Sequential Learning for Bond Risk Premia

Kostas Kalogeropoulos

joint work with
Tomasz Dubiel-Teleszynski
Nikolaos Karouzakis

LSE, June 14, 2022

OUTLINE

1. Dynamic Term Structure Models
2. Forecasting Excess Bond Returns
3. Sequential Learning Scheme
4. Application on US Treasury Bill rates
5. Discussion - Extensions

Outline

Modelling and Forecasting Interest Rates

- Understanding and forecasting the term structure of interest rates is very important in financial markets.
- Central banks - monetary policy, stress testing.
- Insurance corporations, pension funds, university endowments, and other market participants - asset allocation, investment decisions.
- Similarities in terms of modelling with other financial data, such volatility and options, but with tractable expressions.

Resolving the ‘puzzle’

- The standard modelling framework is provided by Dynamic Term Structure Models (DTSMs), e.g. Dai and Singleton (2000); Ang and Piazzesi (2003).

Resolving the 'puzzle'

- The standard modelling framework is provided by Dynamic Term Structure Models (DTSMs), e.g. Dai and Singleton (2000); Ang and Piazzesi (2003).

- It is possible to estimate DTSMs in a time-series context but also from cross-sectional data (term structure).

Resolving the 'puzzle'

- The standard modelling framework is provided by Dynamic Term Structure Models (DTSMs), e.g. Dai and Singleton (2000); Ang and Piazzesi (2003).

- It is possible to estimate DTSMs in a time-series context but also from cross-sectional data (term structure).

- Using only the former results into overly stable long-term predictions, potentially to due to persistence underestimation, known as the 'puzzle' (Bauer, 2018).

Resolving the 'puzzle'

- The standard modelling framework is provided by Dynamic Term Structure Models (DTSMs), e.g. Dai and Singleton (2000); Ang and Piazzesi (2003).
- It is possible to estimate DTSMs in a time-series context but also from cross-sectional data (term structure).
- Using only the former results into overly stable long-term predictions, potentially to due to persistence underestimation, known as the 'puzzle' (Bauer, 2018).
- Several attempts in the literature to link both data streams based on absence of arbitrage.

Resolving the 'puzzle' (cont'd)

- Nevertheless, the 'puzzling' behaviour remains, unless restrictions are placed on risk premia, reflecting the change between the physical and pricing measures.

Resolving the 'puzzle' (cont'd)

- Nevertheless, the 'puzzling' behaviour remains, unless restrictions are placed on risk premia, reflecting the change between the physical and pricing measures.

- Restrictions allow more impact from the cross sectional data to the time series dynamics. This results into higher persistence and more realistic long term variability, resolving the 'puzzle' to some extent.

Resolving the 'puzzle' (cont'd)

- Nevertheless, the 'puzzling' behaviour remains, unless restrictions are placed on risk premia, reflecting the change between the physical and pricing measures.

- Restrictions allow more impact from the cross sectional data to the time series dynamics. This results into higher persistence and more realistic long term variability, resolving the 'puzzle' to some extent.

- From a machine learning viewpoint, this can be cast as a sparsity problem, i.e. looking to set some of the risk-premia parameters to zero.

Our approach

- But even in standard models there may be as many as 12 risk premia parameters corresponding to 2^{12} restriction sets.

Our approach

- But even in standard models there may be as many as 12 risk premia parameters corresponding to 2^{12} restriction sets.
- To navigate thought this space, we adopt a Bayesian approach resembling variable selection with spike and slab priors as in Bauer (2018).

Our approach

- But even in standard models there may be as many as 12 risk premia parameters corresponding to 2^{12} restriction sets.

- To navigate thought this space, we adopt a Bayesian approach resembling variable selection with spike and slab priors as in Bauer (2018).

- In order to address relevant empirical questions on predictability, we embed this framework in a sequential setting, allowing to update estimates and model choices/averaging as new data become available.

Dynamic Term Structure Model

The interest rate is an affine function of N state variables X_t,

$$r_t = \delta_0 + \delta_1' X_t,$$

where X_t under the physical measure $\mathbb{P}$ is defined as

$$X_t - X_{t-1} = \mu^{\mathbb{P}} + \Phi^{\mathbb{P}} X_{t-1} + \Sigma \varepsilon_t^{\mathbb{P}}, \quad \varepsilon_t^{\mathbb{P}} \sim N(0, I_N)$$

Dynamic Term Structure Model

The interest rate is an affine function of N state variables X_t,

$$r_t = \delta_0 + \delta_1' X_t,$$

where X_t under the physical measure $\mathbb{P}$ is defined as

$$X_t - X_{t-1} = \mu^{\mathbb{P}} + \Phi^{\mathbb{P}} X_{t-1} + \Sigma \varepsilon_t^{\mathbb{P}}, \quad \varepsilon_t^{\mathbb{P}} \sim N(0, I_N)$$

Under the essentially-affine specification (Duffee, 2002) of the market prices of risk, $\lambda_t = \Sigma^{-1} (\lambda_0 + \lambda_1 X_t)$, the pricing measure $\mathbb{Q}$ is

$$X_t - X_{t-1} = \mu^{\mathbb{Q}} + \Phi^{\mathbb{Q}} X_{t-1} + \Sigma \varepsilon_t^{\mathbb{Q}}, \quad \varepsilon_t^{\mathbb{Q}} \sim N(0, I_N)$$

where $\mu^{\mathbb{Q}} = \mu - \lambda_0$, $\Phi^{\mathbb{Q}} = \Phi - \lambda_1$.

Observed yields

Taking expectations wrt $\mathbb{Q}$, the time-t price of n-period bond P_t^n is

$$P_t^n = \exp(A_n + B_n' X_t),$$

where A_n and B_n are functions of $(\mu^{\mathbb{Q}}, \Phi^{\mathbb{Q}}, \Sigma)$ and are given from the Riccati recursions (Ang and Piazzesi, 2003).

Observed yields

Taking expectations wrt $\mathbb{Q}$, the time-t price of n-period bond P_t^n is

$$P_t^n = \exp(A_n + B_n' X_t),$$

where A_n and B_n are functions of $(\mu^{\mathbb{Q}}, \Phi^{\mathbb{Q}}, \Sigma)$ and are given from the Riccati recursions (Ang and Piazzesi, 2003).

For y_t being the cross-sectional vector with $y_t^n = -\frac{\log P_t^n}{n}$ $\forall n$, we get

$$y_t = A_{n,X} + B_{n,X} X_t.$$

Observed yields

Taking expectations wrt $\mathbb{Q}$, the time-t price of n-period bond P_t^n is

$$P_t^n = \exp(A_n + B_n' X_t),$$

where A_n and B_n are functions of $(\mu^{\mathbb{Q}}, \Phi^{\mathbb{Q}}, \Sigma)$ and are given from the Riccati recursions (Ang and Piazzesi, 2003).

For y_t being the cross-sectional vector with $y_t^n = -\frac{\log P_t^n}{n}$ $\forall n$, we get

$$y_t = A_{n,X} + B_{n,X} X_t.$$

Proceeding with above data and latent states X_t, identification and estimation of DTSM is challenging (Ang et al., 2007; Chib and Ergashev, 2009).

Canonical setup of Joslin et al. (2011)

X_t is rotated to the first N Principal Components of observed yields $\mathcal{P}_t = Wy_t$. Letting $\mu_{\mathcal{P}}^{\mathbb{P}} = \mu_{\mathcal{P}}^{\mathbb{Q}} + \lambda_{0\mathcal{P}}$, $\Phi_{\mathcal{P}}^{\mathbb{P}} = \Phi_{\mathcal{P}}^{\mathbb{Q}} + \lambda_{1\mathcal{P}}$, gives

$$\mathcal{P}_t - \mathcal{P}_{t-1} = \mu_{\mathcal{P}}^{\mathbb{P}} + \Phi_{\mathcal{P}}^{\mathbb{P}} \mathcal{P}_{t-1} + \Sigma_{\mathcal{P}} \varepsilon_t^{\mathbb{P}}$$

$$\mathcal{P}_t - \mathcal{P}_{t-1} = \mu_{\mathcal{P}}^{\mathbb{Q}} + \Phi_{\mathcal{P}}^{\mathbb{Q}} \mathcal{P}_{t-1} + \Sigma_{\mathcal{P}} \varepsilon_t^{\mathbb{Q}}$$

$$y_t = A_{\mathcal{P}} + B_{\mathcal{P}} \mathcal{P}_t + e_t$$

where $J - N$ of J yields in y_t are observed with $N(0,\ \sigma_e^2)$ errors.

Canonical setup of Joslin et al. (2011)

X_t is rotated to the first N Principal Components of observed yields $\mathcal{P}_t = Wy_t$. Letting $\mu^{\mathbb{P}}_{\mathcal{P}} = \mu^{\mathbb{Q}}_{\mathcal{P}} + \lambda_{0\mathcal{P}}$, $\Phi^{\mathbb{P}}_{\mathcal{P}} = \Phi^{\mathbb{Q}}_{\mathcal{P}} + \lambda_{1\mathcal{P}}$, gives

$$\mathcal{P}_t - \mathcal{P}_{t-1} = \mu^{\mathbb{P}}_{\mathcal{P}} + \Phi^{\mathbb{P}}_{\mathcal{P}}\mathcal{P}_{t-1} + \Sigma_{\mathcal{P}}\varepsilon^{\mathbb{P}}_t$$

$$\mathcal{P}_t - \mathcal{P}_{t-1} = \mu^{\mathbb{Q}}_{\mathcal{P}} + \Phi^{\mathbb{Q}}_{\mathcal{P}}\mathcal{P}_{t-1} + \Sigma_{\mathcal{P}}\varepsilon^{\mathbb{Q}}_t$$

$$y_t = A_{\mathcal{P}} + B_{\mathcal{P}}\mathcal{P}_t + e_t$$

where $J - N$ of J yields in y_t are observed with $N(0,\ \sigma_e^2)$ errors.

Restrictions on $\mu^{\mathbb{Q}}_{\mathcal{P}}, \Phi^{\mathbb{Q}}_{\mathcal{P}}, \delta_0, \delta_1$ for identification. The Joint likelihood is

$$f(Y|\theta) = \left\{\prod_{t=0}^{T} f^{\mathbb{Q}}(y_t|\mathcal{P}_t, k^{\mathbb{Q}}_{\infty}, g^{\mathbb{Q}}, \Sigma_{\mathcal{P}}, \sigma_e^2)\right\} \times \left\{\prod_{t=1}^{T} f^{\mathbb{P}}(\mathcal{P}_t|\mathcal{P}_{t-1}, k^{\mathbb{Q}}_{\infty}, g^{\mathbb{Q}}, \lambda_{0\mathcal{P}}, \lambda_{1\mathcal{P}}, \Sigma_{\mathcal{P}})\right\}$$

Outline

Empirical Questions

- Are DTSMs useful for prediction?
- Identifying a good set of restrictions on $\lambda_{0\mathcal{P}}, \lambda_{1\mathcal{P}}$ is important towards resolving the 'puzzle' but does it translate to improved forecasts?
- If yes to the above, does this predictive ability translate to economic benefits for investors?

Excess Bond Returns

Focus is on forecasting excess bond returns. Letting $p_t^n = \log P_t^n$, they are defined as

$$rx_{t,t+h}^n = p_{t+h}^{n-h} - p_t^n - p_t^h = -(n-h)y_{t+h}^{n-h} + ny_t^n - hy_t^h,$$

i.e. the difference between the h-holding period return of the n-period bond and the h-period yield.

Excess Bond Returns

Focus is on forecasting excess bond returns. Letting $p_t^n = \log P_t^n$, they are defined as

$$rx_{t,t+h}^n = p_{t+h}^{n-h} - p_t^n - p_t^h = -(n-h)y_{t+h}^{n-h} + ny_t^n - hy_t^h,$$

i.e. the difference between the h-holding period return of the n-period bond and the h-period yield.

The excess bond return model-based forecast $\widetilde{rx}_{t,t+h}^n$, based on the forecast $\widetilde{\mathcal{P}}_{t+h}$, is given from

$$\widetilde{rx}_{t,t+h}^n = A_{n-h,\mathcal{P}} - A_{n,\mathcal{P}} + A_{h,\mathcal{P}} + B'_{n-h,\mathcal{P}}\widetilde{\mathcal{P}}_{t+h} - (B_{n,\mathcal{P}} - B_{h,\mathcal{P}})'\mathcal{P}_t,$$

Predictability and Economic Value

To see if predictability translates into economic benefits, we consider a Bayesian investor with power utility preferences

$$U(W_{t+h}) = U(w_t^n, rx_{t+h}^n) = \frac{W_{t+h}^{1-\gamma}}{1-\gamma}$$

where γ is the coefficient of relative risk aversion and

$$W_{t+h} = (1 - w_t^n) \exp(r_t^f) + w_t^n \exp(r_t^f + rx_{t,t+h}^n)$$

where w_t^n is the portfolio weight on the risky n-period bond.

Predictability and Economic Value

To see if predictability translates into economic benefits, we consider a Bayesian investor with power utility preferences

$$U(W_{t+h}) = U(w_t^n, rx_{t+h}^n) = \frac{W_{t+h}^{1-\gamma}}{1-\gamma}$$

where γ is the coefficient of relative risk aversion and

$$W_{t+h} = (1 - w_t^n)\exp(r_t^f) + w_t^n \exp(r_t^f + rx_{t,t+h}^n)$$

where w_t^n is the portfolio weight on the risky n-period bond.

Such an investor rebalances the portfolio at each time t by maximising the expected utility

$$E_t[U(W_{t+h})|x_{1:t}] = \int U(w_t^n, rx_{t+h}^n) f(rx_{t+h}^n) drx_{t+h}^n,$$

where $f(rx_{t+h}^n)$ is a predictive distribution.

Outline

Model and Data Setup

Likelihood $f(Y_{0:t}|\theta)$, based on $Y_{0:t} = (Y_0, Y_1 \ldots, Y_t)$, combined with a prior $\pi(\theta)$, yields the posterior distribution:

$$\pi(\theta|Y_{0:t}) = \frac{1}{m(Y_{0:t})} f(Y_{0:t}|\theta)\pi(\theta).$$

Model and Data Setup

Likelihood $f(Y_{0:t}|\theta)$, based on $Y_{0:t} = (Y_0, Y_1 \ldots, Y_t)$, combined with a prior $\pi(\theta)$, yields the posterior distribution:

$$\pi(\theta|Y_{0:t}) = \frac{1}{m(Y_{0:t})} f(Y_{0:t}|\theta)\pi(\theta).$$

Predictive distribution to assess out-of-sample forecasting performance of the models:

$$f(Y_{t+h}|Y_{0:t}) = \int f(Y_{t+h}|Y_t, \theta)\pi(\theta|Y_{0:t})d\theta.$$

Model and Data Setup

Likelihood $f(Y_{0:t}|\theta)$, based on $Y_{0:t} = (Y_0, Y_1 \ldots, Y_t)$, combined with a prior $\pi(\theta)$, yields the posterior distribution:

$$\pi(\theta|Y_{0:t}) = \frac{1}{m(Y_{0:t})} f(Y_{0:t}|\theta)\pi(\theta).$$

Predictive distribution to assess out-of-sample forecasting performance of the models:

$$f(Y_{t+h}|Y_{0:t}) = \int f(Y_{t+h}|Y_t, \theta)\pi(\theta|Y_{0:t})d\theta.$$

Need the above for many t, e.g. all t after some warm-up period. One option is to use Markov Chain Monte Carlo but it is too costly.

Iterated Batch Importance Sampling (IBIS) (Chopin, 2002; Del Moral et al., 2006)

1. *Initialize* N_θ *particles by drawing independently* $\theta_i \sim \pi(\theta)$ *with importance weights* $\omega_i = 1$, $i = 1, \ldots, N_\theta$.

2. *For* $t, \ldots, T$ *do for all* i*:*

   (a) *Calculate the incremental weights*

   $$u_t(\theta_i) = f(Y_t | Y_{0:t-1}, \theta_i) = f(Y_t | Y_{t-1}, \theta_i)$$

   (b) *Update the importance weights* ω_i *to* $\omega_i u_t(\theta_i)$.

   (c) *If some degeneracy criterion (e.g. ESS(*ω*)) is triggered, perform the following two sub-steps:*

      (i) *Resampling: Sample with replacement* N_θ *times from the set of* θ_i*s according to* ω_i*s. The weights are then reset to one.*

      (ii) *Jittering: Replace* θ_i*s with* $\tilde{\theta}_i$*s by running MCMC chains with each* θ_i *as input and* $\tilde{\theta}_i$ *as output.*

IBIS output

- The set of θ particles can be used to compute expectations with respect to the posterior, $E[g(\theta)|Y_{0:t}]$, for all t using the estimator $\sum_i[\omega_i g(\theta_i)]/\sum_i \omega_i$.

IBIS output

- The set of θ particles can be used to compute expectations with respect to the posterior, $E[g(\theta)|Y_{0:t}]$, for all t using the estimator $\sum_i[\omega_i g(\theta_i)]/\sum_i \omega_i$.
- The θ particles can be transformed into weighted samples from the predictive distribution $f(Y_{t+h}|Y_t)$ for all t.

IBIS output

- The set of θ particles can be used to compute expectations with respect to the posterior, $E[g(\theta)|Y_{0:t}]$, for all t using the estimator $\sum_i[\omega_i g(\theta_i)]/\sum_i \omega_i$.
- The θ particles can be transformed into weighted samples from the predictive distribution $f(Y_{t+h}|Y_t)$ for all t.
- A very useful by-product is the ability to compute the model evidence $m(Y_{0:t}) = f(Y_{0:t})$, for model choice/averaging via its estimator

$$M_t = \frac{1}{\sum_{i=1}^{N_\theta} \omega_i} \sum_{i=1}^{N_\theta} \omega_i u_t(\theta_i).$$

IBIS output

- The set of θ particles can be used to compute expectations with respect to the posterior, $E[g(\theta)|Y_{0:t}]$, for all t using the estimator $\sum_i[\omega_i g(\theta_i)]/\sum_i \omega_i$.
- The θ particles can be transformed into weighted samples from the predictive distribution $f(Y_{t+h}|Y_t)$ for all t.
- A very useful by-product is the ability to compute the model evidence $m(Y_{0:t}) = f(Y_{0:t})$, for model choice/averaging via its estimator

$$M_t = \frac{1}{\sum_{i=1}^{N_\theta} \omega_i} \sum_{i=1}^{N_\theta} \omega_i u_t(\theta_i).$$

- The IBIS output can be used to monitor how estimates of certain parameters evolve as data accumulate.

IBIS output

- The set of θ particles can be used to compute expectations with respect to the posterior, $E[g(\theta)|Y_{0:t}]$, for all t using the estimator $\sum_i [\omega_i g(\theta_i)] / \sum_i \omega_i$.
- The θ particles can be transformed into weighted samples from the predictive distribution $f(Y_{t+h}|Y_t)$ for all t.
- A very useful by-product is the ability to compute the model evidence $m(Y_{0:t}) = f(Y_{0:t})$, for model choice/averaging via its estimator

$$M_t = \frac{1}{\sum_{i=1}^{N_\theta} \omega_i} \sum_{i=1}^{N_\theta} \omega_i u_t(\theta_i).$$

- The IBIS output can be used to monitor how estimates of certain parameters evolve as data accumulate.
- A more robust alternative to MCMC even in offline problems.

Priors

Priors on most parameters were set to relatively flat distributions.

Priors

Priors on most parameters were set to relatively flat distributions.

For the risk premia parameters $\lambda^{\mathcal{P}} = (\lambda^{0\mathcal{P}}, \lambda^{1\mathcal{P}})$, spike-and-slab priors were used, aka as Stochastic Search Variable Selection (SSVS)

$$\lambda_{ij}^{\mathcal{P}} \sim (1-\gamma_{ij})N(0,\tau_{ij}^{(0)}) + \gamma_{ij}N(0,\tau_{ij}^{(1)})$$

where γ_{ij}s are Bernoulli(π) random variables indicating large (free) parameters versus small.

Priors

Priors on most parameters were set to relatively flat distributions.

For the risk premia parameters $\lambda^{\mathcal{P}} = (\lambda^{0\mathcal{P}}, \lambda^{1\mathcal{P}})$, spike-and-slab priors were used, aka as Stochastic Search Variable Selection (SSVS)

$$\lambda_{ij}^{\mathcal{P}} \sim (1 - \gamma_{ij}) N(0, \tau_{ij}^{(0)}) + \gamma_{ij} N(0, \tau_{ij}^{(1)})$$

where γ_{ij}s are Bernoulli(π) random variables indicating large (free) parameters versus small.

$\tau_{ij}^{(1)}$s were set to unit information priors, and $\tau_{ij}^{(0)}$s were set to very small values.

Priors

Priors on most parameters were set to relatively flat distributions.

For the risk premia parameters $\lambda^{\mathcal{P}} = (\lambda^{0\mathcal{P}}, \lambda^{1\mathcal{P}})$, spike-and-slab priors were used, aka as Stochastic Search Variable Selection (SSVS)

$$\lambda_{ij}^{\mathcal{P}} \sim (1 - \gamma_{ij}) N(0, \tau_{ij}^{(0)}) + \gamma_{ij} N(0, \tau_{ij}^{(1)})$$

where γ_{ij}s are Bernoulli(π) random variables indicating large (free) parameters versus small.

$\tau_{ij}^{(1)}$s were set to unit information priors, and $\tau_{ij}^{(0)}$s were set to very small values.

Different approach for π, either fixed to a value such as 0.5 or assigned a Beta prior and estimated by the data.

MCMC and γs

An MCMC algorithm is still required for the jittering step of the IBIS (the most costly one).

MCMC and γs

An MCMC algorithm is still required for the jittering step of the IBIS (the most costly one).

A Gibbs scheme was used splitting the data in different blocks. For some parameters (γs, λs, σ_e^2) the full conditionals are available. For Σ, $\mu_{\mathcal{P}}^{\mathbb{Q}}$ and $\Phi_{\mathcal{P}}^{\mathbb{Q}}$, independence samples were used.

MCMC and γs

An MCMC algorithm is still required for the jittering step of the IBIS (the most costly one).

A Gibbs scheme was used splitting the data in different blocks. For some parameters (γs, λs, σ_e^2) the full conditionals are available. For Σ, $\mu_{\mathcal{P}}^{\mathbb{Q}}$ and $\Phi_{\mathcal{P}}^{\mathbb{Q}}$, independence samples were used.

Samples from the posterior of γs can be used to decide whether each element of $\lambda^{\mathcal{P}}$ should be restricted to 0.

MCMC and γs

An MCMC algorithm is still required for the jittering step of the IBIS (the most costly one).

A Gibbs scheme was used splitting the data in different blocks. For some parameters (γs, λs, σ_e^2) the full conditionals are available. For Σ, $\mu_{\mathcal{P}}^{\mathbb{Q}}$ and $\Phi_{\mathcal{P}}^{\mathbb{Q}}$, independence samples were used.

Samples from the posterior of γs can be used to decide whether each element of $\lambda^{\mathcal{P}}$ should be restricted to 0.

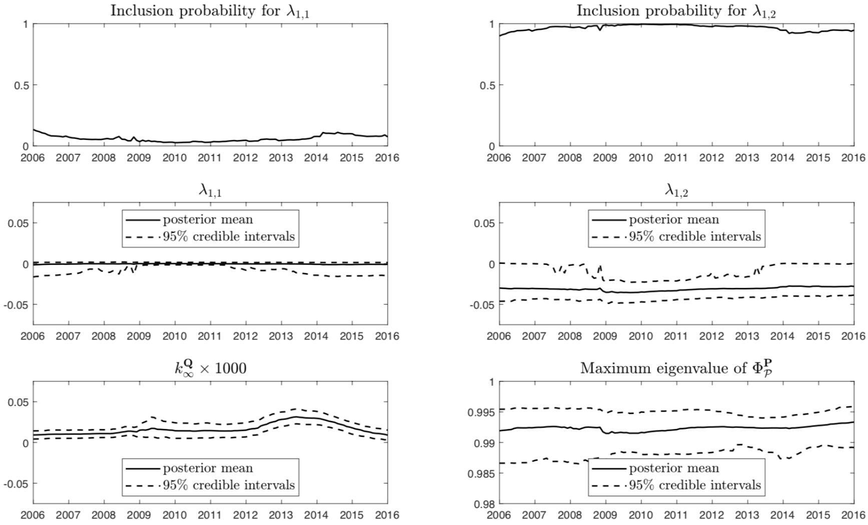
In our approach γs are part of the IBIS and can therefore facilitate sequential model choice and averaging.

Outline

Data and Models

- The data set contains monthly observations of zero-coupon US Treasury yields with maturities of 1-year, 2-year, 3-year, 4-year, 5-year, 7-year, and 10-year
- The period considered is from 1990 to 2016. No predictions are evaluated until 2007 (warm-up period), but this is done for each month afterwards.
- PCA weights are computed based on data up to 2007 and kept fixed afterwards.
- In terms of models, three different SSVS algorithms were used, based on two different prior specifications on π and a third scheme where only two λs were allowed to be non-zero.

Some IBIS Output



Economic Value

Out-of-sample Economic performance of Bond excess return forecasts across investment scenarios: multiple prediction horizons - Period: January 1990 - end of 2016.

h	2Y	3Y	4Y	5Y	7Y	10Y
			M0			
1m	-13.38	-15.71	-14.85	-11.84	-11.51	-16.85
3m	-10.55	-11.77	-10.64	-7.34	-7.73	-9.61
6m	-12.93	-12.74	-11.25	-8.58	-7.78	-7.37
9m	-11.10	-11.45	-10.81	-9.18	-8.32	-6.93
12m	-9.13	-9.62	-9.48	-8.09	-7.34	-5.63
			M1			
1m	4.66**	3.41**	3.77**	3.80**	5.37**	3.59
3m	4.72***	3.65**	3.04**	3.40**	3.04**	2.39
6m	5.18***	3.94***	2.64***	2.67***	2.35**	2.34**
9m	5.82***	4.16***	2.75***	2.36***	1.92**	2.36**
12m	6.37***	5.32***	3.76***	3.22***	2.61***	2.87***
			M5			
1m	2.07	0.73	0.82	1.08	2.35	0.27
3m	4.38***	3.50**	2.77**	3.05*	2.36*	1.62
6m	4.67***	3.42***	2.15**	2.18**	1.72*	1.64
9m	5.26***	3.57***	2.19**	1.88**	1.47	1.88
12m	5.67***	4.64***	3.16***	2.72***	2.21**	2.51**
			M6			
1m	3.70**	2.01	2.05	1.90	3.48*	2.21
3m	4.81***	3.76**	3.14**	3.49**	2.99**	2.21
6m	5.16***	3.94***	2.65***	2.67***	2.27**	2.19*
9m	5.88***	4.19***	2.90***	2.51***	2.08**	2.41**
12m	6.35***	5.36***	3.83***	3.32***	2.77***	2.89**

Outline

Discussion - Extensions

- The development of the sequential version of the SSVS scheme allowed us to establish the predictive ability of DTSMs in cases of extreme risk premia restrictions.

Discussion - Extensions

- The development of the sequential version of the SSVS scheme allowed us to establish the predictive ability of DTSMs in cases of extreme risk premia restrictions.

- Interested to study the method in more challenging settings in terms for model uncertainty.

Discussion - Extensions

- The development of the sequential version of the SSVS scheme allowed us to establish the predictive ability of DTSMs in cases of extreme risk premia restrictions.

- Interested to study the method in more challenging settings in terms for model uncertainty.

- On another project we worked on including a latent unspanned factor representing market environment information.

Discussion - Extensions

- The development of the sequential version of the SSVS scheme allowed us to establish the predictive ability of DTSMs in cases of extreme risk premia restrictions.
- Interested to study the method in more challenging settings in terms for model uncertainty.
- On another project we worked on including a latent unspanned factor representing market environment information.
- Another working projects aims to incroporate unspanned macros as covariates using multiple output Gaussian processes. Potentially interested to check deep Gaussian processes.

References

Ang, A., Dong, S. and Piazzesi, M. (2007), No-arbitrage taylor rules, Working Paper 13448, National Bureau of Economic Research.

Ang, A. and Piazzesi, M. (2003), 'A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables', *Journal of Monetary Economics* **50**(4), 745–787.

Bauer, M. D. (2018), 'Restrictions on risk prices in dynamic term structure models', *Journal of Business & Economic Statistics* **36**(2), 196–211.

Chib, S. and Ergashev, B. (2009), 'Analysis of multifactor affine yield curve models.', *Journal of the American Statistical Association* **104**(488), 1324–1337.

Chopin, N. (2002), 'A sequential particle filter method for static models', *Biometrika* **89**(3), 539–552.

Dai, Q. and Singleton, K. J. (2000), 'Specification analysis of affine term structure models', *The Journal of Finance* **55**(5), 1943–1978.

Del Moral, P., Doucet, A. and Jasra, A. (2006), 'Sequential monte carlo samplers', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68**(3), 411–436.

Duffee, G. R. (2002), 'Term premia and interest rate forecasts in affine models', *The Journal of Finance* **57**(1), 405–443.

Joslin, S., Singleton, K. J. and Zhu, H. (2011), 'A new perspective on Gaussian dynamic term structure models', *Review of Financial Studies* **24**(3), 926–970.