Estimation of non-Gaussian Affine Term Structure Models

Drew D. Creal Jing Cynthia Wu

University of Chicago, Booth School of Business

New York University, Stern School of Business

PRELIMINARY

December 2012

Motivation

Affine term structure models are useful for

- understanding the joint dynamics of the yield curve;
- describing the discount factor for financial markets;
- bridging macroeconomics and finance together;
- informing monetary policy.

Why do we need to go beyond Gaussian models?

- zero lower bound
- time-varying volatility

Estimation of ATSMs

- **Kim and Orphanides (2005):**

  ... the likelihood function seems to have multiple inequivalent local maxima which have similar likelihood values but substantially different implications...

- **Duffee (2002):**

  The QML functions for these models have a large number of local maxima.

- **Ang and Piazzesi (2003):**

  ... difficulties associated with estimating a model with many factors using maximum likelihood when yields are highly persistent.

Literature

Earlier work on estimating non-Gaussian models

- Dai and Singleton (2000), Duffee (2002)
- Collin-Dufresne, Goldstein and Jones (2008/2009), Ait-Sahalia and Kimmel (2010)
- Le, Singleton, and Dai (2010)

Recent work on estimating Gaussian models

- Joslin, Singleton, and Zhu (2011)
- Christensen, Diebold, and Rudebusch (2011)
- Hamilton and Wu (2012)

Contribution

Propose a new estimation approach for non-Gaussian ATSMs.

- Reduce parameter space by concentrating out $\mathbb{P}$ parameters.
- Provide analytical gradient to improve numerical behavior.
- Works for ANY rotation/identification scheme.
- For Gaussian models, this approach generalizes Joslin, Singleton, and Zhu (2011), and Hamilton and Wu (2012).

Extensions:

- Imposing constraints on parameters is straightforward.
- Adding macroeconomic variables is simple.
- Accommodating more general dynamics, e.g. AR(p).

Outline

Outline

Dynamics of the state vector under $\mathbb{P}$

State vector $x_t = (g_t', h_t')'$ has dynamics:

$$
\begin{aligned}
g_{t+1} &= \mu_g + \Phi_g g_t + \Phi_{gh} h_t + \Sigma_{gh}\varepsilon_{h,t+1} + \varepsilon_{g,t+1}, \\
\Sigma_{g,t}\Sigma_{g,t}' &= \Sigma_g \Sigma_g' + \sum_{i=1}^{H} \Sigma_{g,i}\Sigma_{g,i}' h_{it}, \\
h_{i,t+1} &\sim \text{N.C.-Gamma}(\nu_i, \Phi_{h,i}' h_t, \sigma_{h,i}), \qquad i = 1, \dots, H.
\end{aligned}
$$

- $\varepsilon_{g,t+1} \sim \mathrm{N}\left(0, \Sigma_{g,t}\Sigma_{g,t}'\right)$
- g_t and h_t are $G \times 1$ and $H \times 1$ vectors.
- Σ_g and $\Sigma_{g,i}$ are lower triangular.
- h_{it} is a discrete-time Cox Ingersoll Ross (1985) process.

Autoregressive gamma process

The non-Gaussian state variables h_t follow

$$h_{i,t+1}|h_t \sim \text{N.C.-Gamma}(\nu_i, \Phi'_{h,i}h_t, \sigma_{h,i}), \qquad i = 1, \ldots, H.$$

Details

Conditional mean and variance are linear in h_t:

$$\begin{aligned} \mathbb{E}[h_{i,t+1}|h_t] &= \nu_i \sigma_{h,i} + \Phi'_{h,i}h_t \qquad i = 1, \ldots, H \\ \mathbb{V}[h_{i,t+1}|h_t] &= \nu_i \sigma^2_{h,i} + 2\sigma_{h,i}\Phi'_{h,i}h_t \end{aligned}$$

Autoregressive gamma process

Conditional mean and variance are linear in h_{t-1}

$$
\begin{aligned}
\mathbb{E}[h_{i,t+1}|h_t] &= \nu_i \sigma_{h,i} + \Phi'_{h,i} h_t \qquad\qquad i = 1, \ldots, H \\
\mathbb{V}[h_{i,t+1}|h_t] &= \nu_i \sigma_{h,i}^2 + 2\sigma_{h,i} \Phi'_{h,i} h_t
\end{aligned}
$$

Stacked together

$$
\begin{aligned}
\mathbb{E}[h_{t+1}|h_t] &= \mu_h + \Phi_h h_t \\
\mathbb{V}[h_{t+1}|h_t] &= \begin{pmatrix} \nu_1 \sigma_{h,1}^2 + 2\sigma_{h,i} \Phi'_{h,1} h_t & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \nu_H \sigma_{h,H}^2 + 2\sigma_{h,H} \Phi'_{h,H} h_t \end{pmatrix}
\end{aligned}
$$

Note: $\mu_{h,i} = \nu_i \sigma_{h,i}$

Autoregressive gamma process

AG(1) process has the same transition density as the CIR process

$$dh_t = \kappa(\theta - h_t)dt + \sigma\sqrt{h_t}dW_t$$

Details

with the following mapping:

$$\Phi_h = 1 - \kappa\tau \qquad \nu = \frac{2\kappa\theta}{\sigma^2} \qquad \sigma_h = \frac{\sigma^2\tau}{2}$$

- $\nu_i > 1$ is the Feller condition.
- It guarantees that $h_{it} > 0$.

Dynamics of the state vector under $\mathbb{Q}$

$$
\begin{aligned}
g_{t+1} &= \mu_g^{\mathbb{Q}} + \Phi_g^{\mathbb{Q}} g_t + \Phi_{gh}^{\mathbb{Q}} h_t + \Sigma_{gh}\varepsilon_{h,t+1}^{\mathbb{Q}} + \varepsilon_{g,t+1}^{\mathbb{Q}}, \\
\Sigma_{g,t}\Sigma_{g,t}' &= \Sigma_g\Sigma_g' + \sum_{i=1}^{H} \Sigma_{g,i}\Sigma_{g,i}' h_{it}, \\
h_{i,t+1} &\sim \text{N.C.-Gamma}(\nu_i^{\mathbb{Q}}, \Phi_{h,i}^{\prime,\mathbb{Q}} h_t, \sigma_{h,i}), \qquad i = 1, \ldots, H.
\end{aligned}
$$

- $\varepsilon_{g,t+1}^{\mathbb{Q}} \overset{\mathbb{Q}}{\sim} \mathrm{N}\left(0, \Sigma_{g,t}\Sigma_{g,t}'\right)$.
- $\Sigma_g, \Sigma_{gh}, \Sigma_{g,i}, \sigma_{h,i}$ are the same under $\mathbb{P}$ and $\mathbb{Q}$.
- Drift changes but variance does not (approximately).

Bond prices

Short rate is a linear function of the state vector

$$r_t = \delta_0 + \delta_1' x_t$$

Bond prices are exponentially affine

$$P_t^n \;=\; \mathbb{E}_{t-1}^{\mathbb{Q}}\left[\exp\left(-r_t\right) P_{t+1}^{n-1}\right] \;=\; \exp\left(\bar{a}_n + \bar{b}_{n,g}' g_t + \bar{b}_{n,h}' h_t\right)$$

Yields (log-prices) are linear

$$y_t^n \;=\; a_n + b_n' x_t \;=\; a_n + b_{n,g}' g_t + b_{n,h}' h_t$$

with $a_n = -\frac{1}{n}\bar{a}_n$, $b_{n,g} = -\frac{1}{n}\bar{b}_{n,g}$ and $b_{n,h} = -\frac{1}{n}\bar{b}_{n,h}$.

Factor loadings

The bond loadings can be expressed recursively as

$$
\begin{aligned}
\bar{a}_n &= -\delta_0 + \bar{a}_{n-1} - \sum_{i=1}^{H} \nu_i^{\mathbb{Q}} \log\left(1 - \sigma_{h,i}\bar{b}_{n-1,gh,i}\right) + \left(\mu_g^{\mathbb{Q}} - \Sigma_{gh}\mu_h^{\mathbb{Q}}\right)' \bar{b}_{n-1,g} \\
&\quad + \frac{1}{2}\bar{b}_{n-1,g}' \Sigma_g \Sigma_g' \bar{b}_{n-1,g} \\
\bar{b}_{n,g} &= \Phi_g^{\mathbb{Q}\prime} \bar{b}_{n-1,g} - \delta_{1,g} \\
\bar{b}_{n,h} &= \sum_{i=1}^{H} \frac{\bar{b}_{n-1,gh,i}}{1 - \sigma_{h,i}\bar{b}_{n-1,gh,i}} \Phi_{h,i}^{\mathbb{Q}} + \left(\Phi_{gh}^{\mathbb{Q}} - \Sigma_{gh}\Phi_h^{\mathbb{Q}}\right)' \bar{b}_{n-1,g} - \delta_{1,h} \\
&\quad + \frac{1}{2}\left(I_H \otimes \bar{b}_{n-1,g}'\right) \bar{\bar{\Sigma}} \left(\iota_H \otimes \bar{b}_{n-1,g}\right)
\end{aligned}
$$

where $\bar{b}_{n,gh} = \Sigma_{gh}' \bar{b}_{n,g} + \bar{b}_{n,h}$ and with the restriction that $\bar{b}_{n-1,gh,i} < \frac{1}{\sigma_{h,i}}$.

Initial conditions:

$$
\bar{a}_1 = -\delta_0, \quad \bar{b}_{1,g} = -\delta_{1,g}, \quad \bar{b}_{1,h} = -\delta_{1,h}
$$

Observed yields

The measurement equations are

$$\begin{aligned} Y_t^{(1)} &= A_1 + B_1 x_t \\ Y_t^{(2)} &= A_2 + B_2 x_t + \eta_t \qquad \eta_t \sim \mathrm{N}(0, \Omega) \end{aligned}$$

$A_1 = (a_{n_1}, \ldots, a_{n_{N_1}})'$, $A_2 = (a_{n_{N_1+1}}, \ldots, a_{n_{N_1+N_2}})'$, $B_1 = (b_{n_1}, \ldots, b_{n_{N_1}})'$, and $B_2 = (b_{n_{N_1+1}}, \ldots, b_{n_{N_1+N_2}})'$.

- $Y_t^{(1)}$ is a $N_1 \times 1$ vector of yields priced without error.
- $Y_t^{(2)}$ is a $N_2 \times 1$ vector of yields priced with error.
- $N_1 = G + H$.
- Factors are observable: $x_t = B_1^{-1}\left(Y_t^{(1)} - A_1\right)$.

Stochastic discount factor

The SDF is

$$m_{t+1} = \frac{\mathrm{N}\left(g_{t+1}|h_{t+1}, g_t, h_t; \theta^{\mathbb{Q}}\right) \mathrm{NCG}\left(h_{t+1}|h_t; \theta^{\mathbb{Q}}\right)}{\mathrm{N}\left(g_{t+1}|h_{t+1}, g_t, h_t; \theta^{\mathbb{P}}\right) \mathrm{NCG}\left(h_{t+1}|h_t; \theta^{\mathbb{P}}\right)}$$

- The ratio of normals: log-normal
- The ratio of non-central gammas: doubly non-central F

Market price of risk

$$\Lambda_t = \Sigma_{t-1}^{-1}\left(\mathbb{E}^{\mathbb{P}}[x_t|x_{t-1}] - \mathbb{E}^{\mathbb{Q}}[x_t|x_{t-1}]\right)$$

$$\mathbb{E}^{\mathbb{P}}[x_t|x_{t-1}] = \begin{pmatrix} \mu_h \\ \mu_g \end{pmatrix} + \begin{pmatrix} \Phi_h & 0 \\ \Phi_{gh} & \Phi_g \end{pmatrix}\begin{pmatrix} h_{t-1} \\ g_{t-1} \end{pmatrix}$$

$$\mathbb{E}^{\mathbb{Q}}[x_t|x_{t-1}] = \begin{pmatrix} \mu_h^{\mathbb{Q}} \\ \mu_g^{\mathbb{Q}} \end{pmatrix} + \begin{pmatrix} \Phi_h^{\mathbb{Q}} & 0 \\ \Phi_{gh}^{\mathbb{Q}} & \Phi_g^{\mathbb{Q}} \end{pmatrix}\begin{pmatrix} h_{t-1} \\ g_{t-1} \end{pmatrix}$$

$$\Sigma_{t-1} = \begin{pmatrix} I_H & 0 \\ \Sigma_{gh} & I_G \end{pmatrix}\begin{pmatrix} \sqrt{\nu_1\sigma_{h,1}^2 + 2\sigma_{h,1}\Phi_{h,1}' h_{t-1}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Sigma_{g,t-1} \end{pmatrix}$$

Continuous-time limit: $\mathbb{V}[h_{it}|h_{t-1}] = \nu_i\sigma_{h,i}^2 + 2\sigma_{h,i}\Phi_{h,i}' h_{t-1} \approx 2\sigma_{h,i} h_{i,t-1}$

Reduced form: dynamics

$$
\begin{aligned}
Y_{t+1}^{(1)} &= \mu_1^* + \Phi_1^* Y_t^{(1)} \\
&+\Sigma^* \begin{pmatrix} \sqrt{\alpha_1^* + \gamma_1^* Y_t^{(1)}} & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ \vdots & & \sqrt{\alpha_{N_1}^* + \gamma_{N_1}^* Y_t^{(1)}} & 0 \\ 0 & \cdots & 0 & \Sigma_{g,t}^* \end{pmatrix} \begin{pmatrix} \varepsilon_{h,t+1} \\ \varepsilon_{g,t+1} \end{pmatrix}
\end{aligned}
$$

where $\varepsilon_{h,t+1}$ is a standardized NCG r.v. and

$$
\mu_1^* = B_1 \begin{pmatrix} \mu_h \\ \mu_g \end{pmatrix} + (I_{N_1} - \Phi_1^*)\, A_1 \qquad \Phi_1^* = B_1 \begin{pmatrix} \Phi_h & 0 \\ \Phi_{gh} & \Phi_g \end{pmatrix} B_1^{-1}
$$

$$
\Sigma^* = B_1 \begin{pmatrix} I_H & 0 \\ \Sigma_{gh} & I_G \end{pmatrix} \qquad \begin{aligned} \alpha_i^* &= \nu_i \sigma_{h,i}^2 - 2\sigma_{h,i} \Phi_{h,i}' S_{h_i} A_1 \\ \Sigma_g^* &= \Sigma_g \end{aligned} \quad \begin{aligned} \gamma_i^* &= 2\sigma_{h,i} \Phi_{h,i}' S_{h_i} B_1^{-1} \\ \Sigma_{g,i}^* &= \Sigma_{g,i} \end{aligned}
$$

Reduced form: cross-section

$$Y_t^{(2)} = \mu_2^* + \Phi_2^* Y_t^{(1)} + \eta_t, \qquad \eta_t \sim \mathrm{N}\left(0, \Omega^*\right)$$

$$\mu_2^* = A_2 - \Phi_2^* A_1 \qquad \Phi_2^* = B_2 B_1^{-1} \quad \Omega^* = \Omega$$

- Estimation of the reduced form is not as straightforward as Gaussian models.
- QML of reduced form may be possible. Parameters can be concentrated out.
- Hamilton and Wu (2012 JoE)

Conditional moments of observed yields

$$
\begin{aligned}
\mathbb{E}\left[Y_{t+1}^{(1)} | Y_t^{(1)}\right] &= \mu_1^* + \Phi_1^* Y_t^{(1)} \\
\mathbb{V}\left[Y_{t+1}^{(1)} | Y_t^{(1)}\right] &= \Sigma^* \begin{pmatrix} \alpha_1^* + \gamma_1^* Y_t^{(1)} & \cdots & \cdots & 0 \\ \vdots & & \vdots & \vdots \\ \vdots & \cdots & \alpha_H^* + \gamma_H^* Y_t^{(1)} & 0 \\ 0 & \cdots & 0 & \Sigma_t^* \Sigma_t^{*\prime} \end{pmatrix} \Sigma^{*\prime}
\end{aligned}
$$

- Reduced form allows us to calculate moments of *yields*.
- Autocorrelation of $Y_{t+1}^{(1)} | Y_t^{(1)}$ is a VAR(1).
- The conditional variance is a function of a single lag of $Y_t^{(1)}$.
- In an ATSM, autocorrelations of volatility are restricted.

Outline

Model

Estimation

Identifying restrictions

Results

Conclusion

Likelihood function

$$
\begin{aligned}
p\left(Y^{(1)}_{1:T}, Y^{(2)}_{1:T}; \theta\right) &= \left|\det\left(J\right)\right|^{-(T-1)} \prod_{t=2}^{T} p\left(Y^{(2)}_t | x_t; \theta\right) \\
&\quad \prod_{t=2}^{T} p\left(g_t | g_{t-1}, h_t, h_{t-1}; \theta\right) \prod_{t=2}^{T} \prod_{i=1}^{H} p\left(h_{it} | h_{t-1}; \theta\right)
\end{aligned}
$$

▸ Details

- Divide θ into two sub-vectors: $\theta = (\theta_c, \theta_m)$.
- $\theta_c = (\mu_g, \Phi_g, \Phi_{gh}, \Omega)$.
- Given θ_m, we can calculate the bond loadings (A_1, B_1) and

$$
g_t = S_g B_1^{-1}\left(Y^{(1)}_t - A_1\right) \quad h_t = S_h B_1^{-1}\left(Y^{(1)}_t - A_1\right)
$$

- KEY POINT: Concentrate θ_c out of the log-likelihood using GLS.

Can we concentrate out more parameters?

- These are the only parameters that can be "exactly" concentrated out.
- *Approximately* concentrate out μ_h, Φ_h by GLS:

$$h_{t+1} = \mu_h + \Phi_h h_t + \varepsilon_{h,t+1}$$

- *Approximately* concentrate out Σ_{gh} by GLS:

$$g_{t+1} = \mu_g + \Phi_g g_t + \Phi_{gh} h_t + \Sigma_{gh}\varepsilon_{h,t+1} + \Sigma_{g,t}\varepsilon_{g,t+1}$$

- For Gaussian models, *approximately* concentrate out Σ_g using the residuals:

$$g_{t+1} = \mu_g + \Phi_g g_t + \Sigma_g \varepsilon_{g,t+1}$$

Discussion

- Need to invert B_1 to calculate the factors:

$$x_t \ = \ B_1^{-1} \left(Y_t^{(1)} - A_1 \right)$$

- θ_m must satisfy $h_{it} > 0$ for all i and t.
- Need to calculate the GLS/OLS estimator.
- The parameters θ_m must satisfy the restriction $\bar{b}_{n-1,gh,i} < \frac{1}{\sigma_{h,i}}$ in order for bond prices to exist.

Gradients

$$\frac{\partial \ell}{\partial \theta'} = \frac{\partial \ell_{\text{dynamics}}}{\partial \theta'} + \frac{\partial \ell}{\partial A'}\frac{\partial A}{\partial \theta'} + \frac{\partial \ell}{\partial \text{vec}\left(B'\right)'}\frac{\partial \text{vec}\left(B'\right)}{\partial \theta'}$$

- We calculate the gradient analytically.
- Gradients have a simple recursive structure.
- Greatly improves numerical stability.
- Speeds calculations.

Gradients: example # 1

$$\frac{\partial \ell}{\partial \text{vech}\left(\Sigma_g\right)'} = \frac{\partial \ell_{\text{dynamics}}}{\partial \text{vech}\left(\Sigma_g\right)'} + \frac{\partial \ell}{\partial A'} \frac{\partial A}{\partial \text{vech}\left(\Sigma_g\right)'}$$

- The parameter Σ_g enters the dynamics and the bond loadings.
- $\frac{\partial \text{vec}(B')}{\partial \text{vech}(\Sigma_g)'} = 0$.
- The derivative of A w.r.t. Σ_g can be computed recursively.
- Initial condition: $\bar{a}'_{1,\Sigma_g} = 0$.

$$\bar{a}'_{n,\Sigma_g} = \bar{a}'_{n-1,\Sigma_g} + \text{vec}\left(\bar{b}_{n-1,g}\bar{b}'_{n-1,g}\Sigma_g\right)' \mathcal{D}^L_G$$

Gradients: example # 2

$$\frac{\partial \ell}{\partial \mathrm{vec}\left(\Phi_g^{\mathbb{Q}}\right)'} = \frac{\partial \ell}{\partial A'} \frac{\partial A}{\partial \mathrm{vec}\left(\Phi_g^{\mathbb{Q}}\right)'} + \frac{\partial \ell}{\partial \mathrm{vec}\left(B'\right)'} \frac{\partial \mathrm{vec}\left(B'\right)}{\partial \mathrm{vec}\left(\Phi_g^{\mathbb{Q}}\right)'}$$

- The parameter $\Phi_g^{\mathbb{Q}}$ only enters the bond loadings.
- Initial conditions: $\bar{a}'_{1,\Phi_g^{\mathbb{Q}}} = 0 \quad \bar{b}'_{1,\Phi_g^{\mathbb{Q}}} = 0$.

$$\begin{aligned} \bar{a}'_{n,\Phi_g^{\mathbb{Q}}} &= \bar{a}'_{n-1,\Phi_g^{\mathbb{Q}}} + \bar{y}'_{n-1} \bar{b}_{n-1,gh,\Phi_g^{\mathbb{Q}}} + \left(\mu_g^{\mathbb{Q}} - \Sigma_{gh} \mu_h^{\mathbb{Q}}\right)' \bar{b}_{n-1,g,\Phi_g^{\mathbb{Q}}} \\ &\quad + \bar{b}'_{n-1,g} \Sigma_g \Sigma'_g \bar{b}_{n-1,g,\Phi_g^{\mathbb{Q}}} \end{aligned}$$

$$\bar{b}_{n,g,\Phi_g^{\mathbb{Q}}} = \Phi_g^{\mathbb{Q}\prime} \bar{b}_{n-1,g,\Phi_g^{\mathbb{Q}}} + \left(I_G \otimes \bar{b}'_{n-1,g}\right)$$

Extension # 1: macro variables

$$\begin{pmatrix} Y_t^{(1)} \\ Y_t^{(m)} \end{pmatrix} = \begin{pmatrix} A_1 \\ 0 \end{pmatrix} + \begin{pmatrix} B_{1,g} & B_{1,h} & B_{1,m} \\ 0 & 0 & I_M \end{pmatrix} \begin{pmatrix} g_t \\ h_t \\ Y_t^{(m)} \end{pmatrix},$$

$$\begin{pmatrix} g_{t+1} \\ Y_{t+1}^{(m)} \end{pmatrix} = \begin{pmatrix} \mu_g \\ \mu_m \end{pmatrix} + \begin{pmatrix} \Phi_g & \Phi_{gh} & \Phi_{gm} \\ \Phi_{mg} & \Phi_{mh} & \Phi_m \end{pmatrix} \begin{pmatrix} g_t \\ h_t \\ Y_t^{(m)} \end{pmatrix}$$

$$+ \begin{pmatrix} \Sigma_{gh} \\ \Sigma_{mh} \end{pmatrix} \varepsilon_{h,t+1} + \varepsilon_{t+1},$$

$$\varepsilon_{t+1} \sim \mathrm{N}\left(0, \Sigma_t \Sigma_t'\right), \qquad \Sigma_t \Sigma_t' = \Sigma\Sigma' + \sum_{i=1}^{H} \Sigma_i \Sigma_i' h_{it}$$

- Add $M \times 1$ vector of macro variables $Y_t^{(m)}$.
- $\theta_c = (\mu_g, \mu_m, \Phi_g, \Phi_m, \Phi_{mh}, \Phi_{gh}, \Phi_{gm}, \Phi_{mg}, \Omega)$

Extension # 2: estimation under constraints

- It is possible to add linear equality constraints between $\mathbb{P}$ and $\mathbb{Q}$ parameters.
- Concentrating parameters out is constrained least squares.
- Ridge regression of μ_g, Φ_g, Φ_{gh} is straightforward.

Outline

Rotation

$$
\begin{aligned}
r_t &= \delta_0 + \delta_1 x_t & (1) \\
x_t &= \mu + \Phi x_{t-1} + \Sigma \varepsilon_t & (2)
\end{aligned}
$$

Define

$$
\tilde{x}_t = D_0 + D_1 x_t \quad \Rightarrow \quad x_t = D_1^{-1}(\tilde{x}_t - D_0)
$$

Rewrite (1) and (2)

$$
\begin{aligned}
r_t &= \delta_0 + \delta_1 D_1^{-1}(\tilde{x}_t - D_0) \\
D_1^{-1}(\tilde{x}_t - D_0) &= \mu + \Phi D_1^{-1}(\tilde{x}_{t-1} - D_0) + \Sigma \varepsilon_t
\end{aligned}
$$

Equations (1) and (2) are observationally equivalent to:

$$
\begin{aligned}
r_t &= \tilde{\delta}_0 + \tilde{\delta}_1 \tilde{x}_t \\
\tilde{x}_t &= \tilde{\mu} + \tilde{\Phi} \tilde{x}_{t-1} + \tilde{\Sigma} \varepsilon_t
\end{aligned}
$$

Identifying restrictions

Gaussian part

- $\mu_g^{\mathbb{Q}} = 0$
- $\Phi_g^{\mathbb{Q}}$ in ordered Jordan form
- $\Phi_{gh}^{\mathbb{Q}} = 0$
- $\Sigma_g, \Sigma_{g,i}$ lower triangular
- $\delta_{1g} = \iota_G$ a column vector of ones

Non-Gaussian part

- $\delta_{1h,i} = \pm 1 \quad \forall \quad i = 1, \ldots, H$
- $\Phi_h^{\mathbb{Q}}$ in ordered Jordan form
- $h_t > 0 \; \forall \; t$

Local maxima

- $\delta_{1,h}$ is positive or negative?
- Σ_{gh} positive or negative?
- How are the elements across $\Phi_g^{\mathbb{Q}}$ and $\Phi_h^{\mathbb{Q}}$ ordered?
- For non-Gaussian models, you must intentionally find and compare local maxima.

Example 1: $G = 3, H = 0$

Gaussian model

- $\mu_g^{\mathbb{Q}} = 0$
- $\Phi_g^{\mathbb{Q}}$ is diagonal with diagonal elements in descending order
- $\Phi_g^{\mathbb{Q}}$ does/does not have repeated eigenvalues
- Σ_g lower triangular
- $\delta_{1,g} = \iota_G$ a column vector of ones

Example 2: $G = 2, H = 1$

Gaussian part

- $\mu_g^{\mathbb{Q}} = 0$
- $\Phi_g^{\mathbb{Q}}$ is diagonal with diagonal elements in descending order
- $\Phi_g^{\mathbb{Q}}$ does/does not have repeated eigenvalues
- $\Phi_{gh}^{\mathbb{Q}} = 0$
- Σ_g, Σ_{g1} lower triangular
- $\delta_{1g} = \iota_G$ a column vector of ones

Non-Gaussian part

- $\delta_{1h} = 1$
- $h_t > 0 \;\; \forall \;\; t$
- $\nu > 1, \nu^{\mathbb{Q}} > 1$
- $0 < \Phi_h < 1$
- $\bar{b}_{n-1,gh} < \frac{1}{\sigma_h}$

Number of parameters for 3 factor models

[22(10)] Gaussian

- [3]μ_g
- [9]Φ_g
- [0]$\mu_g^{\mathbb{Q}}$
- [3]$\Phi_g^{\mathbb{Q}}$
- [6]Σ_g
- [0]δ_1
- [1]δ_0

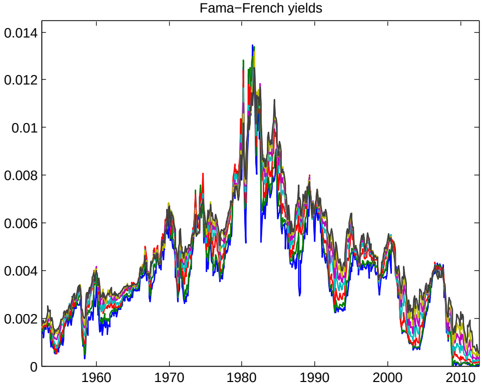
[24(16)]$\mathbb{A}_1(3)$

- [3]$\mu_g + \nu$
- [7]$\Phi_g + \Phi_{gh} + \Phi_h$: minus 2
- [1]$\mu_g^{\mathbb{Q}} + \nu^{\mathbb{Q}}$: plus 1
- [3]$\Phi_g^{\mathbb{Q}} + \Phi_{gh}^{\mathbb{Q}} + \Phi_h^{\mathbb{Q}}$
- [9]$\Sigma_g + \Sigma_{g1} + \Sigma_{gh} + \sigma_h$: plus 3
- [0]δ_1
- [1]δ_0

Comments: Gaussian model has more free parameters in the conditional mean, whereas the non-Gaussian model has more free parameters for the higher moments.

Outline

Data

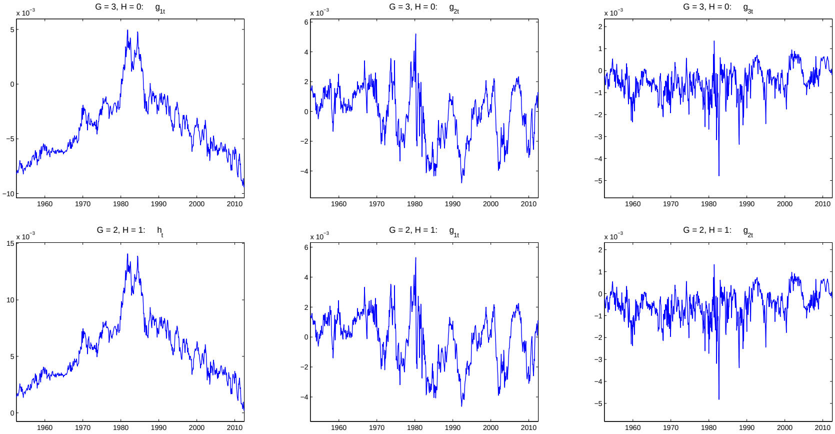


- Fama-Bliss zero-coupon bonds
- $6/1952$ to $6/2012$: $T = 721$ months
- Maturities: 1m, 3m, 1yr, 2yr, 3yr, 4yr, 5yr

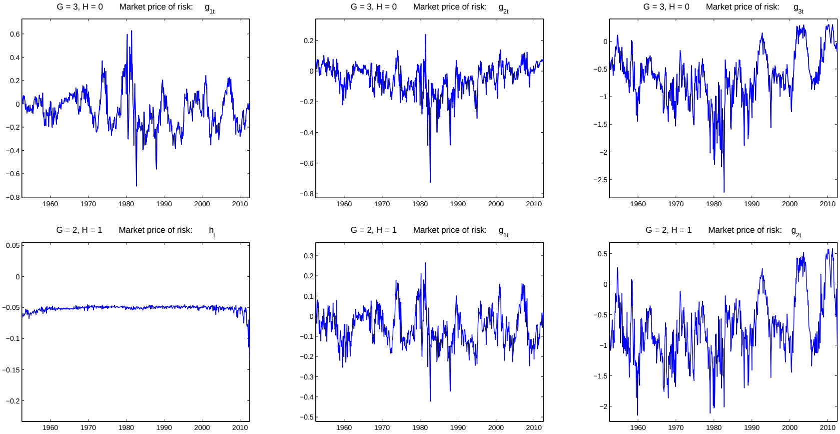
Estimates: 3 factor models

μ_g			μ_h	μ_g		ν
6.97e-05	-4.85e-05	-3.37e-04	2.97e-05	-1.32e-05	3.32e-05	1.934
(6.43e-05)	(4.65e-05)	(6.48e-05)	—	(1.102e-04)	(1.53e-05)	(0.124)
Φ_g			Φ_h			
1.007	0.048	0.067	0.994			
(0.011)	(0.016)	(0.040)	(0.004)			
			Φ_{gh}	Φ_g		
-0.012	0.938	0.0192	0.008	0.984	0.066	
(0.008)	(0.016)	(0.047)	(0.039)	(0.055)	(0.143)	
-0.037	-0.059	0.631	-0.041	-0.073	0.643	
(0.010)	(0.018)	(0.051)	(0.036)	(0.036)	(0.087)	
$\mu_g^{\mathbb{Q}}$			$\mu_h^{\mathbb{Q}}$	$\mu_g^{\mathbb{Q}}$		$\nu^{\mathbb{Q}}$
0	0	0	4.09e-05	0	0	2.637
—	—	—	—	—	—	(0.417)
$\Phi_g^{\mathbb{Q}}$			$\Phi_h^{\mathbb{Q}}$	$\Phi_g^{\mathbb{Q}}$		
0.995	0.954	0.530	0.996	0.951	0.536	
(0.0007)	(0.003)	(0.029)	(0.0009)	(0.003)	(0.033)	
Σ_g			σ_h			
3.99e-04	0	0	1.55e-05			
(2.52e-05)	—	—	(1.60e-06)			
			Σ_{gh}	Σ_g		
-3.09e-04	5.09E-04	0	-0.893	8.20e-09	0	
(3.83e-05)	(3.71E-05)	—	(0.104)	(5.58e-08)	—	
-4.50e-06	-2.52E-04	3.78E-04	0.054	-1.90e-09	1.08e-08	
(9.60e-06)	(2.69E-05)	(2.38E-05)	(0.1004)	(1.23e-07)	(7.33e-08)	
δ_0			δ_0	$\Sigma_{g,1}$		
0.0083			-0.0011	0.0063	0	
(0.0005)			(0.0004)	(0.0005)	—	
				-0.0035	0.0046	
				(0.0003)	(0.0003)	

Factors for 3 factor models



Risk pricing λ_t for 3 factor models



Decomposition

Risk neutral rate

$$\tilde{y}_t^n = \frac{1}{n}\left[r_t + \mathbb{E}\left(r_{t+1}\right) + \ldots + \mathbb{E}\left(r_{t+n-1}\right)\right]$$

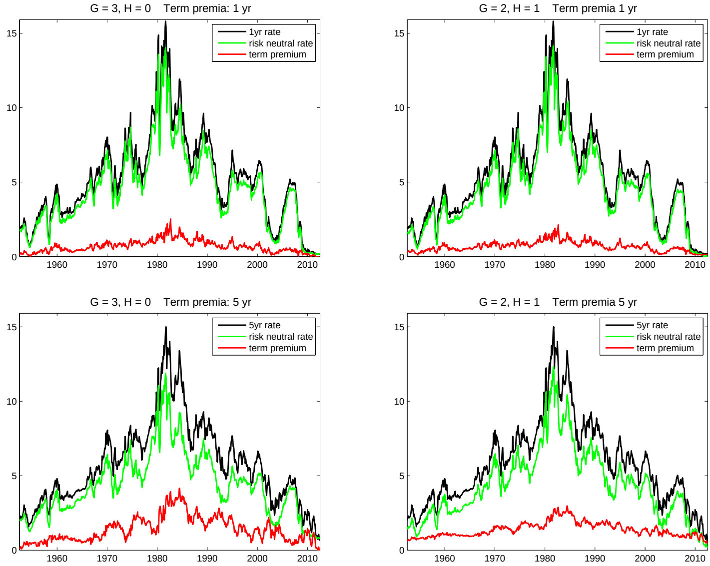
Model implied yield

$$y_t^n = a_n + b_n' x_t$$

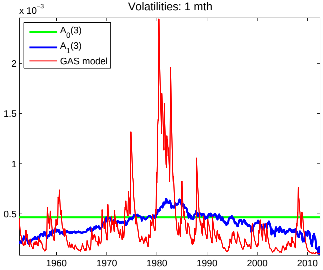
Term premium

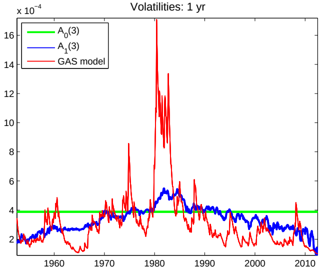
$$tp_t^n = y_t^n - \tilde{y}_t^n$$

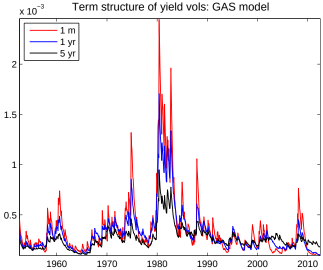
Term premiums: 1 and 5 yrs.

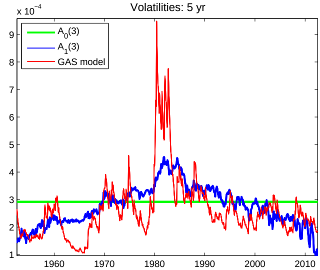


Volatility: Comparing with GAS (Creal, Koopman, Lucas, JAE 2012)









Estimates: 4 factor models

μ_g				μ_h	μ_g			ν		
3.89E-04	-9.73E-04	1.28E-03	-8.19E-04	5.64E-05	3.07E-04	-3.76E-05	-3.10E-04	2.6778		
(1.68E-04)	(2.20E-04)	(2.66E-04)	(3.64E-04)		(3.13E-04)	(3.39E-05)	(3.15E-04)	(2.2549)		
Φ_g				Φ_h						
1.034	0.087	-0.014	0.091	0.990						
(0.038)	(0.058)	(0.016)	(0.034)	(0.008)						
				Φ_{gh}	Φ_g					
-0.078	0.834	0.204	-0.079	0.004	0.867	1.170	0.091			
(0.085)	(0.111)	(0.120)	(0.083)	(0.006)	(0.041)	(1.210)	(0.224)			
0.0890	0.154	0.694	0.1926	0.002	0.015	0.816	0.0028			
(0.112)	(0.177)	(0.174)	(0.148)	(0.003)	(0.003)	(0.106)	(0.017)			
-0.085	-0.141	-0.040	0.546	-0.035	-0.009	-0.124	0.651			
(0.073)	(0.118)	(0.093)	(0.124)	(0.011)	(0.039)	(1.050)	(0.199)			
$\mu_g^{\mathbb{Q}}$				$\mu_h^{\mathbb{Q}}$	$\mu_g^{\mathbb{Q}}$			ν^{Q}		
0	0	0		2.71E-05	0	0	0	1.284		
								(0.338)		
$\Phi_g^{\mathbb{Q}}$				$\Phi_h^{\mathbb{Q}}$	$\Phi_g^{\mathbb{Q}}$					
0.992	0.960	0.876	0.696	0.995	0.912	–	0.702			
(0.003)	(0.013)	(0.033)	(0.043)	(0.002)	(0.015)	–	(0.054)			
Σ_g				σ_h						
6.94E-04	0	0	0	2.11E-05						
(2.29E-04)				(6.28E-06)						
				Σ_{gh}	Σ_g			$\Sigma_{g,1}$		
-1.47E-03	9.77E-04	0	0	0.925	8.37E-04	0	0	1.04E-02	0	0
(3.24E-04)	(4.71E-04)			(0.826)	(5.71E-04)			(3.42E-03)		
1.66E-03	-1.39E-03	8.74E-04	0	-0.217	-9.03E-05	6.85E-13	0	-6.75E-04	8.15E-04	0
(2.24E-04)	(4.43E-04)	(3.23E-04)		(0.057)	(5.25E-05)	(2.24E-05)		(2.97E-04)	(1.01E-04)	
-8.06E-04	5.65E-04	-7.18E-04	4.04E-04	-1.563	-7.96E-04	9.43E-11	4.04E-10	-9.01E-03	1.12E-03	4.68E-03
(3.47E-04)	(1.90E-04)	(3.53E-04)	(2.91E-05)	(0.723)	(5.51E-04)	(9.31E-05)	(3.77E-05)	(3.57E-03)	(7.67E-04)	(3.59E-04)

Conclusion

Methodology: Propose a new estimation approach for non-Gaussian ATSMs.

- Reduce parameter space
- Provide analytical gradient

Applications: Improve performance for any problem with two conditions

- can invert factors from observed yields
- (part of) the factor dynamics has analytical solutions

Caveat and Future directions

Caveat:

- Multiple local maxima due to the nature of the model
- Try to search in each region, and compare likelihood.

What do we use to replace $\mathbb{A}_0(3)$, $\mathbb{A}_1(3)$?

- $\mathbb{A}_1(4)$?, $\mathbb{A}_2(4)$? $\mathbb{A}_2(5)$?
- Unspanned volatility?

Extensions:

- Imposing constraints on parameters is straightforward.
- Adding macroeconomic variables is simple.
- Accommodating more general dynamics, e.g. AR(p).
- Correcting for bias (Bauer, Rudebusch and Wu (JBES 2012)) for the Gaussian dynamics is possible.

Appendix

Vector autoregressive gamma process

Autoregressive gamma process of Gouriéroux and Jasiak (2006)

$$
\begin{aligned}
h_{i,t+1} &\sim \text{Gamma}\left(\nu_i + z_{i,t+1}, \sigma_{h,i}\right) \qquad i = 1, \ldots, H \\
z_{i,t+1} &\sim \text{Poisson}\left(\frac{\Phi'_{h,i} h_t}{\sigma_{h,i}}\right)
\end{aligned}
$$

The transition density is

$$
\begin{aligned}
p(h_{t+1}|h_t;\theta) = & \prod_{i=1}^{H} \left(\frac{h_{i,t+1}}{\Phi_{h,i} h_t}\right)^{\frac{\nu_i - 1}{2}} \exp\left(-\frac{h_{i,t+1} + \Phi'_{h,i} h_t}{\sigma_{h,i}}\right) \\
& \left(\frac{1}{\sigma_{h,i}}\right) I_{\nu_i - 1}\left(2\sqrt{\frac{h_{i,t+1}\Phi'_{h,i} h_t}{\sigma^2_{h,i}}}\right)
\end{aligned}
$$

Back

Convergence to the continuous-time limit

Write the univariate AG(1) as a linear process

$$h_{t+1} \quad = \quad \nu\sigma_h + \Phi_h h_t + \sqrt{\nu\sigma_h^2 + 2\sigma_h \Phi_h h_t}\; \varepsilon_{h,t+1}$$

where ε_{ht} is a standardized NCG r.v. with $\mathbb{E}\left[\varepsilon_{ht}\right] = 0, \mathbb{V}\left[\varepsilon_{ht}\right] = 1$.

Let τ be an interval of time and define

$$\Phi_h \;=\; 1 - \kappa\tau \qquad \nu \;=\; \frac{2\kappa\theta}{\sigma^2} \qquad \sigma_h \;=\; \frac{\sigma^2\tau}{2}$$

The discrete-time process implies

$$\begin{aligned}
\mathbb{E}\left[h_{t+\tau}|h_t\right] &= \nu\sigma_h + \Phi_h h_t = \kappa\theta\tau + (1-\kappa\tau)h_t \\
\mathbb{V}\left[h_{t+\tau}|h_t\right] &= \nu\sigma_h^2 + 2\sigma_h\Phi_h h_t = \frac{\kappa\theta\sigma^2\tau^2}{2} + \sigma^2\tau(1-\kappa\tau)h_t
\end{aligned}$$

Back

Log-likelihood function

$$
\begin{aligned}
\ell_\theta \quad = \quad & \text{CONST} - (T-1)\log\left|\det(B_1)\right| - \frac{T-1}{2}\log|\Omega| - \frac{1}{2}\sum_{t=2}^{T}\text{tr}\left(\Omega^{-1}\eta_t\eta_t'\right) \\
& -\frac{1}{2}\sum_{t=2}^{T}\log\left|\Sigma_{g,t-1}\Sigma_{g,t-1}'\right| - \frac{1}{2}\sum_{t=2}^{T}\text{tr}\left(\left(\Sigma_{g,t-1}\Sigma_{g,t-1}'\right)^{-1}\varepsilon_{gt}\varepsilon_{gt}'\right) \\
& +\sum_{t=2}^{T}\sum_{i=1}^{H}\left[\frac{(\nu_i-1)}{2}\log(h_{it}) - \frac{h_{it}+\Phi_{h,i}'h_{t-1}}{\sigma_{h,i}} - \frac{(\nu_i-1)}{2}\log\left(\Phi_{h,i}'h_{t-1}\right)\right. \\
& \left. -\log(\sigma_{h,i}) + \log\left(I_{\nu_i-1}\left(2\sqrt{\frac{h_{it}\Phi_{h,i}'h_{t-1}}{\sigma_{h,i}^2}}\right)\right)\right]
\end{aligned}
$$

- $\varepsilon_{gt} = g_t - \mu_g - \Phi_g g_{t-1} - \Phi_{gh}h_{t-1} - \Sigma_{gh}\left[h_t - (\mu_h + \Phi_h h_{t-1})\right]$
- Likelihood is (conditionally) quadratic in $\mu_g, \Phi_g, \Phi_{gh}, \Omega$.

Back