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

Daniele Bianchi† Andrea Tamoni‡

**Abstract**

We propose a multi-scale time-series model to gauge the persistence of discount rates. The multiscale framework exploits the low frequency information from observable economic predictors such as the consumption-wealth and dividend-price ratios as a prior to update the distribution of latent expected returns at high frequency. The short-term dynamics for expected returns implied by our framework feature long-range dependence, and are hard to reconcile with low-order auto-regressive models. Finally, we show that our documented long-range dependence in discount rates has first-order effects on forecasting, especially in the long-run.

**Keywords:** Expected Returns, Empirical Asset Pricing, Persistence, Multi-Scale Time-Series, Bayes, MCMC.

*We thank Jaroslav Borovicka, Mike Chernov, Lars A. Lochstoer, Ian Martin, Giorgio Primiceri, Barbara Rossi, Frank Schorfheide, Fabio Trojani, Michela Verardo, Pietro Veronesi, Mike West, and Stan Zin for their helpful comments and suggestions. We also thank seminar participants at the Stockholm School of Economics, London School of Economics, USI-Lugano, Nova School of Business and Economics, Warwick Business School, University of Kent, the CEPR/EABCN Conference: "Asset Prices and the Macro Economy" in Mannheim, the 2016 SFS Cavalcade at Rotman School of Management, the 2016 NBER-NSF Seminar on Bayesian Inference in Econometrics and Statistic at UPenn, the 9th SoFiE annual conference in Hong Kong, the 2016 Barcelona GSE Summer Forum, the 9th ECB Workshop on Forecasting Techniques, and the Workshop on Macroeconomic and Financial Time Series Analysis at Lancaster University.

†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 question of what moves stock prices is of great interest to economists and market participants at least since Shiller (1981). Given the long duration of equity, stock prices are highly sensitive to *persistent* fluctuations in expected returns. We use the term "expected returns" to describe statistical models that use observable economic predictors to extract the latent expectations on future stock returns. Modeling the persistence of expected returns has a long history in the empirical asset pricing literature. A common and widespread practice is to specify persistent transitional dynamics for latent expectations over a short horizon (typically monthly), and then use forward aggregation to infer the medium- and long-run behavior of future expected returns.[1]

The main contribution of this paper is to show that a much stronger form of persistence is required for expected returns to be consistent with the empirical evidence from both short- and long-term predictability of stock returns: our estimated autocorrelations take far longer to decay than the exponential rate associated with low-order ARMA$(p, q)$ specifications. This result is obtained by combining in a coherent way information from observable predictors over alternative horizons. Specifically, we acknowledge the fact that different covariates convey frequency-specific information on the dynamics of discount rates (see Bandi et al., 2017).[2] For example, while the dividend-price ratio is known to reveal long-term tendencies in asset returns, the consumption-wealth ratio shows its strongest forecasting power at the intermediate horizon.

Methodologically, we build upon the multi-scale modeling approach originally proposed by Ferreira and Lee (2007). We generalize and extend their framework along three key dimensions: first, we allow for discount rates at the highest frequency to be latent, as typically assumed in theoretical asset pricing; second, we combine multiple frequency-specific sources of conditioning information in a unifying setting; and third, we document both theoretically and empirically the temporal aggregation prop-

[1] See, e.g. Stambaugh (1999), Barberis (2000), Pastor and Stambaugh (2009), Van Binsbergen and Koijen (2010), and Carvalho et al. (2016) among others.

[2] Throughout the paper we use the terms frequency, horizon, and time scale as synonyms. Also, we use the terms expected returns and discount rates, interchangeably.

erties of multi-scale time-series models. We estimate the model by implementing a Metropolis-within-Gibbs sampling algorithm, which allows to explicitly consider the uncertainty over the latent states and the model parameters.

Our multiscale approach uses the low-frequency information in stock returns predictors as a prior to update the high-frequency distribution of latent expected returns. This approach is novel since it explicitly entails an information updating process from low- to high-frequencies. In this respect, our multi-scale time-series framework attributes a unique role to low-frequency information, which is used to update the dynamics of latent states over a short horizon.

We provide evidence on long-range dependence in expected returns which suggests a key role for small shocks at long lags. This long-range dependence is the feature that yields, upon temporal aggregation, dynamics for discount rates that are consistent with the evidence from standard reduced-form linear predictive regressions. Our documented discount rates dynamics over alternative horizons can be used to restrict the time-series properties of the relevant inter-temporal marginal rate of substitution, and can also guide the search for an economic explanation of movements in aggregate investors' expectations for assets returns. As a matter of fact, although in this paper we take a reduced form approach, it is interesting to observe that our evidence in favour of multi-frequency dynamics in risk premia is supportive of models with heterogeneity across agents and goods. For example, Lochstoer (2009) proposes a 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.

The multi-scale methodology we propose is connected to a recent literature that combines data sampled at different frequencies to improve out-of-sample predictability (see, e.g., the Mixed Data Sampling Regression Models (MIDAS) approach originally proposed in Ghysels et al., 2004, the Heterogeneous Autoregressive model of Realized Volatility (HAR-RV) proposed in Corsi, 2009, the Mixed-frequency forecasting framework of Schorfheide and Song, 2015).

Our evidence of long-memory type of dynamics for expected returns complement the findings in Abadir et al. (2013), who document a form of persistence in various macroeconomic and financial series that cannot be picked up by linear ARIMA models. Whereas Abadir et al. (2013) focus on observable series, our approach yields the autocorrelation function for a latent series using observable proxies at alternative frequencies. Our empirical evidence favouring expected returns dynamics with many lags is also consistent with the findings in Fuster et al. (2010), who show that the hump-shaped dynamics in many macroeconomic and financial variables are difficult to reconcile with low-order models. Fuster et al. (2010) "place some faith in specifications with many lags" based on Monte Carlo analysis suggesting that model selection criteria generally fail to detect statistical models with many lags. Our framework can be interpreted as a tool to detect small positive responses at *large* lags by making direct use of low-frequency information in predictors.[3] Similarly, one can also interpret the methodology we propose as a way to mitigate slight misspecification of finite order autoregressive dynamics for expected returns. In this respect, our paper is close in spirit to Müller and Stock (2011), which discuss mispecified finite order VAR and suggest to use a spectral domain prior to exploit small, possibly long-lag linear predictability beyond the initial autoregressive approximation. Our approach offers a time-domain interpretation of their framework.

The remainder of the paper is organized as follows. Section 2 outlines the model and provide an in-depth discussion of the updating rule from low- to high-frequency. Section 3 describes the estimation strategy. Section 4 presents our empirical analysis. Section 5 concludes.

# 2 Modeling framework

We first show that the dynamics of long-term expected returns implied by the aggregation of low-order ARMA$(p, q)$ specifications are at odds with the evidence from predictive regressions. Let us assume that the latent expected returns, $x_t = E_t\left[r_{t+1}\right]$,

[3]Fuster et al. (2010) document the importance of long-lags using impulse response function. We cannot use their approach since the expected return series is latent.

at time $t = 1, ..., n_x$ evolve as a stationary 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 is Gaussian $p(x_{1:n_x}) = N(0, V_x)$ with $\{V_x\}_{ij} = \sigma_x^2 \phi_x^{|i-j|} / (1 - \phi_x^2)$. We are interested in the long-term implications of this short-run, autoregressive dynamics for expected returns. Let's call $z_s$ the process resulting from the $m$-period aggregation of short-term discount rates, i.e.

$$z_s = m^{-1} \sum_{i=1}^{m} x_{(s-1)m+i}. \tag{2}$$

Eq. (2) implies that the marginal distribution of the long-term discount rates is also Gaussian $z_{1:n_z} \sim N(0, V_z)$ with covariance matrix given by

$$V_z = A V_x A', \tag{3}$$

with $A$ a sparse matrix whose non-zero elements are all $1/m$ and the non-zero elements in row $i$ are those in columns $(i-1)m + 1$ to $im$.[4] Figure 1 shows the first-order autocorrelation implied by $V_z$ for alternative aggregation horizons $m = 1, \ldots, 48$ and $\phi_x$ set to 0.977 as per Table II in Barberis (2000).[5]

[Insert Figure 1 about here]

The figure shows a quickly vanishing memory implied by the short-memory AR(1) process. As we look over longer horizons, the dynamics of the latent state variable

[4]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 $z$ evolve as follows:

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

Hence, the name *multi-scale* time series model: the process $x$ evolves at a fine time-scale (i.e. high frequency), whereas the process $z$ evolves over a coarser time-scale (i.e. lower frequency).

[5]The analysis in this section relies on autocorrelation functions. Therefore the parameter $\sigma_x$ is inconsequential given our assumption of homoscedastic expected returns.

gets closer to an i.i.d process.[6]

However, the serial correlation structure displayed in Figure 1 is at odds with the data. Panel A in Table 2 shows the results from fitting an AR(1) process to the $dp_t$ series temporally aggregated over four-years. The estimate $\hat{\phi}_{dp} = 0.788$ implies an half-live of about $-\log(2)/\log(0.788) \approx 3$ years for the four-year discount rates; this value is in sharp contrast with the much lower autocorrelation (in the order of 0.5, with an half life of about 1-year) implied by the marginal distribution $p(z_{1:z_n}) = N(0, V_z)$. Similarly, the autoregressive coefficient obtained upon fitting an AR(1) process to the consumption-wealth ratio series, $cay_t$, temporally aggregated over one year is $\hat{\phi}_{cay} = 0.512$ (see Table 2–Panel B); once again, this estimates is at odds with the (too high) value implied by $V_z$ at one-year horizon (see Figure 1).[7] As a whole, a highly persistent low-order autoregressive process for high-frequency expected returns implies too little persistence at the long-end and the consequent inability to capture the low-frequency component of expected returns proxied by $dp_t$. Increasing the root of the short-term expected returns comes at the cost of missing the higher-frequency component of expected return captured by $cay_t$.

Let us now assume that we have available a predictor $z_s$ that "imperfectly" captures the dynamics of long-term expected returns. We generalize Eq. (2) as:[8]

$$z_s = m^{-1} \sum_{i=1}^{m} x_{(s-1)m+i} + \epsilon_s, \qquad \text{with} \qquad \epsilon_s \sim N\left(0, \lambda \left(A' V_x A\right)_{11} I\right), \tag{4}$$

where the role of the error term $\epsilon_s$ is to acknowledge the fact that $z_s$ represents an imperfect proxy of the $m$-period latent expected returns. The conditional distribution

[6] The fact that short-memory processes in the form of low-order autoregressive do not generate enough persistence upon aggregation is a general result that can be found, e.g., in Tiao (1972). In particular, Tiao (1972) studies the *size* of the $z$'s process coefficients implied by the aggregation of an ARMA($p$,$q$). He shows that the limiting model for $z$ as $m$ goes to infinity has no autoregressive terms – see Table 2 in Tiao, 1972. Also note that this results is not due to small samples.

[7] Our conclusion holds even after accounting for estimation uncertainty in the persistence of the predictors, see Table 2.

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

of the predictor given the latent state is

$$
\begin{aligned}
p\left(z_{1:n_z}|x_{1:n_x}\right) &= \prod_{s=1}^{n_z} N\left(m^{-1}\sum_{i=1}^{m} x_{(s-1)m+i}, \lambda\left(A'V_xA\right)_{11} I\right), \\
&= N\left(A \cdot x_{1:n_x}, \lambda\left(A'V_xA\right)_{11} I\right),
\end{aligned} \tag{5}
$$

with $I$ as the $n_z-$square identity matrix. Accordingly, we dub $\lambda$ a measure of disagreement *between* the frequency-specific predictors and the aggregate short-term expected returns: when the dynamics of the latter, $Ax_{1:n_x}$, are consistent with those of the predictor $z$ then $\lambda = 0$, and no disagreement across time-scales takes place. From the initial distribution of latent expected returns $p\left(x_{1:n_x}\right)$ and the linkage equation (4), we can obtain the marginal $p\left(z_{1:n_z}\right)$, i.e.

$$
p\left(z_{1:n_z}\right) = \int p\left(z_{1:n_z}|x_{1:n_x}\right) p\left(x_{1:n_x}\right) dx_{1:n_x} = N\left(0, A'V_xA + \lambda\left(A'V_xA\right)_{11} I\right) \ , \tag{6}
$$

However, as discussed above, when $x_{1:n_x}$ follows a persistent low-order ARMA process, like an AR(1), then the serial correlation structure implied by the marginal $p\left(z_{1:n_z}\right)$ is possibly not consistent with the persistence documented by different observable predictors, i.e. the $z_s$ (see again Figure 1).

The idea of our multi-scale model is to use a prior on the marginal distribution of the low-frequency predictor in order to revise the initial beliefs on the dynamics of expected returns. More specifically, let us assume that the $n_z$-dimensional stationary distribution $q\left(z_{1:n_z}\right)$ for the low-frequency signal is well approximated by an AR(1) evolving over a coarse grid,

$$
z_s = \phi_z z_{s-1} + \eta_s, \qquad \text{with} \qquad \eta_s \sim N\left(0, \sigma_z^2\right), \tag{7}
$$

so that its implied stationary distribution is given by $q\left(z_{1:n_z}\right) = N\left(0, Q_z\right)$ with $\left\{Q_z\right\}_{ij} = \sigma_z^2 \phi_z^{|i-j|} / \left(1-\phi_z^2\right)$. Since the dynamics described by $p\left(z_{1:n_z}\right)$ and $q(z_{1:n_z})$ are inconsistent, we look for a revised marginal distribution for $x_{1:n_x}$ that is coherent with the information coming from the low-frequency predictor. The following proposition address the issue of restoring consistency across frequencies:

**Proposition 1.** *Assume that (a) expected returns evolves at high-frequency as in Eq.* (1)*; (b) the link equation takes the form in Eq.* (4)*; (c) the low-frequency predictor evolves as in Eq.* (7)*. If the low-frequency predictor is sufficient to revise the beliefs of expected returns, i.e.* $p\left(x_{1:n_x}|z_{1:n_z}\right) = q\left(x_{1:n_x}|z_{1:n_z}\right)$*, then*

**(i)** *the revised marginal distribution of expected returns is given by*

$$\begin{aligned} q\left(x_{1:n_x}\right) &= \int q\left(x_{1:n_x}, z_{1:n_z}\right) dz_{1:n_z} \\ &= \int p\left(x_{1:n_x}|z_{1:n_z}\right) q\left(z_{1:n_z}\right) dz_{1:n_z} = N\left(0, Q_x\right), \end{aligned} \tag{8}$$

*i.e. the revised distribution of* $x_{1:n_x}$ *is zero-mean normal with covariance matrix*

$$Q_x = V_x - B\left(W - Q_z\right)B' \tag{9}$$

*where* $B = V_x A' W^{-1}$ *and* $W = AV_xA' + \lambda\left(A'V_xA\right)_{11} I$*;*

**(ii)** *the following results hold in the limit*

$$\lim_{\lambda \to 0} AQ_xA' = Q_z\ , \tag{10}$$

$$\lim_{\lambda \to \infty} Q_x = V_x\ . \tag{11}$$

*Proof.* See Appendix A. □

Proposition 1, part (i), provides the revised marginal distribution of high-frequency expected returns that is obtained by incorporating information carried by the low-frequency predictor. Intuitively, $q(x_{1:n_x})$ can be interpreted as a revision of the initial prior $p(x_{1:n_x})$ and the conditional likelihood $p\left(z_{1:n_z}|x_{1:n_x}\right)$. Proposition 1, part (ii), state that the value of $\lambda$ controls how much the low frequency information carried by $q(z_{1:n_z})$ influences the behavior of the revised, high-frequency expected returns. Importantly, when $\lambda$ approaches zero then the covariance structure obtained from aggregating the revised high-frequency process $x_{1:n_x}$, namely $AQ_xA'$, converges to the one of the low-frequency predictor, $Q_z$. In other words consistency is restored when (i) updating takes place as per Proposition 1 *and* (ii) $\lambda$ is small. On the other hand, when $\lambda \to \infty$ the covariance structure of the high-frequency process converges to the one initially assumed and no revision occurs, i.e. $q(x_{1:n_x}) = p(x_{1:n_x})$. We will continue our discussion on the role of the between-scales uncertainty parameter $\lambda$ in Section 2.1.

To summarize, we start from a prior for expected returns $p(x_{1:n_x})$, which upon aggregation delivers dynamics for long-run expected returns, as captured by $AV_xA'$, potentially inconsistent from those proxied by a predictor $z_{1:n_z}$: $AV_xA' \neq Q_z$. We then use the distribution of the predictor $q(z_{1:n_z})$ to reconcile (if $\lambda$ is small) high-frequency dynamics with low-frequency ones. The value of $\lambda$ is an empirical matter and determines the possibility of restoring consistency: in Section 4.2 we discuss in details the estimated values of $\lambda$.

Various comments are in order. First, the multi-scale time series model outlined in Proposition 1 does not require $x_{1:n_x}$ and $z_{1:n_z}$ to follow AR(1) processes. The same framework 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 explicitly compare our estimates with standard persistent AR(1) dynamics which are normally assumed in the empirical finance, particularly asset pricing, literature (see, e.g. Stambaugh, 1999, Barberis, 2000, Pastor and Stambaugh, 2009, Van Binsbergen and Koijen, 2010 and Pástor and Stambaugh, 2012 among others). Second, although tightly parametrized by $\Theta = (\phi_x, \sigma_x^2, \lambda, \phi_z, \sigma_z^2)$, Section 2.1 shows that the multiscale framework is able to generate a variety of autocorrelation structures as encoded by $q(x_{1:n_x})$. Finally, it is key to observe that the model is not built using a bottom-up aggregation from the fine to the coarser levels of resolution, i.e. from (1) to (6) via (4), but rather, the learning scheme goes from the low frequency predictor to the latent higher frequency expected returns, i.e. from Eq. (7) to Eq. (8). This allows us to use information of the predictors at the "right" scale. The following sections show the implications of our modeling approach for the persistence of expected returns and for understanding their future trajectories.

## 2.1 Properties of the model

### 2.1.1 Persistence

The parameter $\lambda$ plays a key role in how information across time-scales is combined and how much the marginal distribution of $x_{1:n_x}$ is revised. To illustrate this point, we study the behavior of the covariance structure of the revised process $x_{1:n_x} \sim q\left(x_{1:n_x}\right)$ as specified in Proposition 1.

Figure 2 exemplifies the role of $\lambda$ in determining the autocorrelation structure of the revised higher-frequency process. In particular, the top panels show the role of $\lambda$ in controlling the decay of the autocorrelation function of the process $x_{1:n_x}$ aggregated over the long-horizon, i.e. $Ax_{1:n_x}$ for $m = 48$.

[Insert Figure 2 about here]

The top left (right) panel shows the autocorrelation of expected returns obtained by fixing the level of between-scales uncertainty at $\lambda = 0.01$ ($\lambda = 10$). The red and blue lines report the theoretical ACF implied by $Q_x$ (see Eq. (9)) and the ACF of the filtered series $x_{1:n_x}$, respectively.[9] Ceteris paribus, for low values of $\lambda$ the model generates long-lasting effects on the dynamics of the high-frequency process which departs from the one that can be obtained by a persistent AR(1) model with $\phi_x = 0.9$ (light blue line). Consistent with the limiting results in Proposition 1, the top right panel shows that if the low-frequency process $z_{1:n_z}$ does not bring sensible information to revise the latent expected returns, i.e. $\lambda \to \infty$, then the persistence of discount rates converges to that of a standard AR(1). To summarize, an increasing $\lambda$ has a natural interpretation in terms of the relative increase in uncertainty due to the lack of agreement across scales.

The structure of the covariance matrix for the revised latent state $Q_x = V_x - B\left(W - Q_z\right)B'$ implies that the persistence of $x_{1:n_x}$ conditioning on $z_{1:n_z}$ is also influenced by the ratio $\sigma_x^2/\sigma_z^2$. However, the bottom panels in Figure 2 shows that effect of the "signal-to-noise" ratio $\sigma_x^2/\sigma_z^2$ on the aggregated series $Ax_{1:n_x}$ is minimal. Indeed,

[9] See Section 3 for a description of our estimation and filtering algorithm.

when fixing $\lambda = 0.01$, the ACF of $Ax_{1:n_x}$ does not change by decreasing (left panel) or increasing (right panel) the ratio $\sigma_x^2/\sigma_z^2$.

As a whole, once aggregated over a 48-periods horizon, the persistence of the expected returns implied by the multi-scale time series model is anywhere higher than a 48-periods theoretical AR(1) dynamics, as far as $\lambda$ is small and regardless of the value of $\sigma_x^2/\sigma_z^2$. Only when the horizon-specific predictor does not convey information, i.e. $\lambda$ is high, the long-lasting effects on the dynamics of the process $x_{1:n_x}$ disappear, and the ACF of the aggregated process collapses to that implied by the classic AR(1) case.

#### 2.1.2 Forecasting

The forecasting for the latent expected returns takes a particular insightful expression. Appendix B shows that, conditional on the one-period ahead prediction for the scale-specific predictors $z_{n_z+1}$, the future expected returns $x_{n_x+1:n_x+m}$ are independent of the observations $x_{1:n_x-1}$ and $z_{n_z}$, i.e. $\mathrm{p}(x_{n_x+1:n_x+m} \mid z_{1:n_z+1}, x_{1:n_x}) = p(x_{n_x+1:n_x+m} \mid z_{n_z+1}, x_{n_x})$. When the multi-scale time series model is built using AR(1) blocks as in Eqs. (1) and (7), the predictive distribution for the higher-frequency latent state becomes $p\left(x_{n_x+1:n_x+m} \mid x_{n_x}, z_{n_z+1}, \Theta\right) \sim N(f_x, F_x)$, with the conditional mean equal to

$$f_x = r + m^{-1} R 1 \left(m^{-2} 1' R 1 + \lambda \left(A' V_x A\right)_{11} I\right)^{-1} \left(z_{n_z+1} - m^{-1} 1' r\right) . \qquad (12)$$

where $r = x_{n_x}\left(\phi_x, \ldots, \phi_x^m\right)$ and $R_{ij} = \sigma_x^2 \phi_x^{|i-j|} \frac{1-\phi_x^{2\min(i,j)}}{1-\phi_x^2}$ denote the $m$-step ahead predictive mean vector and covariance matrix for the non-revised AR(1) process, respectively. The $m$-steps ahead forecasts $x_{n_x+1:n_x+m}$ are generated conditional on $z_{n_z+1} \sim p\left(z_{n_z+1} \mid x_{1:n_x}, z_{1:n_z}, \Theta\right) = N(f_z, F_z)$ where

$$f_z = F_z \left[\sigma_z^{-2} \phi_z z_{n_z} + m^{-1} \left(m^{-2} 1' R 1 + \lambda \left(A' V_x A\right)_{11} I\right)^{-1} 1' r - P_{n_z}^{-1} p_{n_z}\right] , \qquad (13)$$

and $F_z = \left[\sigma_z^{-2} + \left(m^{-2} 1' R 1 + \lambda \left(A' V_x A\right)_{11} I\right)^{-1} - P_{n_z}^{-1}\right]^{-1}$. Notice that $p_{n_z}$ and $P_{n_z}$ are the predictive mean and covariance implied by the standard predictive distribution

$p(z_{n_z+1} \mid z_{1:n_z}, \Theta)$.

Equation (12) highlights the unique role assigned by the multi-scale time-series model to low-frequency information. More precisely, the forecast of future expected returns is the outcome of both the compounding effect of standard AR(1) prediction, i.e. $r$, and the revision due to the information in the scale-specific predictor, i.e. $(z_{n_z+1} - m^{-1}1'r)$. Besides controlling the persistence of the latent state, $\lambda$ plays also a key role in determining the contribution of information at multiple scales for predicting future expected returns. For $\lambda \to \infty$, the future path of expected returns collapse to the standard iterated AR(1) case, $f_x = r$. Analogously, when $\sigma_z^2$ and $P_{n_z}$ are much larger than $(m^{-2}1'R1 + \lambda (A'V_xA)_{11} I)$, the forecast for the coarse level is given by $f_z = m^{-1}1'r$. In all other scenarios, the multi-scale time series framework implies that forecasts at low frequency, i.e. $z_{n_z+1}$, are going to affect future expectations.

# 3 Estimation strategy

We propose a Bayesian approach to estimate the model. This allows to specify economically motivated prior distributions for the parameters of interests and also explicitly address uncertainty over the latent states and the model parameters. In the following we describe the estimation strategy for the case in which we have two observable predictors. A major aspect of the model is that conditional on the latent expected returns $x_{1:n_x}$, the dynamics of predictors corresponding to different frequencies are conditionally independent. This greatly simplifies the sampling scheme for the joint posterior distribution of the parameters.

## 3.1 Posterior Simulation

For the *ith* frequency-specific predictors, we specify an independent normal-inverse-gamma marginal prior: $\phi_{zi} \sim N(m_{\phi_{zi}}, M_{\phi_{zi}})$, $\sigma_{zi}^2 \sim IG(\nu_{\sigma_{zi}}/2, \nu_{\sigma_{zi}} s_{\sigma_{zi}}/2)$. Draws from the posteriors of the scale-specific predictors will be sampled independently across time-scales since they are conditionally orthogonal. As far as the latent state is concerned, we assume $\phi_x \sim N(m_{\phi_x}, M_{\phi_x})$, $\sigma_x^2 \sim IG(\nu_{\sigma_x}/2, \nu_{\sigma_x} s_{\sigma_x}/2)$. Finally, we specify an inverse-gamma for the marginal prior of $\lambda_i \sim IG\left(\nu_{\lambda i}/2, \nu_{\lambda i} s_{\lambda i}/2\right)$.

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 of the parameters and latent states we follow a data augmentation principle, which relies on the complete likelihood function.[10]

Let the observable predictors consist of two horizon-specific covariates $z_{1:n_z} = \left(z^1_{1:n_{z1}}, z^2_{1:n_{z2}}\right)$, each with $n_{zi}$ observations. For the ease of exposition, in the following we simply refer to $x$ and $z$ as the latent state and the set of predictors. The collection of parameters is defined as $\theta = (\sigma^2, \phi_x, \sigma^2_x, \lambda, \phi_z, \sigma^2_z)$ with $\lambda = (\lambda_1, \lambda_2)$, $\phi_z = (\phi_{z1}, \phi_{z2})$ and $\sigma^2_z = (\sigma^2_{z1}, \sigma^2_{z2})$. The full conditional distribution for the parameters of the expected returns and linking equation can be written as

$$\begin{aligned} p\left(\phi_x, \sigma^2_x, \lambda | z, x\right) &\propto p\left(x | z, \phi_x, \lambda, \sigma^2_x\right) p\left(\phi_x, \sigma^2_x, \lambda\right), \\ &= \frac{p\left(z | x, \phi_x, \sigma^2_x, \lambda\right) p\left(x | \phi_x, \sigma^2_x\right)}{p\left(z | \phi_x, \sigma^2_x, \lambda\right)} p\left(\phi_x, \sigma^2_x, \lambda\right), \end{aligned} \tag{14}$$

where the conditional distribution $p\left(z | x, \phi_x, \sigma^2_x, \lambda\right)$ is the product of the ones of the scale-specific predictors given their independence conditional on $x$. The conditional distribution for the parameters of the frequency-specific predictors can be defined as

$$p\left(\phi_z, \sigma^2_z | z\right) \propto q\left(z | \phi_z, \sigma^2_z\right) p\left(\phi_z, \sigma^2_z\right),$$

Posterior simulation is implemented by adapting the algorithm proposed by Ferreira et al. (2006), which we extend and generalize to the case in which there are multiple observable predictors and a latent expected returns at the higher frequency.

### Step 1. Drawing the expected returns

To draw the time series of expected returns we use a forward filtering, backward sampling (FFBS) estimation scheme. Although the marginal distribution of the scale-specific predictors is substituted by $q\left(z | \phi_z, \sigma^2_z\right)$, we can exploit the linkage equation

[10] We simulate 50,000 draws from the full conditional distributions, discard 20,000 initial draws as burn-in sample, and keep one out of ten draws in order to increase sampling efficiency.

(4) to extract expected returns on observable predictors. More precisely, we can write a state-space dynamic linear model in which frequency-specific predictors represent the observation equation and the initial dynamics of expected returns as the state equation;

$$
\begin{aligned}
z_s &= F x_s + \epsilon_{z,s}, & \epsilon_{z,s} &\sim N\left(0, \Sigma_{zz}\right), \\
x_s &= G x_{s-1} + \epsilon_{x,s}, & \epsilon_{x,s} &\sim N\left(0, \Sigma_{xx}\right),
\end{aligned} \tag{15}
$$

where $z_s = (z_{s1}, z_{s2})$, $x_s' = \left(x_{(s-1)m+1}, ..., x_{sm}\right)$, $\Sigma_{xx,ij} = \sigma_x^2 \phi_x^{|i-j|} \left(1 - \phi_x^{2\min(i,j)}\right) / (1 - \phi_x^2)$, $\Sigma_{zz} = diag\left(\tau_1, \tau_2\right)$ 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 $\iota_m$ an $m$-dimensional vector of ones. For identification purposes, scale-specific predictors and expected returns are uncorrelated to each other;

$$
\begin{bmatrix} \epsilon_{z1,s} \\ \epsilon_{z1,s} \\ \epsilon_{x,s} \end{bmatrix} \sim N\left(0, \Sigma\right), \qquad \text{with} \qquad \Sigma = \begin{bmatrix} \tau_1 & 0 & 0 \\ 0 & \tau_2 & 0 \\ 0 & 0 & \Sigma_{xx} \end{bmatrix} \tag{16}
$$

with $\tau_i = \lambda_i \left(F' V_x F\right)_{ii} I$ with $I$ as the $n_{zi}-$square identity matrix. Conditionally on the parameters, a standard version of the FFBS algorithm can then be implemented (see West and Harrison, 1997, Chapter 4, for additional details). Notice at each iteration the backward sampling generates a posterior distribution of the revised expected returns which is numerically equivalent to the closed-form expression shown in Eq. (A.19). In an unreported set of simulations we show that in fact a draw from the closed-form of $Q_x$ and posterior draws from FFBS are indistinguishable in large samples, with small deviations in small samples (results are available upon request).

### Step 2. Sampling the parameters for the expected returns

Posterior distributions for the parameters of the expected returns are not available in closed form, therefore we implement a Metropolis-Hastings (MH) step. For each draw $g = 1, ..., G$, the proposal for $\sigma_{zi}^2$ is generated from a uniform distribution $\sigma_{zi}^{2(g)} \sim U\left(\sigma_z^{2(g-1)}/\delta_{\sigma_{zi}}, \sigma_{zi}^{2(g-1)} \cdot \delta_{\sigma_{zi}}\right)$ where $\delta_{\sigma_{zi}} > 1$ is a tuning parameter that must be fixed to ensure a good trade-off between efficiency and accuracy in approximate draws from the joint posterior distribution. The acceptance probability is computed as

$$\pi\left(\sigma_x^{2(g)}, \sigma_x^{2(g-1)}\right) = \min\left(1, \frac{\sigma_x^{2(g-1)} IG\left(\sigma_z^{2(g)}|\nu_{\sigma_x}^*/2, \nu_{\sigma_x}^* s_{\sigma_x}^*/2\right) p\left(z|\phi_x^{(g-1)}, \lambda^{(g-1)}, \sigma_x^{2(g-1)}\right)}{\sigma_x^{2(g)} IG\left(\sigma_z^{2(g-1)}|\nu_{\sigma_x}^*/2, \nu_{\sigma_x}^* s_{\sigma_x}^*/2\right) p\left(z|\phi_x^{(g-1)}, \lambda^{(g-1)}, \sigma_x^{2(g)}\right)}\right),$$

with

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

with $n_z = n_{z1} + n_{z2}$ and $\tau = (\tau_1, \tau_2)$. The marginal probability $p\left(z|\phi_x, \lambda, \sigma_x^2\right)$ can be efficiently evaluated by using the Kalman filter recursion proposed in Step 1 $p\left(z|\phi_x, \lambda, \sigma_x^2\right) = \prod_{s=1}^{n_z} p\left(z_s|z_{s-1}, \phi_x, \lambda, \sigma_x^2\right)$.

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 use as 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 $\delta_{\phi_x}$ has to be tuned to ensure convergence and accuracy in the posterior approximation. The acceptance probability is computed as

$$\pi\left(\phi_x^{(g)},\phi_x^{(g-1)}\right)=\min\left(1,\frac{\phi_x^{(g-1)}p\left(z|x,\phi_x^{(g)},\lambda^{(g-1)},\sigma_x^{2(g)}\right)N\left(\phi_x^{2(g)}|m^*_{\phi_x},M^*_{\phi_x}\right)}{\phi_x^{(g)}p\left(z|x,\phi_x^{(g-1)},\lambda^{(g-1)},\sigma_x^{2(g)}\right)N\left(\phi_x^{2(g-1)}|m^*_{\phi_x},M^*_{\phi_x}\right)}\right.$$
$$\left.\times\frac{p\left(z|\phi_x^{(g-1)},\lambda^{(g-1)},\sigma_x^{2(g)}\right)}{p\left(z|\phi_x^{(g)},\lambda^{(g-1)},\sigma_x^{2(g)}\right)}\right),$$

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

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

**Step 4. Sampling the between-scales agreement**

The parameter $\lambda_i$ has the natural interpretation in terms of the relative lack of agreement between the *ith* scale-specific observable predictor and the latent discount rates. Let $F_i$ correspond to the block of $F$ specific to the *ith* predictors, the marginal distribution of the scale-specific predictor can be defined as $p\left(z^i_{1:n_{zi}}\right)=N\left(0,F_iV_xF_i'+\lambda_iF_iV_xF_i'\right)$. We exploit the relationship with the parameters of the expected returns since, conditional on $\phi_x$ and $\sigma_x^2$, $\lambda_i$ can be recovered from $\tau_i=\lambda_i\left(F_iV_xF_i'\right)_{11}$.

We generate a proposal $\lambda_i^{(g)}\sim U\left(\max\left(0,\lambda_i^{(g-1)}-\delta_\lambda\right),\min\left(10,\lambda_i^{(g-1)}+\delta_\lambda\right)\right)$ where $\delta_\lambda$ has to be tuned to ensure convergence and effectiveness in the posterior approximation. Such proposal implies a given $\tau_i^{(g)}=\lambda_i^{(g)}\left(F_iV_x^{(g)}F_i'\right)_{11}$. The acceptance probability is computed as

$$\pi\left(\tau_i^{(g)},\tau_i^{(g-1)}\right)=\min\left(1,\frac{\tau_i^{(g-1)}IG\left(\tau_i^{(g)}|\nu^*_{\tau_i}/2,\nu^*_{\tau_i}s^*_{\tau_i}/2\right)p\left(z^i_{1:n_{zi}}|\phi_x^{(g)},\tau_i^{(g-1)},\sigma_x^{2(g)}\right)}{\tau_i^{(g)}IG\left(\tau_i^{(g-1)}|\nu^*_{\tau_i}/2,\nu^*_{\tau_i}s^*_{\tau_i}/2\right)p\left(z^i_{1:n_{zi}}|\phi_x^{(m)},\tau_i^{(g)},\sigma_x^{2(g)}\right)}\right),$$

with

$$\nu^*_{\tau_i} = \nu_{\tau_i} + n_{zi} - 1,$$
$$\nu^*_{\tau_i} s^*_{\tau_i} = \nu_{\tau_i} s_{\tau_i} + \left(z^i_{1:n_{zi}} - F_i x_{1:n_x}\right)' \left(z^i_{1:n_{zi}} - F_i x_{1:n_x}\right),$$

Once, $\tau_i^{(g)}$ is sampled, we can back out $\lambda_i^{(g)}$ conditional on $V_x^{(g)}$. The independence across scales allows to simply compute $\tau_i^{(g)}$ separately for each scale-specific predictor.[11]

**Step 5. Sampling the parameters for scale-specific predictors**

For the observable scale-specific predictors the posterior distribution is available in closed form. For example, if $z_{1:n_z}$ follows an ARMA process then posterior draws 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 as priors are conjugate;

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

with

$$M^*_{\phi_{zi}} = \left(M^{-1}_{\phi_{zi}} + \sigma^{-2}_{zi} z^{i\prime}_{1:n_{zi}-1} z^i_{1:n_{zi}-1}\right), \qquad m^*_{\phi_{zi}} = M^*_{\phi_{zi}}\left(M^{-1}_{\phi_{zi}} m_{\phi_{zi}} + \sigma^{-2}_{zi} z^{i\prime}_{1:n_{zi}-1} z^i_{2:n_{zi}}\right),$$
$$\nu^*_{\phi_{zi}} = \nu_{\phi_{zi}} + n_{zi} - 1, \qquad \nu^*_{\phi_{zi}} s^*_{\phi_{zi}} = \nu_{\phi_{zi}} s_{\phi_{zi}} + \sum_{j=2}^{n_{zi}} \left(z_{ij} - \phi_{zi} z_{ij-1}\right)^2,$$

The functional form of the posterior does not change if we consider $K$ independent observable scale processes. Independence across scales allows to update them separately.

[11] The tuning parameters are set to $\delta_{\phi_x} = 0.01, \delta_{\sigma^2_x} = 1.5$ and $\delta_\lambda = 0.01$. Under this tuning the acceptance rates are 59%, 20%, 29% and 49% for $\phi_x, \sigma^2_x, \lambda_{dp}$ and $\lambda_{cay}$, respectively.

# 4 Empirical analysis

We impose uninformative priors for the $\phi_x$ parameter centered at 0.9 and flat on most of the $(-1, 1)$ range, with prior mass tailing off near $\pm 1$ to ensure stationarity. We also considered alternative priors calibrations 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 predictors and latent discount rates we fix the initial degrees of freedom equal to two-percent of the sample sizes.

As far as the between-scale uncertainty is concerned, we start with a benchmark prior calibration that assumes a moderate level of discrepancy between information across different scales and expected returns. We also consider three alternative settings. In the first case, we assume that observable predictors do not bring any information on extracting the latent path of expectations although we keep such prior belief vague. Similarly, in the second case we assume that predictors do not contain additional information to understand expected returns, but now we also increase the precision of the prior. In the third case, we assume instead that predictors almost perfectly reveal the underlying process of expected returns. In the analysis that follows, we provide posterior estimates for each of the alternative prior specification detailed above.

The latent expected returns are extracted at the monthly frequency to mimic a typical time horizon for forecasting and investment decision making. The sample period is 1952:01-2013:12. We focus on two well-known predictors as conditioning information, the log dividend-yield, i.e. $dp_t$, and the (Markov-Switching) consumption-wealth ratio, i.e. $cay_t$ proposed in Bianchi et al. (2016). Our choice 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 see Campbell and Mankiw (1989) and Lettau and Ludvigson (2001). Also, $cay_t$ and $dp_t$ are arguably used as benchmark in modeling returns predictability and found to capture returns predictability at different frequency of observations (see, e.g. Greenwood and Shleifer, 2014). Our framework would be applicable to a broader range of predictors provided there is a minimum amount of information needed for estimation purposes.

### 4.1 The choice of the coarsening parameter $m$

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 scale at which predictors operates in order to set $m$ accordingly in the aggregation scheme and, thus, in the link equation. We compare 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 and we resort to numerical methods and approximate a consistent estimate of the marginal likelihood via importance sampling (see, e.g. Gelfand and Dey (1994) and Newton and Raftery (1994)).

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 favour 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 substantially higher marginal likelihood. We assume throughout the rest of the empirical analysis that the mid-frequency fluctuations in 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. This is consistent with Cochrane (2008) who shows that, in an AR(1) latent variable model with different speeds of adjustment for the expected return and the expected dividend growth, a weighted average of past dividend yields can be used to tease out more information about expected returns than is present in the current dividend yield.

## 4.2 Persistence of expected returns

Panel A of Table 2 reports posterior summaries for the persistence $\phi_z$ and the conditional variance $\sigma_z^2$ for both frequency-specific predictors, $z = \{cay, dp\}$. These estimates are obtained under the benchmark uninformative priors. Posterior estimates of these parameters under alternative prior specifications are available in a separate online appendix.

[Insert Table 2 about here]

The magnitude of the posterior mean of $\phi_z$ and $\sigma_z^2$, for $z_t = \{cay_t, dp_t\}$ shows that our predictors exhibit different degrees of persistence and volatility. The fact that $\hat{\phi}_{dp} \gg \hat{\phi}_{cay}^4$ highlights that each predictor conveys scale-specific information: temporally aggregating the yearly information in $cay_t$ to a four-year horizon would miss most of the persistent fluctuations that $dp_t$ brings about. 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, although a clear cut comparison is non-trivial being processes sampled at different frequencies.

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 $\lambda_{dp}, \lambda_{cay}$. Under our benchmark uninformative prior, the posterior mean for the between-level uncertainty parameters are $\hat{\lambda}_{dp} = 0.013$ and $\hat{\lambda}_{cay} = 0.427$. These low values of $\hat{\lambda}_{dp}$ and $\hat{\lambda}_{cay}$ imply that the initial AR(1) process for expected returns is substantially revised after observing the mid-to-low frequency information conveyed by $cay_t$ and $dp_t$.

Figures 3 displays the ACF of the monthly expected returns (Panel A), as well as the ones 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 an ARMA(2,1) process fitted on the filtered expected return series. The ARMA(2,1) represents a sensible benchmark since it can be thought of as the process resulting from summing two (independent) AR(1) with different persistence roots; thus, an ARMA(2,1) has the potential to capture fluctuations at multiple frequencies. Also, the ARMA(2,1) has the same number of parameters as our model.

[Insert Figures 3 about here]

Although the ARMA(2,1) process for the monthly series seems to be appropriate to describe the first 8 lags, Panel A shows that the series of expected returns extracted from the frequency-specific consumption-wealth and log dividend-price processes exhibits strong dependence at long lags. The long-memory-type of persistence of expected returns remains unmatched as we go farther in investigating the impact of past shocks. Panels B and C highlight this fact by aggregating our extracted expected returns and comparing its ACF to that implied by the aggregation of an ARMA(2,1) process.

The Figure suggests that our dynamics for short-term expected returns manifests long-range dependence that reconcile both the very high-persistence of long-term expected returns as proxied by financial ratios, and the modest persistence of medium term discount rate fluctuations proxied by the consumption-wealth ratio. Indeed, the autocorrelation of the estimated expected returns over a one-year horizon is about 0.6 a value close to that obtained from fitting an AR(1) process to the consumption-wealth ratio series (see Panel A in Table 2). Analogously, the first order autocorrelation of the latent discount rates over an horizon of 4-year is about 0.8, a value that matches the results from fitting an AR(1) process to the dividend-price ratio series (see again Panel A in Table 2). Thus, differently from a low-order ARMA$(p, q)$ process that delivers lower persistence as the aggregation horizon increases (see Figure 1), our multi-scale framework is able to detect long lags in the dynamics; in turn, these long lags play a key role in determining the persistence of the process at long aggregation horizons.

Figure 4 reports the ACFs across alternative prior specifications for the parameters. The top left panel shows that the autocorrelation of the extracted expected returns is not sensitive to the prior specification of the parameter $\phi_x$. In particular the dependence at long lags is present independently of the prior specification. This holds upon aggregation over a 4-year horizon (top right panel). The bottom left panel shows the ACFs for alternative priors on the conditional variance of expected returns $\sigma_x^2$. The pattern is consistent for different prior specifications except for the case with strong beliefs on $\sigma_x^2 \approx 0$ (solid line with squares). 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 supportive of long-term fluctuations in discount rates as shown in the bottom right panel for the 4-year aggregated expected returns.

## 4.3 Implications for forecasting expected returns

We test the ability of our multi-scale time-series model to capture future fluctuations in the latent expected returns. More specifically, we compare the $m$-step ahead predictive performance of our multi-scale model against a persistent AR(1) and a set of alternative model specifications which arguably capture persistence heterogeneity. The conditional predictive for the one-block ahead forecast is obtained as outlined in Eq. (13). Actual forecasts for future expected returns are then computed for each block by integrating out parameter uncertainty. The marginal predictive distribution is simulated by using an importance sampling estimator (see, e.g. Geweke, 2005), in which draws from the posterior distribution of the parameters are obtained using the entire sample of observable predictors and latent expected returns.[12] Multi-period forecasts can be obtained by iterating $l$ times the one-block ahead and the one-step ahead predictive distributions of both the predictor and the expected returns, respectively.

As a preliminary in-sample comparison with existing research we compare graph-

[12]The predictive density of future expected returns can be alternatively approximated with an harmonic mean estimator; see Gelfand and Dey (1994), or the truncated elliptical mean estimator in Sims et al. (2008).

ically the $m$-step ahead forecast obtained using our multiscale time series process against the one obtained from a standard AR(1), with the persistence parameter calibrated to a high value of 0.98, as commonly used in the literature (see, e.g., Table II in Barberis (2000)).[13] Figure 5 shows the model-implied future expected returns for the horizon $h = 48$ months, for the years 1956-1959, 1980-1983, 1988-1991, and 2008-2011 which represents the period across the great financial crisis.

[Insert Figures 5 about here]

More specifically, at $t = 1955, 1979, 1987$ and $t = 2007$ we compute the future expected returns forecasts for the period $h = 1, ..., 48$, which coincides with the coarsening window in Eq. (4). 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 seems to 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 5 suggests that by incorporating information at multiple horizons has implications for understanding future expected returns. We now formally investigate the out-of-sample performance of the multi-scale by comparing the forecasting against three different alternative specifications: (1) an AR(1) process (see, e.g., Barberis, 2000); (2) an heterogeneous persistence process where the heterogeneity is generated by summing two independent (at all leads and lags) AR(1) processes with different roots; (3) an AR(1) model with time-varying mean (the mean is further modeled as a persistent Markov process).

Given the latent nature of expect returns, a full-blown comparison based on observables is not easily obtained. In this respect, we compare each of the three speci-

[13] We also compute the forecasts based on OLS estimates of an AR(1) fitted on filtered expected returns. Conclusions are almost identical to the case where we fix the persistence to 0.98.

fication fitted on the latent expected returns series extracted using the entire sample period. We rely on the relative predictive Root Mean Squared Error (RMSE) which allows to investigate the performance of alternative methodologies compared to our multi-scale model in terms of point forecasts. The benchmark model is our multi-scale time series model. For each alternative specification, we compute the relative value as the ratio of the RMSE implied by the alternative specification over the benchmark, so that a value greater than one indicates that our multi-scale model improves upon the alternative specification. Gneiting (2011) showed that RMSE 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 posterior mean of the predictive distribution derived above. Panel A of Table 3 shows the relative RMSE for fixed forecasting horizons $h = 12, 24, 36, 48$, both in-sample and out-of-sample. For the in-sample analysis we use the whole sample to estimate the parameters of the alternative specifications, while in the out-of-sample setting we cut the last four-year of observations, and compare the corresponding future projections with the ex-post estimates one would have obtained by using the whole sample.

[Insert Table 3 about here]

The relative RMSE shows that the point forecast error generated by the multi-scale model is lower across longer horizons $h = 24, 36, 48$ and comparable at $h = 12$ to the AR(1) or the sum of two AR(1). When compared with the persistent AR(1) model, the use of information at different time scales reduces the root mean squared prediction error for $h = 48$ by around 7%. Similarly, the 48-step ahead forecasts obtained from the sum of two independent AR(1) processes generate a prediction error which is almost 6% higher than our multi-scale time series model. As far as the model with a time-varying mean is concerned, the multi-scale model delivers a substantially lower average square loss across all horizons. Interestingly, while comparable to other benchmarks (in particular the AR(1) and the sum of two AR(1)s) at short horizons, the out-performance of our model increases with $h$. This result is driven by the fact that forecasts at the coarse scale for the predictors guide the forecasts at the fine scale for expected returns, as shown in Eq. (12). More specifically, the term $(z_{n_z+1} - m^{-1}1'r)$ acts as an error correction term induced by low frequency information on the high-frequency dynamics; this error correction term reduces the prediction error especially at the intermediate-to-long prediction horizons. Panel B of Table 3 shows the out-of-sample results. The multi-scale model improves upon alternative specifications especially at longer horizon although somewhat under-perform in the short-term in a mean squared sense.

For the sake of completeness we also investigate the performance of our model throughout the term-structure of expected returns, namely the recursive in-sample and out-of-sample RMSE for $h = 1 : 48$.

[Insert Figure 6 about here]

The top panel of Figure 6 shows the in-sample estimates. The top panel makes clear that the multi-scale model compares favourably in the short-term and sensibly improves upon alternative specifications as the predictive horizon increases. The in-sample exercise is really pseudo-out-of-sample, meaning that although parameters are estimated using the whole sample period, conditional forecasts are generated recursively at each time $s = 1, \ldots, n_z$. The RMSE is then averaged out across the $n_z - 1$ blocks of $m$ predicted expected returns.

The bottom panel shows the out-of-sample estimates. This panel shows that the ability of the multi-scale model to capture future expected returns is retained in the long-run, although it sensibly decreases in the shorter term. The short-term under-performance in the truly out-of-sample exercise is possibly due to our specific choice of leaving out the period that coincides with the aftermath of the great financial crisis. This choice has been made on purpose to make the comparison as effective as possible against our methodology. Also, the under-performance at short horizons is somehow expected as we do not include any predictor with a short-term predictive power as conditioning variables. In fact, in a separate calculation we show that by computing the recursive relative RMSE starting for the one-year ahead, i.e. $h = 12 : 48$, our benchmark model compares favourably across all horizons.

# 5 Concluding remarks

We define expected returns as the outcome of an updating process based on medium- and long-term information. Empirically, we show that accounting for the low- to mid-frequency information conveyed by standard predictors, allows to identify long-lasting effects in the dynamics of high-frequency expected returns. Importantly we show that low-order ARMA models have hard time in capturing these effects. We also show that combining information at multiple horizons to extract the series of short-term expected returns has relevant implications for forecasting and, thus, it may have first order importance for investment decision purposes. We have quantified these effects along several dimensions.

From a methodological point of view, our multi-scale time series model provides a parsimonious framework to combine information across horizons to extract the dynamics of short-term expected returns. Within this framework, we also show how to form expectations over multiple horizons. Thus, our econometric modeling framework and our empirical analysis speak to a large financial economics literature that tries to understand the role of expectations formation in financial markets.

We conclude by emphasizing that, although in this paper we have focused on the dynamics of expected stock returns, our multi-scale time series framework is relevant in any setting where it is important to incorporate information available at different (typically longer) horizons to make inference on the process of interest evolving over short-horizon. For instance, our forecasting dynamics might be useful to government and budget office that need to forecast the future path of GDP over the next quarters given long-term productivity forecasts and/or fiscal budget constraints, and to central banks that need to incorporate information about the medium- and long-term dynamics of output and inflation to understand the evolution of interest rates over short-term. We view these as promising avenues for future research.

# References


Abadir, K. M., G. Caggiano, and G. Talmain (2013). Nelson–Plosser revisited: The ACF approach. *Journal of Econometrics 175*(1), 22–34.

Bandi, F., P. Benoit, A. Tamoni, and C. Tebaldi (2017). 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.

Bianchi, F., M. Lettau, and S. C. Ludvigson (2016, August). Monetary Policy and Asset Valuation: Evidence From a Markov-Switching cay. NBER Working Papers 22572, National Bureau of Economic Research, Inc.

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

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

Carvalho, C. M., H. F. Lopes, and R. E. McCulloch (2016). On the long run volatility of stocks.

Chib, S. and E. Greenberg (1994). Bayes inference in regression models with ARMA(p,q) errors. *Journal of Econometrics 64*(1-2), 183–206.

Cochrane, J. H. (2008). State-Space vs. VAR models for Stock Returns. Manuscript, Chicago Booth.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*.

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

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

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

Fuster, A., B. Hebert, and D. Laibson (2010). Natural Expectations and Macroeconomic Fluctuations. *Journal of Economic Perspectives 24*(4), 67–84.

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

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

Ghysels, E., P. Santa-Clara, and R. Valkanov (2004). The {MIDAS} touch: Mixed data sampling regression models. *Working Paper UNC and UCLA*.

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

Greenwood, R. and A. Shleifer (2014). Expectations of returns and expected returns. *Review of Financial Studies 27*(3), 714–746.

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.

Lettau, M. and S. Ludvigson (2001, 06). Consumption, aggregate wealth, and expected stock returns. *Journal of Finance 56*(3), 815–849.

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

Müller, U. K. and J. H. Stock (2011, January). Forecasts in a Slightly Misspecified Finite Order VAR. NBER Working Papers 16714, National Bureau of Economic Research, Inc.

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

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

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

Schorfheide, F. and D. Song (2015). Real-time forecasting with a mixed-frequency VAR. *Journal of Business & Economic Statistics 33*(3), 366–380.

Shiller, R. J. (1981). Do stock prices move too much to be justified by subsequent changes in dividends? *The American Economic Review 71*(3), 421–436.

Sims, C. A., D. F. Waggoner, and T. Zha (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.

Tiao, G. C. (1972). Asymptotic behaviour of temporal aggregates of time series. *Biometrika 59*(3), 525–531.

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

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

# Appendix

## A Proof of Proposition 1

*Proof:* (i) The assumption that in the revision of beliefs the low-frequency predictor is sufficient as conditioning information implies 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 (4), 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 $z_{1:n_z}$ (see Jeffrey, 1957, Jeffrey, 1965 and Diaconis and Zabell, 1982). The conditional distribution $p\left(x_{1:n_x}|z_{1:n_z}\right) \propto p\left(x_{1:n_x}\right) p\left(z_{1:n_z}|x_{1:n_x}\right)$ is given by a standard linear projection

$$x_{1:n_x}|z_{1:n_z} \sim N\left(Bz_{1:n_z}, V_x - BWB'\right) \tag{A.17}$$

where $B = V_x A' W^{-1}$ and $W = AV_xA' + \lambda\left(A'V_xA\right)_{11} I$. The joint multi-scale model for the low- and high-frequency series is

$$\begin{aligned} q\left(x_{1:n_x}, z_{1:n_z}\right) &= q\left(x_{1:n_x}|z_{1:n_z}\right) q\left(z_{1:n_z}\right) \\ &= N\left(Bz_{1:n_z}, V_x - BWB'\right) N\left(0, Q_z\right), \end{aligned} \tag{A.18}$$

By integrating (A.18) with respect to the low-frequency predictor we obtain the marginal distribution for the latent expected returns at high frequency, namely

$$q\left(x_{1:n_x}\right) = \int q\left(x_{1:n_x}, z_{1:n_z}\right) dz_{1:n_z} = N\left(0, Q_x\right), \tag{A.19}$$

where $Q_x = V_x - B\left(W - Q_z\right) B'$. This concludes the first part of the proof[14]
(ii) To prove Eq. (10) in main text use the fact that as $\lambda \to 0$, $W = AV_xA' + \lambda\left(A'V_xA\right)_{11} I = AV_xA'$ and, thus, $B = V_xA'\left(AV_xA'\right)^{-1}$. Plugging these expressions into $Q_x$, see Eq. (9), yields the result. Analogously, to prove Eq. (11) it is enough to use the fact that $\lim_{\lambda\to\infty} W^{-1} = 0$.[15] Also, observe that these limiting results 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 (4) and stationary and invertible ARMA processes as building blocks.

## B Forecasting within the multiscale framework

We want to show that conditional on $x_{1:n_x}$ at the fine level and on the future observation $z_{n_z+1}$ at the coarse level, the future observations at the fine level $x_{n_x+1:n_x+m}$ are independent of the observations at the fine level up to time $n_x - 1$ and of the observations at the coarse level up to time $n_z$.

First, we use the fact that $q(x_{1:n_x} \mid z_{1:n_z}, \phi_x, \sigma_x^2, \lambda) = p(x_{1:n_x} \mid z_{1:n_z}, \phi_x, \sigma_x^2, \lambda)$ , the conditional

[14]See Ferreira and Lee, 2007 Ch. 10 for a related example on Markov random fields.
[15]See also Theorem 11.1 and Theorem 11.2 in Ferreira and Lee, 2007.

distribution of the fine level given the coarse level is not revised by Jeffrey's rule:[16]

$$
\begin{aligned}
q(x_{1:n_x} \mid z_{1:n_z}) &= p(x_{1:n_x} \mid z_{1:n_z}) \\
&\propto p(z_{1:n_z} \mid x_{1:n_x}) p(x_{1:n_x}) \\
&\propto \underbrace{p(x_1)\left[\prod_{i=2}^{n_x} p(x_i \mid x_{i-1})\right]}_{=p(x_{1:n_x})} \times \underbrace{\left[\prod_{j=1}^{n_z} p(z_j \mid x_{mj-m+1:mj})\right]}_{=p(z_{1:n_z}|x_{1:n_x})} \qquad (\star)
\end{aligned}
$$

where we use the Markovian structure of $x$ to write $p(x_{1:n_x})$ as a product of $p(x_i \mid x_{i-1})$.[17] Therefore

$$
\begin{aligned}
p(x_{n_x-m+1:n_x} \mid z_{1:n_z}, x_{1:n_x-m}) &= \frac{p(x_{n_x-m+1:n_x}, x_{1:n_x-m} \mid z_{1:n_z})}{p(x_{1:n_x-m} \mid z_{1:n_z})} \\
&\propto p(x_{1:n_x} \mid z_{1:n_z})
\end{aligned}
$$

Now using Eq.($\star$), and disregarding all the terms $x_{1:n_x-m}$ (i.e. considering only those $x_i$s for $i \geq (n_x - m + 1)$) we finally have that:

$$
p(x_{n_x-m+1:n_x} \mid z_{1:n_z}, x_{1:n_x-m}) \propto \left[\prod_{i=n_x-m+1}^{n_x} p(x_i \mid x_{i-1})\right] \times [p(z_{n_z} \mid x_{n_x-m+1:n_x})]
$$

i.e.

$$
p(x_{n_x-m+1:n_x} \mid z_{1:n_z}, x_{1:n_x-m}) = p(x_{n_x-m+1:n_x} \mid z_{n_z}, x_{n_x-m})
$$

By mathematical induction one has that

$$
p(x_{n_x+1:n_x+m} \mid z_{1:n_z+1}, x_{1:n_x}) = p(x_{n_x+1:n_x+m} \mid z_{n_z+1}, x_{n_x})
$$

Specifying this expression to the case when the multiscale model is constructed with AR(1) building blocks, one obtains Eq.(10) in the main text.

[16] The steps here mimic the proof of Theorems 11.7 and 11.8 Ferreira and Lee, 2007.

[17] If the building block is an AR(2) instead of an AR(1) we are going to re-define $\tilde{x}_{n_x} = [x_{n_x},\ x_{n_x-1}]$ to accommodate a first order Markov process. Analogously for higher order ARMA(p,q) processes.

Figure 1: Aggregation dynamics for expected returns

This figure shows the persistence properties of the long-term returns resulting from the $m$-period aggregation of short-term discount rates. The blue line with circles shows the first-order autocorrelation as implied by the marginal covariance matrix $V_z = A'V_xA$ for different aggregation windows $m = 1, \ldots, 48$. Short-term discount rates are assumed to follow a persistent AR(1) process. The parameter $\phi_x$ of the expected returns is set to 0.977 as per Table II in Barberis (2000).

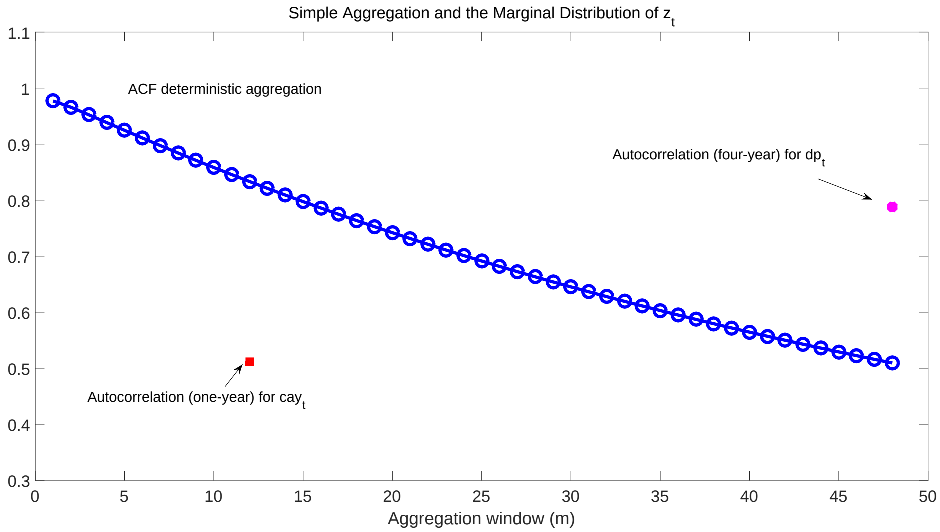


Figure 2: Persistence of Multi-scale vs AR(1): Four-year aggregation

Multi-scale time series Vs AR(1): autocorrelation functions of four-year aggregated series. This figure shows the theoretical autocorrelation function for a simulated multi-scale time series (red line with circles) aggregated over four years, the ACF for the series extracted by conditioning on the lower-frequency series $z$ and aggregated over four-years (blue line with diamonds), and the theoretical four-year ACF 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 $n_x = 720$ (say, monthly) observations, and we let $m = 48$ in all cases.

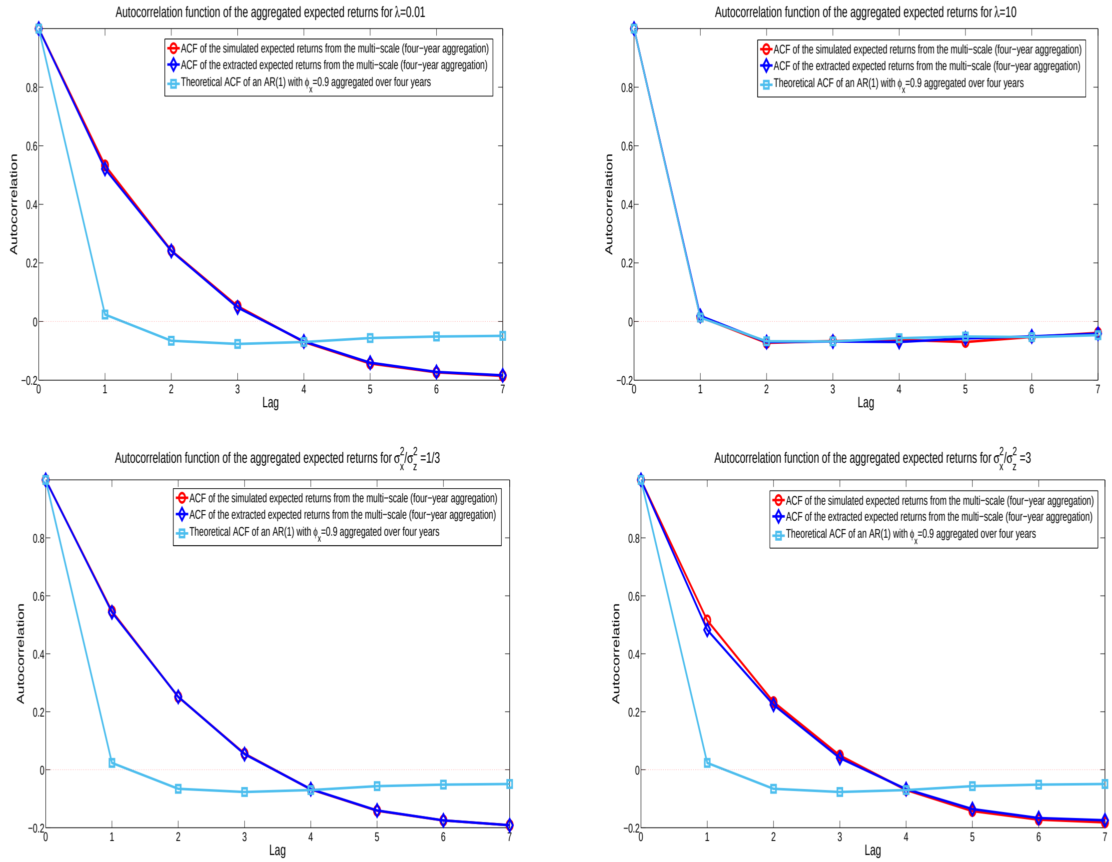


Figure 3: Persistence from a Multi-Scale Vs. ARMA(2,1)

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 (fitted to the filtered return series) and suitably aggregated in Panel B and C.

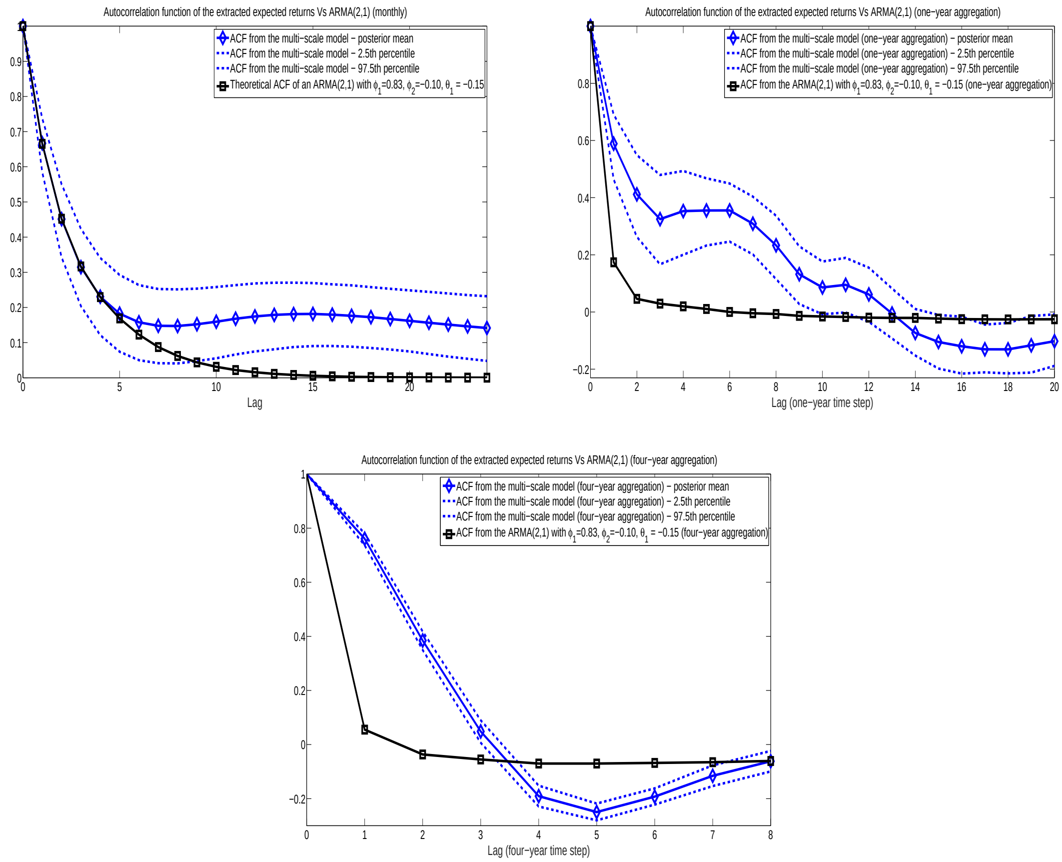


Figure 4: Persistence of Latent State: Prior Sensitivity

This figure reports the autocorrelation functions for alternative prior specifications of the autoregressive parameter $\phi_x$ (top panel) and the conditional variance $\sigma_x^2$ (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. The sample period for the expected returns is 1952:01-2013:12.

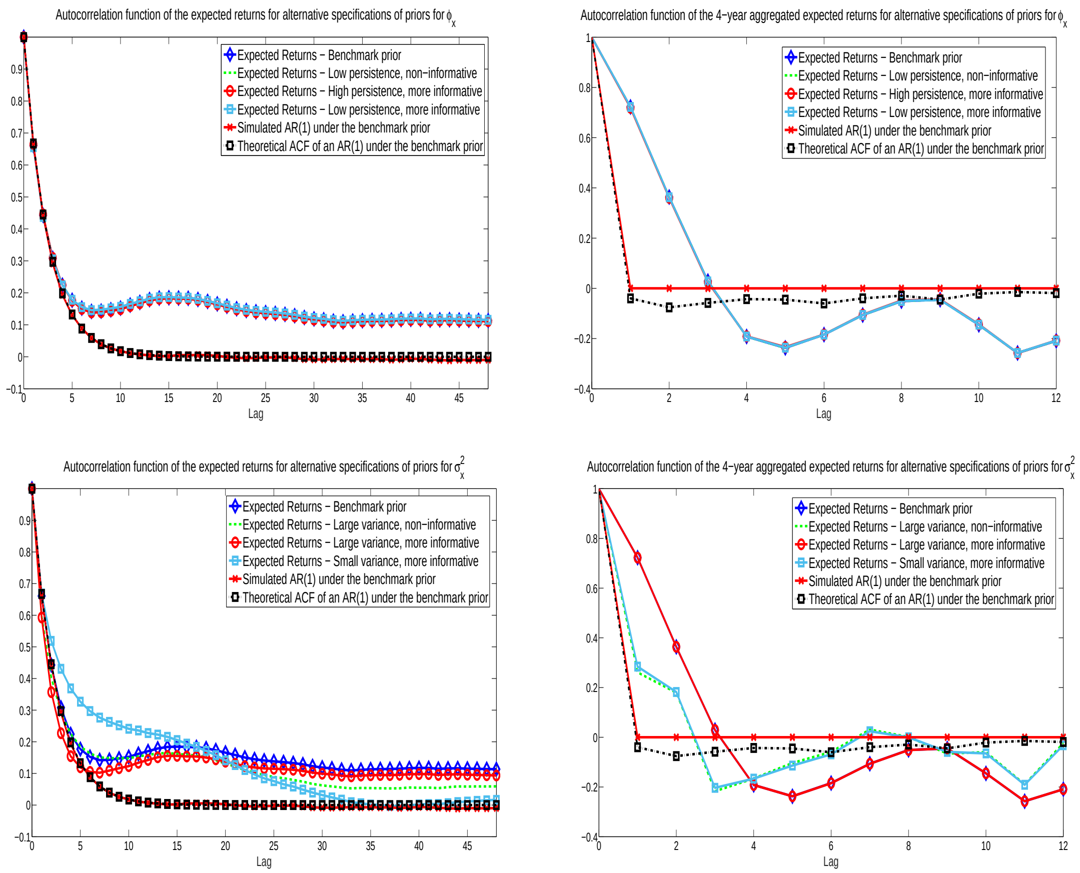


Figure 5: Multi-scale vs AR(1)-iterated forecasts: Empirical Evidence

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 obtained fixing the autoregressive parameter to $\phi_x = 0.98$ (red line with circles), and our multi-scale-based forecast (magenta line with circles). The sample period for the expected returns is 1952:01-2013:12.

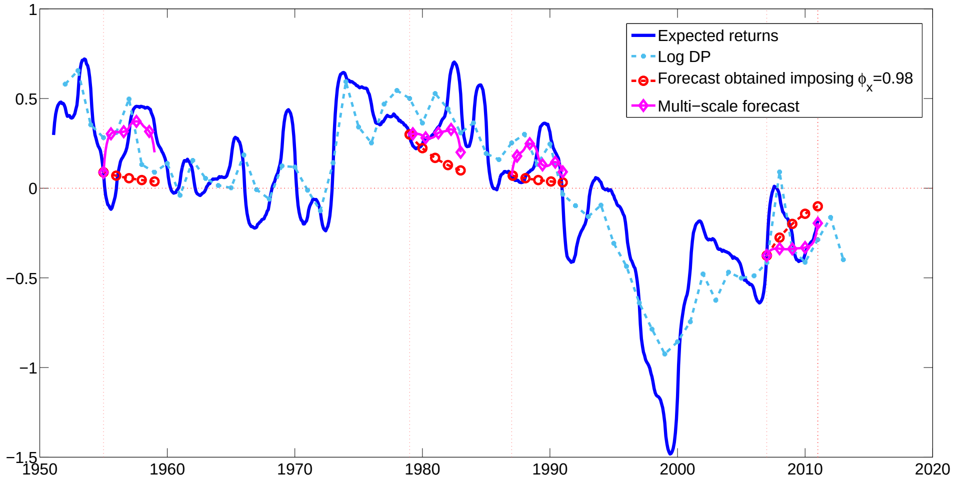
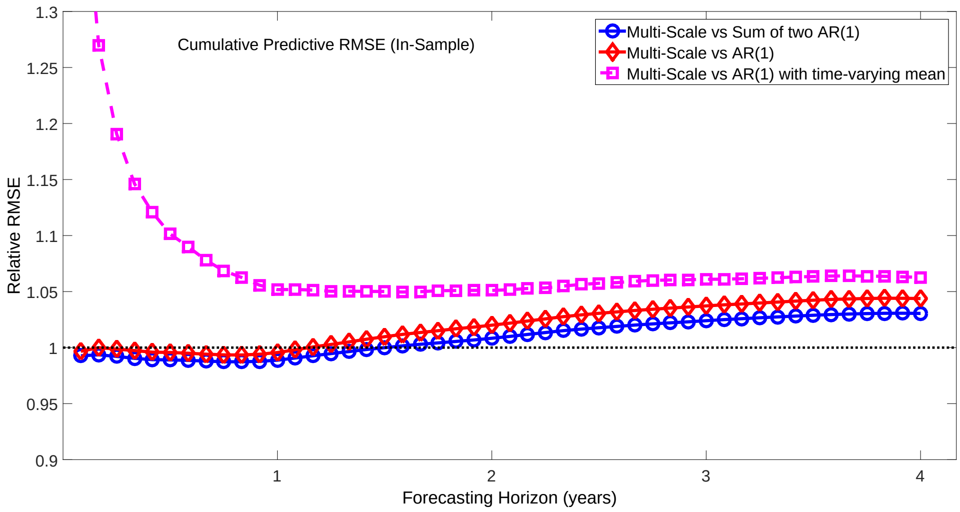


Figure 6: Forecasting Future Expected Returns: Recursive Relative RMSE

This figure shows the recursive relative Root Mean Squared Error (RMSE) for alternative model specifications. In all cases, the benchmark is our multi-scale time-series model so that values higher than one indicates that the multi-scale model improves upon the alternative specification. The (magenta line with squares) compares the AR(1) with stationary time-varying mean and the multi-scale model. The (red line with diamonds) compares a standard AR(1) and our benchmark model. Finally, the (blue line with circles) makes a comparison between the sum of two AR(1) with different persistence and our multi-scale model. The recursive relative RMSE is computed by using the marginal predictive distribution of future expected returns. The sample period is 1952:01-2013:12. **Panel A**: shows the in-sample results computed by conditioning on the posterior estimates obtained using the entire sample information. **Panel B**: show the out-of-sample results in which parameters are estimated by cutting the last four years of scale specific predictors, i.e. last data point for $dp_t$ and last four data points for $cay_t$. The resulting 48-month ahead forecasts of expected returns are compared to the ex-post estimates obtained by using the full sample.



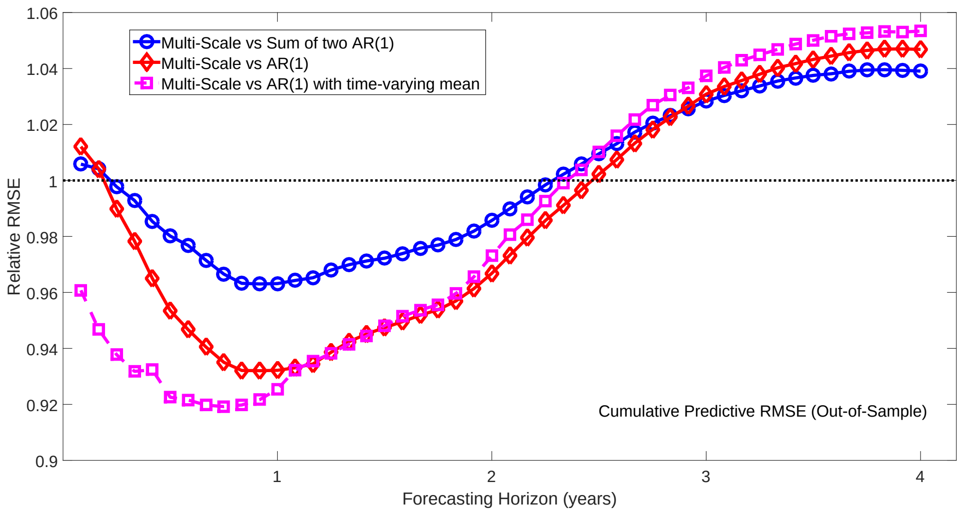



Table 1: Inference on the Coarsening Windows

This table reports the (log of) marginal likelihood computed from the multi-scale time series model with different specifications of the size of the coarsening window $m$. 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. **Panel A:** (log) Marginal likelihood for the multi-scale model with the log dividend-price ratio as predictor. **Panel B:** (log) Marginal likelihood for the multi-scale model with the consumption-wealth variable $CAY$ introduced by Lettau and Ludvigson (2001).

**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 $x_{1:n_x}$. The sample period is 1952:01-2013:12.

**Panel A:** Observable Predictors

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

**Panel B:** Latent Expected Returns

| Predictor | Parameter | Mean | Median | $2.5_{th}$ | $97.5_{th}$ |
|---|---|---|---|---|---|
| | $\phi_x$ | 0.667 | 0.667 | 0.592 | 0.746 |
| | $\sigma_x^2$ | 2.401 | 2.362 | 1.683 | 3.371 |
| $z_t = dp_t$ | $\lambda_{dp}$ | 0.013 | 0.012 | 0.007 | 0.019 |
| $z_t = cay_t$ | $\lambda_{cay}$ | 0.427 | 0.421 | 0.301 | 0.573 |

Table 3: Forecasting Future Expected Returns: RMSE $h$-Step Ahead

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, with the forecasts obtained from a sum of two independent (at all leads and lags) AR(1) processes, and with the forecasts obtained from an AR(1) model for which the mean is stationary and is allowed to vary stochastically over time. 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). Direct multi-step forecasts are produced monthly for an horizon of $h = 48$ months. For ease of exposition the table reports the results only for the $h = 12, 24, 36, 48$ months ahead (non-recursive performance). **Panel A:** In-Sample Relative Root Mean Squared Errors obtained from the marginal predictive mean. We report the ratios between the alternative specifications and the multi-scale model. **Panel B:** Out-of-Sample Relative Root Mean Squared Errors obtained from the marginal predictive mean. The sample period is 1952:01-2013:12.

**Panel A:** In-Sample Relative Root Mean Squared Error (RMSE)

| | Forecasting Horizon (months) | | | |
|---|---|---|---|---|
| Model | $h = 12$ | $h = 24$ | $h = 36$ | $h = 48$ |
| AR(1) | 0.998 | 1.045 | 1.060 | 1.068 |
| Sum of two AR(1) | 0.993 | 1.024 | 1.046 | 1.056 |
| AR(1) with time-varying mean | 1.223 | 1.057 | 1.053 | 1.051 |

**Panel B:** Out-of-Sample Relative Root Mean Squared Error (RMSE)

| | Forecasting Horizon (months) | | | |
|---|---|---|---|---|
| Model | $h = 12$ | $h = 24$ | $h = 36$ | $h = 48$ |
| AR(1) | 1.012 | 0.944 | 1.132 | 1.126 |
| Sum of two AR(1) | 1.006 | 0.978 | 1.096 | 1.101 |
| AR(1) with time-varying mean | 0.961 | 1.011 | 1.169 | 1.141 |