Fixed Income and Credit Risk

Lecture 5

Professor	Assistant	Program
Fulvio Pegoraro	Roberto Marfè	MSc. Finance

Fall Semester 2012

UNIL | Université de Lausanne

Fixed Income and Credit Risk

Lecture 5 - Part I

Empirical Analysis of Gaussian

Affine Term Structure Models

Outline of Lecture 5 - Part I

5.1 An Empirical Analysis of Gaussian $ATSMs$

5.1.1 Description of the Data

5.1.2 Estimated Models

5.1.3 Estimation Method

5.1.4 Results for the AR(p) Factor-Based Term Structure Models

5.1.5 Results for the bivariate VAR(p) Factor-Based Term Structure Models

5.1.6 In-sample Fit of the Yield Curve

5.1.7 Do these models explain the Violation of the EHT ?

5.2 Alternative Estimation Procedures for Gaussian $ATSMs$

5.2.1 MLE through the "Inversion Procedure"

5.2.2 MLE through Kalman Filter Recursions

5.2.3 The Adrian, Crump and Moench (2012, JFE) Approach

5.3 Ang and Piazzesi (2003, JME)

5.3.1 Purpose of the paper

5.3.2 Main results

5.3.3 Setup

5.3.4 Estimation Procedure

5.3.5 Results

5.1 An Empirical Analysis of Gaussian $ATSMs$

5.1.1 Description of the Data

- □ The CRSP data set on the U. S. term structure of interest rates (treasury zero-coupon bond yields), that we consider in the following application, covers the period from June 1964 to December 1995 and contains 379 monthly observations for each of the nine maturities : 1, 3, 6 and 9 months and 1, 2, 3, 4 and 5 years.

- □ Summary statistics about the above mentioned (annualized) yields are presented in Table 1 :

Table 1 : Summary Statistics on U. S. Monthly Yields from June 1964 to December 1995.

ACF(k) indicates the empirical autocorrelation between yields $R(t,h)$ and $R(t-k,h)$, with h and k expressed on a monthly basis.

Maturity	1-m	3-m	6-m	9-m	1-yr	2-yr	3-yr	4-yr	5-yr
Mean	0.0645	0.0672	0.0694	0.0709	0.0713	0.0734	0.0750	0.0762	0.0769
Std. Dev.	0.0265	0.0271	0.0270	0.0269	0.0260	0.0252	0.0244	0.0240	0.0237
Skewness	1.2111	1.2118	1.1518	1.1013	1.0307	0.9778	0.9615	0.9263	0.8791
Kurtosis	4.5902	4.5237	4.3147	4.1605	3.9098	3.6612	3.5897	3.5063	3.3531
Minimum	0.0265	0.0277	0.0287	0.0299	0.0311	0.0366	0.0387	0.0397	0.0398
Maximum	0.1640	0.1612	0.1655	0.1644	0.1581	0.1564	0.1556	0.1582	0.1500
ACF(1)	0.9587	0.9731	0.9747	0.9745	0.9727	0.9780	0.9797	0.9802	0.9822
ACF(5)	0.8288	0.8531	0.8579	0.8588	0.8604	0.8783	0.8915	0.8986	0.9053
ACF(10)	0.7278	0.7590	0.7691	0.7699	0.7683	0.7885	0.8021	0.8075	0.8212
ACF(20)	0.4303	0.4631	0.4880	0.4996	0.5156	0.5742	0.6051	0.6193	0.6431
ACF(30)	0.2548	0.2682	0.3016	0.3213	0.3518	0.4358	0.4725	0.4994	0.5187
ACF(40)	0.1362	0.1415	0.1677	0.1853	0.2160	0.3056	0.3427	0.3780	0.3961

☐ The term structure of ZCB yields is, on average:

- upward sloping
- and the yields with larger standard deviation, positive skewness and kurtosis are those with shorter maturities.
- Moreover, yields are **highly autocorrelated** with a persistence which is increasing with the time to maturity.

5.1.2 Estimated Models

- In the present empirical analysis we follow an endogenous approach, given that it gives several important advantages coming from the observations we have about the factor, that is, the short rate in the scalar case, or yields at different maturities in the multivariate framework.

- First we are able to detect stylized facts giving us the possibility to justify the AR(p) model we propose for the historical dynamics of (x_t) : indeed, a large empirical literature on bond yields show that interest rates have an historical multi-lag dynamics [see, among the others, Hamilton (1989), Christiansen and Lund (2003), Cochrane and Piazzesi (2005)].

- Second, observations about the Gaussian-distributed factor lead to an exact maximum likelihood estimation of historical parameters: in this way, we are able to test hypotheses using likelihood ratio statistics, and rank the models in terms of various information criteria.

- Finally, the difference between directly observed and estimated factor values determine model residuals that can be used to derive various diagnostic criteria.

5.1.3 Estimation Method

☐ The methodology we follow to estimate the parameters of the endogenous Multi-Lag term structure models is based on a consistent two-step procedure.

☐ In the first step, thanks to observations on the K-dimensional endogenous factor (x_t), we estimate the $[K(1+Kp)+(K(K+1)/2)]$-dimensional vector of parameters $\Theta_{\mathbb{P}} = [\nu', vec(\Phi)', \ vech(\Sigma\Sigma')']'$, characterizing the historical dynamics (x_t), by Maximum Likelihood (ML).

☐ In the case of a Gaussian VAR(p) process, the ML estimator coincides with the OLS estimator.

OLS Estimation of a Gaussian VAR(p) process with observable factor

□ Notation: $\mathbf{X} := (x_1, \ldots, x_T)$ is (K, T) matrix of observations; $B := (\nu, \Phi_1, \ldots, \Phi_p)$ is $(K, Kp + 1)$ matrix of parameters;

□ $Z_t := \begin{bmatrix} 1 \\ x_t \\ x_{t-1} \\ \vdots \\ x_{t-p+1} \end{bmatrix}$, $\mathbf{Z} := (Z_0, \ldots, Z_{T-1})$ is $((Kp + 1), T)$ matrix.

□ $U := (\varepsilon_1, \ldots, \varepsilon_T)$ is (K, T) matrix. We, thus, can write $\mathbf{X} = B\,\mathbf{Z} + U$.

□ $\mathbf{x} := vec(X)$ is $(TK, 1)$ vector, $\beta := vec(B)$ is $(K^2p + K, 1)$ vector.

□ OLS estimator : $vec(\widehat{B}) = \widehat{\beta} = vec(\mathbf{X}\,\mathbf{Z}'(\mathbf{Z}\,\mathbf{Z}')^{-1})$;

□ Given that $\Omega = E(\varepsilon_t\,\varepsilon_t')$, we estimate this matrix by:

$$\widehat{\Omega} \;=\; \frac{1}{T}\sum_{t=1}^{T}\widehat{\varepsilon}_t\,\widehat{\varepsilon}_t' = \frac{1}{T}\widehat{U}\,\widehat{U}' = \frac{1}{T}(\mathbf{X}-\widehat{B}\,\mathbf{Z})\,(\mathbf{X}-\widehat{B}\,\mathbf{Z}) = \frac{1}{T}\mathbf{X}\,(I_T - \mathbf{Z}'(\mathbf{Z}\,\mathbf{Z}')^{-1}\,\mathbf{Z})\,\mathbf{X}'\,.$$

□ How do we select the number of lags p (VAR order selection) in the VAR(p) model ?

$a)$ minimizing the Forecast Mean Square Error we obtain a criterion called *Final Prediction Error (FPE)*:

$$FPE(p) = \left[\frac{T+Kp+1}{T-Kp-1}\right]^K det(\widehat{\Omega}(p))\,.$$

b) *Akaike's Information Criterion (AIC)*: $AIC(p) = \ln det(\widehat{\Omega}(p)) + \frac{2\,p\,K^2}{T}$;

c) *Hannan-Quinn Criterion (HQ)*: $HQ(p) = \ln det(\widehat{\Omega}(p)) + \frac{2 \ln \ln T}{T} pK^2$;

d) *Schwarz Information Criterion (SC)*: $SC(p) = \ln det(\widehat{\Omega}(p)) + \frac{\ln T}{T} pK^2$;

- the selected AR order p is the one minimizing the criterion.

- Small sample comparisons: $p(SC) \leq p(AIC)$ if $T \geq 8$; $p(SC) \leq p(HQ)$ for all T; $p(HQ) \leq p(AIC)$ if $T \geq 16$.

□ In the second step, using observations on yields with maturities different from those used in the first step and for a given estimates of $vech(\Sigma\Sigma')$, we estimate the $[K(1+Kp)]$-dimensional vector of parameters $\Theta_{\mathbb{Q}} = [(\nu^*)', vec(\Phi^*)']'$, characterizing the risk-neutral dynamics of (x_t), by minimizing the sum of squared fitting errors between the observed and theoretical yields.

□ In other words, in this second step and for a given $\widehat{\Theta}_{\mathbb{P}}$, we estimate $(\gamma_o, \widetilde{\Gamma})$.

□ More precisely, in the **scalar case**, we estimate $\Theta_{\mathbb{Q}}$ by nonlinear lest squares (**NLLS**), while, in the **multivariate case**, these parameters are estimated by **Constrained NLLS**.

- The constraints are imposed to satisfy internal consistency conditions on (C_h, D_h) implied by the absence of arbitrage opportunity principle [see Lecture 4 and next slides].

- Given the complete set of 9 maturities of our data base, and given a number m of yields used to estimate the vector of historical parameters $\Theta_{\mathbb{P}}$, we denote by H_m^* the set of remaining maturities used to estimate the vector of risk-neutral parameters $\Theta_{\mathbb{Q}}$.

- In the AR(p) Factor-Based case, x_t is the one-month yield to maturity $R(t, t+1m) = r_t$ expressed at a monthly frequency.

□ In the bivariate VAR(p) Factor-Based case the factor is given by :

$$x_t = [R(t, t+1m), R(t, t+60m) - R(t, t+1m)]',$$

where $[R(t, t+60m) - R(t, t+1m)]$ is the spread at date t between the five-year and one-month yield to maturity, expressed at a monthly frequency.

□ The NLLS estimator for the AR(p) case, is determined by :

$$\begin{cases} \widehat{\Theta}_{\mathbb{Q}} = Arg\min_{\Theta_{\mathbb{Q}}} S^2(\Theta_{\mathbb{Q}}), \\ S^2(\Theta_{\mathbb{Q}}) = \sum_{t=p}^{T} \sum_{h \in H_1^*} [R^o(t, t+h) - R(t, t+h)]^2, \end{cases} \tag{1}$$

given the set H_1^* of maturities used to estimate the risk-neutral parameters; $R^o(t, t+h)$ is the observed yield, while $R(t, t+h)$ is the model-implied one.

□ The constrained NLLS estimator, in our bivariate model specification, is given by :

$$\begin{cases} \widehat{\Theta}_{\mathbb{Q}} = Arg\,\min_{\Theta_{\mathbb{Q}}} S^2(\Theta_{\mathbb{Q}}) \\ S^2(\Theta_{\mathbb{Q}}) = \sum_{t=p}^{T} \sum_{h \in H_2^*} [R^o(t,t+h) - R(t,t+h)]^2, \\ \text{s. t.} \ \sum_{t=p}^{T} [R^o(t,t+60m) - R(t,t+60m)]^2 = 0\,, \end{cases} \quad (2)$$

□ The constraint in the minimization program (2) guarantees the absence of arbitrage opportunity on the five-year yield to maturity.

5.1.4 Results for the AR(p) Factor-Based Term Structure Models

☐ The maximum value of the mean Log-Likelihood and the values of the estimated vector of parameters $\Theta_{\mathbb{P}} = (\nu, \varphi_1, \ldots, \varphi_p, \sigma)'$ of the AR(p) Factor-Based Term Structure models, for $p \in \{1, \ldots, 6\}$, are presented in Tables 2 and 3 [the t-values are given in parenthesis].

☐ We denote with $mlogL$ the mean log-Likelihood of the AR(p) model : $mlogL = logL(\Theta_{\mathbb{P}}|x_1, \ldots, x_{T-p})/(T-p)$.

☐ The Akaike Information Criterion (AIC) (for ranking among models) is given by $2mlogL - (2k/(T-p))$, with k denoting the dimension of $\Theta_{\mathbb{P}}$.

Table 2 : AR(p) Factor-Based Term Structure models. Maximum value of the mean Log-Likelihood, AIC and parameter estimates of ν and σ. (**) denotes a parameter significant at 0.05; (*) denotes a parameter significant at 0.1.

	AR(1)	AR(2)	AR(3)	AR(4)	AR(5)	AR(6)
$mlogL$	5.95657	5.95868	5.96082	5.96134	5.97224	5.97092
AIC	11.8973	11.8961	11.8950	11.8907	11.9071	11.8990
ν	0.00023	0.00021	0.00023	0.00021	0.00019	0.00019
	[2.6725] **	[2.4822] **	[2.6598] **	[2.4761] **	[2.1571] **	[2.1262] **
σ^2	0.00000039	0.00000039	0.00000039	0.00000039	0.00000038	0.00000038
	[13.7483] **	[13.7301] **	[13.7118] **	[13.6937] **	[13.6754] **	[13.6571] **

Table 3 : AR(p) Factor-Based Term Structure models Parameter estimates of $(\varphi_1, \ldots, \varphi_p)$.

(**) denotes a parameter significant at 0.05; (*) denotes a parameter significant at 0.1.

	AR(1)	AR(2)	AR(3)	AR(4)	AR(5)	AR(6)
φ_1	0.9580 ** [65.5620]	0.8798 ** [17.2393]	0.8861 ** [17.1525]	0.8912 ** [17.1688]	0.8814 ** [17.1628]	0.8806 ** [16.9714]
φ_2		0.0811 [1.5938]	0.1547 ** [2.2869]	0.1456 ** [2.0843]	0.1672 ** [2.4260]	0.1675 ** [2.3885]
φ_3			−0.0829 * [−1.6459]	−0.1372 * [−1.9204]	−0.1595 ** [−2.3048]	−0.1586 ** [−2.2623]
φ_4				0.0608 [1.1455]	−0.0790 [−1.1788]	−0.0798 [−1.1240]
φ_5					0.1557 ** [3.1048]	0.1510 ** [2.4443]
φ_6						0.0053 [0.1232]

- An examination of the above displayed parameter estimates show, first of all, that the historical dynamics of the (one-month to maturity) short rate is not Markovian of order one, given that, in the AR(5) and AR(6) specifications, the parameters $(\varphi_1, \varphi_2, \varphi_3, \varphi_5)$ are always significative.

- The minimum value of the mean NLLS criterion $[S^2(\widehat{\Theta}_{\mathbb{Q}})/T^*]$ and the values of the estimated vector of risk-neutral parameters $\Theta_{\mathbb{Q}} = (\nu^*, \varphi_1^*, \ldots, \varphi_p^*)$, with $p \in \{1, \ldots, 6\}$, are presented in Tables 5 and 6 [the t-values are given in parenthesis]. We also rank the models in terms of the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE).

Table 4 : AR(p) Factor-Based Term Structure models. Minimum value of the mean NLLS criterion, RMSE, MAE and parameter estimates of ν^*. (**) denotes a parameter significant at 0.05; (*) denotes a parameter significant at 0.1.

	AR(1)	AR(2)	AR(3)	AR(4)	AR(5)	AR(6)
$S^2(\widehat{\Theta}_{\mathbb{Q}})/T^*$	0.00000054	0.00000051	0.00000050	0.00000048	0.00000047	0.00000046
RMSE	0.000736	0.000716	0.000709	0.000696	0.000687	0.000679
MAE	0.000530	0.000526	0.000528	0.000524	0.000517	0.000509
ν^*	0.000110	0.000151	0.000152	0.000148	0.000148	0.000152
	[33.2526] **	[22.6031] **	[22.9266] **	[22.9794] **	[22.7051] **	[22.4479] **

Table 5 : AR(p) Factor-Based Term Structure models. Parameter estimates of $(\varphi_1^*, \ldots, \varphi_p^*)$.

(**) denotes a parameter significant at 0.05; (*) denotes a parameter significant at 0.1.

	AR(1)	AR(2)	AR(3)	AR(4)	AR(5)	AR(6)
φ_1^*	0.9899 ** [1877]	0.5076 ** [9.6003]	0.7333 ** [14.2703]	0.7758 ** [15.4922]	0.7382 ** [14.2057]	0.7037 ** [13.3209]
φ_2^*		0.4788 ** [9.1313]	−0.0299 [−0.4132]	0.2291 ** [2.8931]	0.2947 ** [3.6124]	0.2998 ** [3.6802]
φ_3^*			0.2832 ** [7.5221]	−0.3860 ** [−5.3681]	−0.1600 ** [−2.0834]	−0.1069 [−1.3898]
φ_4^*				0.3685 ** [10.2233]	−0.1977 ** [−2.6864]	0.0123 [0.1609]
φ_5^*					0.3126 ** [8.4180]	−0.2173 ** [−2.9386]
φ_6^*						0.2961 ** [7.7697]

5.1.5 Results for the bivariate VAR(p) Factor-Based Term Structure Models

☐ As in the scalar case, we present the maximum value of the mean Log-Likelihood and the values of the estimated vector of parameters $\Theta_{\mathbb{P}} = [\nu', vec(\Phi)', vech(\Sigma\Sigma')']'$ of the bivariate VAR(p) Factor-Based Term Structure models, for an AR order $p = 1$ and $p = 2$.

☐ These results are presented in Tables 6 and 7. We have also estimated the historical parameters of the above mentioned bivariate VAR(p) model, for p larger than 2, but the AIC criterion has indicated the first two AR orders as the preferred ones.

Table 6 : VAR(p) Factor-Based Term Structure models. (**) denotes a parameter significant at 0.05; (*) denotes a parameter significant at 0.1.

	VAR(1)	VAR(2)
$mlogL$	12.6403	12.6837
AIC	25.2330	25.2984
ν_1	0.000065 [0.5856]	0.000132 [1.2262]
ν_2	0.000080 [0.8157]	0.000026 [0.2701]
σ_1^2	0.00000039 [5.94750] **	0.00000036 [6.02614] **
σ_{21}	-0.00000028 [-6.0995] **	-0.00000026 [-6.2100] **
σ_2^2	0.00000030 [7.6713] **	0.00000028 [8.0731] **

Table 7 : VAR(1) and VAR(2) Factor-Based Term Structure models. Parameter estimates of (φ_1, φ_2). (**) denotes a parameter significant at 0.05; (*) denotes a parameter significant at 0.1.

	VAR(1)		VAR(2)	
Φ_1	0.9742	0.0719	1.3318	0.6207
	[59.8835] **	[2.2174] **	[15.0111] **	[7.0095] **
	0.0091	0.8769	−0.2744	0.4353
	[0.6388]	[30.7835] **	[-3.4988] **	[5.5601] **
Φ_2			−0.3648	−0.5762
			[-3.6117] **	[-5.8201] **
			0.2893	0.4642
			[3.2397] **	[5.3020] **

- If we consider the parameter estimates of Tables 6 and 7, we observe that the joint historical dynamics of short rate and spread is not Markovian of order one, given that, in the VAR(2) specification, the parameters in the second autoregressive matrix φ_2 are significantly different from zero.

- Moreover, the AIC indicates this model as the preferred one. Table 6 shows also that the constant term $(\nu_1, \nu_2)'$ is not significative for both AR orders.

- We present the minimum value of the mean NLLS criterion $[S^2(\widehat{\Theta}_{\mathbb{Q}})/T^*]$ and the values of the estimated vector of risk-neutral parameters $\Theta_{\mathbb{Q}} = [(\nu^*)', vec(\Phi^*)']'$, for the bivariate VAR(1) and VAR(2) Factor-Based Term Structure models, in Tables 8 and 9.

Table 8 : VAR(p) Factor-Based Term Structure models. Minimum value of the mean NLLS criterion, RMSE, MAE and parameter estimates of (ν_1^*, ν_2^*). (**) denotes a parameter significant at 0.05; (*) denotes a parameter significant at 0.1.

	VAR(1)	VAR(2)
$S^2(\widehat{\theta}_{\mathbb{Q}})/T^*$	0.00000009	0.00000008
RMSE	0.000297	0.000283
MAE	0.000208	0.000198
ν_1^*	-0.000058	-0.000055
	$[-6.6459]$ **	$[-4.9423]$ **
ν_2^*	0.000072	0.000071
	[5.7860] **	[4.5783] **

Table 9 : VAR(1) and VAR(2) Factor-Based Term Structure models. Parameter estimates of (Φ_1^*, Φ_2^*). (**) denotes a parameter significant at 0.05; (*) denotes a parameter significant at 0.1.

	VAR(1)		VAR(2)	
Φ_1^*	1.0131	0.1105	1.3154	0.6020
	[805.8869] **	[34.5743] **	[28.4716] **	[9.5120] **
	-0.0156	0.9072	-0.2528	0.4142
	[-8.6611] **	[203.2978] **	[-3.5778] **	[4.2509] **
Φ_2^*			-0.3004	-0.4890
			[-6.5177] **	[-7.8923] **
			0.2342	0.4839
			[3.3244] **	[5.0769] **

☐ We find that, also in this bivariate risk-neutral (pricing) framework, the lagged values of the short rate and spread play an important role in the model specification. One may observe the significativity of all risk-neutral AR coefficients in the VAR(2) specification.

☐ In other words, a VAR(2) specification for the historical and risk-neutral dynamics of the factor driving term structure shapes, lead to propose a bivariate term structure model which is able to fit yields to maturity better than the VAR(1) and AR(p) specification.

5.1.6 In-sample fit of the Yield Curve

- Summary of in-sample fit performances, using $RMSE$ and MAE.

- The bivariate setting strongly dominate the scalar one, regardless the number of lags.

- In the bivariate setting, the introduction of an additional lag (marginally) improves the fitting performance.

	AR(6)	VAR(1)	VAR(2)
$RMSE$	0.000679	0.000297	0.000283
MAE	0.000509	0.000208	0.000198

5.1.7 Do these models explain the Violation of the EHT ?

☐ Short Horizon Expectation Hypothesis Tests : lags are useful !

Short Horizon	$m = 3$ months	$m = 6$ months	$m = 9$ months
$h = 6$ months	-0.6942 (0.2533)		
2-Factor VAR(1)	0.5828 (0.3485)		
2-Factor VAR(2)	-0.3800 (0.3837)		
$h = 9$ months	-0.8863 (0.3238)	-0.4023 (0.2429)	
2-Factor VAR(1)	0.4133 (0.3469)	0.4722 (0.2693)	
2-Factor VAR(2)	-0.5480 (0.3960)	-0.3890 (0.3182)	
$h = 12$ months	-1.3226 (0.3530)	-0.7867 (0.2381)	-0.4371 (0.1312)
2-Factor VAR(1)	0.2454 (0.3486)	0.3187 (0.2710)	0.3796 (0.2430)
2-Factor VAR(2)	-0.6935 (0.4069)	-0.5272 (0.3248)	-0.3675 (0.2930)

□ Long Horizon Expectation Hypothesis Tests: some problem !

Long Horizon	$m = 1$ year	$m = 2$ years	$m = 3$ years
$h = 4$ years	-1.8078 (0.2981)	-0.8380 (0.2889)	-0.0421 (0.2682)
2-Factor VAR(1)	-0.8569 (0.3536)	-0.0085 (0.3414)	0.8626 (0.3514)
2-Factor VAR(2)	-1.4088 (0.4084)	-0.2338 (0.3864)	0.9368 (0.3843)
$h = 5$ years	-1.7470 (0.3291)	-0.9720 (0.3199)	-0.2378 (0.3283)
2-Factor VAR(1)	-1.1444 (0.4102)	-0.0033 (0.3953)	1.1279 (0.3970)
2-Factor VAR(2)	-1.6686 (0.4635)	-0.2112 (0.4373)	1.2060 (0.4267)

□ Which "directions" should we follow to improve the empirical performances of a given model ?

i) Adding **new factors** (latent and/or observable) or **sources of non-linearities** (stochastic volatility, jumps, switching of regimes) able to explain the strong persistence in yields [see Dai, Singleton and Yang (2007, RFS), Monfort and Pegoraro (2007, JFEC) and Gourieroux, Monfort, Pegoraro and Renne (2012)].

ii) Estimating model parameters in a way coherent with interest rates persistence [see Jardet, Monfort and Pegoraro (2012, JBF)].

5.2 Alternative Estimation Procedures for Gaussian $ATSMs$

5.2.1 MLE through the "Inversion Procedure"

□ Let us consider a Gaussian VAR(1) Factor-Based term structure model in which the latent factor (x_t) is K-dimensional. Let us consider, at date t, K yields (among the M in the data base) that we organize in the vector $R_t^K = [R(t, t + h_1), \ldots, R(t, t + h_K)]'$.

□ Now, the affine relation between this vector of yields and the factor x_t can be written in the following way:

$$R_t^K = \mathcal{C}_K \, x_t + \mathcal{D}_K \,, \;\; \mathcal{C}_K = \mathcal{C}_K(\theta) \,, \; \mathcal{D}_K = \mathcal{D}_K(\theta) \,, \; \theta = (\theta^{\mathbb{P}}, \theta^{\mathbb{Q}})$$

$$\text{where } \mathcal{C}_K = \begin{bmatrix} -\dfrac{c_{1,h_1}}{h_1} & \cdots & -\dfrac{c_{K,h_1}}{h_1} \\ \vdots & \ddots & \vdots \\ -\dfrac{c_{1,h_K}}{h_K} & \cdots & -\dfrac{c_{K,h_K}}{h_K} \end{bmatrix} , \;\; \text{and} \;\; \mathcal{D}_K = \begin{bmatrix} -\dfrac{d_{h_1}}{h_1} \\ \vdots \\ -\dfrac{d_{h_K}}{h_K} \end{bmatrix}$$

- it is a linear system of K equations in K unknowns (the scalar variables in x_t).

- Given the observed yields $R_K(t)$, we can easily solve for x_t and write:

$$x_t = \mathcal{C}_K^{-1}\left[R_t^K - \mathcal{D}_K\right].$$

- Given that the conditional p.d.f. $f(x_{t+1}\,|\,x_t)$ is known (it is the p.d.f. of K-dimensional conditional Gaussian process with conditional mean $E_t[x_{t+1}] = \nu + \Phi x_t$ and conditional variance $V_t[x_{t+1}] = \Omega$), we have that (exercise):

$$f(R_{t+1}^K\,|\,R_t^K) = \frac{1}{\det(\mathcal{C}_K)}\,f(x_{t+1}\,|\,x_t)\,.$$

☐ Given the set of observations at times $\{1, \ldots, T\}$, the log-Likelihood function is given by:

$$\mathcal{L}(\theta) = \sum_{t=1}^{T} \log f(R_t^K \,|\, R_{t-1}^K)\,,$$

assuming $f(R_1^K \,|\, R_0^K) = f(R_0^K)\,,$ i.e., the marginal density.

☐ The Maximum Likelihood Estimator (MLE) is : $\theta^{ML} = ArgMax_\theta \, \mathcal{L}(\theta)$ [**Pearson and Sun (1994)**].

☐ Here we have assumed that K **yields are observed without errors** $\rightarrow$ in reality they are reconstructed by interpolation/fitting techniques.

☐ Moreover, we have to decide, in our data base, **which yields (residual maturities)** are observed without errors.

□ In small sample (quarterly observations), different results (estimates) are obtained when different maturities are used.

□ Chen and Scott (1993) tackle this problem assuming that additional yields are observed with errors.

□ Let us assume that $M-K$ additional yields are measured with errors, besides the K observed without errors:

$$R_t^{M-K} = \mathcal{C}_{M-K}\, x_t + \mathcal{D}_{M-K} + \eta_t\,,$$

$$R_t^{M-K} = [R(t,t+h_{K+1}),\ldots,R(t,t+h_M)]'\,,$$

where the conditional distribution of the measurement errors (η_t) is known and given by $h(\eta_t \,|\, \eta_{t-1})$. Moreover, $\eta_t \perp R_t^{M-K}$, $\eta_t \perp R_t^{K}$.

☐ Then, it is possible to prove that (exercise) the Log-Likelihood function is given by:

$$\mathcal{L}^*(\theta) = \mathcal{L}(\theta) + \sum_{t=1}^{T} \log h(\eta_t \,|\, \eta_{t-1}) \,,$$

assuming $h(\eta_1 \,|\, \eta_0) = h(\eta_1)$, i.e., the marginal density.

☐ The Maximum Likelihood Estimator (MLE) is : $\theta^{ML} = ArgMax_{\theta} \, \mathcal{L}^*(\theta)$ [**Chen and Scott (1993)**].

☐ We can not apply the two above mentioned estimation procedures if $p > 1$, given that the inversion of the yield-to-maturity formula (using observed yields) provides at two subsequent dates two different values for the same scalar factor.

5.2.2 MLE through Kalman Filter recursions

□ If we assume that all yields are observed with errors, the Gaussian VAR(1)-based ATSM can be written in a State Space form:

$$R_t^M = \mathcal{C}_M(\theta)\, x_t + \mathcal{D}_M(\theta) + \eta_t\,, \quad \eta_t \sim IIN(0,Q)\,, \quad (Measurement\ \ Equation)\,,$$

$$x_t = \nu + \Phi\, x_{t-1} + \varepsilon_t\,, \quad \varepsilon_t \sim IIN(0,R)\,, \quad (Transition\ \ Equation)\,,$$

$$\eta_t \perp \varepsilon_t\,.$$

□ R_t^M is the $(M \times 1)$ vector of **observed** variables (observed yields);

□ x_t is the $(K \times 1)$ vector of **unobserved** factors (latent factors).

□ Unknown vector of parameters we have to estimate is $\theta = (\theta_{\mathbb{P}}, \theta_{\mathbb{Q}})'$.

- Statistical Inference:
  - estimate θ by MLE;
  - estimate the unobserved latent factors x_t (filtering);
- The **Kalman Filter (KF)** is a recursive algorithm consisting of a **prediction** and **update** step.
- It is a Linear Gaussian State Space Model and, thus, parameters can be efficiently estimated by Maximum Likelihood with the (exact !) **Likelihood function calculated** by the Kalman Filter. KF is optimal in MSE sense.
- Notation: $\underline{R_t^M} = (R_t^M, R_{t-1}^M, \ldots, R_1^M)$ (date-t information set);

☐ Some definitions:

– Let $x_{t|t-1} := E\left[x_t \,|\, \underline{R^M_{t-1}}\right] = \nu + \Phi\, x_{t-1|t-1}$ be the best linear predictor of x_t given the history of observable until $t-1$;

– Let $R^M_{t|t-1} := E\left[R^M_t \,|\, \underline{R^M_{t-1}}\right] = \mathcal{C}_M\, x_{t|t-1} + \mathcal{D}_M$ be the best linear predictor of R^M_t given $\underline{R^M_{t-1}}$;

– Let $x_{t|t} := E\left[x_t \,|\, \underline{R^M_t}\right]$ be the best linear predictor of x_t given the history of observable until t;

☐ What is the purpose of the Kalman Filter ?

- Let us assume we have $x_{t|t-1}$ and $R^M_{t|t-1}$.

- We observe a new R^M_t.

- We need to obtain $x_{t|t}$.

- Note that $x_{t+1|t} = \nu + \Phi\, x_{t|t}$ and $R^M_{t+1|t} = \mathcal{C}_M\, x_{t+1|t} + \mathcal{D}_M$, so we can go back to the first step and wait for R^M_{t+1}.

- So, the key question is how to obtain $x_{t|t}$ from $x_{t|t-1}$ and R^M_t.

- ☐ Let us assume we adopt the following equation to get $x_{t|t}$ from $x_{t|t-1}$ and R_t^M:

$$x_{t|t} = x_{t|t-1} + \mathcal{K}_t \left(R_t^M - R_{t|t-1}^M\right) = x_{t|t-1} + \mathcal{K}_t \left(R_t^M - \mathcal{C}_M \, x_{t|t-1} - \mathcal{D}_M\right),$$

- ☐ This is formula has a probabilistic justification (to follow)

- ☐ What is $\mathcal{K}_t$? It is the **Kalman filter gain** and it measures how much we update $x_{t|t-1}$ as a function of the error we make in predicting R_t^M.

- ☐ How do we find optimal $\mathcal{K}_t$? The KF is about how to build $\mathcal{K}_t$ such that we optimally update $x_{t|t}$ from $x_{t|t-1}$ and R_t^M.

☐ Some additional definitions:

– Let $\Sigma_{t|t-1} := E\left[(x_t - x_{t|t-1})\,(x_t - x_{t|t-1})' \,|\, \underline{R^M_{t-1}}\right]$ be the predicting error variance-covariance matrix of x_t given the history of observable until $t-1$.

– Let $\Omega_{t|t-1} := E\left[(R^M_t - R^M_{t|t-1})\,(R^M_t - R^M_{t|t-1})' \,|\, \underline{R^M_{t-1}}\right]$ be the predicting error variance covariance matrix of R^M_t given the history of observable until $t-1$.

– Let $\Sigma_{t|t} := E\left[(x_t - x_{t|t})\,(x_t - x_{t|t})' \,|\, \underline{R^M_t}\right]$ be the predicting error variance covariance matrix of x_t given the history of observable until t.

☐ Finding the optimal $\mathcal{K}_t$:

– We search for $\mathcal{K}_t$ such that $\rightarrow$ Min $\Sigma_{t|t}$.

– It can be shown that, if it is the case:

$$\mathcal{K}_t = \Sigma_{t|t-1}\,\mathcal{C}_M\,\left(\mathcal{C}'_M \Sigma_{t|t-1} \mathcal{C}_M + Q\right)^{-1},$$

– we will provide some intuition later.

☐ Given $\Sigma_{t|t-1}$, R_t^M and $x_{t|t-1}$, we can now set the Kalman Filter algorithm.

☐ Given $\Sigma_{t|t-1}$:

$$\Omega_{t|t-1} = \mathcal{C}_M' \, \Sigma_{t|t-1} \, \mathcal{C}_M + Q$$

and

$$E\left[(R_t^M - R_{t|t-1}^M) \, (x_t - x_{t|t-1})' \, | \, \underline{R_{t-1}^M} \right] = \mathcal{C}_M' \, \Sigma_{t|t-1}$$

☐ Given $\Sigma_{t|t-1}$, we can also compute:

$$\mathcal{K}_t = \Sigma_{t|t-1} \, \mathcal{C}_M \, \left(\mathcal{C}_M' \Sigma_{t|t-1} \mathcal{C}_M + Q \right)^{-1} = \Sigma_{t|t-1} \, \mathcal{C}_M \, \Omega_{t|t-1}^{-1}$$

☐ Given $x_{t|t-1}$:

$$R_{t|t-1}^M = \mathcal{C}_M \, x_{t|t-1} + \mathcal{D}_M$$

□ Once we have $\Sigma_{t|t-1}$, R_t^M, $x_{t|t-1}$ and $\mathcal{K}_t$, we compute:

$$x_{t|t} = x_{t|t-1} + \mathcal{K}_t\,(R_t^M - R_{t|t-1}^M) = x_{t|t-1} + \mathcal{K}_t\,(R_t^M - \mathcal{C}_M\, x_{t|t-1} - \mathcal{D}_M)\,,$$

□ and

$$\Sigma_{t|t} = E\left[(x_t - x_{t|t})\,(x_t - x_{t|t})' \,|\, \underline{R_t^M}\right] = \Sigma_{t|t-1} - \mathcal{K}_t\, \mathcal{C}_M\, \Sigma_{t|t-1}$$

where we exploit the fact that $x_t - x_{t|t} = x_t - x_{t|t-1} - \mathcal{K}_t\,(R_t^M - \mathcal{C}_M\, x_{t|t-1} - \mathcal{D}_M)$.

□ Given $\Sigma_{t|t}$, we compute:

$$\Sigma_{t+1|t} = \Phi\, \Sigma_{t|t}\, \Phi' + R$$

- ☐ Given $x_{t|t}$, we can compute:

$$x_{t+1|t} = \Phi\, x_{t|t}\,,$$

$$R^M_{t+1|t} = \mathcal{C}_M\, x_{t+1|t} + \mathcal{D}_M$$

- ☐ Therefore, from $x_{t|t-1}$, $\Sigma_{t|t-1}$ and R^M_t we compute $x_{t|t}$ and $\Sigma_{t|t}$.

- ☐ We also compute $R^M_{t|t-1}$ and $\Omega_{t|t-1}$. Why ?

- To calculate the likelihood function of $\underline{R^M_T} = (R^M_T, R^M_{T-1}, \ldots, R^M_1)$ (to follow).

- This estimation methodology is adapted also to the case $p > 1$ (companion form).

□ The Kalman Filter Algorithm: A Review

- We start with $x_{t|t-1}$, $\Sigma_{t|t-1}$ and we observe R_t^M. Then:

$$\Omega_{t|t-1} = \mathcal{C}_M' \, \Sigma_{t|t-1} \, \mathcal{C}_M + Q$$

$$R_{t|t-1}^M = \mathcal{C}_M \, x_{t|t-1} + \mathcal{D}_M \, .$$

- **Filtering Step:**

$$\mathcal{K}_t = \Sigma_{t|t-1} \, \mathcal{C}_M \, \left(\mathcal{C}_M' \Sigma_{t|t-1} \mathcal{C}_M + Q\right)^{-1} = \Sigma_{t|t-1} \, \mathcal{C}_M \, \Omega_{t|t-1}^{-1}$$

$$\Sigma_{t|t} = \Sigma_{t|t-1} - \mathcal{K}_t \, \mathcal{C}_M \, \Sigma_{t|t-1}$$

$$x_{t|t} = x_{t|t-1} + \mathcal{K}_t \left(R_t^M - \mathcal{C}_M \, x_{t|t-1} - \mathcal{D}_M\right),$$

- **Prediction Step:**

$$x_{t+1|t} = \nu + \Phi \, x_{t|t} \, ,$$

$$\Sigma_{t+1|t} = \Phi \, \Sigma_{t|t} \, \Phi' + R \, .$$

□ Some intuition about the optimal $\mathcal{K}_t = \Sigma_{t|t-1} \, \mathcal{C}_M \, \left(\mathcal{C}_M' \Sigma_{t|t-1} \mathcal{C}_M + Q\right)^{-1}$

- As we have seen before, we can write $\mathcal{K}_t = \Sigma_{t|t-1} \, \mathcal{C}_M \, \Omega_{t|t-1}^{-1}$

- If we have made a big mistake in forecasting $x_{t|t-1}$ using the past information (i.e. $\Sigma_{t|t-1}$ large) we give a lot of weight to the new information ($\mathcal{K}_t$ large).

- If the new information is noise (Q large) we give a lot of weight to the old prediction ($\mathcal{K}_t$ small).

☐ An important step in the Kalman filter is to set the initial conditions

☐ Initial conditions:

1. $x_{1|0}$

2. $\Sigma_{1|0}$

☐ How do we fix them ? Since we consider only stable system (stationary VAR dynamics) the standard approach is to set $x_{1|0} = E(x_t)$ (marginal mean) and $\Sigma_{1|0} = V(x_t)$ (marginal variance).

□ Writing the Log-Likelihood Function

- We want to write (to calculate) the likelihood function of $\underline{R_T^M} = (R_T^M, R_{T-1}^M, \ldots, R_1^M)$:

$$\begin{aligned}\mathcal{L}(\theta) &= \ln f(R_T^M, R_{T-1}^M, \ldots, R_1^M \mid \theta) = \sum_{t=1}^{T} \ln f(R_t^M \mid \underline{R_{t-1}^M}; \theta) \\ &= -\sum_{t=1}^{T} \left[\frac{N}{2} \ln 2\pi + \frac{1}{2} \ln |\Omega_{t|t-1}| + \frac{1}{2} \sum_{t=1}^{T} \upsilon_t \, \Omega_{t|t-1}^{-1} \, \upsilon_t \right]\end{aligned}$$

- where:

$$\begin{aligned}\upsilon_t &= R_t^M - R_{t|t-1}^M = R_t^M - \mathcal{C}_M \, x_{t|t-1} - \mathcal{D}_M \\ \Omega_{t|t-1} &= \mathcal{C}_M' \, \Sigma_{t|t-1} \, \mathcal{C}_M + Q \, .\end{aligned}$$

- Remember: KF calculates $\mathcal{L}(\theta)$ while its maximization is obtained through a numerical algorithm (BFGS, BHHH, ...) and provides the MLE $\widehat{\theta}_T$.

5.2.3 The Adrian, Crump and Moench (2012, JFE) Approach

□ A Gaussian ATSM with VAR(1) K-dimensional factor x_t:

$$x_{t+1} \;=\; \nu + \Phi x_t + \Sigma \varepsilon_{t+1}\,,$$

- and with exponential-affine SDF ($\Gamma_t = \gamma_o + \gamma\, x_t$):

$$M_{t,t+1} = \exp\left[-r_t + \Gamma_t' \varepsilon_{t+1} - \frac{1}{2}\Gamma_t'\Gamma_t\right]\,,$$

- has a one-period geometric bond return following:

$$\rho(t+1,T) \;=\; r_t - \tfrac{1}{2}\,\omega(t+1,T)'\omega(t+1,T) \,+\, \omega(t+1,T)'\Gamma_t - \omega(t+1,T)'\,\varepsilon_{t+1}\,,$$

where $\omega(t+1,T) = -(\Sigma' C_{T-t-1})$ is an K-dimensional vector.

☐ Adrian, Crump and Moench (2012) exploit the fact that the one-period excess bond return:

$$rx_{t+1}^{(h-1)} := \log B(t+1, t+h) - \log B(t, t+h) - r_t$$

- is conditionally Gaussian and linear in (γ_o, γ)
- in order to make their estimation computationally fast, even for a large number of factors.
- Let us present their approach in the following slides.

☐ The Adrian, Crump and Moench (2012) approach:

- Given the VAR(1) factor x_t, the SDF $M_{t,t+1} = \exp\left[-r_t - \Gamma_t' \varepsilon_{t+1} - \frac{1}{2}\Gamma_t'\Gamma_t\right]$ and the excess bond return $rx_{t+1}^{(h-1)}$, from $B(t, t+h) = E_t[M_{t,t+1}\, B(t+1, t+h)]$ we find:

$$1 = E_t\left[\exp\left(rx_{t+1}^{(h-1)} - \frac{1}{2}\,\Gamma_t'\Gamma_t - \Gamma_t'\varepsilon_t\right)\right]$$

- Under the assumption that $\{rx_{t+1}^{(h-1)}, \varepsilon_{t+1}\}$ are jointly normally distributed:

$$\begin{aligned} E_t\left[rx_{t+1}^{(h-1)}\right] &= Cov_t\left(rx_{t+1}^{(h-1)}, \varepsilon_{t+1}'\Gamma_t\right) - \frac{1}{2}\,Var_t\left(rx_{t+1}^{(h-1)}\right) \\ &= Cov_t\left(rx_{t+1}^{(h-1)}, \varepsilon_{t+1}'\right)\,(\gamma_o + \gamma\, x_t) - \frac{1}{2}\,Var_t\left(rx_{t+1}^{(h-1)}\right) \\ &= \beta_t^{(h-1)\prime}(\gamma_o + \gamma\, x_t) - \frac{1}{2}\,Var_t\left(rx_{t+1}^{(h-1)}\right)\;, \end{aligned}$$

where $\beta_t^{(h-1)\prime} := Cov_t\left(rx_{t+1}^{(h-1)}, \varepsilon_{t+1}'\right)$.

- Now, we can always decompose $rx_{t+1}^{(h-1)}$ into an expected component and an unexpected one:

$$rx_{t+1}^{(h-1)} = E_t\left[rx_{t+1}^{(h-1)}\right] + \left(rx_{t+1}^{(h-1)} - E_t\left[rx_{t+1}^{(h-1)}\right]\right)$$

- and the latter can be further decomposed into a component conditionally correlated with ε_{t+1} and another that is conditionally orthogonal to ε_{t+1}:

$$rx_{t+1}^{(h-1)} - E_t\left[rx_{t+1}^{(h-1)}\right] = \beta_t^{(h-1)\prime}\varepsilon_{t+1} + e_{t+1}^{(h-1)}$$

where the return pricing errors $e_{t+1}^{(h-1)}$ are conditionally i.i.d. with variance σ^2.

- Given that $Var_t\left(rx_{t+1}^{(h-1)}\right) = \beta_t^{(h-1)\prime}\beta_t^{(h-1)} + \sigma^2$ can thus write:

$$rx_{t+1}^{(h-1)} = \beta_t^{(h-1)\prime}(\gamma_o + \gamma\, x_t) - \frac{1}{2}\left(\beta_t^{(h-1)\prime}\beta_t^{(h-1)} + \sigma^2\right) + \beta_t^{(h-1)\prime}\varepsilon_{t+1} + e_{t+1}^{(h-1)}$$

- In their baseline model they assume x_t is observable and made of a linear combination of yields, such as principal components. They estimate model parameters using holding period returns based on the same set of yields. Per construction, this implies $\beta_t = \beta$.

- Stacking the system across $h \in \{2, \ldots, H\}$ maturities and $t \in \{1, \ldots, T-1\}$ time periods, we rewrite it as:

$$rx = \beta' \left(\gamma_o \, \mathbf{1}'_{T-1} + \gamma \, X_{-}\right) - \frac{1}{2} \left(B^* vec(I_K) + \sigma^2 \, \mathbf{1}_{H-1}\right) \mathbf{1}'_{T-1} + \beta' \mathcal{E} + E$$

- where rx is a $(H-1, T-1)$ matrix of excess returns,

  $\beta = \left[\beta^{(2)} \,|\, \beta^{(3)} \,|\, \ldots \,|\, \beta^{(H)}\right]$ is a $(K, H-1)$ matrix of factor loadings

- $\mathbf{1}_\ell$ is an ℓ-dimensional vector of ones;

- $X_- = [X_1, \ldots, X_{T-1}]$ is a $(K, T-1)$ matrix of lagged pricing factors;

- $B^* = \left[vec(\beta^{(2)}\beta^{(2)\prime}) \,|\, vec(\beta^{(3)}\beta^{(3)\prime}) \,|\, \ldots \,|\, vec(\beta^{(H)}\beta^{(H)\prime})\right]'$ is a $(H-1, K^2)$ matrix;

- $\mathcal{E}$ is a $(K, T-1)$ matrix of normalized residuals $\{\widehat{\varepsilon_t}\}$,

  and E is a $(H-1, T-1)$ matrix of $\{\widehat{e_t^{(h)}}\}$ residuals.

□ Estimation Procedure:

1^{st}) We estimate $\theta_{\mathbb{P}} = (\nu, \Phi, \Sigma)$ by OLS, and then we build $\mathcal{E}$ from $\{\widehat{\varepsilon}_t\}$

2^{nd}) Run the following regression:

$$\begin{aligned} rx &= \mathbf{a}\,\mathbf{1}'_{T-1} + \beta'\mathcal{E} + \mathbf{c}X_{-} + E \\ \mathbf{a} &= \beta'\gamma_o - \frac{1}{2}\left(B^{*}vec(I_K) + \sigma^2\,\mathbf{1}_{H-1}\right) \\ \mathbf{c} &= \beta'\gamma \end{aligned}$$

in order to obtain:

$$\left[\widehat{\mathbf{a}}\,|\,\widehat{\beta}'\,|\,\widehat{\mathbf{c}}\right] = rx\,\widetilde{Z}'(\widetilde{Z}\,\widetilde{Z}')^{-1} \;\; \texttt{where} \;\; \widetilde{Z} = \left[\mathbf{1}_{T-1}\,|\,\widehat{\mathcal{E}}'\,|\,X'_{-}\right]',$$

we collect the associated residuals in $\widehat{E}$, in order to calculate $\widehat{\sigma}^2 = Tr(\widehat{E}\,\widehat{E}')/(H-1)(T-1)$, and we calculate $\widehat{B}^{*}$ from $\widehat{\beta}'$.

3^{rd}) Given the previously estimated parameters, we easily find:

$$\widehat{\gamma} = \left(\widehat{\beta}\widehat{\beta}'\right)^{-1} \widehat{\beta}\,\widehat{\mathbf{c}}$$

and from $\widehat{\beta}'\gamma_o = \widehat{\mathbf{a}} + \frac{1}{2}\left(\widehat{B}^* vec(I_K) + \widehat{\sigma}^2\,\mathbf{1}_{H-1}\right)$ we retrieve:

$$\widehat{\gamma}_o = \left(\widehat{\beta}\widehat{\beta}'\right)^{-1} \widehat{\beta}\left[\widehat{\mathbf{a}} + \frac{1}{2}\left(\widehat{B}^* vec(I_K) + \widehat{\sigma}^2\,\mathbf{1}_{H-1}\right)\right].$$

- From the estimated model parameters, we can generate e zero-coupon yield curve using the recursive equations (C_h, D_h) and an identification between $\beta^{(h)}$ and C_h is easily found (exercise).

- Possible generalization to $p > 1$.

5.3 Ang and Piazzesi (2003, JME)

5.3.1 Purpose of the paper

- The authors describe the (particular!) joint dynamics of bond yields and macroeconomic variables in a discrete-time Gaussian Vector Autoregression setting, where "causality" and no-arbitrage restrictions are imposed in order to guarantee the **theoretical** and **empirical tractability** of the model.

- Using an affine discrete-time term structure model with inflation and economic growth factors, along with latent variables, they investigate how macro variables affect bond prices and the dynamics of the yield curve.

5.3.2 Main results

- They find that:
  - the **forecasting performance** of a Gaussian VAR **improves** when **no-arbitrage restrictions are imposed**
  - and that (no-arbitrage) models **with macro factors forecast better** than models with only unobservable factors.
- Variance decompositions show that **macro factors explain up to** 85% of the variation in bond yields (over short and middle maturities).

- **Macro factors** primarily explain movements at the **short end and middle** of the yield curve.

- **Unobservable factors** still account for most of the movement at the **long end** of the yield curve.

- They observe (monthly) yields of 1, 3, 12, 36 and 60 months to maturity (1, 12 and 60 observed without errors) from 1952:06 to 2000:12.

- Macro-variables are observed from 1952:01 to 2000:12. These variables are divided in two groups.

☐ The first group consists of various **inflation measures** which are based on the CPI (consumer price index), the PPI (producer price index) of finished goods, and spot market commodity prices ($PCOM$).

☐ The second group contains variables that capture **real activity**: the index of Help Wanted Advertising in Newspapers ($HELP$), unemployment (UE), the growth rate of employment ($EMPLOY$) and the growth rate of industrial production (IP).

- This list of variables includes most variables that have been used in monthly VARs in the macro literature. Among these variables, $PCOM$ and $HELP$ are traditionally thought of as leading indicators of inflation and real activity, respectively.

- All growth rates (including inflation) are measured as the difference in logs of the index at time t and $(t-12)$; t in months.

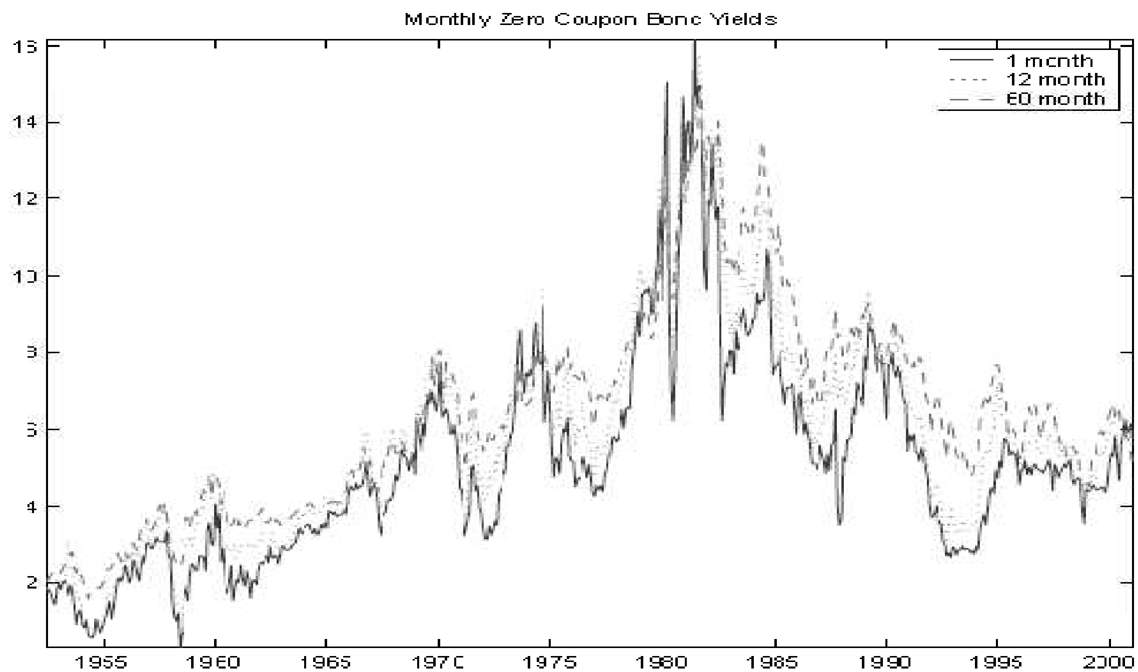

Monthly Zero Coupon Bond Yields
1 month
12 month
60 month
16
14
12
10
8
6
4
2
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000

Table 1
Summary statistics of data

	Central moments				Autocorrelations		
	Mean	Stdev	Skew	Kurt	Lag 1	Lag 2	Lag 3
1 mth	5.1316	2.7399	1.0756	4.6425	0.9716	0.9453	0.9323
3 mth	5.4815	2.8550	1.0704	4.5543	0.9815	0.9606	0.9419
12 mth	5.8849	2.8445	0.8523	3.8856	0.9824	0.9626	0.9457
36 mth	6.2241	2.7643	0.7424	3.5090	0.9875	0.9739	0.9620
60 mth	6.4015	2.7264	0.6838	3.2719	0.9892	0.9782	0.9687
CPI	3.8612	2.8733	1.2709	4.3655	0.9931	0.9847	0.9738
PCOM	0.9425	11.2974	1.0352	5.0273	0.9684	0.9162	0.8600
PPI	3.0590	3.6325	1.4436	4.9218	0.9863	0.9705	0.9521
HELP	66.7517	22.0257	0.1490	1.8665	0.9944	0.9900	0.9830
EMPLOY	1.6594	1.5282	0.4690	3.2534	0.9378	0.8954	0.8410
IP	3.4717	5.3697	0.5578	3.6592	0.9599	0.8889	0.7972
UE	5.7344	1.5650	0.4924	3.2413	0.9906	0.9777	0.9595

The 1, 3, 12, 36 and 60 month yields are annual zero coupon bond yields from the Fama Bliss CRSP bond files. The inflation measures CPI, PCOM and PPI refer to CPI inflation, spot market commodity price inflation, and PPI (Finished Goods) inflation respectively. We calculate the inflation measure at time t using $\log(P_t/P_{t-12})$ where P_t is the inflation index. The real activity measures HELP, EMPLOY, IP and UE refer to the Index of Help Wanted Advertising in Newspapers, the growth rate of employment, the growth rate in industrial production and the unemployment rate respectively. The growth rate in employment and industrial production are calculated using $\log(I_t/I_{t-12})$ where I_t is the employment or industrial production index. For the macro variables, the sample period is 1952:01 to 2000:12. For the bond yields, the sample period is 1952:06 to 2000:12.

5.3.3 Setup

Historical Factor Dynamics

- The multivariate FACTOR (the information used by the investor to price bonds) is denoted:

$$
\begin{aligned}
X_t &= (X_t^{o\prime}, X_t^{u\prime})', \text{ where} \\
X_t^{o\prime} &= (f_t^{o\prime}, {f_{t-1}^{o}}', \ldots, {f_{t-p+1}^{o}}')', \text{ the MACRO factors}, \\
X_t^{u\prime} &= f_t^{u}, \text{ the LATENT factors}.
\end{aligned}
$$

- f_t^o denotes a bivariate process of macro factors following a Gaussian VAR(p) process with $p = 12$ (monthly observations):

$$
\begin{aligned}
f_t^o &= \rho_1 f_{t-1}^o + \ldots + \rho_{12} f_{t-12}^o + \Omega u_t^o, u_t^o \sim IIN(0, I), \\
\Omega &= (2 \times 2) \text{ lower triangular}, (\rho_1, \ldots, \rho_{12}) \ (2 \times 2) \text{ full AR matrices}.
\end{aligned}
$$

□ f_t^u denotes a trivariate process of latent factors following a Gaussian VAR(1) process:

$$\begin{aligned} f_t^u &= \rho f_{t-1}^u + u_t^u \,, u_t^u \sim IIN(0, I)\,, \\ &\rho \text{ is } (3 \times 3) \text{ lower triangular AR matrix}. \end{aligned}$$

□ $u_t^u \perp u_t^o$

□ If we define $F_t = (f_t^{o\prime}, f_t^{u\prime})'$ (5-dimensional vector) we can represent the joint dynamics of the macro and latent factors in the following way:

$$\begin{aligned} F_t &= \Phi_1 F_{t-1} + \ldots + \Phi_{12} F_{t-12} + \theta u_t \,, u_t = (u_t^{o\prime}, u_t^{u\prime})' \sim IIN(0, I)\,, \\ &\text{the coefficients of } (\Phi_2, \ldots, \Phi_{12}) \text{ corresponding to } f_t^u \\ &\text{are set equal to zero. More precisely} \rightarrow \end{aligned}$$

□ The AR matrices are specified in the following way:

$$\Phi_1 = \begin{bmatrix} \Phi^o_{1,11} & 0_{2\times 3} \\ 0_{3\times 2} & \Phi^u_{1,22} \end{bmatrix}, \;\; \Phi_j = \begin{bmatrix} \Phi^o_{j,11} & 0_{2\times 3} \\ 0_{3\times 2} & 0_{2\times 3} \end{bmatrix}, \;\; \forall j \in \{2, \ldots, 12\}.$$

□ We observe that ALL autoregressive matrices are block-diagonal. The authors assume "lagged independence" between macro and latent factors : f_t^o does not Granger-cause f_t^u (and vice versa).

□ Lower-right corners of Φ_j, $\forall j \in \{2, \ldots, 12\}$, are equal to zero ($\Phi^u_{j,22} = 0$) $\rightarrow$ because of the assumption $f_t^u \sim VAR(1)$.

□ the matrix θ is such that:

$$\theta = \left[\begin{array}{ll} \theta_1^o & 0_{2\times 3} \\ 0_{3\times 2} & \theta_2^u \end{array} \right], \text{ where } \theta_1^o = I_{3\times 3}\,, \;\; \theta_1^u = \Omega \text{ lower triangular}\,.$$

□ The (historical) dynamics of $X_t = (X_t^{o\prime}, X_t^{u\prime})'$ is:

$$X_t \;\; = \;\; \mu + \Phi X_{t-1} + \Sigma \varepsilon_t\,, \varepsilon_t = (u_t^{o\prime}, 0, \ldots, 0, u_t^{u\prime})'\,.$$

□ What $f_t^o = (f_t^{o,1}, f_t^{o,2})'$ is ?

- $f_t^{o,1} =$ 1^{st} Principal Component (71% of explained variance) extracted from an "inflation group" variables, with "inflation group" variables = CPI inflation, PCOM (spot mkt commodity price inflation, PPI inflation.)

- $f_t^{o,2} =$ 1^{st} Principal Component (52% of explained variance) extracted from a "real activity group" variables, with "real activity group" variables = HELP, Unemployment, employment growth rate, industrial prod growth rate.)

Table 2
Principal component analysis

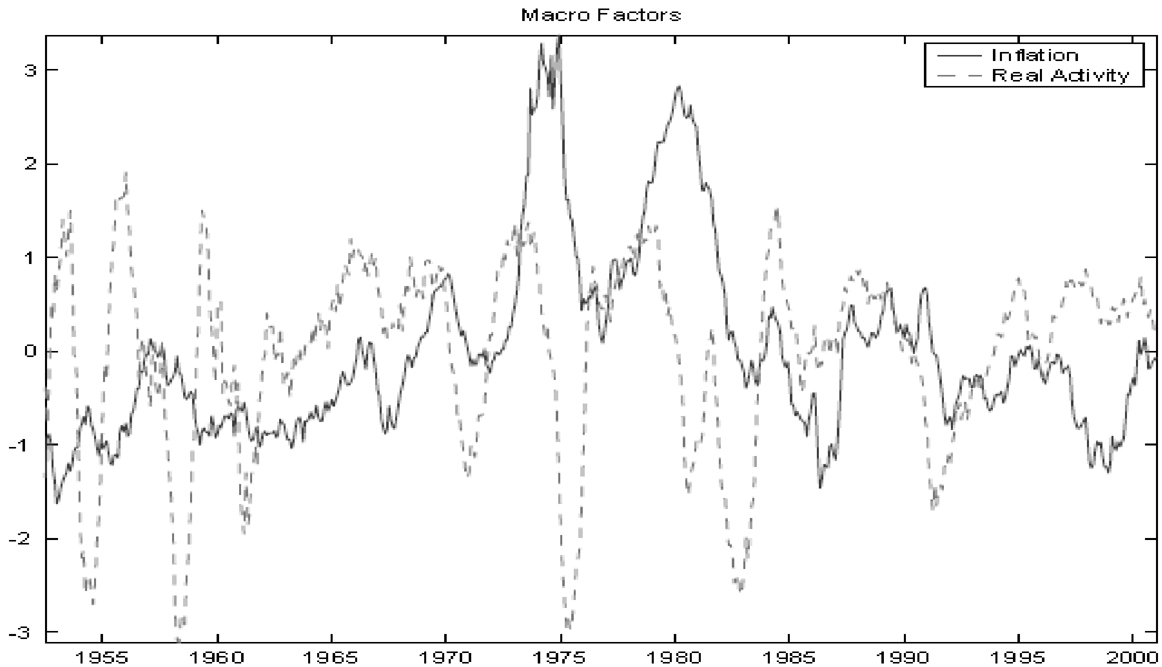
	Principal components: inflation		
	1st	2nd	3rd
CPI	0.6343	0.3674	0.6802
PCOM	0.4031	0.9080	0.1145
PPI	0.6597	0.2015	0.7240
% variance explained	0.7143	0.9775	1.0000

	Principal components: real activity			
	1st	2nd	3rd	4th
HELP	0.3204	0.7365	0.5300	0.2719
UE	0.3597	0.6283	0.6871	0.0612
EMPLOY	0.6330	0.1648	0.2444	0.7158
IP	0.6060	0.1886	0.4327	0.6403
% variance explained	0.5202	0.7946	0.9518	1.0000

We take the three (four) macro variables representing inflation (real activity) and normalize them to zero mean and unit variance. For each group i, the normalized data Z_t^i follows the 1 factor model:

$$Z_t^i = Cf_t^{o,i} + \varepsilon_t^i,$$

where C is the factor loading vector, $E(f_t^{o,i}) = 0$, $\text{cov}(f_t^{o,i}) = I$, $E(\varepsilon_t^i) = 0$, and $\text{cov}(\varepsilon_t^i) = \Gamma$, where Γ is a diagonal matrix. The columns titled "principal components" list the principal components corresponding to the first to smallest eigenvalue. The % variance explained for the nth principal component gives the cumulative proportion of the variance explained by the first up to the nth eigenvalue. IP refers to the growth in industrial production, CPI to CPI inflation, PCOM to commodity price inflation and PPI to PPI inflation, HELP refers to the Index of Help Wanted Advertising in Newspapers, UE to the unemployment rate, EMPLOY to the growth in employment. The sample period is 1952:01 to 2000:12


Macro Factors
Inflation
Real Activity
3
2
1
0
-1
-2
-3
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000

☐ What $f_t^u = (f_t^{u,1}, f_t^{u,2}, f_t^{u,3})'$ is ?

- Extracted from 3 yields assumed to be perfectly observed (inverting the affine yield-to-maturity formula): $R(t, t+1)$, $R(t, t+12)$ and $R(t, t+60)$.

- They (classically $\rightarrow$ Litterman and Scheinkman (1991)) act as a:

  - LEVEL FACTOR $= f_t^{u,1} \rightarrow$ correlation of 0.92 with $[R(t,t+1)+R(t,t+12)+R(t,t+60)]/3$ (called "level transformation");

  - SLOPE FACTOR $= f_t^{u,2} \rightarrow$ correlation of 0.58 with $[R(t,t+60)-R(t,t+1)]$;

  - CURVATURE FACTOR $= f_t^{u,3} \rightarrow$ correlation of 0.77 with $[R(t,t+1)-2R(t,t+12)+R(t,t+60)]$.

Short Rate Historical Dynamics

□ The authors assume that:

$$r_t = \delta_0 + \delta'_{11} X_t^o + \delta'_{12} X_t^u = \delta_0 + \delta'_1 X_t\,, \text{ with } X_t^o \perp X_t^u\,.$$

□ If δ_1 is constrained to depend on just contemporaneous values, then we have the *classical Taylor rule* given that $v_t = \delta'_{12} X_t^u$ can be interpreted as an "orthogonal (monetary policy) shock". It is named "Macro Model".

□ If δ_1 is unconstrained: introducing lagged values, they hope to catch relevant information to forecast inflation or output. They interpret that specification as a *forward-looking version of the Taylor rule* ("Macro Lag Model").

No-Arbitrage, Stochastic Discount Factors and Pricing Formulas

- To develop the affine term structure model, they use the assumption of no-arbitrage (Harrison and Kreps, 1979) to guarantee the existence of a (positive and not unique in general) Stochastic Discount Factor $M_{t,t+1}$ (or pricing kernel) such that the price of any asset V_t that does not pay any dividend at $t+1$ satisfies:

$$V_t = E_t^{\mathbb{P}}[M_{t,t+1}V_{t+1}] \, .$$

- If we consider at t a zero-coupon bond maturing at $t+1$ we have:

$$B(t,t+1) = E_t^{\mathbb{P}}[M_{t,t+1}] = \exp(-r_t) \, .$$

- More generally, for any payoff V_{t+h} at $t+h$, we have :

$$\begin{aligned} V_t &= E_t^{\mathbb{P}}[M_{t,t+h} V_{t+h}] \,, \\ &= E_t^{\mathbb{P}}[M_{t,t+1} \dots M_{t+h-1,t+h} V_{t+h}] \,. \end{aligned}$$

- Now, it is well known from asset pricing theory that, under the absence of arbitrage opportunity, there exist equivalent (to $\mathbb{P}$) probability measures under which asset prices, evaluated with respect to some numeraire N_t, are martingales.

- A numeraire is defined as a non-dividend-paying price process $N = (N_t,\, t \geq 0)$ with $N_0 = 1$. In other words, N is a stochastic process such that, for every $T > t$:

$$N_t = E_t^{\mathbb{P}}[M_{t,T} N_T] \,, \text{ and } E_0^{\mathbb{P}}[M_{0,T} N_T] = 1 \,, \text{ where}$$

$$M_{t,T} = M_{t,t+1} \cdot \ldots \cdot M_{T-1,T} \,.$$

□ The process $N^* = (N_t M_{0,t},\ t \geq 0)$ is therefore a $\mathbb{P}$-martingale with unitary value in $t = 0$, and if $\mathbb{Q}$ is the probability (equivalent to $\mathbb{P}$) defined by the sequence of conditional densities:

$$\frac{d\mathbb{Q}_{t,t+1}}{d\mathbb{P}_{t,t+1}} = \frac{N_{t+1} M_{t,t+1}}{N_t} > 0\,, \quad E_t^{\mathbb{P}}\left[\frac{d\mathbb{Q}_{t,t+1}}{d\mathbb{P}_{t,t+1}}\right] = 1$$

then, a price process V_t is such that V_t / N_t is a $\mathbb{Q}$-martingale:

$$V_t = E_t^{\mathbb{P}}[M_{t,t+1} V_{t+1}] \Longleftrightarrow \frac{N_t}{N_t} V_t = E_t^{\mathbb{P}}\left[\frac{N_{t+1}}{N_{t+1}} M_{t,t+1} V_{t+1}\right]$$

□ thus:

$$\frac{V_t}{N_t} = E_t^{\mathbb{P}}\left[\frac{N_{t+1} M_{t,t+1}}{N_t} \frac{V_{t+1}}{N_{t+1}}\right] = E_t^{\mathbb{Q}}\left[\frac{V_{t+1}}{N_{t+1}}\right]\,.$$

☐ If we consider as numeraire the money-market account $N_t = \exp(r_0 + \ldots + r_{t-1}) = (A_{0,t})^{-1}$, where $A_{0,t} = E_0(M_{0,1}) \cdots E_{t-1}(M_{t-1,t})$, the associated equivalent probability $\mathbb{Q}_{t,t+1}$ has a one-period conditional density, with respect to $\mathbb{P}_{t,t+1}$, given by :

$$\frac{d\mathbb{Q}_{t,t+1}}{d\mathbb{P}_{t,t+1}} = \frac{A_{0,t} M_{t,t+1}}{A_{0,t+1}} = \frac{M_{t,t+1}}{E_t(M_{t,t+1})} .$$

and it is called *Risk-Neutral* probability measure.

☐ This means that the pricing formula $V_t = E_t^{\mathbb{P}}[M_{t,t+1} V_{t+1}]$ can be written:

$$V_t = E_t^{\mathbb{P}} \left[\frac{M_{t,t+1}}{E_t^{\mathbb{P}}[M_{t,t+1}]} E_t^{\mathbb{P}}[M_{t,t+1}] V_{t+1} \right] = E_t^{\mathbb{Q}}[\exp(-r_t) V_{t+1}] ,$$

where r_t is the $(t, t+1)$ short rate, known in t

☐ In a general $(T-t)$-period horizon, the conditional (to I_t) density of the risk-neutral probability $\mathbb{Q}_{t,T}$ with respect to the historical probability $\mathbb{P}_{t,T}$ is given by:

$$\frac{d\mathbb{Q}_{t,T}}{d\mathbb{P}_{t,T}} = \frac{M_{t,t+1} \cdot \ldots \cdot M_{T-1,T}}{E_t(M_{t,t+1}) \cdot \ldots \cdot E_{T-1}(M_{T-1,T})},$$

☐ More generally, for any payoff V_{t+h} at $t+h$, we have :

$$V_t = E_t^{\mathbb{Q}}[\exp(-r_t - \ldots - r_{t+h-1})V_{t+h}],$$

☐ In the case of a ZCB maturing at $t+h$ we have:

$$\begin{aligned} B(t,t+h) &= E_t^{\mathbb{P}}[M_{t,t+h}] = E_t^{\mathbb{P}}[M_{t,t+1}B(t+1,t+h)] \\ &= E_t^{\mathbb{Q}}[\exp(-r_t - \ldots - r_{t+h-1})] = E_t^{\mathbb{Q}}[\exp(-r_t)B(t+1,t+h)], \end{aligned}$$

The Exponential-Affine Stochastic Discount Factor and the Yield Curve

- The authors assume the following exponential-affine (in X_t) Stochastic Discount Factor:

$$M_{t,t+1} = \exp\left[-r_t - \lambda_t' \varepsilon_{t+1} - \frac{1}{2}\lambda_t'\lambda_t\right] = \exp\left[-\delta_0 - \delta_1' X_t - \lambda_t' \varepsilon_{t+1} - \frac{1}{2}\lambda_t'\lambda_t\right],$$

where $\lambda_t = \lambda_0 + \lambda_1 X_t$.

- From $B(t,t+h) = E_t^{\mathbb{P}}[M_{t,t+1} B(t+1,t+h)]$ we obtain:

$$B(t,t+h) = \exp\left[A_h + B_h' X_t\right], \text{ where } A_h \text{ and } B_h \text{ are:}$$

$$A_{h+1} = A_h + B_h'(\mu - \Sigma\lambda_0) + \frac{1}{2} B_h' \Sigma\Sigma' B_h - \delta_0, \text{ where } A_1 = -\delta_0$$

$$B_{h+1}' = B_h'(\Phi - \Sigma\lambda_1) - \delta_1, \text{ where } B_1 = -\delta_1$$

☐ The yield-to-maturity formula (the affine yield curve) is therefore:

$$R(t, t+h) = -\frac{1}{h} \ln[B(t, t+h)] = -\frac{A_h}{h} - \frac{B_h' X_t}{h},$$

☐ Parameters in λ_0 and λ_1 associated to lagged macro-variables are set equal to zero. This means that, they consider:

$$\lambda_t = \lambda_0 + \lambda_1 [f_t^{o\prime}, f_t^{u\prime}]' .$$

☐ In addition, they assume the (5×5) matrix λ_1 block-diagonal.

5.3.4 Estimation Procedure

☐ They estimate three type of models : a) $X_t = f_t^u$ (only latent variables); b) "Macro Model" (no lags); c) "Macro Lag Model";

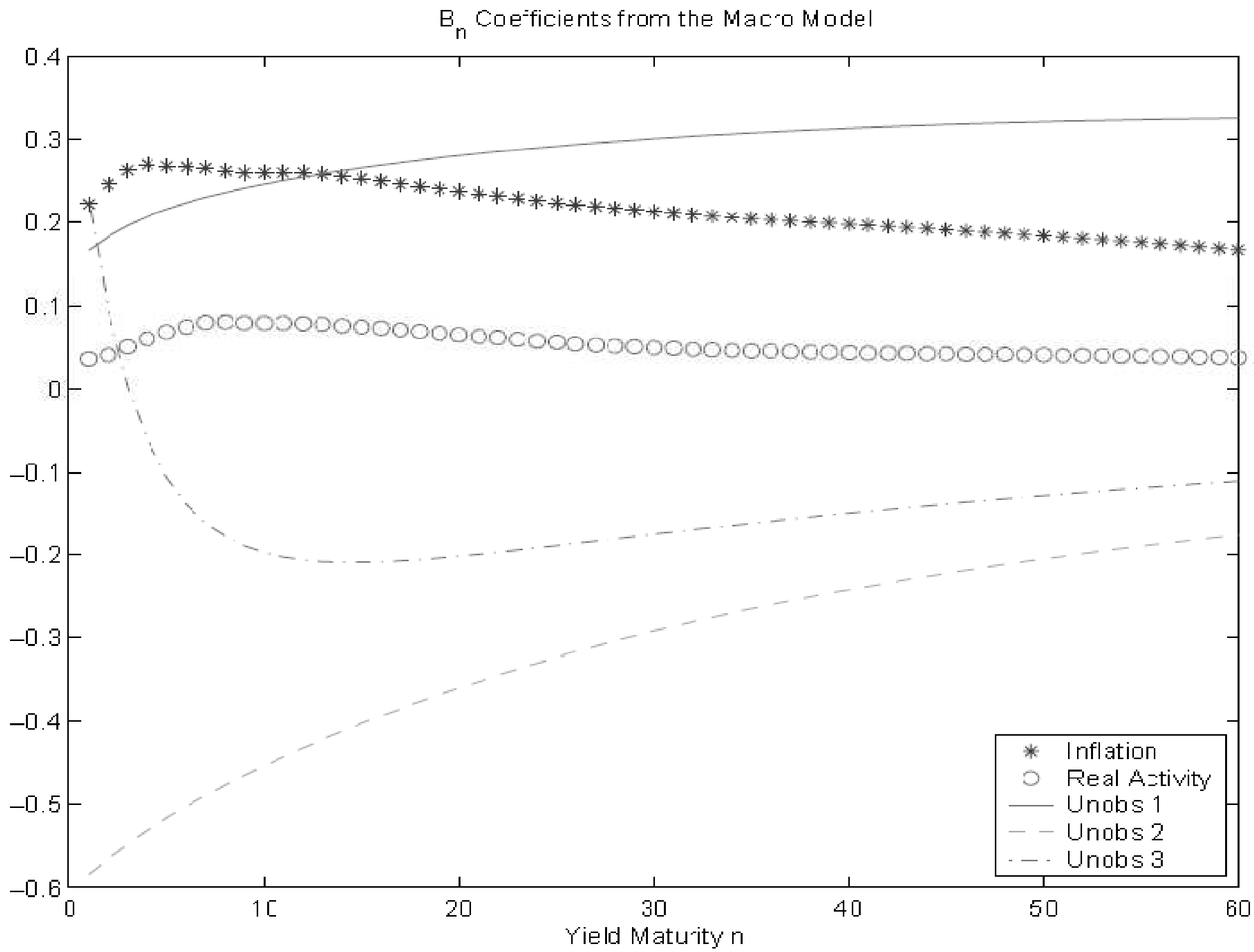
☐ They follow a two-step consistent estimation procedure, which is adapted to forecast with models characterized by several parameters.

☐ They first estimate by OLS the parameters in the VAR(12) dynamics of f_t^o, and the parameters (δ_0, δ_{11}) in the short rate dynamics (exploiting the fact that $X_t^u \perp X_t^o$).

- Then, keeping fixed that parameters to the estimated values, they estimate the remaining parameters, namely $(\lambda_0, \lambda_1, \delta_{12})$, using the Chen-Scott (1993) inversion procedure.

- Let us take a look now to what happens to yield curve factor loadings associated to latent and macro variables.

- In other words: do the three latent factors keep their role of level, slope and curvature ? How macro variables affect the yield curve shape ?

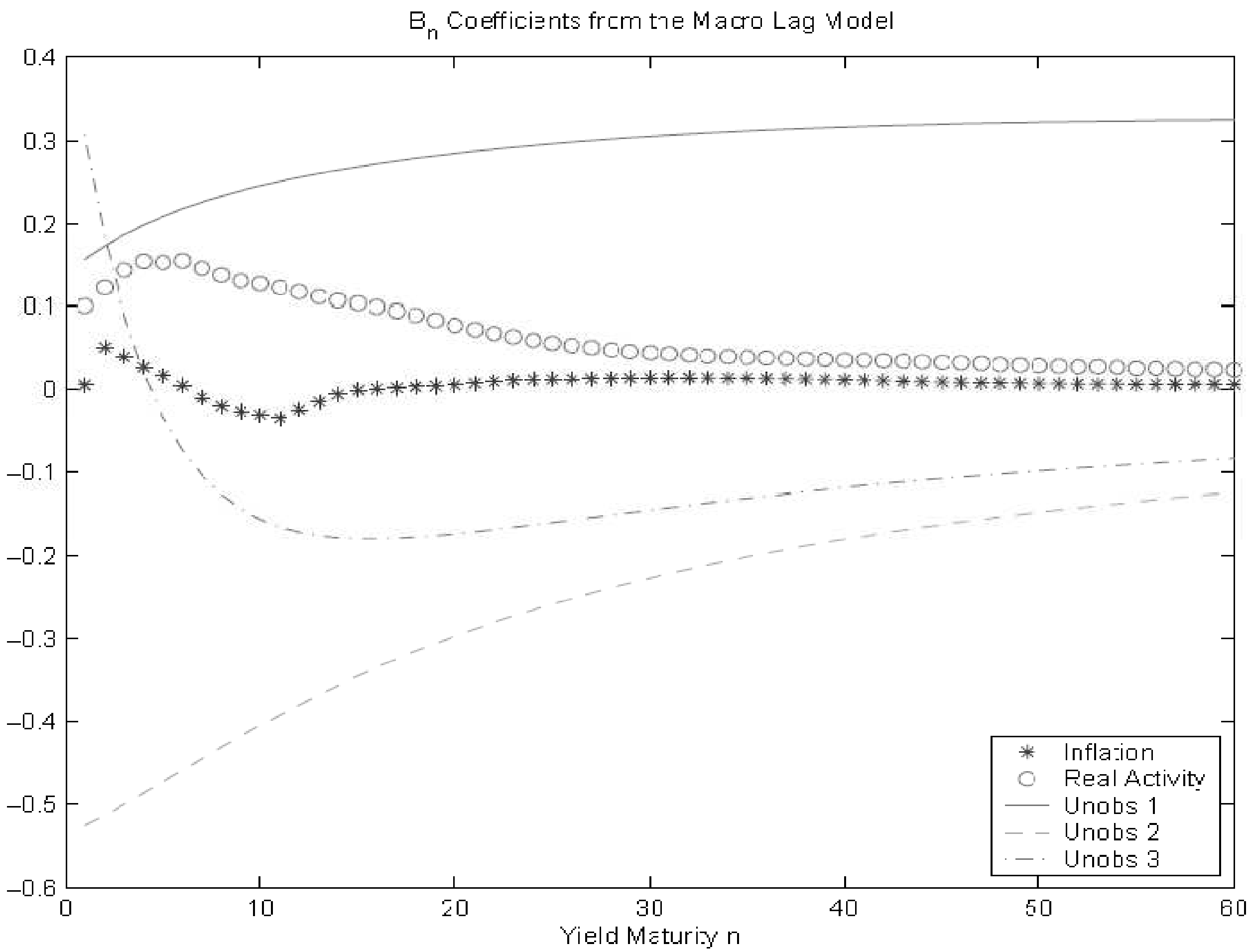


B_n Coefficients from the Macro Model
0.4
0.3
0.2
0.1
0
−0.1
−0.2
−0.3
−0.4
−0.5
−0.6
0
10
20
30
40
50
60
Yield Maturity n
Inflation
Real Activity
Unobs 1
Unobs 2
Unobs 3

- We observe that, in the "Macro Model", the three latent factors keep their role of LEVEL, SLOPE and (almost) CURVATURE.

- Inflation and Real Activity factors (at date t) affect yields (at date t) almost uniformly over the maturity spectrum.

- What about the "Macro Lag Model" ?


B_n Coefficients from the Macro Lag Model
0.4
0.3
0.2
0.1
0
−0.1
−0.2
−0.3
−0.4
−0.5
−0.6
0
10
20
30
40
50
60
Yield Maturity n
Inflation
Real Activity
Unobs 1
Unobs 2
Unobs 3

☐ Now, in the "Macro Lag Model", Inflation and Real Activity (contemporaneous) factors have little impact on the yield curve for $h > 30$ months.

☐ Thus, "Macro Model" and "Macro Lag Model" imply different impact of macro factors on the yield curve.

☐ The Forward Looking Taylor Rule (with lags) $r_t = \delta_0 + \delta_1' X_t$ show that lags are important in determining yield variations

$\rightarrow$ Contemporaneous shocks have less of an impact on the yields.

☐ **Question 1 :** are no-arbitrage restrictions and/or macro variables useful for yields out-of-sample forecasts ?

☐ **Question 2 :** Are the three latent factors (level, slope, curvature) explained/linked to the macro variables (inflation and economic activity) ?

5.3.5 Results

- ☐ Imposing no-arbitrage restrictions improves yields out-of-sample forecasts w.r.t. a VAR ("Yield only No-arbitrage ATSM" dominates "Yield only VAR model").

- ☐ These forecasts can be further improved incorporating macro factors into the Yield only No-arbitrage ATSM ("Macro Model" dominates "Macro Lag Model").

- ☐ They find that macro factors explain a significant portion (up to 85%) of movements in the short and middle parts of the yield curve, but explain only around 40% of movements at the long end of the yield curve.

- ☐ The effects of inflation shocks are strongest at the short end of the yield curve.

- ☐ A significant proportion of the ”level” and ”slope” factors are attributed to macro factors, particularly to inflation. However, the level effect qualitatively survives largely intact when macro factors are added to a term structure model.

Table 11
Comparison of Yields-Only and macro factors

Dependent variable	Independent variables					
	Inflation	Real activity	Unobs 1	Unobs 2	Unobs 3	Adj R^2
Panel A: Regressions on macro factors						
Unobs 1	0.4625	0.0726				0.2180
"level"	(0.0735)	(0.0860)				
Unobs 2	0.6707	0.1890				0.4902
"spread"	(0.0716)	(0.0611)				
Unobs 3	0.0498	0.1794				0.0343
"curvature"	(0.0629)	(0.0714)				
Panel B: Regressions on factors from macro model						
Unobs 1	0.1118	0.0307	0.9507	0.0174	0.0038	0.9971
	(0.0054)	(0.0056)	(0.0055)	(0.0056)	(0.0047)	
Unobs 2	0.9364	0.1026	0.0199	0.7624	0.0279	0.9981
	(0.0037)	(0.0037)	(0.0042)	(0.0032)	(0.0029)	
Unobs 3	0.0427	0.1238	0.1656	0.1455	0.9071	0.9256
	(0.0262)	(0.0260)	(0.0289)	(0.0241)	(0.0233)	
Panel C: Regressions on factors from macro lag model						
Unobs 1	0.0580	0.0207	1.0248	0.0035	0.0058	0.9979
	(0.0049)	(0.0040)	(0.0044)	(0.0047)	(0.0036)	
Unobs 2	0.7069	0.1132	0.2955	0.5700	0.1306	0.8715
	(0.0393)	(0.0313)	(0.0356)	(0.0376)	(0.0315)	
Unobs 3	0.1112	0.0081	0.2059	0.0228	0.8119	0.7470
	(0.0458)	(0.0386)	(0.0507)	(0.0365)	(0.0424)	

Regressions of the latent factors from the Yields-Only model with only latent factors (dependent variables) onto the macro factors and latent factors from the Macro and Macro Lag model (independent variables). All factors are normalized, and standard errors, produced using 3 Newey West (1987) lags, are in parentheses. Panel A lists coefficients from a regression of the Yields-Only latent factors onto only macro factors. Panel B lists coefficients from a regression of Yields-Only latent factors on the macro and latent factors from the Macro model with only contemporaneous inflation and real activity in the short rate equation. Panel C lists coefficients from a regression of Yields-Only latent factors on the macro and latent factors from the Macro Lag model with contemporaneous inflation and real activity and 11 lags of inflation and real activity in the short rate equation.

Fixed Income and Credit Risk

Lecture 5 - Part II

Discrete - Time Univariate

Positive Term Structure Models

Outline of Lecture 5 - Part II

5.5.3 The Affine Positive Term Structure of Interest Rates

5.5.4 Risk-Neutral Dynamics

5.6 Univariate ARG(p) Factor-Based Term Structure Models

5.6.1 Historical Dynamics

5.6.2 The Stochastic Discount Factor

5.6.3 The Affine Positive Term Structure of Interest Rates

5.6.4 Risk-Neutral Dynamics

5.4 The Autoregressive Gamma of order p Process

5.4.1 The Non-centered Gamma Distribution

☐ We say that the positive random variable Y is **Gamma** with parameters $\nu > 0$ and $\mu > 0$, i.e. $Y \sim \gamma(\nu, \mu)$, if and only if its probability density function is:

$$f_Y(y\,;\,\nu,\mu) = \frac{\exp(-y/\mu)\, y^{\nu-1}}{\Gamma(\nu)\,\mu^{\nu}}\, \mathbb{I}_{\{y>0\}}\,,$$

where $\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt\,,\ \ x \in \mathbb{C}\,,\ \ Re(x) > 0\,,$

$\Gamma(x) = \Gamma(x-1)\,(x-1)\,;\ \ \Gamma(x) = (x-1)!$ if x is a positive integer.

☐ ν is the shape (or degree of freedom) parameter, μ is the scale parameter.

□ We have that:

- $E[Y] = \nu\,\mu$ and $V[Y] = \nu\,\mu^2$ (**mean** and **variance**);
- $E[\exp(uY)] = \left(\dfrac{1}{1 - u\,\mu}\right)^{\nu}$, for $u < 1/\mu$ (**Laplace transform**);
- $Y \sim \gamma(\nu,\mu) \iff \dfrac{Y}{\mu} \sim \gamma(\nu, 1)$ (**scaling**).

□ We say that the positive random variable Y is **Non-centered Gamma** with parameters $\nu > 0$, $\beta > 0$ and $\mu > 0$, i.e. $Y \sim \widetilde{\gamma}(\nu,\beta,\mu)$, if and only if there exists a random variable $Z \sim \mathcal{P}(\beta)$ such that:

$$\begin{cases} \dfrac{Y}{\mu}\,|\,Z \sim \gamma(\nu + Z, 1)\,, \;\; \nu > 0\,, \\[2ex] Z \sim \mathcal{P}(\beta)\,, \;\; \beta > 0\,, \mu > 0\,, \end{cases} \iff \begin{cases} Y\,|\,Z \sim \gamma(\nu + Z, \mu)\,, \;\; \nu > 0\,, \\[2ex] Z \sim \mathcal{P}(\beta)\,, \;\; \beta > 0\,, \mu > 0\,, \end{cases}$$

where β is the non-centrality parameter.

☐ Let us remember that a discrete non-negative random variable Z is Poisson with parameter $\theta > 0$, i.e. $Z \sim \mathcal{P}(\theta)$ $(\theta > 0)$, if and only if:

$$\mathbb{P}[Z = z] = \frac{\exp(-\beta)\,\beta^z}{z!}\,, \;\; z \in \{0, 1, 2, \ldots\}\,,$$

$$E[Z] = V[Z] = \beta\,,$$

$$E[\exp(uZ)] = \exp[\beta\,(e^u - 1)]\,.$$

☐ The p.d.f. of $Y \sim \widetilde{\gamma}(\nu, \beta, \mu)$ is given by:

$$\begin{aligned} f_Y(y\,;\,\nu, \beta, \mu) &= \sum_{z=0}^{+\infty} f_Y(y \,|\, Z = z\,;\,\nu, \mu) \times f_Z(z\,;\,\beta) \\ &= \sum_{z=0}^{+\infty} \left[\frac{\exp(-y/\mu)\,y^{\nu+z-1}}{\Gamma(\nu+z)\,\mu^{\nu+z}} \times \frac{\exp(-\beta)\beta^z}{z\,!} \right] \mathbb{I}_{\{y>0\}}\,. \end{aligned}$$

☐ We have that:

- $E[Y] = \nu\,\mu + \beta\,\mu$ and $V[Y] = \nu\,\mu^2 + 2\mu^2\beta$ (**mean** and **variance**);

- $E[\exp(uY)] = \exp\left[-\nu \log(1 - u\,\mu) + \beta \frac{u\,\mu}{1 - u\,\mu}\right]$, for $u < 1/\mu$

  (**Laplace transform**);

- exercise!

5.4.2 The Autoregressive Gamma of order 1 Process

- The Autoregressive Gamma of order one [ARG(1)] process $\{x_t\}$ (say) is the exact discrete-time equivalent of the square-root process introduce in the continuous-time term structure literature by Cox, Ingersoll and Ross (1985). This (positive valued) process can be defined as:

$$
\begin{aligned}
&\frac{x_{t+1}}{\mu}|z_{t+1} \sim \gamma(\nu + z_{t+1}, 1)\,, \;\; \nu > 0\,, \\
&z_{t+1}|x_t \sim \mathcal{P}(\rho x_t/\mu)\,, \;\; \rho > 0\,, \mu > 0\,, \rho = \beta\,\mu
\end{aligned}
\tag{3}
$$

- where $\gamma(.)$ denotes the Gamma distribution, μ is the scale parameter, ν is the degree of freedom, ρ is the correlation (AR) parameter, and z_t is the mixing variable.

□ This means that the conditional probability density function $f(x_{t+1} \,|\, x_t; \mu, \nu, \rho)$ (say) of the ARG(1) process is the following mixture of Gamma densities with Poisson weights:

$$f(x_{t+1} \,|\, x_t; \mu, \nu, \rho) = \sum_{k=0}^{+\infty} \left[\frac{1}{\mu} \frac{e^{-\frac{x_{t+1}}{\mu}} \left(\frac{x_{t+1}}{\mu}\right)^{\nu+k-1}}{\Gamma(\nu+k)} \times \frac{\left(\frac{\rho x_t}{\mu}\right)^{k}}{k!} e^{-\frac{\rho x_t}{\mu}} \right] \mathbb{I}_{\{x_{t+1}>0\}} \,,$$

where $\rho > 0 \,, \mu > 0 \,, \nu > 0 \,,$ (4)

□ Its conditional Laplace transform has the following exponential-affine (in x_t) form [see Gourieroux and Jasiak (2006) for details; exercise!]:

$$E\left[\exp(u x_{t+1}) \,|\, \underline{x_t}\right] = \exp\left[\frac{\rho u}{1 - u\mu} x_t - \nu \log(1 - u\mu)\right] . \quad (5)$$

☐ The conditional mean and variance are respectively given by

- $E(x_{t+1}\,|\,x_t) = \nu\mu + \rho x_t$,
- and $V(x_{t+1}\,|\,x_t) = \nu\mu^2 + 2\mu\rho x_t$.
- exercise!

☐ Consequently, the process $\{x_t\}$ has the following weak AR(1) representation:

$$x_{t+1} = \nu\mu + \rho x_t + \varepsilon_{t+1}\,, \tag{6}$$

☐ where $\{\varepsilon_t\}$ is a conditionally heteroskedastic martingale difference

($\Rightarrow$ $E(\varepsilon_{t+1}\,|\,\varepsilon_t) = 0$), whose conditional variance is $V(\varepsilon_{t+1}\,|\,\varepsilon_t) = \nu\mu^2 + 2\mu\rho x_t$ (exercise!).

☐ The process is stationary (of second order) if and only if $\rho < 1$.

☐ In this case, the process $\{\varepsilon_t\}$ has finite unconditional variance given by:

$V(\varepsilon_t) = \nu\mu^2 + 2\nu\mu^2 \frac{\rho}{1-\rho}$ (exercise!).

☐ The unconditional mean and variance of $\{x_t\}$ are respectively given by:

- $E(x_t) = \dfrac{\nu\mu}{1-\rho}$,
- and $V(x_t) = \dfrac{\nu\mu^2}{(1-\rho)^2}$.
- exercise!

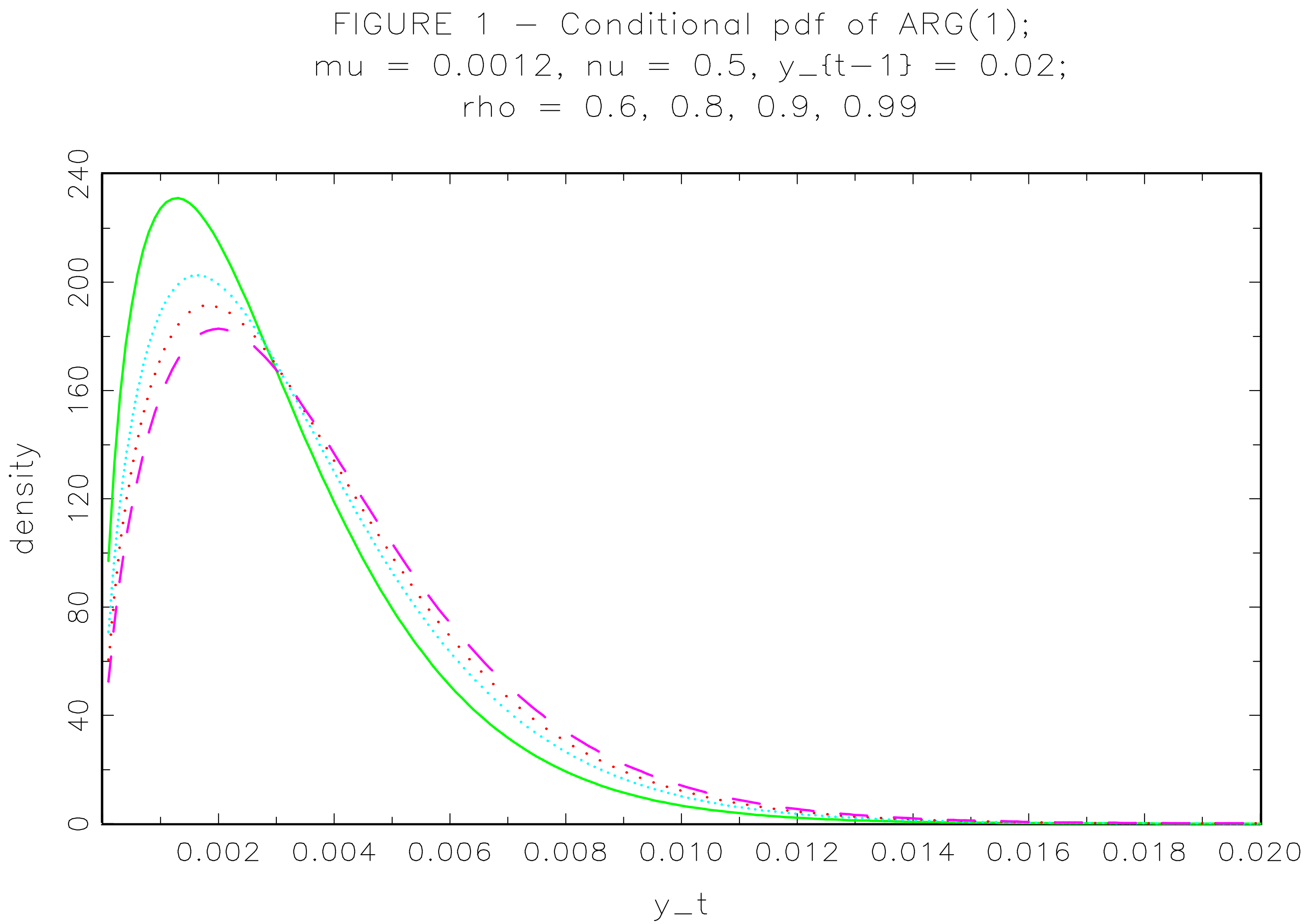
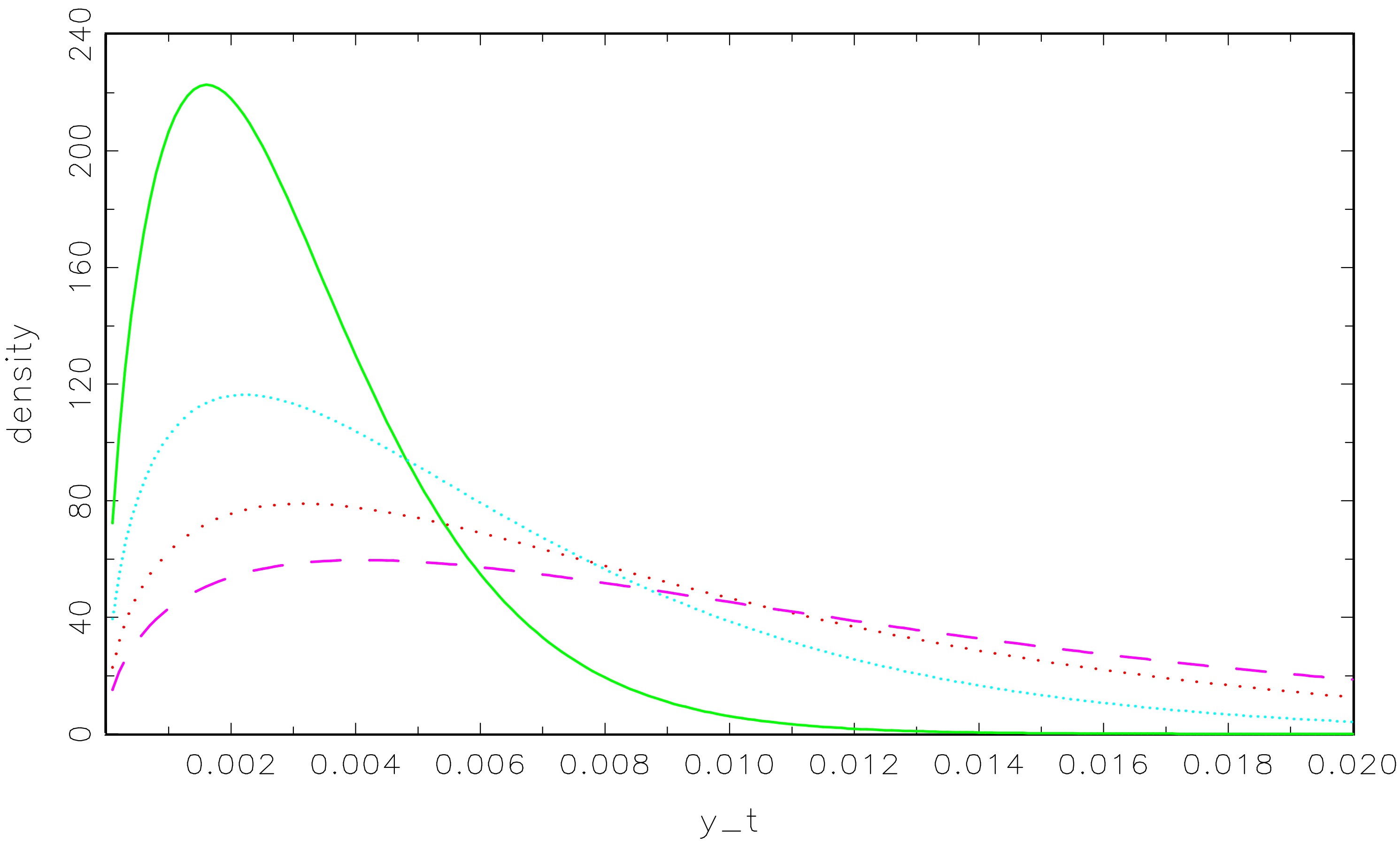


FIGURE 1 – Conditional pdf of ARG(1);
mu = 0.0012, nu = 0.5, y_{t-1} = 0.02;
rho = 0.6, 0.8, 0.9, 0.99
density
0
40
80
120
160
200
240
0.002
0.004
0.006
0.008
0.010
0.012
0.014
0.016
0.018
0.020
y_t

FIGURE 2 – Conditional pdf of ARG(1);
rho = 0.8, nu = 0.5, y_{t−1} = 0.02;
mu = 0.001, 0.005, 0.01, 0.05



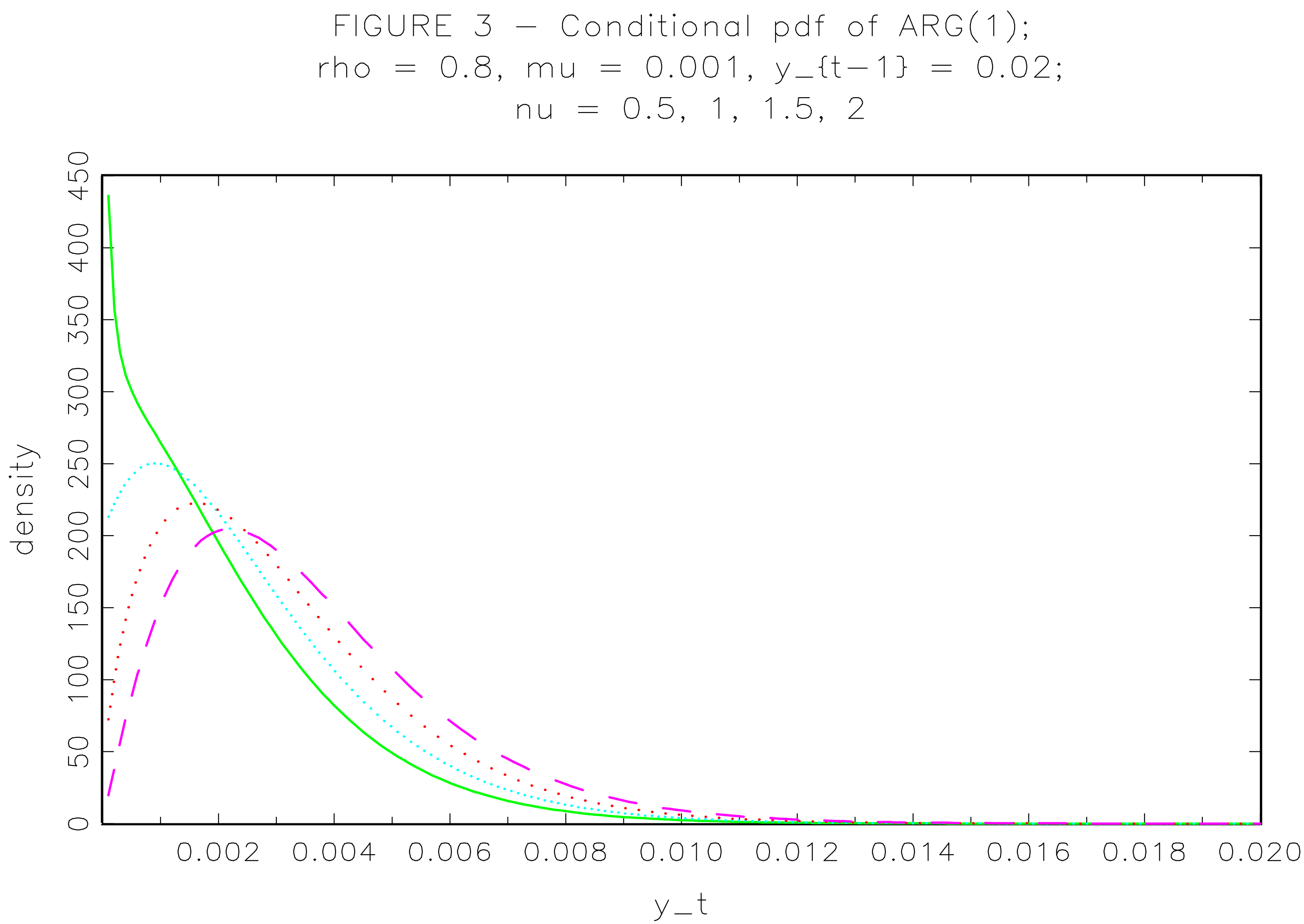


FIGURE 3 – Conditional pdf of ARG(1);
rho = 0.8, mu = 0.001, y_{t−1} = 0.02;
nu = 0.5, 1, 1.5, 2
density
0
50
100
150
200
250
300
350
400
450
0.002
0.004
0.006
0.008
0.010
0.012
0.014
0.016
0.018
0.020
y_t

5.4.3 The Autoregressive Gamma of order p Process

□ The Autoregressive Gamma of order p [ARG(p)] process can be defined as:

$$\frac{x_{t+1}}{\mu}|z_{t+1} \sim \gamma(\nu + z_{t+1}, 1)\,, \;\; \nu > 0\,,$$
$$z_{t+1}|\underline{x_t} \sim \mathcal{P}\left(\frac{\rho_1 x_t + \ldots + \rho_p x_{t-p+1}}{\mu}\right)\,, \;\; \rho_i = \beta_i\,\mu\,, \;\; i \in \{1, \ldots, p\}\,. \tag{7}$$

□ With the notation $X_t = (x_t, \ldots, x_{t-p+1})'$ and $\rho = (\rho_1, \ldots, \rho_p)'$ we have that the conditional Laplace transform of the ARG(p) process is (exercise!):

$$E\left[\exp(u x_{t+1})\,|\,\underline{x_t}\right] = \exp\left[\frac{u}{1-u\mu}\rho' X_t - \nu \log(1-u\mu)\right]\,, \tag{8}$$

☐ and the p.d.f. $f(x_{t+1} \,|\, X_t; \mu, \nu, \rho)$ (say) is given by:

$$f(x_{t+1} \,|\, X_t; \mu, \nu, \rho) = \sum_{k=0}^{+\infty} \left[\frac{1}{\mu} \frac{e^{-\frac{x_{t+1}}{\mu}} \left(\frac{x_{t+1}}{\mu}\right)^{\nu+k-1}}{\Gamma(\nu+k)} \times \frac{\left(\frac{\rho' X_t}{\mu}\right)^k}{k!} e^{-\frac{\rho' X_t}{\mu}} \right] \mathbb{I}_{\{x_{t+1}>0\}} \,. \tag{9}$$

☐ It easily seen that the conditional mean and variance of x_{t+1}, given $\underline{x_t}$, are respectively given by

- $E(x_{t+1} \,|\, \underline{x_t}) = \nu\mu + \rho' X_t$
- and $V(x_{t+1} \,|\, \underline{x_t}) = \nu\mu^2 + 2\mu\rho' X_t$.
- exercise!

☐ This means that, the ARG(p) process $\{x_t\}$ has the weak AR(p) representation:

$$x_{t+1} = \nu\mu + \rho' X_t + \xi_{t+1}\,, \tag{10}$$

where $\{\xi_t\}$ is a conditionally heteroskedastic martingale difference, whose conditional variance is $V(\xi_{t+1} \,|\, \underline{\xi_t}) = \nu\mu^2 + 2\mu\rho' X_t$ (exercise!).

☐ The process $\{x_t\}$ is stationary if and only if $\rho' e < 1$ [where $e = (1, \ldots, 1) \in \mathbb{R}^p$].

☐ In this case, $\{\xi_t\}$ has finite unconditional variance given by:

$V(\xi_t) = \nu\mu^2 + 2\nu\mu^2 \dfrac{\rho' e}{1 - \rho' e}$ (exercise!).

☐ The unconditional mean of $\{x_t\}$ is given by $E(x_t) = \dfrac{\nu\mu}{1 - \rho' e}$ (exercise!).

5.5 Univariate ARG(1) Factor-Based Term Structure Models

5.5.1 The Historical Dynamics

- We consider our discrete-time economy between dates 0 and T.

- x_t is our **factor or a state vector**, and it may be observable, partially observable or unobservable by the econometrician.

- Gaussian VAR(p) ATSMs do not (theoretically) guarantee that the yield-to-maturity formula generate positive yields for any date t, residual maturity h, any parameter values and realization of the factor x_t.

☐ We are going to see now that, if assume that the factor x_t follows an Autoregressive Gamma of order p Process (with a well specified SDF), then :

- the term structure of interest rates will be **affine** in the factor x_t (or X_t if $p > 1$).
- any model implied yield-to-maturity will be **strictly positive**.

☐ For ease of exposition (and for reason of time) we will consider only the case of a **scalar** and **latent** factor.

☐ In this section we consider $p = 1$ and then, in the next one, we will assume $p > 1$.

☐ Let us assume that the scalar latent factor x_t has a dynamics, under the historical probability $\mathbb{P}$, described by an ARG(1) process.

☐ This means that, under $\mathbb{P}$, the Laplace transform of x_{t+1}, conditionally to $\underline{x_t}$, is given by:

$$\begin{aligned} E\left[\exp(u x_{t+1}) \mid \underline{x_t}\right] &= \exp\left[\frac{\rho u}{1 - u\mu} x_t - \nu \log(1 - u\mu)\right], \\ &= \exp\left[a(u;\rho,\mu)\, x_t + b(u;\nu,\mu)\right]. \end{aligned}$$

☐ We have seen in the previous sections that this process has the following weak positive AR(1) representation:

$$x_{t+1} = \nu\mu + \rho x_t + \varepsilon_{t+1},$$

□ where $\{\varepsilon_t\}$ is a conditionally heteroskedastic martingale difference:

- $\Rightarrow E(\varepsilon_{t+1} \,|\, \varepsilon_t) = 0$

- whose conditional variance is $V(\varepsilon_{t+1} \,|\, \varepsilon_t) = \nu\mu^2 + 2\mu\rho x_t$,

□ and whose conditional Laplace transform is given by:

$$\begin{aligned} E\left[\exp(u\varepsilon_{t+1}) \,|\, \underline{\varepsilon_t}\right] &= E\left\{\exp\left[u(x_{t+1} - \nu\mu - \rho\, x_t) \,|\, \underline{x_t}\right]\right\} , \\ &= \exp\left[a(u;\rho,\mu)\, x_t + b(u;\nu,\mu) - u\,(\nu\,\mu + \rho\, x_t)\right] , \\ &= \exp\left[(a(u;\rho,\mu) - u\rho)\, x_t + b(u;\nu,\mu) - u\,\nu\,\mu\right] . \end{aligned}$$

□ This result is going to be useful in the construction of our one-period exponential-affine SDF $M_{t,t+1}$.

5.5.2 The Stochastic Discount Factor

☐ The one-period SDF $M_{t,t+1}$ is assumed to be given by:

$$
\begin{aligned}
M_{t,t+1} &= \exp\left[-\beta - \alpha x_t + \Gamma_t\, \varepsilon_{t+1}\right] \\
&\qquad \times \exp\left[-a\left(\Gamma_t;\rho,\mu\right)\, x_t - b\left(\Gamma_t;\nu,\mu\right) + \Gamma_t\left(\nu\,\mu + \rho\, x_t\right)\right]
\end{aligned}
$$

☐ with stochastic risk-correction coefficient given by $\Gamma_t = \gamma_o + \gamma\, x_t$.

☐ It is built in such a way that:

– $\dfrac{d\,\mathbb{Q}_{t,t+1}}{d\,\mathbb{P}_{t,t+1}} = \dfrac{M_{t,t+1}}{E_t[M_{t,t+1}]}$ is a density : $\dfrac{M_{t,t+1}}{E_t[M_{t,t+1}]} > 0$ and $E_t\left[\dfrac{M_{t,t+1}}{E_t[M_{t,t+1}]}\right] = 1$;

– the no-arbitrage restriction is explicitly satisfied : $E_t[M_{t,t+1}] = \exp(-\,r_t)$ if and only if $r_t = \beta + \alpha\, x_t$.

☐ **A useful Lemma** - Let us consider the functions:

$$a(u;\rho,\mu) = \frac{\rho u}{1-u\mu} \text{ and } b(u;\nu,\mu) = -\nu \log(1-u\mu)\,;$$

then, we have:

☐ **Lemma :**

$$a(u+g;\rho,\mu) - a(g;\rho,\mu) = a(u;\rho^*,\mu^*)$$

$$b(u+g;\nu,\mu) - b(g;\nu,\mu) = b(u;\nu,\mu^*)$$

$$\text{with } \rho^* = \frac{\rho}{(1-g\mu)^2},\ \mu^* = \frac{\mu}{1-g\mu}\,,$$

[Proof : exercise] and we will consider the case $g = \Gamma_t$.

5.5.3 The Affine Positive Term Structure of Interest Rates

□ The price at date t of the zero-coupon bond with time to maturity h is :

$$B(t, t+h) = \exp(c_h x_t + d_h)\,, \;\; h \geq 1\,,$$

□ where c_h and d_h satisfies, for $h \geq 1$, the recursive equations:

$$\begin{cases} c_h &= -\alpha + [a(c_{h-1} + \Gamma_t; \rho, \mu) - a(\Gamma_t; \rho, \mu)] \\ &= -\alpha + a(c_{h-1}; \rho^*, \mu^*)\,, \\ d_h &= -\beta + [b(c_{h-1} + \Gamma_t; \nu, \mu) - b(\Gamma_t; \nu, \mu)] + d_{h-1} \\ &= -\beta + b(c_{h-1}; \nu, \mu^*) + d_{h-1}\,, \end{cases}$$

□ with initial conditions $c_0 = 0, d_0 = 0$ (or $c_1 = -\alpha, d_1 = -\beta$). If $x_t = r_t$, then $c_1 = -1$ and $d_1 = 0$.

- The (continuously compounded) affine term structure of interest rates is:

$$R(t, t+h) = -\frac{1}{h} \log B(t, t+h) = -\frac{c_h}{h} x_t - \frac{d_h}{h}, \quad h \geq 1,$$

- **positivity of the yields :** Since $r_t = R(t, t+1) = \beta + \alpha x_t$, and since x_t is a positive process, the short rate process will be positive as soon as β and α are nonnegative.

- The positivity of r_t implies that of $R(t, t+h)$, at any date t and time to maturity h, because $R(t, h) = -\frac{1}{h} \log E_t^{\mathbb{Q}} [\exp(-r_t - \ldots - r_{t+h-1})]$.

- This is the discrete-time equivalent of the (continuous-time affine) Cox-Ingersoll-Ross (1985) model.

5.5.4 The Positive Risk-Neutral Dynamics

☐ The risk-neutral Laplace transform of x_{t+1}, conditionally to $\underline{x_t}$, is given by:

$$
\begin{aligned}
E_t^{\mathbb{Q}}[\exp(u x_{t+1})] &= E_t\left[\frac{M_{t,t+1}}{E_t(M_{t,t+1})}\exp(u x_{t+1})\right] \\
&= \exp\{[a(u+\Gamma_t;\rho,\mu) - a(\Gamma_t;\rho,\mu)]\, x_t + [b(u+\Gamma_t;\nu,\mu) - b(\Gamma_t;\nu,\mu)]\} \\
&= \exp[a(u;\rho^*,\mu^*)\, x_t + b(u;\nu,\mu^*)]
\end{aligned}
$$

☐ Under the risk-neutral probability $\mathbb{Q}$, x_{t+1} is a positive weak AR(1) process of the following type:

$$
x_{t+1} \;=\; \nu\,\mu^* + \rho^* x_t + \eta_{t+1}\,,
$$

☐ with $\rho^* = \dfrac{\rho}{(1-\Gamma_t\,\mu)^2} > 0$ and $\mu^* = \dfrac{\mu}{1-\Gamma_t\,\mu} > 0$, and where η_{t+1} is such that $E(\eta_{t+1}\,|\,\eta_t) = 0$ and $V(\eta_{t+1}\,|\,\eta_t) = \nu(\mu^*)^2 + 2\mu^*\rho^* x_t$.

5.6 Univariate ARG(p) Factor-Based Term Structure Models

5.6.1 The Historical Dynamics

□ Let us assume that the scalar latent factor x_t has a dynamics, under the historical probability $\mathbb{P}$, described by an ARG(p) process.

□ This means that, under $\mathbb{P}$, the Laplace transform of x_{t+1}, conditionally to $\underline{x_t}$, is given by:

$$\begin{aligned} E\left[\exp(u x_{t+1}) \,|\, \underline{x_t}\right] &= \exp\left[\frac{u}{1-u\mu}\left(\rho_1\, x_t + \ldots + \rho_p\, x_{t-p+1}\right) - \nu \log(1-u\mu)\right] , \\ &= \exp\left[\frac{u}{1-u\mu}\,\rho' X_t - \nu \log(1-u\mu)\right] , \\ &= \exp\left[a(u;\rho,\mu)\,' X_t + b(u;\nu,\mu)\right] . \end{aligned}$$

□ We seen in the previous sections that this process has the following weak positive AR(p) representation:

$$x_{t+1} = \nu\mu + \rho' X_t + \varepsilon_{t+1} \, ,$$

□ where $\{\varepsilon_t\}$ is a conditionally heteroskedastic martingale difference:

– $\Rightarrow E(\varepsilon_{t+1} \,|\, \varepsilon_t) = 0$

– whose conditional variance is $V(\varepsilon_{t+1} \,|\, \varepsilon_t) = \nu\mu^2 + 2\mu\rho' X_t$,

□ and whose conditional Laplace transform is given by:

$$\begin{aligned} E\left[\exp(u\varepsilon_{t+1}) \,|\, \underline{\varepsilon_t}\right] &= E\left\{\exp\left[u(x_{t+1} - \nu\mu - \rho'\, X_t) \,|\, \underline{x_t}\right]\right\} \, , \\ &= \exp\left[a(u;\rho,\mu)'\, X_t + b(u;\nu,\mu) - u\,(\nu\,\mu + \rho'\, X_t)\right] \, , \\ &= \exp\left[(a(u;\rho,\mu) - u\rho)'\, X_t + b(u;\nu,\mu) - u\,\nu\,\mu\right] \, . \end{aligned}$$

5.6.2 The Stochastic Discount Factor

☐ The one-period SDF $M_{t,t+1}$ is assumed to be given by:

$$\begin{aligned} M_{t,t+1} &= \exp\left[-\beta - \alpha' X_t + \Gamma_t \, \varepsilon_{t+1}\right] \\ &\quad \times \exp\left[-a\left(\Gamma_t; \rho, \mu\right)' X_t - b\left(\Gamma_t; \nu, \mu\right) + \Gamma_t \left(\nu\, \mu + \rho' X_t\right)\right] \end{aligned}$$

☐ with stochastic risk-correction coefficient given by $\Gamma_t = \gamma_o + \gamma' X_t$.

☐ It is built in such a way that:

– $\dfrac{d\,\mathbb{Q}_{t,t+1}}{d\,\mathbb{P}_{t,t+1}} = \dfrac{M_{t,t+1}}{E_t[M_{t,t+1}]}$ is a density : $\dfrac{M_{t,t+1}}{E_t[M_{t,t+1}]} > 0$ and $E_t\left[\dfrac{M_{t,t+1}}{E_t[M_{t,t+1}]}\right] = 1$;

– the no-arbitrage restriction is explicitly satisfied : $E_t[M_{t,t+1}] = \exp(-r_t)$ if and only if $r_t = \beta + \alpha' X_t$.

□ **A "generalization" of the Lemma** - Let us consider the functions:

$$a(u;\rho,\mu) = \frac{u}{1-u\mu}\,\rho \;\text{ and }\; b(u;\nu,\mu) = -\nu\log(1-u\mu)\,;$$

then, we have:

□ **Lemma :**

$$a(u+g;\rho,\mu) - a(g;\rho,\mu) = a(u;\rho^*,\mu^*)$$

$$b(u+g;\nu,\mu) - b(g;\nu,\mu) = b(u;\nu,\mu^*)$$

$$\text{with } \rho^* = \frac{1}{(1-g\mu)^2}\,\rho,\;\; \mu^* = \frac{\mu}{1-g\mu}\,,$$

and we will consider the case $g = \Gamma_t$.

5.6.3 The Affine Positive Term Structure of Interest Rates

☐ The price at date t of the zero-coupon bond with time to maturity h is :

$$B(t, t+h) = \exp(c_h' X_t + d_h)\,, \;\; h \geq 1\,,$$

☐ where c_h and d_h satisfies, for $h \geq 1$, the recursive equations:

$$\begin{cases} c_h & = & -\alpha + \left[a(c_{1,h-1} + \Gamma_t; \rho, \mu) - a(\Gamma_t; \rho, \mu)\right] + \bar{c}_{h-1} \\ & = & -\alpha + a(c_{1,h-1}; \rho^*, \mu^*) + \bar{c}_{h-1}\,, \\ d_h & = & -\beta + \left[b(c_{1,h-1} + \Gamma_t; \nu, \mu) - b(\Gamma_t; \nu, \mu)\right] + d_{h-1} \\ & = & -\beta + b(c_{1,h-1}; \nu, \mu^*) + d_{h-1}\,, \end{cases}$$

☐ where $\bar{c}_{h-1} = (c_{2,h-1}, \ldots, c_{p,h-1}, 0)'$, and with initial conditions $c_0 = 0, d_0 = 0$ (or $c_1 = -\alpha, d_1 = -\beta$). If $x_t = r_t$, then $c_1 = -e_1$ and $d_1 = 0$.

□ The affine positive term structure of interest rates is given by:

$$R(t,t+h) = -\frac{1}{h}\log B(t,t+h) = -\frac{c_h'}{h}X_t - \frac{d_h}{h}\,, \quad h \geq 1\,,$$

□ **positivity of the yields :** Since $r_t = R(t,t+1) = \beta + \alpha' X_t$, and since x_t is a positive process, the short rate process will be positive as soon as β and α are nonnegative.

□ The positivity of r_t implies that of $R(t,t+h)$, at any date t and time to maturity h, because $R(t,t+h) = -\frac{1}{h}\log E_t^{\mathbb{Q}}[\exp(-r_t - \ldots\ -r_{t+h-1})]$.

□ This is the discrete-time multiple lags generalization of the (continuous-time affine) Cox-Ingersoll-Ross (1985) model.

5.6.4 The Positive Risk-Neutral Dynamics

☐ The risk-neutral Laplace transform of x_{t+1}, conditionally to $\underline{x_t}$, is given by:

$$
\begin{aligned}
E_t^{\mathbb{Q}}[\exp(ux_{t+1})] &= E_t\left[\frac{M_{t,t+1}}{E_t(M_{t,t+1})}\exp(ux_{t+1})\right] \\
&= \exp\left\{[a(u+\Gamma_t;\rho,\mu) - a(\Gamma_t;\rho,\mu)]'\,X_t + [b(u+\Gamma_t;\nu,\mu) - b(\Gamma_t;\nu,\mu)]\right\} \\
&= \exp\left[a(u;\rho^*,\mu^*)'\,X_t + b(u;\nu,\mu^*)\right]
\end{aligned}
$$

☐ Under the risk-neutral probability $\mathbb{Q}$, x_{t+1} is a positive weak AR(1) process of the following type:

$$x_{t+1} = \nu\,\mu^* + \rho^{*\prime}X_t + \eta_{t+1}\,,$$

☐ with $\rho^* = \dfrac{1}{(1-\Gamma_t\,\mu)^2}\,\rho > 0$ and $\mu^* = \dfrac{\mu}{1-\Gamma_t\,\mu} > 0$, and where η_{t+1} is such that $E(\eta_{t+1}\,|\,\eta_t) = 0$ and $V(\eta_{t+1}\,|\,\eta_t) = \nu(\mu^*)^2 + 2\mu^*\rho^{*\prime}X_t$.