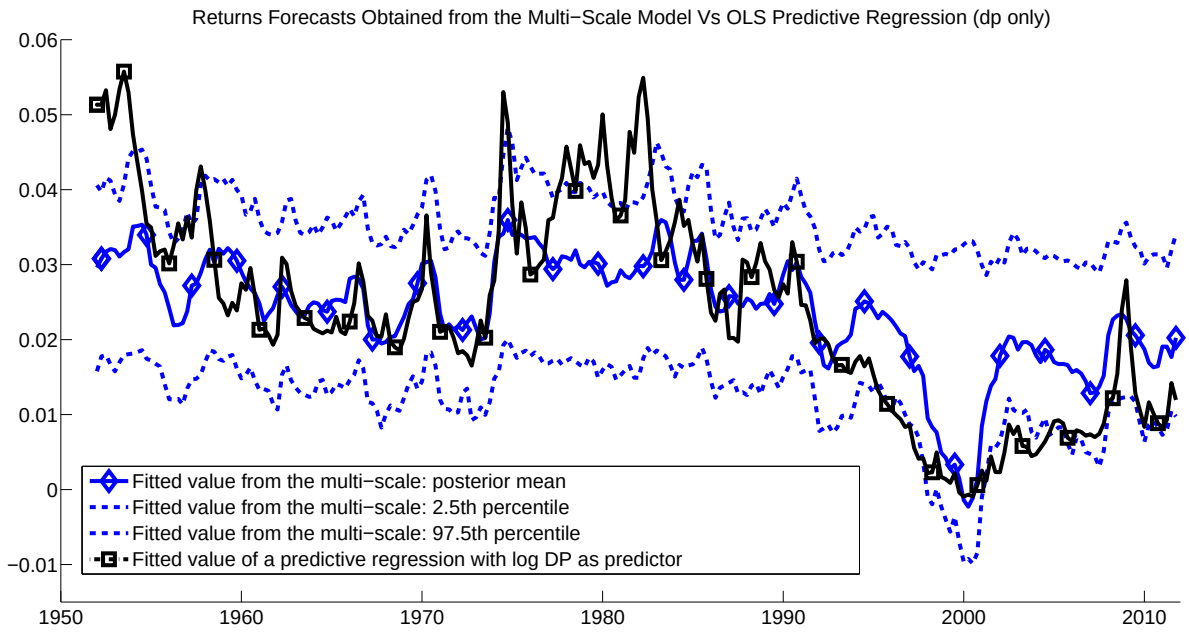


Figure 12. Forecasts from expected returns: Comparison with OLS predictive regressions.

Comparison of forecasts with OLS predictive regressions. This figure shows the forecasts of future realized returns implied by our expected returns extracted by using jointly both the annual consumption-wealth ratio ($m = 12$) and the four-year log dividend-price ($m = 48$), in comparison to the ones obtained by standard OLS predictive regressions. In order to compare our estimates with OLS, the model-implied expected return process are time-aggregated from monthly to quarterly frequency. We run simple regressions of quarterly returns onto our extracted series, on the dividend-price ratio, and on the consumption-wealth ratio. We also consider a multiple regression of returns onto the dividend-price ratio and the consumption-wealth ratio. The Figure displays the mean of the fitted values (solid line with diamonds) along with the 95% confidence intervals (dashed lines). Panel A compares our expected return series with the OLS fitted value from the log-dividend price ratio regression (black line with squares). Panel B compares our expected return series with the OLS fitted values from the consumption-wealth ratio (red line with circles) as well as with the fitted value from a multiple regression of quarterly returns onto CAY and log DP (green line with triangles).



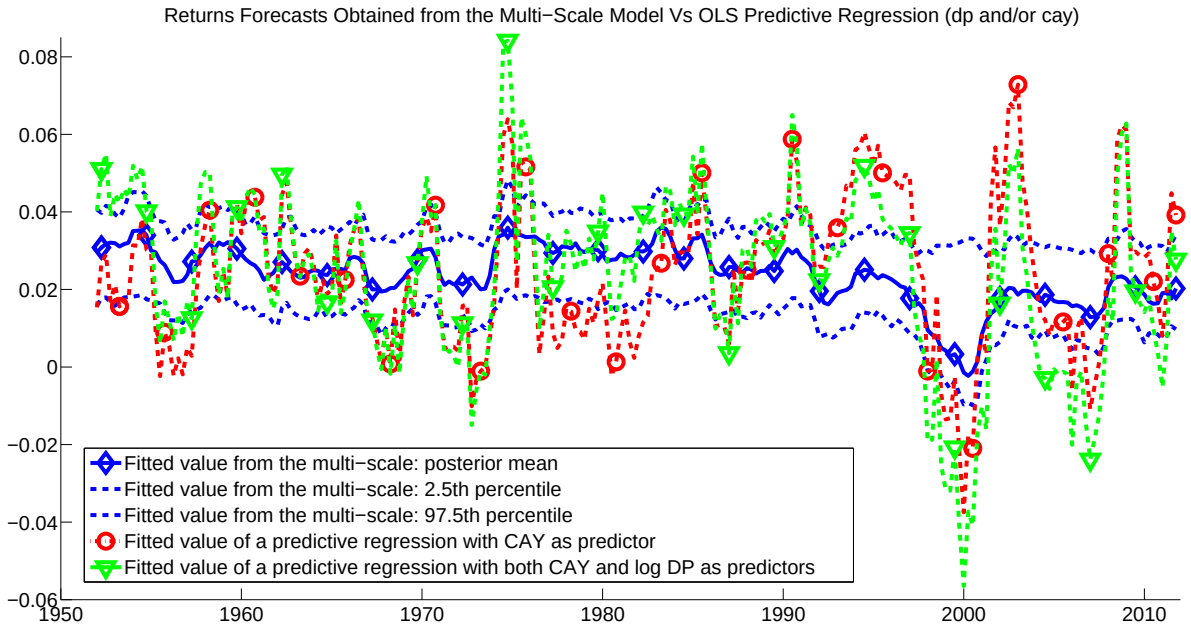


Figure 13. Multi-scale vs AR(1)-iterated forecasts: Empirical Evidence

Comparing forecasts. This figure shows the posterior means of the filtered expected returns (blue solid line), the demeaned log dividend-price ratio (light-blue dashed line with circles), the forecasts based on OLS estimates of an AR(1) fitted on filtered expected returns (green line with circles), the forecasts obtained fixing the autoregressive parameter to $\phi_x = 0.98$ (red line with circles), and our multi-scale-based forecast (magenta line with circles). Forecasts are computed as the posterior predictive mean values obtained from a sample of 30,000 draws out of 50,000 simulations, storing every other draw (see Appendix B for more details). The sample period for the expected returns is 1952:01-2013:12.

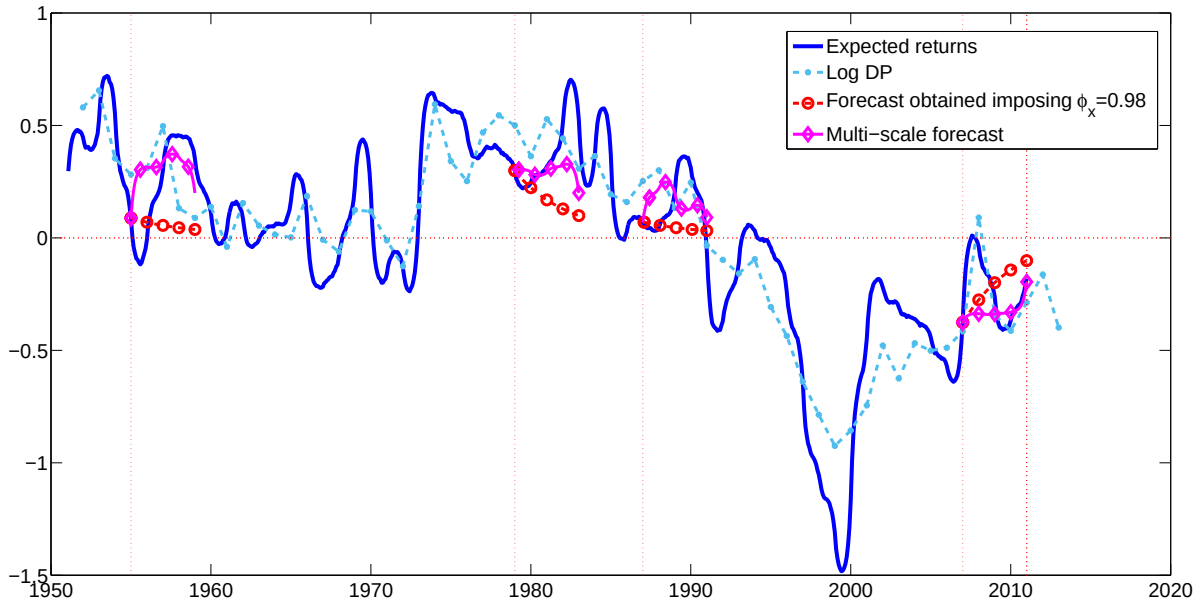
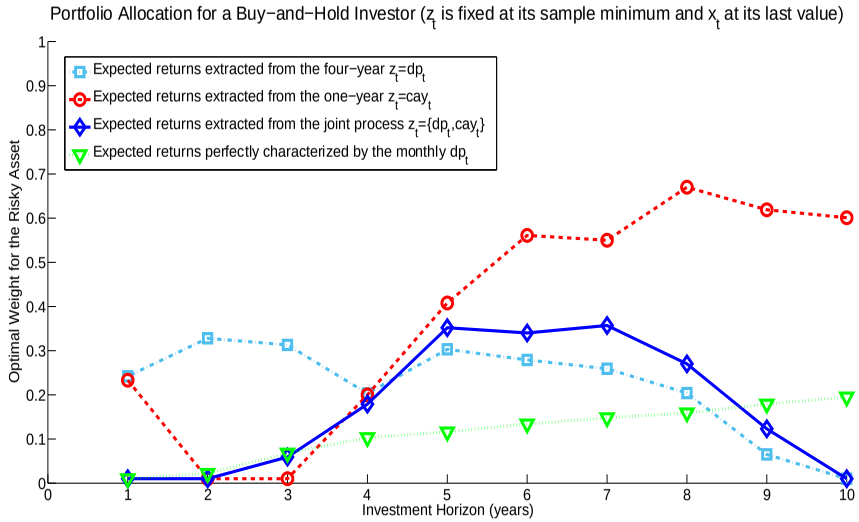
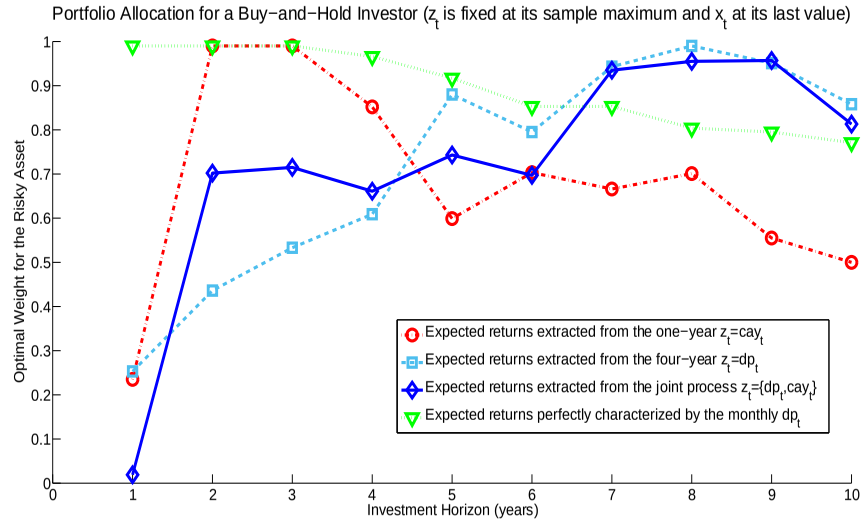
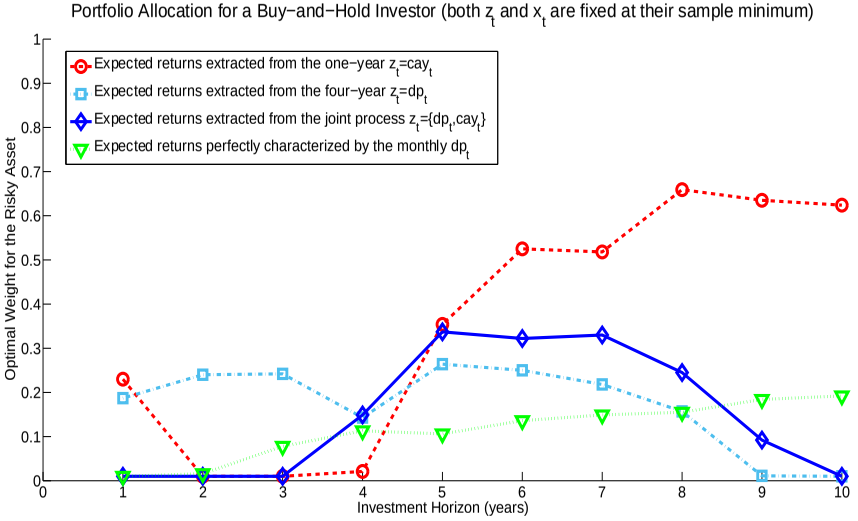


Figure 14. Optimal Allocation for a Buy-and-Hold Strategy.

Optimal allocation. This figure shows the optimal asset allocation for a buy-and-hold strategy of an investor with a relative risk aversion $\gamma = 5$ and investment horizon $h = 120$. The blue line with diamond markers represents the allocation obtained by using the expected returns extracted from our multi-scale time series model conditioning on both the four-year log dividend-price ($m = 48$) and the annual consumption-wealth ratio ($m = 12$). The light-blue line with square markers is the allocation obtained by using the expected returns extracted from the four-year log dividend-price ratio. The red-circled line represents the optimal allocation obtained by using the expected returns extracted conditioning only on the one-year consumption-wealth ratio. The green line with squared markers represents the optimal allocation obtained by using the observable monthly log dividend-price ratio as the only predictor. Top left (right) panel shows the results obtained by fixing the last value of the lower-frequency predictors to their minimum (maximum) and taking the last sample value for the expected returns. Bottom left (right) shows the results obtained by fixing both the last value of the lower-frequency predictors and the expected returns to their minimum (maximum). Estimates of the parameters and the expected returns are obtained from a sample of 30,000 draws out of 50,000 simulations. The sample period used for parameters estimates is 1952:01-2013:12.







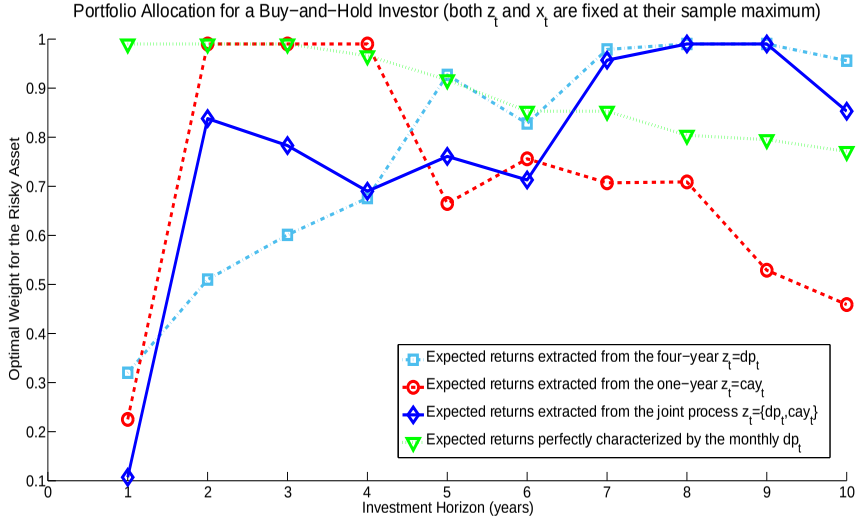
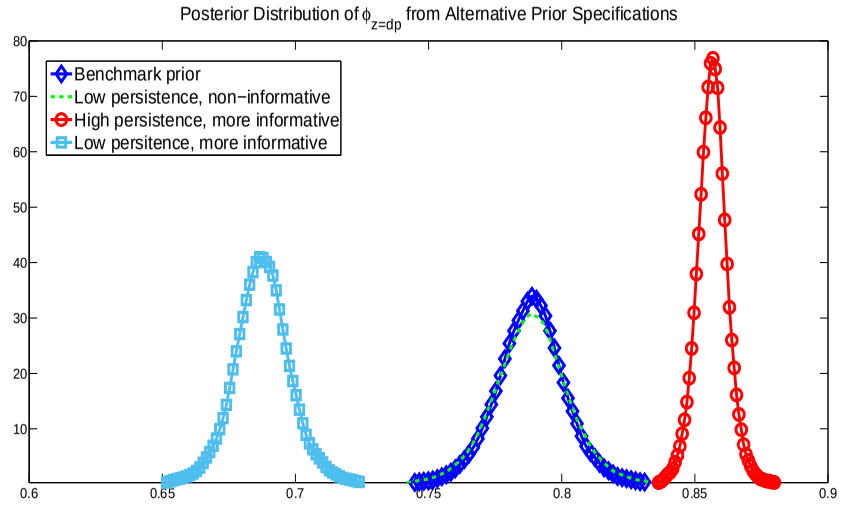
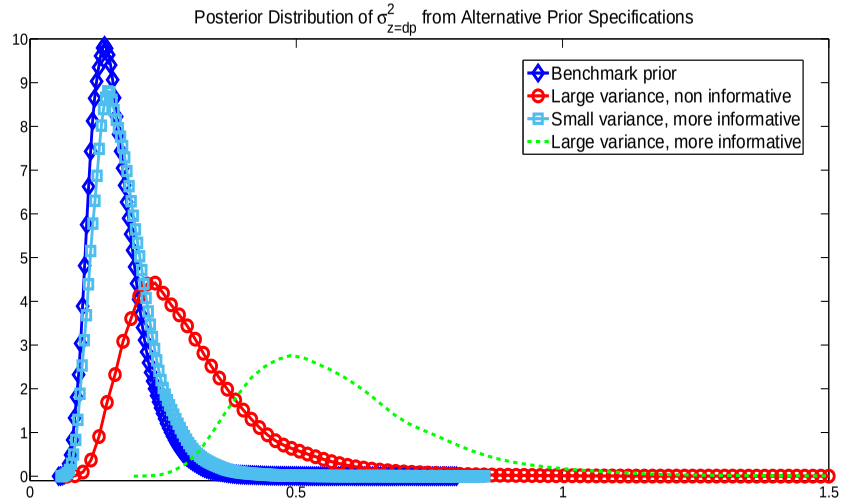
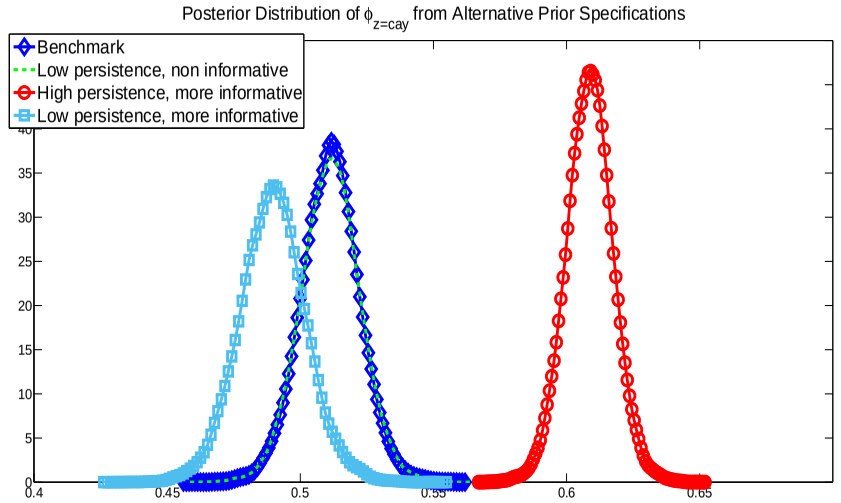


Figure B.1. Posterior Distributions of the Parameters of Scale-Specific Predictors

Posterior Summaries. This figure reports the posterior distributions of the parameters for the observable scale-specific predictors, namely, the four-year log dividend-price and the one-year consumption-wealth variable *CAY* introduced by Lettau and Ludvigson (2001). Top (bottom) left panel shows the posterior distribution of the persistence parameter for the log dividend-price (consumption-wealth ratio). Right panels show the posterior distribution of the conditional variances for both predictors. The blue line with diamond markers shows the posterior distribution obtained from our benchmark prior specification. The light-blue squared and red-circled lines represent the posterior distribution obtained under tighter prior on either high or low persistence. Finally, the green-dashed line reports the results obtained for an uninformative prior on low persistence. Estimates are obtained from a sample of 30,000 draws out of 50,000 simulations, storing every other draw. The sample period is 1952:01-2013:12.







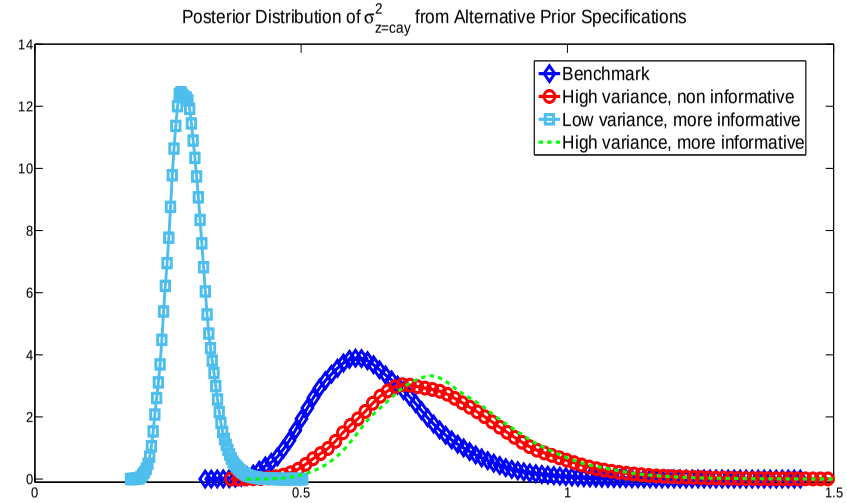


Table 1. Inference on the Coarsening Windows

Marginal likelihood validation. This table reports summary statistics about the marginal likelihood computed from the multi-scale time series model with different specifications of the size of the coarsening window in the linking equation (3). The marginal likelihood for each model is computed as the harmonic mean of the conditional likelihood evaluated for each draw of the parameters from the full conditionals (see, e.g. Gelfand and Dey 1994a and Newton and Raftery 1994). **Panel A:** Marginal likelihood for the multi-scale model with the log dividend-price ratio as predictor. **Panel B:** Marginal likelihood for the multi-scale model with the consumption-wealth variable *CAY* introduced by Lettau and Ludvigson (2001). Draws from the predictive distribution are obtained from a sample of 30,000 draws out of 50,000 simulations, with a thinning size of 20 (see Appendix A for details on the estimation algorithm). The sample period is 1952:01-2013:12.

Panel A: Marginal Likelihood for $z_t = dp_t$

Predictor	$m = 24$	$m = 48$
$z_t = dp_t$	-69.08	-49.81

Panel B: Marginal Likelihood for $z_t = cay_t$

Predictor	$m = 12$	$m = 24$
$z_t = cay_t$	-96.74	-111.57

Table 2. Parameters Posterior Estimates

This table reports summary statistics of the posterior estimates for the parameters of the observable predictors and the filtered expected returns. The latter is extracted on the basis of the joint process for the log dividend-price and the consumption-wealth variable *CAY* introduced by Lettau and Ludvigson (2001). **Panel A:** Posterior summaries of the dynamics for $z_{1:n_{z[j]}}$ for $j = ldp, cay$. **Panel B:** Posterior summaries for the dynamics of the latent expected returns $x_{1:n_x}$. Estimates are obtained from a sample of 30,000 draws out of 50,000 simulations. The sample period is 1952:01-2013:12.

Panel A: Observable Predictors

Predictor	Parameter	Mean	Median	2.5_{th}	97.5_{th}
$z_t = dp_t$	ϕ_z	0.788	0.789	0.762	0.815
	σ_z^2	0.162	0.154	0.093	0.279
$z_t = cay_t$	ϕ_z	0.512	0.511	0.489	0.532
	σ_z^2	0.640	0.627	0.461	0.885

Panel B: Latent Expected Returns

Predictor	Parameter	Mean	Median	2.5_{th}	97.5_{th}
	ϕ_x	0.667	0.667	0.592	0.746
	σ_x^2	2.401	2.362	1.683	3.371
$z_t = dp_t$	λ_{dp}	0.013	0.012	0.007	0.019
$z_t = cay_t$	λ_{cay}	0.427	0.421	0.301	0.573

Table 3. Forecasting Accuracy of Future Expected Returns

Forecasting accuracy measures. This table reports summary statistics about the forecasting accuracy of future expected returns obtained from our multi-scale time series model. The forecasting performance of the model is compared with the forecasts obtained from a simple AR(1) fitted on the extracted expected returns and the predictions obtained by imposing the AR(1) estimates to generate high persistence and low volatility, consistent with existing literature. The latent series of expected returns and corresponding forecasts are obtained from the joint process of the log dividend-price and the consumption-wealth variable CAY introduced by Lettau and Ludvigson (2001). Forecasts are produced monthly with an horizon of $h = 48$ months. For the ease of exposition the table reports the results for $h = 12, 24, 36, 48$ months. **Panel A:** Mean Squared Errors obtained from the marginal predictive mean. **Panel B:** Log predictive scores obtained from the marginal predictive likelihood. Draws from the predictive distribution are obtained from a sample of 30,000 draws out of 50,000 simulations, with a thinning size of 20. The sample period is 1952:01-2013:12.

Panel A: Mean Squared Error

	Forecasting Horizon (months)			
Model	$h = 12$	$h = 24$	$h = 36$	$h = 48$
AR(1) unrestricted	5.290	4.082	4.262	4.877
AR(1) restricted $\phi_x = 0.98$	6.895	5.273	5.101	5.271
Multi-scale	5.104	3.749	3.796	4.278

Panel B: Log-Predictive Score

	Forecasting Horizon (months)			
Model	$h = 12$	$h = 24$	$h = 36$	$h = 48$
AR(1) unrestricted	-26.974	-27.627	-27.835	-28.434
AR(1) restricted $\phi_x = 0.98$	-301.05	-40.677	-61.599	-28.136
Multi-scale	-26.587	-26.474	-26.603	-27.158

Table B.1. Convergence Diagnostics

The table summarizes the convergence results for the posterior draws for the parameters of the latent expected returns extracted on the basis of the joint process for the log dividend-yield and the consumption-wealth ratio *cay* introduced by Lettau and Ludvigson (2001). The sample period is 1952:01-2013:12. Estimates are obtained from a sample of 30,000 draws out of 50,000 simulations. In order to reduce the amount of auto-correlation across draws we keep one every other draw from the posterior and discard the rest. **Panel A:** Test statistics of the Geweke (1992) convergence test for both the whole sample including burn-in (third column) and the sample of draws excluding burn-in (last column). The asymptotic variance for the convergence test statistics is computed from the spectral density of the series of draws evaluated at frequency zero, and by using a bandwith of 4% of the sample sizes considered. **Panel B:** Effective sample size and corresponding inefficiency factor for each parameter.

Panel A: Convergence Test

Predictor	Parameter	All Draws		w/o Burn-in	
		t-stat	p-value	t-stat	p-value
	ϕ_x	-3.436	0.000	-1.069	0.286
	σ_x^2	-3.932	0.000	1.881	0.062
$z_t = dp_t$	λ_{ldp}	12.16	0.000	0.012	0.991
$z_t = cay_t$	λ_{cay}	9.919	0.000	-1.331	0.182

Panel B: Effective Sample Size

Predictor	Parameter	Effective Sample Size	Inefficiency Factor
	ϕ_x	836.2	59.79
	σ_x^2	1,040	48.17
$z_t = dp_t$	λ_{ldp}	796.4	62.78
$z_t = cay_t$	λ_{cay}	822.8	60.77

References

Amemiya, T. and Wu, R. Y. (1972). The effect of aggregation on prediction in the autoregressive model. *Journal of the American Statistical Association*, 67(339):pp. 628–632.

Ang, A. and Bekaert, G. (2007). Stock return predictability: Is it there? *Review of Financial studies*, 20(3):651–707.

Avramov, D. (2004). Stock return predictability and asset pricing models. *Review of Financial Studies*, 17(3):699–738.

Bandi, F., Benoit, P., Tamoni, A., and Tebaldi, C. (2015). The scale of predictability. *Working Paper*.

Barberis, N. (2000). Investing for the long-run when returns are predictable. *The Journal of Finance*, 55(1):225–264.

Bollerslev, T. and Wright, J. H. (2000). Semiparametric estimation of long-memory volatility dependencies: The role of high-frequency data. *Journal of Econometrics*, 98(1):81–106.

Calvet, L. E. and Fisher, A. J. (2007). Multifrequency news and stock returns. *Journal of Financial Economics*, 86(1):178–212.

Campbell, J. and Mankiw, N. G. (1989). Consumption, income and interest rates: Reinterpreting the time series evidence. In *NBER Macroeconomics Annual 1989, Volume 4*, pages 185–246. National Bureau of Economic Research, Inc.

Campbell, J. and Viceira, L. (2005). The term structure of the risk-return trade-off. *Financial Analyst Journal*, 61(1):34–44.

Campbell, J. Y. and Shiller, R. J. (1988). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, 1(3):195–228.

Campbell, J. Y. and Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies*, 21(4):1509–1531.

Carter, C. and Kohn, R. (1994). On gibbs sampling for state-space models. *Biometrika*, 3(81):541–553.

Casella, G. and Robert, C. P. (2004). *Monte Carlo Statistical Methods*. Springer Verlag, New York.

Chib, S. (1995). Marginal likelihood from the gibbs output. *Journal of the American Statistical Association*, 90(432):1313–1321.

Chib, S. and Greenberg, E. (1994). Bayes inference in regression models with arma(p,q) errors. *Journal of Econometrics*, 64:183–206.

Chib, S. and Jeliazkov, I. (2001). Marginal likelihood from the metropolis-hastings output. *Journal of the American Statistical Association*, 96(453):270–281.

Cohen, L. and Frazzini, A. (2008). Economic links and predictable returns. *The Journal of Finance*, 63(4):1977–2011.

Cowles, M. K. and Carlin, B. P. (1996). Markov chain monte carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association*, 91(434):883–904.

Dangl, T. and Halling, M. (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economics*, 106(1):157–181.

Diaconis, P. and Zabell, S. L. (1982). Updating subjective probability. *Journal of the American Statistical Association*, 77(380):822–830.

Engle, R. F. (1974). Band spectrum regression. *International Economic Review*, 15(1):1–11.

Ferreira, M. and Lee, H. (2007). *Multiscale Modeling: A Bayesian Perspective*. Springer Verlag, New York.

Ferreira, M., West, M., Lee, H., and Higdon, D. (2006). Multi-scale and hidden resolution time series models. *Bayesian Analysis*, 1(4):947–968.

Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, 15:183–202.

Gelfand, A. E. and Dey, D. K. (1994a). Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 501–514.

Gelfand, A. E. and Dey, D. K. (1994b). Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 501–514.

Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions*, 6:721–741.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. *Bayesian Statistics 4*.

Geweke, J. (2005). *Contemporary Bayesian Econometrics and Statistics*. John Wiley, Hoboken.

Geweke, J. and Amisano, G. (2012). Prediction and mispecified models. *American Economic Review*, 102:482–486.

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494):746–762.

Goyal, A. and Welch, I. (2003). Predicting the equity premium with dividend ratios. *Management Science*, 49(5):639–654.

Jeffrey, R. (1957). Contributions to the theory of inductive probability. *PhD Thesis, Princeton University*.

Jeffrey, R. (1965). *The Logic of Decision*. McGraw-Hill, New York.

Kandel, S. and Stambaugh, R. F. (1996). On the predictability of stock returns: An asset-allocation perspective. *The Journal of Finance*, 51(2):385–424.

Lanne, M. (2002). Testing the predictability of stock returns. *Review of Economics and Statistics*, 84(3):407–415.

Lettau, M. and Ludvigson, S. (2001). Consumption, aggregate wealth, and expected stock returns. *Journal of Finance*, 56(3):815–849.

Lettau, M. and Van Nieuwerburgh, S. (2008). Reconciling the return predictability evidence. *Review of Financial Studies*, 21(4):1607–1652.

Lochstoer, L. A. (2009). Expected Returns and the Business Cycle: Heterogeneous Goods and Time-Varying Risk Aversion. *Review of Financial Studies*, 22(12):5251–5294.

Newton, M. and Raftery, A. (1994). Approximate bayesian inference by the weighted likelihood bootstrap. *Journal of the Royal Statistical Society Series B. v56*, 56:1–48.

Pastor, L. and Stambaugh, R. F. (2009). Predictive systems: Living with imperfect predictors. *The Journal of Finance*, 64(4):1583–1628.

Pástor, L. and Stambaugh, R. F. (2012). Are stocks really less volatile in the long run? *The Journal of Finance*, 67(2):431–478.

Raftery, A. E. and Lewis, S. (1992). How many iterations in the gibbs sampler. *Bayesian statistics*, 4(2):763–773.

Roberts, G. and Sahu, S. (1997). Updating schemes, covariance structure, blocking and parametrisation for the gibbs sampler. *Journal of the Royal Statistical Society*, 59:291 – 318.

Sims, C. A., Waggoner, D. F., and Zha, T. (2008). Methods for inference in large multiple-equation markov-switching models. *Journal of Econometrics*, 146(2):255–274.

Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics*, 54(3):375–421.

Tanner, A. and Wong, W. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82(398):528–540.

Van Binsbergen, Jules, H. and Koijen, R. S. J. (2010). Predictive regressions: A present-value approach. *The Journal of Finance*, 65(4):1439–1471.

West, M. and Harrison, J. (1997). *Bayesian forecasting and dynamics models*. Springer.

Appendix

A The Gibbs sampler

A major aspect of our model is that conditional on the latent expected returns $x_{1:n_x}$, the dynamics of predictors corresponding to different levels of resolution and the predictive regression are conditionally independent, greatly simplifying the sampling scheme for the joint posterior distribution of the parameters. In order to find a Bayesian estimation of the parameters and the latent hidden resolution process we follow a data augmentation principle which relies on the complete likelihood function, that is the product of the data and hidden processes, given the structural parameters. In this section we first specify the prior structure, which is mostly conjugate, and then we provide a detailed explanation of the posterior simulation.

A.1 Prior specification

For the Bayesian inference to work, we need to specify the prior distributions for the model parameters. For a given $x_t, t = 1, ..., n_x$ the prior structure is conjugate. As far as the dynamics of the observable scale-specific predictors $z_{1:n_z}$ is concerned, we start from a standard conjugate normal-inverse-gamma-2 independent prior on the persistence parameter and the conditional variance

$$\phi_z \sim N\left(m_{\phi_z}, M_{\phi_z}\right), \qquad \sigma_z^2 \sim IG\left(\nu_{\sigma_z}/2, \nu_{\sigma_z} s_{\sigma_z}/2\right)$$

with m_{ϕ_z}, M_{ϕ_z} the location and scale hyper-parameters of the AR(1) coefficient, and $\nu_{\sigma_z}, \nu_{\sigma_z} s_{\sigma_z}$ the degrees of freedom and scale hyper-parameters for the conditional variance σ_z^2. Notice that, by using two conditionally orthogonal low-frequency forecasting variables does not change our prior specification for $\theta_z = \left(\phi_z, \sigma_z^2\right)$ as draws from the posteriors will be sampled independently across scales. Similarly, for the fine level process we start from a classic independent normal-inverse-gamma-2 prior on the persistence parameter and the conditional variance

$$\phi_x \sim N\left(m_{\phi_x}, M_{\phi_x}\right), \qquad \sigma_x^2 \sim IG\left(\nu_{\sigma_x}/2, \nu_{\sigma_x} s_{\sigma_x}/2\right)$$

with m_{ϕ_x}, M_{ϕ_x} the location and scale hyper-parameters of the AR(1) coefficient, and $\nu_{\sigma_x}, \nu_{\sigma_x} s_{\sigma_x}$ the degrees of freedom and scale hyper-parameters for the conditional variance σ_x^2. As far as the between scales uncertainty is concerned we specify a standard inverse-gamma-2 prior;

$$\lambda \sim IG\left(\nu_\lambda/2, \nu_\lambda s_\lambda/2\right), \tag{A.16}$$

$\nu_{\sigma_\lambda}, \nu_{\sigma_\lambda} s_{\sigma_\lambda}$ the degrees of freedom and scale hyper-parameters for the conditional variance for the uncertainty between levels λ. Finally, given the independence with observable predictors, and conditional on the expected returns, for the predictive regression we specify a conjugate independent Normal-Inverse-Gamma-2 prior structure;

$$\alpha, \beta \sim N\left(m_\beta, M_\beta\right), \qquad \sigma^2 \sim IG\left(\nu_\beta/2, \nu_\beta s_\beta/2\right),$$

with m_β, M_β the location and scale hyper-parameters of the regression coefficients, and $\nu_\beta, \nu_\beta s_\beta$ the degrees of freedom and scale hyper-parameters for the conditional variance if the idiosyncratic error σ^2. Notice that, the functional form of the predictive regression does not change across different investment horizons. In that respect the prior structure is not scale-specific.

A.2 Posterior simulation

Although we use a conjugate prior setting, the joint posterior distribution of structural parameters and the latent expected returns is not available in closed form. In order to draw from the posterior distribution we follow a data augmentation principle which relies on the complete likelihood function (e.g. Tanner and Wong 1987). Let

expected returns consist of n_x observations $\mathbf{x}_{1:n_x}$, and the observable scale-specific predictors consist of $n_{z[j]}$ observations $\mathbf{z}^j_{1:n_{z[j]}}$ for $j = 1, 2, ..., K$ scale-specific predictors. Stock excess returns are assumed observable at the same frequency of latent expected returns and defined as $\mathbf{y}_{1:n_y}$.[35] The collection of parameters is defined as $\theta = \left(\alpha, \beta, \sigma^2, \phi_x, \sigma^2_x, \boldsymbol{\lambda}, \boldsymbol{\phi}_z, \boldsymbol{\sigma}^2_z\right)$ with $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_K)$, $\boldsymbol{\phi}_z = \left(\phi_{z[1]}, \cdots, \phi_{z[K]}\right)$ and $\boldsymbol{\sigma}^2_z = \left(\sigma^2_{z[1]}, \cdots, \phi_{z[K]}\right)$. Given the AR(1) dynamics imposed at different scales, and the corresponding linking equation, the complete likelihood can now be defined as

$$\begin{aligned} p\left(\mathbf{y}_{1:n_y}, \mathbf{x}_{1:n_x}, \mathbf{z}_{1:n_z} | \alpha, \beta, \phi_x, \lambda, \sigma^2, \sigma^2_x\right) &= p\left(\mathbf{y}_{1:n_y} | \mathbf{x}_{1:n_x}, \alpha, \beta, \sigma^2\right) p\left(\mathbf{z}_{1:n_z}, \mathbf{x}_{1:n_x} | \phi_x, \lambda, \sigma^2_x\right), \\ &= p\left(\mathbf{y}_{1:n_y} | \mathbf{x}_{1:n_x}, \alpha, \beta, \sigma^2\right) p\left(\mathbf{x}_{1:n_x} | \mathbf{z}_{1:n_z}, \phi_x, \lambda, \sigma^2_x\right) q\left(\mathbf{z}_{1:n_z} | \phi_z, \sigma^2_z\right), \end{aligned}$$

with $\mathbf{z}_{1:n_z} = \left(\mathbf{z}^1_{1:n_{z[1]}}, \cdots, \mathbf{z}^K_{1:n_{z[K]}}\right)$ the set of observable predictors at different scales, and $q\left(\mathbf{z}_{1:n_z} | \boldsymbol{\phi}_z, \boldsymbol{\sigma}^2_z\right) = \prod_{j=1}^K q\left(\mathbf{z}^j_{1:n_{z[j]}} | \phi_{z[j]}, \sigma^2_{z[j]}\right)$ that derives from the scale-specific AR(1) dynamics of $\mathbf{z}^j_{1:n_{z[j]}}$ which is super-imposed on expected returns. Consistency across distributions is ensured by the Jeffrey's rule on conditioning information (see Jeffrey 1957, Jeffrey 1965 and Diaconis and Zabell 1982 for more details).

A major aspect of our model is that conditional on the hidden fine scale $x_{1:n_x}$, the observable coarse scale and the predictive regression are conditionally independent, greatly simplifying the sampling scheme for the joint posterior distribution of the parameters. The full conditional distribution of the parameters for the scale dependent predictive regression is defined as;

$$p\left(\alpha, \beta, \sigma^2 | \mathbf{y}_{1:n_y}, \mathbf{x}_{1:n_x}\right) \propto p\left(\mathbf{y}_{1:n_y} | \mathbf{x}_{1:n_x}, \alpha, \beta, \sigma^2\right) p\left(\alpha, \beta, \sigma^2\right), \tag{A.17}$$

while for the parameters of the fine scale and linking equation the full conditional can be written as

$$\begin{aligned} p\left(\phi_x, \sigma^2_x, \lambda | \mathbf{z}_{1:n_z}, \mathbf{x}_{1:n_x}\right) &\propto p\left(\mathbf{x}_{1:n_x} | \mathbf{z}_{1:n_z}, \phi_x, \lambda, \sigma^2_x\right) p\left(\phi_x, \sigma^2_x, \lambda\right), \\ &= \frac{p\left(\mathbf{z}_{1:n_z} | \mathbf{x}_{1:n_x}, \phi_x, \sigma^2_x, \lambda\right) p\left(\mathbf{x}_{1:n_x} | \phi_x, \sigma^2_x\right)}{p\left(\mathbf{z}_{1:n_z} | \phi_x, \sigma^2_x, \lambda\right)} p\left(\phi_x, \sigma^2_x, \lambda\right), \end{aligned} \tag{A.18}$$

Finally, given the conditional independence across scales, the parameters of the observable coarse level can be sampled from the following distribution

$$p\left(\phi_z, \sigma^2_z | \mathbf{z}_{1:n_z}\right) \propto q\left(\mathbf{z}_{1:n_z} | \phi_z, \sigma^2_z\right) p\left(\phi_z, \sigma^2_z\right),$$

The parameters for both the predictive regression and the dynamics of $\mathbf{z}_{1:n_z}$ can be easily updated using standard conjugate priors. Posterior simulation is implemented through a Gibbs sampler algorithm (Geman and Geman 1984). We propose a collapsed multi-move Gibbs sampling algorithm (see e.g. Roberts and Sahu 1997, Casella and Robert 2004, Ferreira et al. 2006 and Ferreira and Lee 2007), where expected returns and the parameters are sampled in blocks. More specifically we combine forward filtering backward sampling (see Frühwirth-Schnatter 1994 and Carter and Kohn 1994) to extract the expected returns from scale-heterogenous forecasting variables, a Metropolis-Hasting step for direct inference on a set of parameters for which conjugacy is not obtainable and standard Bayesian updating for the parameters of the predictive regression and the observable scale-specific predictors. At each iteration the Gibbs sampler sequentially cycles through the following steps:

1. Draw $\mathbf{x}_{1:n_x}$ conditional on $\sigma^2_x, \phi_x, \lambda$, and $\mathbf{z}_{1:n_z}$.
2. Draw σ^2_x conditional on ϕ_x, λ, $\mathbf{z}_{1:n_z}$, $\mathbf{x}_{1:n_x}$.
3. Draw ϕ_x conditional on σ^2_x, λ, $\mathbf{z}_{1:n_z}$, $\mathbf{x}_{1:n_x}$.
4. Draw λ_j conditional on σ^2_x, ϕ_x, $\mathbf{z}^j_{1:n_{z[j]}}$, $\mathbf{x}_{1:n_x}$ for $j = 1, \cdots, K$.
5. Draw $\sigma^2_{z[j]}$ conditional on $\phi_{z[j]}$, and $\mathbf{z}^j_{1:n_{z[j]}}$ for $j = 1, \cdots, K$.

[35] Assuming stock returns and the expected returns are on the same scale in the initial predictive regression does not represent a limit of the model. In fact, our modeling setting is general and the estimation procedure can be easily adapted. The reason why we keep consistency in the scales for stock excess returns and investors' expectations is to gain interpretability in comparison to existing research on predictability and optimal portfolio allocation (see, e.g. Barberis 2000 and Pastor and Stambaugh 2009).

6. Draw $\phi_{z[j]}$ conditional on $\sigma^2_{z[j]}$, and $\mathbf{z}^j_{1:n_{z[j]}}$ for $j = 1, \cdots, K$.
7. Draw σ^2 conditional on $\mathbf{y}_{1:n_y}$, $\mathbf{x}_{1:n_x}$, α and β.
8. Draw α, β conditional on $\mathbf{y}_{1:n_y}$, $\mathbf{x}_{1:n_x}$, and σ^2.

Steps 7-8 are collapsed since α, β, σ^2 are conditionally independent on $\mathbf{z}_{1:n_z}$ conditioning on latent expected returns $\mathbf{x}_{1:n_x}$. In order to reduce the amount of auto-correlation across draws we keep one every other draw from the posterior and discard the rest.

Step 1. Drawing the expected returns

To draw the time series of expected returns we use a forward filtering, backward sampling (FFBS) estimation scheme (see Carter and Kohn 1994, Frühwirth-Schnatter 1994 and West and Harrison 1997 for more details). Although the marginal distribution of the scale-specific predictors is substituted by $q\left(\mathbf{z}_{1:n_z}|\phi_z, \sigma^2_z\right)$, we can exploit the linkage equation (3) to extract expected returns on observable predictors. More precisely, we can rewrite (1)-(3) as a states-space model in which scale-specific predictors characterize the observables equation and investors' expectations are latent. For the ease of exposition we write the recursion for $K = 2$, any generalization to $K > 2$ would readily apply;

$$\begin{aligned} \mathbf{z}_s &= F\mathbf{x}_s + \epsilon_{z,s}, \qquad \epsilon_{z,s} \sim N\left(0, \Sigma_{zz}\right), \\ \mathbf{x}_s &= G\mathbf{x}_{s-1} + \epsilon_{x,s}, \qquad \epsilon_{x,s} \sim N\left(0, \Sigma_{xx}\right), \end{aligned} \tag{A.19}$$

where $\mathbf{z}_s = \left(z_{s[1]}, z_{s[2]}\right)$, $\mathbf{x}'_s = \left(x_{(s-1)m+1}, ..., x_{sm}\right)$, $\Sigma_{xx,ij} = \sigma^2_x \phi_x^{|i-j|}\left(1 - \phi_x^{2\min(i,j)}\right) / \left(1 - \phi^2_x\right)$, $\Sigma_{zz} = diag\left(\tau_1, \tau_2\right)$ and

$$G = \begin{pmatrix} 0 & \cdots & 0 & \phi_x \\ 0 & \cdots & 0 & \phi_x^2 \\ \vdots & & \vdots & \vdots \\ 0 & \cdots & 0 & \phi_x^m \end{pmatrix}, \qquad F = \begin{pmatrix} m^{-1}\iota_m \\ I_{m/m_1} \otimes m_1^{-1}\iota_{m_1} \end{pmatrix},$$

with $m > m_1$ the window of coarsening non-overlapping averages for the two observable predictors, I_{m/m_1} an identity matrix of size $m/m_1 \in \mathbb{R}$ and ι_m an m-dimensional vector of ones. For identification purposes, exogenous across scale-specific predictors and expected returns are uncorrelated to each other;

$$\begin{bmatrix} \epsilon_{z[1],s} \\ \epsilon_{z[2],s} \\ \epsilon_{x,s} \end{bmatrix} \sim N\left(0, \Sigma\right), \qquad \text{with} \qquad \Sigma = \begin{bmatrix} \tau_1 & 0 & 0 \\ 0 & \tau_2 & 0 \\ 0 & 0 & \Sigma_{xx} \end{bmatrix} \tag{A.20}$$

with $\tau_j = \lambda_j \left(F' V_x F\right)_{jj} I$ with I as the $n_{z[1]}-$square identity matrix and λ_j measuring the *between* resolutions uncertainty between predictor at the scale j and expected returns.

Filtering

The first stage follows a standard Kalman filtering recursion. Define the cumulative information about the predictor at the scale s as $D_s = D_{s-1} \cap \{z_s\}$, $s = 1, ..., n_z$, and the conditional moments as

$$\begin{aligned} a_s &= E\left[\mathbf{x}_s|D_{s-1}\right], \qquad f_s = E\left[z_s|D_{s-1}\right], \qquad m_s = E\left[\mathbf{x}_s|D_s\right], \\ R_s &= Var\left[\mathbf{x}_s|D_{s-1}\right], \qquad Q_s = Var\left[z_s|D_{s-1}\right], \qquad C_s = Var\left[\mathbf{x}_s|D_s\right], \end{aligned}$$

Conditioning on the parameters is assumed throughout the description of the FFBS setting but is omitted for the ease of exposition. We impose an uninformative initial prior $\mathbf{x}_0|D_0 \sim N\left(m_0, C_0\right)$ with $m_0 = \mathbf{0}_{m\times 1}$ an $m-$dimensional vector of zeros, and $C_0 = 1000 I_m$ the prior covariance matrix. At time $s = 1$ the filtering

recursion starts with a propagation step

$$\mathbf{x}_1|D_0 \sim N\left(a_1, R_1\right),$$

with

$$\begin{aligned} a_1 &= E\left[\mathbf{x}_1|D_0\right] = GE\left[\mathbf{x}_0|D_0\right] = Gm_0, \\ R_1 &= Var\left[\mathbf{x}_1|D_0\right] = GVar\left[\mathbf{x}_0|D_0\right]G' + \Sigma_{xx} = GC_0G' + \Sigma_{xx}\ . \end{aligned}$$

The second prediction step for the observables is defined as

$$z_1|D_0 \sim N\left(f_1, Q_1\right),$$

with

$$\begin{aligned} f_1 &= E\left[z_1|D_0\right] = E\left[F\mathbf{x}_1|D_0\right] = FE\left[\mathbf{x}_1|D_0\right] = Fa_1, \\ Q_1 &= Var\left[z_1|D_0\right] = FR_1F' + \Sigma_{zz}\ . \end{aligned}$$

The last step is updating the beliefs once information about predictors is available at time $s = 1$. The filtering distribution is defined as

$$\mathbf{x}_1|D_1 \sim N\left(m_1, C_1\right),$$

with

$$\begin{aligned} m_1 &= E\left[\mathbf{x}_1|D_1\right] = a_1 + A_1e_1, \\ C_1 &= Var\left[\mathbf{x}_1|D_1\right] = R_1 - A_1Q_1A_1', \end{aligned}$$

where the Kalman gain $A_1 = R_1F'Q_1^{-1}$, and $e_1 = z_1 - f_1$. Continuing in this fashion, we find that all conditional densities are normally distributed, and we obtain all the required moments for $s = 2, ..., n_y$:

$$\begin{aligned} a_s &= Gm_{s-1}, \\ R_s &= GC_{s-1}G' + \Sigma_{xx}, \\ f_s &= Fa_s, \end{aligned}$$

$$\begin{aligned} Q_s &= FR_sF' + \Sigma_{zz}, \\ m_s &= a_s + A_se_s, \\ C_s &= R_s - A_sQ_sA_s', \\ A_s &= R_sF'Q_s^{-1}, \\ e_s &= z_s - f_s, \end{aligned}$$

Sampling

We want to draw the hidden states $\mathbf{x}_s, s = 1, \cdots, n_z$ conditional on the available information D_{n_z} for the long-run predictor. The backward-sampling approach relies on the Markov property of the evolution of the underlying state

$$p\left(\mathbf{x}_0, \mathbf{x}_1, \cdots, \mathbf{x}_{n_z}|D_{n_z}\right) = p\left(\mathbf{x}_{n_z}|D_{n_z}\right)p\left(\mathbf{x}_{n_z-1}|\mathbf{x}_{n_z}, D_{n_z-1},\right)\cdots p\left(\mathbf{x}_0|\mathbf{x}_1, D_1,\right),$$

We first sample $\mathbf{x}_{n_z}$ from $p\left(\mathbf{x}_{n_z}|D_{n_z}\right)$ from the last step of the Kalman filtering above. Then, from $s = n_z - 1, n_z - 2, \cdots, 1, 0$, we sample $\mathbf{x}_s$ from the conditional density $p\left(\mathbf{x}_s|\mathbf{x}_{s+1}, D_s,\right)$. To obtain the backward recursion,

first recall that

$$\mathbf{x}_s|D_{s-1} \sim N\left(a_s, R_s\right), \qquad \text{with} \qquad a_s = Gm_{s-1}, \qquad \text{and} \qquad R_s = GC_{s-1}G' + \Sigma_{xx},$$

and

$$\mathbf{x}_s|D_s \sim N\left(m_s, C_s\right), \qquad \text{with} \qquad m_s = a_s + A_s e_s, \qquad \text{and} \qquad C_s = R_s - A_s Q_s A_s',$$

where $A_s = R_s F' Q_s^{-1}$,, and $e_s = z_s - f_s$. Then the backward recursion can be defined as

$$\mathbf{x}_s|\mathbf{x}_{s+1}, D_s \sim N\left(h_s, H_s\right), \tag{A.21}$$

with

$$\begin{aligned} h_s &= m_s + B_s\left(\mathbf{x}_{s+1} - a_{s+1}\right), \\ H_s &= C_s - B_s R_{s+1} B_s', \end{aligned}$$

and $B_s = C_s G' R_{s+1}^{-1}$ (see West and Harrison 1997, Chapter 4, for additional details).

Step 2 and 3. Sampling the parameters for the expected returns

The parameters of the fine level and the link equation can not be generated by standard Bayesian updating as priors are not conjugate and posterior distributions are not available in closed form for sampling. To overcome this problem we approximate the posterior samples using Metropolis-Hastings (MH) proposals. Once a proposal is generated, the full conditional (A.18) is used to compute the acceptance probability.

For the simulation step $g = 1, ..., G$, the proposal for σ_z^2 is generated from a uniform distribution $\sigma_z^{2(m)} \sim U\left(\sigma_z^{2(m-1)}/\delta_{\sigma_z}, \sigma_z^{2(m-1)} \cdot \delta_{\sigma_z}\right)$ where $\delta_{\sigma_z} > 1$ is a tuning parameter that must be fixed to ensure a good trade-off between the convergence of the Markov chain and the ability of the algorithm to search within the parameter space. The acceptance probability is computed as

$$\begin{aligned} \pi\left(\sigma_x^{2(g)}, \sigma_x^{2(g-1)}\right) &= \min\left(1, \frac{\sigma_x^{2(g-1)} p\left(\mathbf{z}_{1:n_z}|\mathbf{x}_{1:n_x}, \phi_x^{(g-1)}, \lambda^{(g-1)}, \sigma_x^{2(g)}\right) p\left(\mathbf{x}_{1:n_x}|\phi_x^{(g-1)}, \sigma_x^{2(g)}\right) p\left(\sigma_x^{2(g)}\right)}{\sigma_x^{2(g)} p\left(\mathbf{z}_{1:n_z}|\mathbf{x}_{1:n_x}, \phi_x^{(g-1)}, \lambda^{(g-1)}, \sigma_x^{2(g-1)}\right) p\left(\mathbf{x}_{1:n_x}|\phi_x^{(g-1)}, \sigma_x^{2(g-1)}\right) p\left(\sigma_x^{2(g-1)}\right)} \times \right. \\ &\left. \times \frac{p\left(\mathbf{z}_{1:n_z}|\phi_x^{(g-1)}, \lambda^{(g-1)}, \sigma_x^{2(g-1)}\right)}{p\left(\mathbf{z}_{1:n_z}|\phi_x^{(g-1)}, \lambda^{(g-1)}, \sigma_x^{2(g)}\right)}\right), \\ &= \min\left(1, \frac{\sigma_x^{2(g-1)} IG\left(\sigma_z^{2(g)}|\nu_{\sigma_x}^*/2, \nu_{\sigma_x}^* s_{\sigma_x}^*/2\right) p\left(\mathbf{z}_{1:n_z}|\phi_x^{(g-1)}, \lambda^{(g-1)}, \sigma_x^{2(g-1)}\right)}{\sigma_x^{2(g)} IG\left(\sigma_z^{2(g-1)}|\nu_{\sigma_x}^*/2, \nu_{\sigma_x}^* s_{\sigma_x}^*/2\right) p\left(\mathbf{z}_{1:n_z}|\phi_x^{(g-1)}, \lambda^{(g-1)}, \sigma_x^{2(g)}\right)}\right), \end{aligned}$$

with

$$\begin{aligned} \nu_{\sigma_x}^* &= \nu_{\sigma_x} + n_x + n_z - 1, \\ \nu_{\sigma_x}^* s_{\sigma_x}^* &= \nu_{\sigma_x} s_{\sigma_x} + \sum_{i=2}^{n_x}\left(x_i - \phi_x^{(g-1)} x_{i-1}\right)^2 + \left(\mathbf{z}_{1:n_z} - F\mathbf{x}_{1:n_x}\right)'\left(\mathbf{z}_{1:n_z} - F\mathbf{x}_{1:n_x}\right)/\boldsymbol{\tau}^{(g-1)\prime} \end{aligned}$$

with $n_z = n_{z[1]} + n_{z[2]}$, $\tau = (\tau_1, \tau_2)$, $\nu_{\sigma_\lambda}, \nu_{\sigma_\lambda} s_{\sigma_\lambda}$ the degrees of freedom and scale hyper-parameters of an Inverse-Gamma distribution. The marginal probability $p\left(\mathbf{z}_{1:n_z}|\phi_x, \lambda, \sigma_x^2\right)$ can be efficiently evaluated by using the Kalman filter recursion proposed in Step 1 above. In fact, $p\left(\mathbf{z}_{1:n_z}|\phi_x, \lambda, \sigma_x^2\right) = \prod_{s=1}^{n_z} p\left(z_s|z_{s-1}, \phi_x, \lambda, \sigma_x^2\right)$. To ensure stationarity, the proposal for the AR(1) parameter of the hidden fine scale process must have probability mass entirely inside the unit circle. We generate a proposal $\phi_x^{(g)} \sim U\left(\max\left(-1, \phi_x^{(g-1)} - \delta_{\phi_x}\right), \min\left(1, \phi_x^{(g-1)} + \delta_{\phi_x}\right)\right)$ where δ_{ϕ_x} has to be tuned to ensure convergence and effectiveness in the posterior approximation. The acceptance

probability is computed as

$$\pi\left(\phi_x^{(g)},\phi_x^{(g-1)}\right)=\min\left(1,\frac{\phi_x^{(g-1)}p\left(\mathbf{z}_{1:n_z}|\mathbf{x}_{1:n_x},\phi_x^{(g)},\lambda^{(g-1)},\sigma_x^{2(g)}\right)p\left(\mathbf{x}_{1:n_x}|\phi_x^{(g)},\sigma_x^{2(g)}\right)p\left(\phi_x^{(g)}\right)}{\phi_x^{(g)}p\left(\mathbf{z}_{1:n_z}|\mathbf{x}_{1:n_x},\phi_x^{(g-1)},\lambda^{(g-1)},\sigma_x^{2(g)}\right)p\left(\mathbf{x}_{1:n_x}|\phi_x^{(g-1)},\sigma_x^{2(g)}\right)p\left(\phi_x^{(g-1)}\right)}\times\right.$$
$$\left.\times\frac{p\left(\mathbf{z}_{1:n_z}|\phi_x^{(g-1)},\lambda^{(g-1)},\sigma_x^{2(g)}\right)}{p\left(\mathbf{z}_{1:n_z}|\phi_x^{(g)},\lambda^{(g-1)},\sigma_x^{2(g)}\right)}\right),$$
$$=\min\left(1,\frac{\phi_x^{(g-1)}p\left(\mathbf{z}_{1:n_z}|\mathbf{x}_{1:n_x},\phi_x^{(g)},\lambda^{(g-1)},\sigma_x^{2(g)}\right)N\left(\phi_x^{2(g)}|m_{\phi_x}^*,M_{\phi_x}^*\right)}{\phi_x^{(g)}p\left(\mathbf{z}_{1:n_z}|\mathbf{x}_{1:n_x},\phi_x^{(g-1)},\lambda^{(g-1)},\sigma_x^{2(g)}\right)N\left(\phi_x^{2(g-1)}|m_{\phi_x}^*,M_{\phi_x}^*\right)}\right.$$
$$\left.\times\frac{p\left(\mathbf{z}_{1:n_z}|\phi_x^{(g-1)},\lambda^{(g-1)},\sigma_x^{2(g)}\right)}{p\left(\mathbf{z}_{1:n_z}|\phi_x^{(g)},\lambda^{(g-1)},\sigma_x^{2(g)}\right)}\right),$$

The conditional likelihood in both the numerator and the denominator is computed numerically. The hyper-parameters of the truncated Gaussian are updated as

$$M_{\phi_x}^*=\left(M_{\phi_x}^{-1}+\mathbf{x}_{1:n_x}'\mathbf{x}_{1:n_x}/\sigma_x^{2(g)}\right)^{-1},\qquad m_{\phi_x}^*=M_{\phi_x}^*\left(M_{\phi_x}^{-1}m_{\phi_x}+\mathbf{x}_{1:n_x-1}'\mathbf{x}_{2:n_x}/\sigma_x^{2(g)}\right)^{-1}$$

Step 4. Sampling the between-scales uncertainty

The parameter λ_j has the natural interpretation in terms of the relative increase in uncertainty due to lack of agreement between the *jth* scale-specific observable predictor and investors' expectations. Let F_j correspond to the block of F specific to the *jth* predictors, the between-scales uncertainty nature of λ_j can be seen from the marginal distribution $p\left(z_{1:n_{z[j]}}^j\right)=N\left(0,F_jV_xF_j'+\lambda_jF_jV_xF_j'\right)$. The parameter λ_j enters in the model dynamics in highly non-linear and complicated. Luckily we can exploit its relationship with the parameters of the expected returns since, conditional on ϕ_x and σ_x^2, λ_j can be directly recovered from $\tau_j=\lambda_j\left(F_jV_xF_j'\right)_{11}$. For the simulation step $g=1,...,G$, we generate a proposal $\lambda_j^{(g)}\sim U\left(\max\left(0,\lambda_j^{(g-1)}-\delta_\lambda\right),\min\left(10,\lambda_j^{(g-1)}+\delta_\lambda\right)\right)$ where δ_λ has to be tuned to ensure convergence and effectiveness in the posterior approximation. Given the simulated parameters of the fine level at the *gth* iteration, such proposal implies a given $\tau_j^{(g)}=\lambda_j^{(g)}\left(F_jV_x^{(g)}F_j'\right)_{11}$ with $\left\{V_x^{(g)}\right\}_{ij}=\sigma_x^{2(g)}\phi_x^{|i-j|(g)}/\left(1-\phi_x^{2(g)}\right)$. The acceptance probability is computed as

$$\pi\left(\lambda_j^{(g)},\lambda_j^{(g-1)}\right)=\min\left(1,\frac{\tau_j^{(g-1)}p\left(\mathbf{z}_{1:n_{z[j]}}^j|\mathbf{x}_{1:n_x},\phi_x^{(g)},\lambda_j^{(g)},\sigma_x^{2(g)}\right)p\left(\tau_j^{(g)}\right)}{\tau_j^{(g)}p\left(\mathbf{z}_{1:n_{z[j]}}^j|\mathbf{x}_{1:n_x},\phi_x^{(g)},\lambda^{(g-1)},\sigma_x^{2(g)}\right)p\left(\tau_j^{(g-1)}\right)}\times\right.$$
$$\left.\times\frac{p\left(\mathbf{z}_{1:n_{z[j]}}^j|\phi_x^{(g)},\lambda_j^{(g-1)},\sigma_x^{2(g)}\right)}{p\left(\mathbf{z}_{1:n_{z[j]}}^j|\phi_x^{(g)},\lambda_j^{(g)},\sigma_x^{2(g)}\right)}\right),$$
$$=\min\left(1,\frac{\tau_j^{(g-1)}IG\left(\tau_j^{(g)}|\nu_{\lambda_j}^*/2,\nu_{\lambda_j}^*s_{\lambda_j}^*/2\right)p\left(\mathbf{z}_{1:n_{z[j]}}^j|\phi_x^{(g)},\lambda_j^{(g-1)},\sigma_x^{2(g)}\right)}{\tau_j^{(g)}IG\left(\tau_j^{(g-1)}|\nu_{\lambda_j}^*/2,\nu_{\lambda_j}^*s_{\lambda_j}^*/2\right)p\left(\mathbf{z}_{1:n_{z[j]}}^j|\phi_x^{(m)},\lambda_j^{(g)},\sigma_x^{2(g)}\right)}\right),$$

with

$$\nu_{\lambda_j}^*=\nu_{\lambda_j}+n_{z[j]}-1,$$
$$\nu_{\lambda_j}^*s_{\lambda_j}^*=\nu_{\lambda_j}s_{\lambda_j}+\left(\mathbf{z}_{1:n_{z[j]}}^j-F_j\mathbf{x}_{1:n_x}\right)'\left(\mathbf{z}_{1:n_{z[j]}}^j-F_j\mathbf{x}_{1:n_x}\right),$$

Notice that the independence across scales allows to simply compute $\lambda_j^{(g)}$ separately for $j=1,...,K$ by updating posterior on the basis of predictor-specific shocks $\mathbf{z}_{1:n_{z[j]}}^j-F_j\mathbf{x}_{1:n_x}$.

Step 5 and 6. Sampling the parameters for scale-specific predictors

For the observable scale-specific predictors, updating is generally simple as priors are conjugate. For example, if the coarse level follows an ARMA process then the coarse level parameters can be generated with the procedure proposed by Chib and Greenberg (1994). In the case of an AR(1) the full conditional distribution is readily available by standard Bayesian updating. Given the priors specified above (*need to specify priors*) posterior draws for each predictor $j = 1, ..., K$ are defined by;

$$\phi_{z[j]}|\sigma^2_{z[j]}, \mathbf{z}^j_{1:n_{z[j]}} \sim N\left(m^*_{\phi_{z[j]}}, M^*_{\phi_{z[j]}}\right), \qquad \sigma^2_{z[j]}|\phi_{z[j]}, \mathbf{z}^j_{1:n_{z[j]}} \sim IG\left(\nu^*_{\phi_{z[j]}}/2, \nu^*_{\phi_{z[j]}} s^*_{\phi_{z[j]}}/2\right),$$

with

$$M^*_{\phi_{z[j]}} = \left(M^{-1}_{\phi_{z[j]}} + \sigma^{-2}_{z[j]} \mathbf{z}^{j\prime}_{1:n_{z[j]}-1} \mathbf{z}^j_{1:n_{z[j]}-1}\right), \qquad m^*_{\phi_{z[j]}} = M^*_{\phi_{z[j]}}\left(M^{-1}_{\phi_{z[j]}} m_{\phi_{z[j]}} + \sigma^{-2}_{z[j]} \mathbf{z}^{j\prime}_{1:n_{z[j]}-1} \mathbf{z}^j_{2:n_{z[j]}}\right),$$

$$\nu^*_{\phi_{z[j]}} = \nu_{\phi_{z[j]}} + n_{z[j]} - 1, \qquad \nu^*_{\phi_{z[j]}} s^*_{\phi_{z[j]}} = \nu_{\phi_{z[j]}} s_{\phi_{z[j]}} + \sum_{i=2}^{n_{z[j]}} \left(z_{i[j]} - \phi_{z[j]} z_{i-1[j]}\right)^2,$$

The functional form of the posterior does not change if we consider N independent observable scale processes. Independence across scales allows to update separately each posterior. As such, we can generate scale-dependent posterior distribution for the autoregressive coefficients.

Step 7 and 8. Sampling the Parameters for the Predictive Regression

Conditional on the expected returns, the parameters of the predictive regression can be sampled independently by using standard conjugate analysis. For the simple case of a one-month investment horizon the full conditional distributions for the parameters take the form

$$\alpha, \beta|\sigma^2, \mathbf{y}_{1:n_y}, \mathbf{x}_{1:n_x} \sim N\left(m^*_\beta, M^*_\beta\right), \qquad \sigma^2|\alpha, \beta, \mathbf{y}_{1:n_y}, \mathbf{x}_{1:n_x} \sim IG\left(\nu^*_\beta/2, \nu^*_\beta s^*_\beta/2\right),$$

with

$$M^*_\beta = \left(M^{-1}_\beta + \sigma^{-2}\mathbf{X}'\mathbf{X}\right), \qquad m^*_\beta = M^*_\beta\left(M^{-1}_\beta m_\beta + \sigma^{-2}\mathbf{X}'\mathbf{y}_{2:n_y}\right),$$

$$\nu^*_\beta = \nu_\beta + n_y - 1, \qquad \nu^*_\beta s^*_\beta = \nu_\beta s_\beta + \sum_{i=2}^{n_y}\left(y_i - \alpha - \beta x_{i-1}\right)^2,$$

with $\mathbf{X} = [\iota, \mathbf{x}_{1:n_x-1}]$ and ι an $(n_x - 1)$-dimensional vector of ones. The functional form of the posterior does not change with the investment horizon. However, the updated hyper-parameters will differ given different scales imply different information. As such, we can generate scale-dependent posterior distribution for both intercepts and betas.

B Prior sensitivity analysis for the scale-specific predictors

In this section we report the posterior estimates for the parameters of the scale-specific predictors computed by using the alternative prior specifications outlined in Section 4. Figure B.1 shows the results. The left panel shows the posterior distributions for the persistence parameters. The blue line with diamond markers shows posterior estimates computed from the benchmark marginal prior which entails high persistence with a low degree of confidence. The location of the prior does not play a significant role. In fact, imposing a dispersed prior with low persistence (dashed green line) does not change posterior estimates.

[Insert Figure B.1 about here]

On the other hand, if one has strong reasons to believe that predictors show the same level of persistence across scales, posterior estimates deviate from the non-informative case, although in an economically negligible way. In particular, the difference between posterior means in the non-informative and informative cases equals to

0.05 for the log dividend-price, and 0.10 for the consumption-wealth ratio. Importantly, such differences do not imply dramatic changes in the autocorrelation structure of the predictors. As far as the conditional variances are concerned, the right panel shows that, under alternative priors, posterior distributions cannot be sensibly distinguished, although for the one-year predictor cay_t a highly confident prior on low volatility has a relevant effect on the posterior.

C MCMC convergence analysis

In order to effectively claim that our MCMC simulation make a good representation of the true joint distribution of the parameters we must rely on the assumption that draws from the full conditionals converge to samples obtained from the true distribution. This motivates the use of convergence diagnostics that aims at investigate the ergodic properties of the draws from the MCMC. In this section we report the results of a convergence analysis based on sub-samples of draws from the full conditionals (see Geweke 1992, Raftery and Lewis 1992 and Cowles and Carlin 1996 for more details). As a main test we compare the location of the sampled parameter on two different intervals across the chain. The underlying idea is that if the mean values of the parameter across two distant sub-samples of the Markov chain have similar locations in the state space, we can assume that draws from these sub-samples come from the same distribution. If values of the parameter $\theta^{(i)}$ are computed after each iteration of the Gibbs sampler, then the sequence of simulated values can be seen as a time series. Geweke (1992) suggests that the nature of the MCMC process imply the existence of a spectral density $S_\theta(0)$ for such sequence that has not discontinuity at frequency zero. If this assumption is met, then the estimator for $E[\theta]$;

$$\hat{\theta} = \frac{1}{M}\sum_{i=1}^{M}\theta^{(i)},$$

has asymptotic variance equal to $S_\theta(0)/M$. Let us consider two sub-samples of the Markov chain as $A = \{i; 1 \leq i \leq M_A\}$, $B = \{i; M_B \leq i \leq M\}$. The sample specific estimators for the expected value of θ are defined as

$$\hat{\theta}_A = \frac{1}{M_A}\sum_{i\in A}\theta^{(i)}, \qquad \hat{\theta}_B = \frac{1}{M - M_B + 1}\sum_{i\in B}\theta^{(i)},$$

Geweke (1992) convergence diagnostic after M iterations is calculated by taking the difference $\hat{\theta}_A - \hat{\theta}_B$ and dividing by the asymptotic standard error of the difference. If the ratios M_A/M and M_B/M are held fixed and $1 < M_A < M_B < M$ and $M_A + M_B < M$, then by central limit theorem, the distribution of this ratio approaches a standard Normal as M tends to infinity;

$$Z_n = \frac{\hat{\theta}_A - \hat{\theta}_B}{\sqrt{\frac{1}{M_A}\hat{S}_\theta^A(0) + \frac{1}{M - M_B + 1}\hat{S}_\theta^B(0)}} \quad \longrightarrow \quad N(0,1) \qquad \text{for} \qquad M \quad \longrightarrow \quad \infty,$$

with $\hat{S}_\theta^A(0)$ and $\hat{S}_\theta^B(0)$ the estimated asymptotic variance for $\hat{\theta}_A$ and $\hat{\theta}_B$, respectively. Geweke (1992) suggests to take $M_A = .1M$ and $M_B = .5M$. The null hypothesis that the chain is converging is rejected if $|Z_n|$ is large. Panel A of Table B.1 shows the results obtained by either including or excluding the initial burn-in sample. This is done to check convergence after the initial $M^\star = 20,000$ draws are discarded. Notice we report the results of the parameters for the expected returns as posterior can be derived analytically and does not required, by definition, convergence diagnostics.

[Insert Table B.1 about here]

In computing the asymptotic variance we use a bandwith of 4% of the sample sizes considered. As we would expect without disregarding the initial burni-in sample the null hypothesis of convergence is strongly rejected for all parameters. By excluding an initial pre-sample of 20,000 draws convergence is reached. The null hypothesis of convergence cannot be rejected at the 10% significance level for both the persistence parameter and the between-levels uncertainty, while cannot be rejected at the 5% for the conditional variance of expected returns. All in all, convergence diagnostics show that after disregarding the burn-in sample we can sensibly assume that draws from

the full conditional are a safe approximation of the true underlying distribution. For the sake of completeness we also compute the Effective Sample Size (ESS henceforth). The latter tells us how much information can we actually use to estimate a single parameter, given the properties of the Markov chain. Indeed, if there is correlation between successive draws of $\theta^{(i)}$, then we might expect that our sample of draws has not revealed as much information of the posterior distribution of the parameter as we could have had by having i.i.d. draws. Let $\rho_k = corr\left(\theta^{(i)}, \theta^{(i-k)}\right)$ be the correlation between two draws of the same chain at distance $k > 1$, then the ESS can be computed as;

$$M_{EFF} = \frac{M}{1 + 2\sum_{k=1}^{\infty} \rho_k},$$

In all our estimates we reduce the amount of correlation by keeping one every other draw from the posterior distribution and discard the rest. As a by-product of the EFF we can compute the inefficiency factor, which simply consists of the denominator in M_{EFF} above, and measures how far posterior draws are far from an i.i.d. sample. Panel A of Table B.1 shows the values for the ESS and the inefficiency factor for the 30,000 retained draws out of 50,000 simulations. Not surprisingly, last column shows that posterior draws are far from being i.i.d. Despite thinning of the chain, the auto-correlation of the draws is still quite high. However, the ESS is sufficiently high to provide sensible estimates of the model parameters. For instance, our estimation algorithm guarantees that we have more than 800 meaningful draws to estimate the posterior mean of the persistence parameter ϕ_x.