Affine term structure models

A. Intro to Gaussian affine term structure models
B. Estimation by minimum chi square (Hamilton and Wu)
C. Estimation by OLS (Adrian, Moench, and Crump)
D. Dynamic Nelson-Siegel model (Christensen, Diebold, and Rudebusch)
E. Small-sample bias (Bauer, Rudebusch, and Wu)

A. Intro to Gaussian affine term structure models

P_{nt} = price at t of pure-discount n-period bond

$P_{nt} = E_t\{M_{t+1}P_{n-1,t+1}\}$

One approach: specify M_{t+1} and derive bond prices.

Today: reverse engineer– start with convenient empirical model of risk and then figure out what M_{t+1} this requires.

Later: will give a partial equilibrium example which would imply this M_{t+1}.

Suppose there is an $r \times 1$ vector $\boldsymbol{\xi}_t$ of possibly unobserved factors that summarize everything that matters for determining interest rates.

Suppose log of any bond price is affine function of these factors:

$$p_{nt} = \alpha_n + \boldsymbol{\beta}_n' \boldsymbol{\xi}_t$$

e.g., $r = 3$ (level, slope, curvature).

Conjecture that factors follow a first-order homoskedastic Gaussian VAR:

$$\boldsymbol{\xi}_{t+1} = \mathbf{c} + \boldsymbol{\rho}\boldsymbol{\xi}_t + \boldsymbol{\Sigma}\mathbf{u}_{t+1}$$

$$\mathbf{u}_{t+1} \sim \text{i.i.d. } N(\mathbf{0}, \mathbf{I}_r)$$

$\boldsymbol{\Sigma}$ summarizes unpredictability of factors and risk premia should be functions of $\boldsymbol{\Sigma}$.

Thus for Ω_t the information set at t,

$$\boldsymbol{\xi}_{t+1}|\Omega_t \sim N(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}\boldsymbol{\Sigma}')$$

$$\boldsymbol{\mu}_t = E_t\boldsymbol{\xi}_{t+1} = \mathbf{c} + \boldsymbol{\rho}\boldsymbol{\xi}_t$$

Consider asset that pays $\xi_{i,t+1}$ dollars next period. How much would you pay for each of these assets $i = 1,..,r$ today? If risk neutral, price would be $e^{-r_t}\mu_{it}$.

If risk averse, maybe I would only pay $e^{-r_t}\mu_{it}^{Q}$

$$\underset{(r\times 1)}{\boldsymbol{\mu}_t^{Q}} = \underset{(r\times 1)}{\boldsymbol{\mu}_t} - \underset{(r\times r)}{\boldsymbol{\Sigma}} \underset{(r\times 1)}{\boldsymbol{\lambda}_t}$$

$\lambda_{it} =$ price of factor i risk

If $\lambda_{it} = 0$, act as if column i of $\boldsymbol{\Sigma} = \mathbf{0}$

(uncertainty about factor i does

not affect price of any security).

True distribution of factors (sometimes called "historical distribution" or "P measure")

$$\boldsymbol{\xi}_{t+1}|\Omega_t \overset{P}{\sim} N(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}\boldsymbol{\Sigma}')$$

Risk-averse investors behave the same way as a risk-neutral investor would if that person believed the distribution was instead the "Q measure" or "risk-neutral distribution"

$$\boldsymbol{\xi}_{t+1}|\Omega_t \overset{Q}{\sim} N(\boldsymbol{\mu}_t^Q, \boldsymbol{\Sigma}\boldsymbol{\Sigma}')$$

$$\boldsymbol{\mu}_t^Q = \boldsymbol{\mu}_t - \boldsymbol{\Sigma}\boldsymbol{\lambda}_t$$

Question: what pricing kernel

M_{t+1} would imply this?

$$e^{-r_t}\boldsymbol{\mu}_t^Q = E_t\{M_{t+1}\boldsymbol{\xi}_{t+1}\}$$
$$= \int M_{t+1}\boldsymbol{\xi}_{t+1}\phi(\boldsymbol{\xi}_{t+1};\boldsymbol{\mu}_t,\boldsymbol{\Sigma\Sigma}')d\boldsymbol{\xi}_{t+1}$$

We would obtain desired answer if

$$M_{t+1}\phi(\boldsymbol{\xi}_{t+1};\boldsymbol{\mu}_t,\boldsymbol{\Sigma\Sigma}') = e^{-r_t}\phi(\boldsymbol{\xi}_{t+1};\boldsymbol{\mu}_t^Q,\boldsymbol{\Sigma\Sigma}')$$

$$\phi(\boldsymbol{\xi}_{t+1};\boldsymbol{\mu}_t^Q,\boldsymbol{\Sigma}\boldsymbol{\Sigma}') = \frac{1}{(2\pi)^{r/2}|\boldsymbol{\Sigma}|}\exp\left[-\frac{(\boldsymbol{\xi}_{t+1}-\boldsymbol{\mu}_t^Q)'(\boldsymbol{\Sigma}\boldsymbol{\Sigma}')^{-1}(\boldsymbol{\xi}_{t+1}-\boldsymbol{\mu}_t^Q)}{2}\right]$$

$$\begin{aligned}
(\boldsymbol{\xi}_{t+1}-\boldsymbol{\mu}_t^Q)'(\boldsymbol{\Sigma}\boldsymbol{\Sigma}')^{-1}(\boldsymbol{\xi}_{t+1}-\boldsymbol{\mu}_t^Q) \\
&= (\boldsymbol{\xi}_{t+1}-\boldsymbol{\mu}_t+\boldsymbol{\Sigma}\boldsymbol{\lambda}_t)'(\boldsymbol{\Sigma}\boldsymbol{\Sigma}')^{-1}(\boldsymbol{\xi}_{t+1}-\boldsymbol{\mu}_t+\boldsymbol{\Sigma}\boldsymbol{\lambda}_t) \\
&= (\boldsymbol{\xi}_{t+1}-\boldsymbol{\mu}_t)'(\boldsymbol{\Sigma}\boldsymbol{\Sigma}')^{-1}(\boldsymbol{\xi}_{t+1}-\boldsymbol{\mu}_t) \\
&\quad +\boldsymbol{\lambda}_t'\boldsymbol{\lambda}_t+2\boldsymbol{\lambda}_t'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\xi}_{t+1}-\boldsymbol{\mu}_t)
\end{aligned}$$

Or since

$$\boldsymbol{\xi}_{t+1} = \mathbf{c} + \boldsymbol{\rho}\boldsymbol{\xi}_t + \boldsymbol{\Sigma}\mathbf{u}_{t+1}$$

$$\boldsymbol{\Sigma}^{-1}(\boldsymbol{\xi}_{t+1} - \boldsymbol{\mu}_t) = \mathbf{u}_{t+1}$$

$$\phi(\boldsymbol{\xi}_{t+1}; \boldsymbol{\mu}_t^Q, \boldsymbol{\Sigma}\boldsymbol{\Sigma}')$$

$$= \phi(\boldsymbol{\xi}_{t+1}; \boldsymbol{\mu}_t, \boldsymbol{\Sigma}\boldsymbol{\Sigma}') \exp[-(1/2)\boldsymbol{\lambda}_t'\boldsymbol{\lambda}_t - \boldsymbol{\lambda}_t'\mathbf{u}_{t+1}]$$

Conclusion:

$$\phi(\boldsymbol{\xi}_{t+1};\boldsymbol{\mu}_t^Q,\boldsymbol{\Sigma\Sigma}')$$

$$= \phi(\boldsymbol{\xi}_{t+1};\boldsymbol{\mu}_t,\boldsymbol{\Sigma\Sigma}')\exp\left[-\frac{\boldsymbol{\lambda}_t'\boldsymbol{\lambda}_t+2\boldsymbol{\lambda}_t'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\xi}_{t+1}-\boldsymbol{\mu}_t)}{2}\right]$$

Summary:

$$M_{t+1}\phi(\boldsymbol{\xi}_{t+1};\boldsymbol{\mu}_t,\boldsymbol{\Sigma\Sigma}') = e^{-r_t}\phi(\boldsymbol{\xi}_{t+1};\boldsymbol{\mu}_t^Q,\boldsymbol{\Sigma\Sigma}')$$

calls for specifying

$$M_{t+1} = \exp[-r_t - (1/2)\boldsymbol{\lambda}_t'\boldsymbol{\lambda}_t - \boldsymbol{\lambda}_t'\mathbf{u}_{t+1}]$$

Suppose we further conjecture that price of risk is also affine function:

$$\underset{(r\times1)}{\boldsymbol{\lambda}_t} = \underset{(r\times1)}{\boldsymbol{\lambda}} + \underset{(r\times r)}{\boldsymbol{\Lambda}} \underset{(r\times1)}{\boldsymbol{\xi}_t}$$

Then

$$
\begin{aligned}
\boldsymbol{\mu}_t^Q &= \boldsymbol{\mu}_t - \boldsymbol{\Sigma}\boldsymbol{\lambda}_t \\
&= \mathbf{c} + \boldsymbol{\rho}\boldsymbol{\xi}_t - \boldsymbol{\Sigma}\boldsymbol{\lambda} - \boldsymbol{\Sigma}\boldsymbol{\Lambda}\boldsymbol{\xi}_t \\
&= \mathbf{c}^Q + \boldsymbol{\rho}^Q\boldsymbol{\xi}_t
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{c}^Q &= \mathbf{c} - \boldsymbol{\Sigma}\boldsymbol{\lambda} \\
\boldsymbol{\rho}^Q &= \boldsymbol{\rho} - \boldsymbol{\Sigma}\boldsymbol{\Lambda}
\end{aligned}
$$

P-measure dynamics:

$$\boldsymbol{\xi}_{t+1} = \mathbf{c} + \boldsymbol{\rho}\boldsymbol{\xi}_t + \boldsymbol{\Sigma}\mathbf{u}_{t+1}$$

$$\mathbf{u}_{t+1} \overset{P}{\sim} N(\mathbf{0}, \mathbf{I}_r)$$

Q-measure dynamics:

$$\boldsymbol{\xi}_{t+1} = \mathbf{c}^Q + \boldsymbol{\rho}^Q\boldsymbol{\xi}_t + \boldsymbol{\Sigma}\mathbf{u}_{t+1}^Q$$

$$\mathbf{u}_{t+1}^Q \overset{Q}{\sim} N(\mathbf{0}, \mathbf{I}_r)$$

Investors act the way a risk-neutral investor would who thought the factors follow the Q-measure distribution, that is,

$$\begin{aligned} P_{nt} &= E_t^Q[e^{-r_t} P_{n-1,t+1}] \\ &= E_t[M_{t+1} P_{n-1,t+1}] \end{aligned}$$

Recall that if

$$z \sim N(\mu, \sigma^2)$$

then

$$E(e^z) = \exp(\mu + \sigma^2/2)$$

Thus if $p_{nt} = \alpha_n + \boldsymbol{\beta}_n' \boldsymbol{\xi}_t$, we require

$$e^{p_{nt}} = E_t^Q[e^{-r_t} e^{p_{n-1,t+1}}]$$

$$\exp[\alpha_n + \boldsymbol{\beta}_n' \boldsymbol{\xi}_t] =$$

$$\exp\left[-r_t + \alpha_{n-1} + \boldsymbol{\beta}_{n-1}' \boldsymbol{\mu}_t^Q + (1/2)\boldsymbol{\beta}_{n-1}' \boldsymbol{\Sigma}\boldsymbol{\Sigma}' \boldsymbol{\beta}_{n-1}\right]$$

$$\exp[\alpha_n + \boldsymbol{\beta}_n' \boldsymbol{\xi}_t] =$$

$$\exp\left[-r_t + \alpha_{n-1} + \boldsymbol{\beta}_{n-1}' \boldsymbol{\mu}_t^Q + (1/2)\boldsymbol{\beta}_{n-1}' \boldsymbol{\Sigma\Sigma}' \boldsymbol{\beta}_{n-1}\right]$$

Or since

$$-r_t = p_{1t} = \alpha_1 + \boldsymbol{\beta}_1' \boldsymbol{\xi}_t$$

$$\boldsymbol{\mu}_t^Q = \mathbf{c}^Q + \boldsymbol{\rho}^Q \boldsymbol{\xi}_t$$

we require

$$\alpha_n + \boldsymbol{\beta}_n' \boldsymbol{\xi}_t = \alpha_1 + \boldsymbol{\beta}_1' \boldsymbol{\xi}_t + \alpha_{n-1}$$

$$+ \boldsymbol{\beta}_{n-1}' (\mathbf{c}^Q + \boldsymbol{\rho}^Q \boldsymbol{\xi}_t) + (1/2)\boldsymbol{\beta}_{n-1}' \boldsymbol{\Sigma\Sigma}' \boldsymbol{\beta}_{n-1}$$

$$\alpha_n + \boldsymbol{\beta}_n' \boldsymbol{\xi}_t = \alpha_1 + \boldsymbol{\beta}_1' \boldsymbol{\xi}_t + \alpha_{n-1} + \boldsymbol{\beta}_{n-1}' (\mathbf{c}^Q + \boldsymbol{\rho}^Q \boldsymbol{\xi}_t) + (1/2) \boldsymbol{\beta}_{n-1}' \boldsymbol{\Sigma\Sigma}' \boldsymbol{\beta}_{n-1}$$

$$\boldsymbol{\beta}_n' = \boldsymbol{\beta}_{n-1}' \boldsymbol{\rho}^Q + \boldsymbol{\beta}_1'$$

$$\alpha_n = \alpha_{n-1} + \boldsymbol{\beta}_{n-1}' \mathbf{c}^Q + (1/2) \boldsymbol{\beta}_{n-1}' \boldsymbol{\Sigma\Sigma}' \boldsymbol{\beta}_{n-1} + \alpha_1$$

Given $\mathbf{c}^Q, \boldsymbol{\rho}^Q, \alpha_1, \boldsymbol{\beta}_1, \boldsymbol{\Sigma}$ we can calculate the log of the price of any bond p_{nt}.

If $\mathbf{c}^Q = \mathbf{c}$ and $\boldsymbol{\rho}^Q = \boldsymbol{\rho}$ this would correspond to the expectations hypothesis of the term structure.

Gives us a way of summarizing dynamics of yield curve in terms of separate contributions of risk premia λ_t and expectations $E_t(r_{t+j})$.

B. Estimation by minimum chi square (Hamilton and Wu)

Model implies

$$y_{nt} = a_n + \mathbf{b}_n' \boldsymbol{\xi}_t$$

for yield on any maturity n and $\boldsymbol{\xi}_t$ an $(r \times 1)$ vector. If number of observed yields $> r,$ system is stochastically singular.

One solution: assume any observed yield n differs from model prediction by measurement or specification error:

$$y_{nt} = a_n + \mathbf{b}_n' \boldsymbol{\xi}_t + \varepsilon_{nt}$$

Collect system for observed yields in a vector

$$\underset{(N\times 1)}{\mathbf{y}_t} = (y_{n_1,t}, y_{n_2,t}, \ldots, y_{n_N,t})'$$

$$\mathbf{y}_t = \mathbf{a} + \mathbf{B}\boldsymbol{\xi}_t + \boldsymbol{\varepsilon}_t$$

$$\boldsymbol{\xi}_t = \mathbf{c} + \boldsymbol{\rho}\boldsymbol{\xi}_{t-1} + \mathbf{v}_t$$

This is state-space system

observation equation: $\mathbf{y}_t$

state equation: $\boldsymbol{\xi}_t$

parameters $\mathbf{a}$ and $\mathbf{B}$ are highly nonlinear functions of $\{\mathbf{c}, \boldsymbol{\rho}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \boldsymbol{\Lambda}, \alpha_1, \boldsymbol{\beta}_1\}$.

Alternative popular approach: assume that model holds exactly for r of the observed y_{nt}

$$\begin{bmatrix} y_{6t} \\ y_{24,t} \\ y_{120,t} \end{bmatrix} = \begin{bmatrix} a_6 \\ a_{24} \\ a_{120} \end{bmatrix} + \begin{bmatrix} \mathbf{b}_6' \\ \mathbf{b}_{24}' \\ \mathbf{b}_{120}' \end{bmatrix} \boldsymbol{\xi}_t$$

$$a_n = -\alpha_n / n$$

$$\mathbf{b}_n = -\boldsymbol{\beta}_n / n$$

$$\boldsymbol{\beta}_n' = \boldsymbol{\beta}_{n-1}' \boldsymbol{\rho}^Q + \boldsymbol{\beta}_1'$$

$$\alpha_n = \alpha_{n-1} + \boldsymbol{\beta}_{n-1}' \mathbf{c}^Q + (1/2) \boldsymbol{\beta}_{n-1}' \boldsymbol{\Sigma\Sigma}' \boldsymbol{\beta}_{n-1} + \alpha_1$$

$$\mathbf{y}_{1t} = \mathbf{a}_1 + \mathbf{B}_1 \boldsymbol{\xi}_t$$

For other yields, model holds with error

$$\mathbf{y}_{2t} = (y_{3t}, y_{12,t}, y_{36,t}, y_{60,t}, y_{84,t})'$$

$$\mathbf{y}_{2t} = \mathbf{a}_2 + \mathbf{B}_2 \boldsymbol{\xi}_t + \boldsymbol{\varepsilon}_{2t}$$

$$\mathbf{y}_{1t} = \mathbf{a}_1 + \mathbf{B}_1\boldsymbol{\xi}_t$$

$$\boldsymbol{\xi}_t = \mathbf{c} + \boldsymbol{\rho}\boldsymbol{\xi}_{t-1} + \mathbf{v}_t$$

$$\Rightarrow \mathbf{y}_{1t} = \mathbf{a}_1^* + \boldsymbol{\Phi}_{11}^*\mathbf{y}_{1,t-1} + \boldsymbol{\varepsilon}_{1t}$$

$$\mathbf{a}_1^* = \mathbf{a}_1 + \mathbf{B}_1\mathbf{c} - \mathbf{B}_1\boldsymbol{\rho}\mathbf{B}_1^{-1}\mathbf{a}_1$$

$$\boldsymbol{\Phi}_{11}^* = \mathbf{B}_1\boldsymbol{\rho}\mathbf{B}_1^{-1}$$

$$\boldsymbol{\varepsilon}_{1t} = \mathbf{B}_1\mathbf{v}_t$$

$\Rightarrow \mathbf{y}_{1t}$ follows VAR(1)

$$\mathbf{y}_{2t} = \mathbf{a}_2 + \mathbf{B}_2\boldsymbol{\xi}_t + \boldsymbol{\varepsilon}_{2t}$$

$$\mathbf{y}_{1t} = \mathbf{a}_1 + \mathbf{B}_1\boldsymbol{\xi}_t$$

$$\Rightarrow \mathbf{y}_{2t} = \mathbf{a}_2^* + \boldsymbol{\Phi}_{21}^*\mathbf{y}_{1t} + \boldsymbol{\varepsilon}_{2t}$$

$$\mathbf{a}_2^* = \mathbf{a}_2 - \mathbf{B}_2\mathbf{B}_1^{-1}\mathbf{a}_1$$

$$\boldsymbol{\Phi}_{21}^* = \mathbf{B}_2\mathbf{B}_1^{-1}$$

$$\mathbf{y}_{1t} = \mathbf{a}_1^* + \mathbf{\Phi}_{11}^* \mathbf{y}_{1,t-1} + \boldsymbol{\varepsilon}_{1t}$$

$$\mathbf{y}_{2t} = \mathbf{a}_2^* + \mathbf{\Phi}_{21}^* \mathbf{y}_{1t} + \boldsymbol{\varepsilon}_{2t}$$

$\Rightarrow \mathbf{y}_t = (\mathbf{y}_{1t}', \mathbf{y}_{2t}')'$ follows restricted VAR(1) whose coefficients are nonlinear functions of $\{\mathbf{c}, \boldsymbol{\rho}, \mathbf{\Sigma}, \boldsymbol{\lambda}, \mathbf{\Lambda}, \alpha_1, \boldsymbol{\beta}_1\}$.

Can first estimate unrestricted parameters $\{\mathbf{a}_1^*, \mathbf{\Phi}_{11}^*, \mathbf{a}_2^*, \mathbf{\Phi}_{21}^*, \mathbf{\Sigma}^*\}$ by OLS, then find structural parameters that make predicted values close to observed (minimize chi square statistic for test that restrictions are valid). Asymptotically equivalent to MLE, but simpler.

Can easily generalize above to suppose that there are r linear combinations of $\mathbf{y}_t$ for which model holds without error:

$$\underset{(r\times 1)}{\mathbf{y}_{1t}} = \underset{(r\times N)}{\mathbf{H}} \underset{(N\times 1)}{\mathbf{y}_t}$$

$$\underset{(r\times 1)}{\mathbf{y}_{1t}} = \underset{(r\times N)}{\mathbf{H}} \underset{(N\times 1)}{\mathbf{a}} + \underset{(r\times N)}{\mathbf{H}} \underset{(N\times r)}{\mathbf{B}} \boldsymbol{\xi}_t$$

$$\Rightarrow \mathbf{y}_{1t} = \mathbf{a}_1^* + \boldsymbol{\Phi}_{11}^* \mathbf{y}_{1,t-1} + \boldsymbol{\varepsilon}_{1t}$$

e.g., $\mathbf{y}_{1t}$ = first r principal components of $\mathbf{y}_t$ ($\mathbf{H}'$ = first r eigenvectors of $T^{-1}\sum_{t=1}^{T}(\mathbf{y}_t - \bar{\mathbf{y}})(\mathbf{y}_t - \bar{\mathbf{y}})'$)

Note the model as written is unidentified.

If $\boldsymbol{\xi}_t \rightarrow \boldsymbol{\xi}_t + \mathbf{q}$ and $\boldsymbol{\lambda} \rightarrow \boldsymbol{\lambda} + \boldsymbol{\Sigma}\mathbf{q}$, the model would be observationally identical

Same if $\boldsymbol{\xi}_t \rightarrow \mathbf{Q}\boldsymbol{\xi}_t,\ \boldsymbol{\rho} \rightarrow \mathbf{Q}\boldsymbol{\rho}\mathbf{Q}^{-1}, \mathbf{b}_1' \rightarrow \mathbf{b}_1'\mathbf{Q}^{-1}$

(1) Sample normalization:

$\mathbf{c} = \mathbf{0}$

$\boldsymbol{\rho}^Q$ lower triangular

$\rho_{ii}^Q \geq \rho_{jj}^Q$ for $i < j$

$\boldsymbol{\Sigma} = \mathbf{I}_r$

elements of $\mathbf{b}_1$ nonpositive

This normalization is internally inconsistent.

If observe $\boldsymbol{\xi}_t = \mathbf{H}\mathbf{y}_t$ directly and $\mathbf{y}_t = \mathbf{a} + \mathbf{B}\boldsymbol{\xi}_t$, then $\boldsymbol{\xi}_t = \mathbf{H}\mathbf{a} + \mathbf{H}\mathbf{B}\boldsymbol{\xi}_t$ requiring $\mathbf{H}\mathbf{a} = \mathbf{0}$ and $\mathbf{H}\mathbf{B} = \mathbf{I}_r$.

Upside: in practice $\mathbf{H}\mathbf{a}$ turns out to be close to $\mathbf{0}$ and $\mathbf{H}\mathbf{B}$ close to $\mathbf{I}_r$ even without imposing.

(2) Joslin, Singleton Zhu normalization

(Rev Financial Studies, 2011) as

implemented by Hamilton-Wu

(J. Econometrics, 2014):

unknown parameters are $\{\mathbf{c}, \boldsymbol{\rho}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \alpha_1\}$

$\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_r)'$ are eigenvalues of $\boldsymbol{\rho}^Q$.

$$\underset{(1\times1)}{\omega_n(x)} = n^{-1}\sum_{j=0}^{n} x^j$$

$$\underset{(r\times N)}{\mathbf{K}(\boldsymbol{\gamma})} = \begin{bmatrix} \omega_{n_1}(\gamma_1) & \omega_{n_2}(\gamma_1) & \cdots & \omega_{n_N}(\gamma_1) \\ \omega_{n_1}(\gamma_2) & \omega_{n_2}(\gamma_2) & \cdots & \omega_{n_N}(\gamma_2) \\ \vdots & \vdots & \cdots & \vdots \\ \omega_{n_1}(\gamma_r) & \omega_{n_2}(\gamma_r) & \cdots & \omega_{n_N}(\gamma_r) \end{bmatrix}$$

$$\underset{(r\times r)}{\mathbf{V}(\boldsymbol{\gamma})} = \begin{bmatrix} \gamma_1 & \cdots & 0 \\ \vdots & \cdots & \vdots \\ 0 & \cdots & \gamma_r \end{bmatrix}$$

$$\underset{(r\times r)}{\boldsymbol{\rho}^{Q\prime}} = [\mathbf{K}(\boldsymbol{\gamma})\mathbf{H}']^{-1}[\mathbf{V}(\boldsymbol{\gamma})][\mathbf{K}(\boldsymbol{\gamma})\mathbf{H}']$$

$\boldsymbol{\beta}_1 = [\mathbf{K}(\boldsymbol{\gamma})\mathbf{H}']^{-1}\mathbf{1}_r$ for $\mathbf{1}_r' = (1, 1, \ldots, 1)$

Then $\mathbf{HB} = \mathbf{I}_r$.

A related calculation for the intercepts guarantees $\mathbf{Ha} = \mathbf{0}$.

Benefits: $\mathbf{c}, \boldsymbol{\rho}, \boldsymbol{\Sigma}$ estimated by simple OLS:

$$\boldsymbol{\xi}_t = \mathbf{c} + \boldsymbol{\rho}\boldsymbol{\xi}_{t-1} + \boldsymbol{\varepsilon}_{1t}$$

$$E(\boldsymbol{\varepsilon}_{1t}\boldsymbol{\varepsilon}_{1t}') = \boldsymbol{\Sigma}\boldsymbol{\Sigma}'$$

$\boldsymbol{\gamma}, \boldsymbol{\Sigma}_2, \alpha_1$ estimated by MCS on

$$\mathbf{y}_{2t} = \mathbf{a}_2^* + \boldsymbol{\Phi}_{21}^*\boldsymbol{\xi}_t + \boldsymbol{\varepsilon}_{2t}$$

$$E(\boldsymbol{\varepsilon}_{2t}\boldsymbol{\varepsilon}_{2t}') = \boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_2'$$

C. Estimation by OLS

Simpler approach (Adrian, Crump, and Moench, JFE 2013):

Don't impose any restrictions, get everything by OLS.

$\boldsymbol{\xi}_t$ = first $r = 5$ principal components of observed set of yields y_{nt} for $n \in \mathbb{N}$

$\mathbb{N} = \{6m, 12m, 18m, 24m, 30m, 36m, 42m, 48m, 54m, 60m, 7y, 10y\}$

$\boldsymbol{\xi}_{t+1} = \mathbf{c} + \boldsymbol{\rho}\boldsymbol{\xi}_t + \mathbf{v}_{t+1}$

can estimate by OLS

log price of n-period bond:

$$p_{nt} = \alpha_n + \boldsymbol{\beta}_n' \boldsymbol{\xi}_t$$

excess return of n-period bond:

$$\begin{aligned} x_{n-1,t+1} &= p_{n-1,t+1} - p_{nt} - r_t \\ &= p_{n-1,t+1} - p_{nt} + p_{1t} \\ &= \alpha_{n-1} + \boldsymbol{\beta}_{n-1}' (\mathbf{c} + \boldsymbol{\rho} \boldsymbol{\xi}_t + \boldsymbol{\Sigma} \mathbf{u}_{t+1}) - \alpha_n \\ &\quad - \boldsymbol{\beta}_n' \boldsymbol{\xi}_t + \alpha_1 + \boldsymbol{\beta}_1' \boldsymbol{\xi}_t \end{aligned}$$

$$
\begin{aligned}
x_{n-1,t+1} &= \alpha_{n-1} + \boldsymbol{\beta}'_{n-1}(\mathbf{c} + \boldsymbol{\rho}\boldsymbol{\xi}_t + \mathbf{v}_{t+1}) - \alpha_n \\
&\quad - \boldsymbol{\beta}'_n \boldsymbol{\xi}_t + \alpha_1 + \boldsymbol{\beta}'_1 \boldsymbol{\xi}_t \\
&= a_{n-1} + \mathbf{b}'_{n-1}\mathbf{v}_{t+1} + \mathbf{c}'_{n-1}\boldsymbol{\xi}_t
\end{aligned}
$$

$$
\begin{aligned}
a_{n-1} &= \alpha_{n-1} + \boldsymbol{\beta}'_{n-1}\mathbf{c} - \alpha_n + \alpha_1 \\
\mathbf{b}'_{n-1} &= \boldsymbol{\beta}'_{n-1} \\
\mathbf{c}'_{n-1} &= \boldsymbol{\beta}'_{n-1}\boldsymbol{\rho} - \boldsymbol{\beta}'_n + \boldsymbol{\beta}'_1
\end{aligned}
$$

$$\mathbf{b}'_{n-1} = \boldsymbol{\beta}'_{n-1}$$

$$\mathbf{c}'_{n-1} = \boldsymbol{\beta}'_{n-1}\boldsymbol{\rho} - \boldsymbol{\beta}'_{n} + \boldsymbol{\beta}'_{1}$$

Predicted coefficients: affine model implies

$$\boldsymbol{\beta}'_{n} = \boldsymbol{\beta}'_{n-1}(\boldsymbol{\rho} - \boldsymbol{\Sigma\Lambda}) + \boldsymbol{\beta}'_{1}$$

$$\Rightarrow \mathbf{c}'_{n-1} = \mathbf{b}'_{n-1}\boldsymbol{\Sigma\Lambda}$$

$$\Rightarrow \mathbf{c}'_{n-1} = \mathbf{b}'_{n-1}\boldsymbol{\Sigma}\boldsymbol{\Lambda}$$

Proposal: estimate

$$x_{n-1,t+1} = a_{n-1} + \mathbf{b}'_{n-1}\hat{\mathbf{v}}_{t+1} + \mathbf{c}'_{n-1}\boldsymbol{\xi}_t + e_{n-1,t+1}$$

for $\hat{\mathbf{v}}_{t+1} = \boldsymbol{\xi}_{t+1} - \hat{\mathbf{c}} - \hat{\boldsymbol{\rho}}\boldsymbol{\xi}_t$

by unrestricted OLS separately for each $n \in \mathbb{N}$

Then estimate $\underset{(r\times r)}{\tilde{\boldsymbol{\Lambda}}} = \boldsymbol{\Sigma}\boldsymbol{\Lambda}$ to minimize

sum of squared discrepancies between

$\hat{\mathbf{c}}_{n-1}$ and $\tilde{\boldsymbol{\Lambda}}'\hat{\mathbf{b}}_{n-1}$ across n

$$\Rightarrow \tilde{\boldsymbol{\Lambda}}' = \left(\sum_{n\in\mathbb{N}} \hat{\mathbf{c}}_{n-1}\hat{\mathbf{b}}_{n-1}'\right)\left(\sum_{n\in\mathbb{N}} \hat{\mathbf{b}}_{n-1}\hat{\mathbf{b}}_{n-1}'\right)^{-1}$$

Predicted intercepts:

$$a_{n-1} = \alpha_{n-1} + \boldsymbol{\beta}'_{n-1}\mathbf{c} - \alpha_n + \alpha_1$$

Affine model implies

$$\alpha_n = \alpha_{n-1} + \boldsymbol{\beta}'_{n-1}(\mathbf{c} - \boldsymbol{\Sigma\lambda}) + \alpha_1$$
$$+ (1/2)(\boldsymbol{\beta}'_{n-1}\boldsymbol{\Sigma\Sigma}'\boldsymbol{\beta}_{n-1} + \sigma^2)$$
$$\Rightarrow a_{n-1} = \boldsymbol{\beta}'_{n-1}\boldsymbol{\Sigma\lambda} - (1/2)(\boldsymbol{\beta}'_{n-1}\boldsymbol{\Sigma\Sigma}'\boldsymbol{\beta}_{n-1} + \sigma^2)$$

$$\Rightarrow \; a_{n-1} = \boldsymbol{\beta}'_{n-1}\boldsymbol{\Sigma}\boldsymbol{\lambda} - (1/2)(\boldsymbol{\beta}'_{n-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\boldsymbol{\beta}_{n-1} + \sigma^2)$$

Let $\tilde{\boldsymbol{\lambda}} = \boldsymbol{\Sigma}\boldsymbol{\lambda}$

$$\tilde{a}_{n-1} = \hat{a}_{n-1} + (1/2)(\hat{\mathbf{b}}'_{n-1}\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{\Sigma}}'\hat{\mathbf{b}}_{n-1} + \hat{\sigma}^2)$$

Estimate $\tilde{\boldsymbol{\lambda}}$ by minimizing difference between

$\tilde{a}_{n-1}$ and $\tilde{\boldsymbol{\lambda}}'\hat{\mathbf{b}}_{n-1}$

$$\Rightarrow \; \tilde{\boldsymbol{\lambda}} = \left(\sum_{n\in\mathbb{N}} \hat{\mathbf{b}}_{n-1}\hat{\mathbf{b}}'_{n-1}\right)^{-1}\left(\sum_{n\in\mathbb{N}} \hat{\mathbf{b}}_{n-1}\tilde{a}_{n-1}\right)$$

Summary: we have now estimated

$\mathbf{c}, \boldsymbol{\rho}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}, \boldsymbol{\lambda}, \sigma^2$ using only OLS regressions.

Can estimate α_1 and $\boldsymbol{\beta}_1$ by OLS regression

$$r_t = -\alpha_1 - \boldsymbol{\beta}_1' \boldsymbol{\xi}_t + e_t^{(1)}$$

And then calculate α_n and $\boldsymbol{\beta}_n$ by recursion:

$$\boldsymbol{\beta}_n' = \boldsymbol{\beta}_{n-1}' (\boldsymbol{\rho} - \boldsymbol{\Sigma\Lambda}) + \boldsymbol{\beta}_1'$$

$$\alpha_n = \alpha_{n-1} + \boldsymbol{\beta}_{n-1}' (\mathbf{c} - \boldsymbol{\Sigma\lambda}) + \alpha_1$$
$$+ (1/2)(\boldsymbol{\beta}_{n-1}' \boldsymbol{\Sigma\Sigma}' \boldsymbol{\beta}_{n-1} + \sigma^2)$$

From these we can calculate the predicted yield on any bond

$$y_{nt} = -n^{-1}(\alpha_n + \boldsymbol{\beta}_n' \boldsymbol{\xi}_t).$$

We can then redo the recursions setting $\boldsymbol{\lambda} = \mathbf{0}$ and $\boldsymbol{\Lambda} = \mathbf{0}$ to get predicted yields if investors were risk neutral

$$\boldsymbol{\beta}_n^{RF\prime} = \boldsymbol{\beta}_{n-1}^{RF\prime}\boldsymbol{\rho} + \boldsymbol{\beta}_1^{\prime}$$

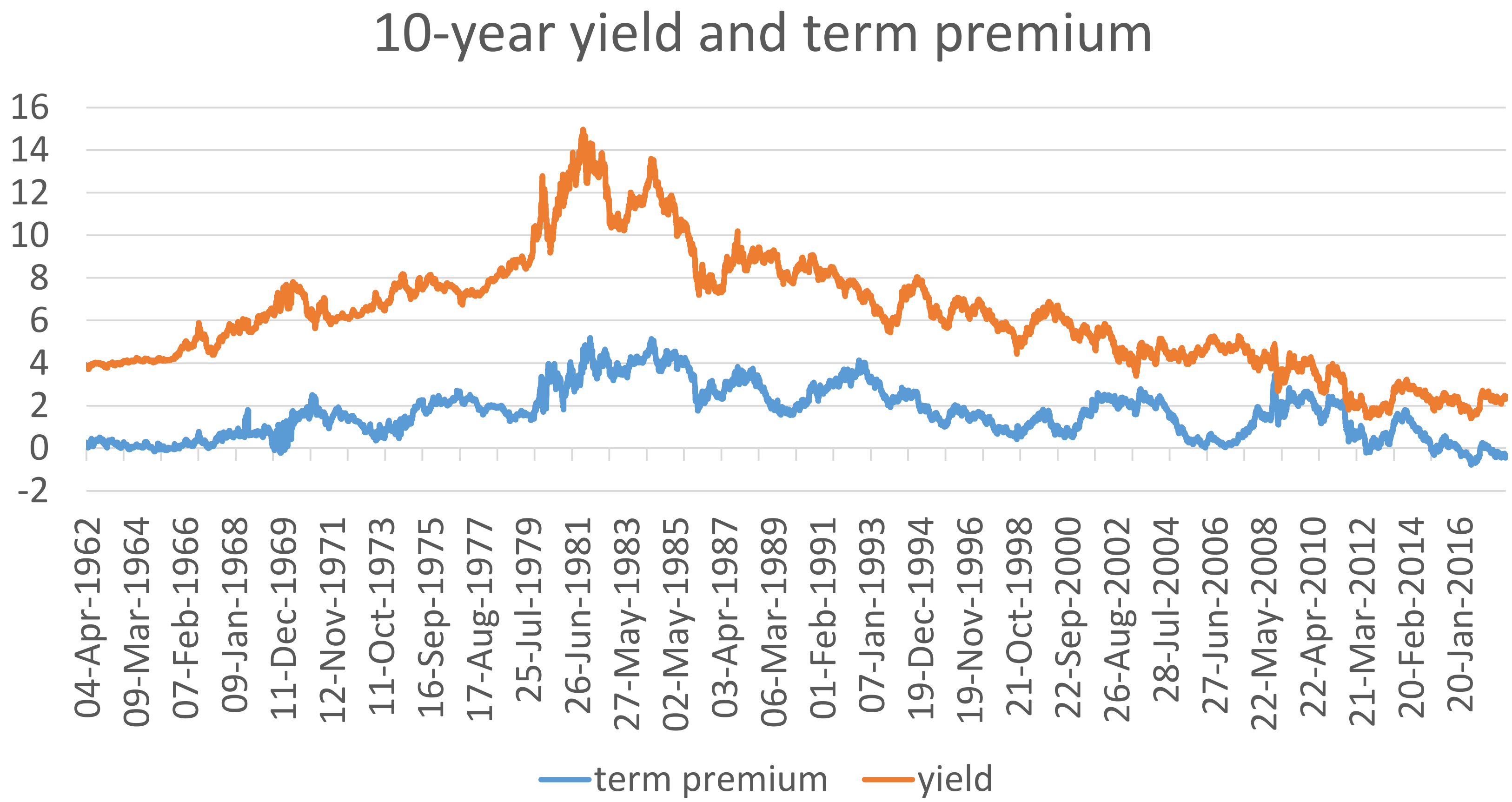
$$\alpha_n^{RF} = \alpha_{n-1}^{RF} + \boldsymbol{\beta}_{n-1}^{RF\prime}\mathbf{c} + \alpha_1$$

$$+ (1/2)(\boldsymbol{\beta}_{n-1}^{\prime}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{\prime}\boldsymbol{\beta}_{n-1} + \sigma^2)$$

$$y_{nt}^{RF} = -n^{-1}(\alpha_n^{RF} + \boldsymbol{\beta}_n^{RF\prime}\boldsymbol{\xi}_t)$$

Can calculate the risk premium as the difference $y_{nt} - y_{nt}^{RF}$

Updated daily at https://www.newyorkfed.org/medialibrary/media/research/data_indicators/ACMTermPremium.xls



D. Dynamic Nelson-Siegel model (Christensen, Diebold, and Rudebusch)

Recall that we often summarize forward curve at date t using function such as

$$f_{nt} = \beta_{0t} + \beta_{1t}\exp(-n/\tau_{1t}) \\ + \beta_{2t}(n/\tau_{1t})\exp(-n/\tau_{1t}) \\ + \beta_{3t}(n/\tau_{2t})\exp(-n/\tau_{2t})$$

Nelson-Siegel: Describe forward rate of any maturity n as smooth function of three magnitudes at t (level, slope, and curvature).

Consider following special case of GATSM normalization and additional restrictions (Bauer, 2011, following Christensen, Diebold, and Rudebusch):

$$\rho^Q = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \gamma & 1-\gamma \\ 0 & 0 & \gamma \end{bmatrix}$$

has eigenvalues $(1, \gamma, \gamma)$

$$\mathbf{c}^Q = \mathbf{0}$$

$$\mathbf{b}_1 = (1, 1, 0)'$$

Implications:

$$\boldsymbol{\xi}_{t+1} = \mathbf{c}^Q + \boldsymbol{\rho}^Q \boldsymbol{\xi}_t + \mathbf{v}_{t+1}^Q$$

$$r_t = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \boldsymbol{\xi}_t$$

$$f_{nt} = \mathbf{b}_1' E_t^Q \boldsymbol{\xi}_{t+n} = \mathbf{b}_1' (\boldsymbol{\rho}^Q)^n \boldsymbol{\xi}_t$$

$$= \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \gamma^n & n\gamma^{n-1}(1-\gamma) \\ 0 & 0 & \gamma^n \end{bmatrix} \begin{bmatrix} \xi_{1t} \\ \xi_{2t} \\ \xi_{3t} \end{bmatrix}$$

$$= \xi_{1t} + \gamma^n \xi_{2t} + n\gamma^{n-1}(1-\gamma)\xi_{3t}$$

$$f_{nt} = \xi_{1t} + \gamma^n \xi_{2t} + n\gamma^{n-1}(1-\gamma)\xi_{3t}$$

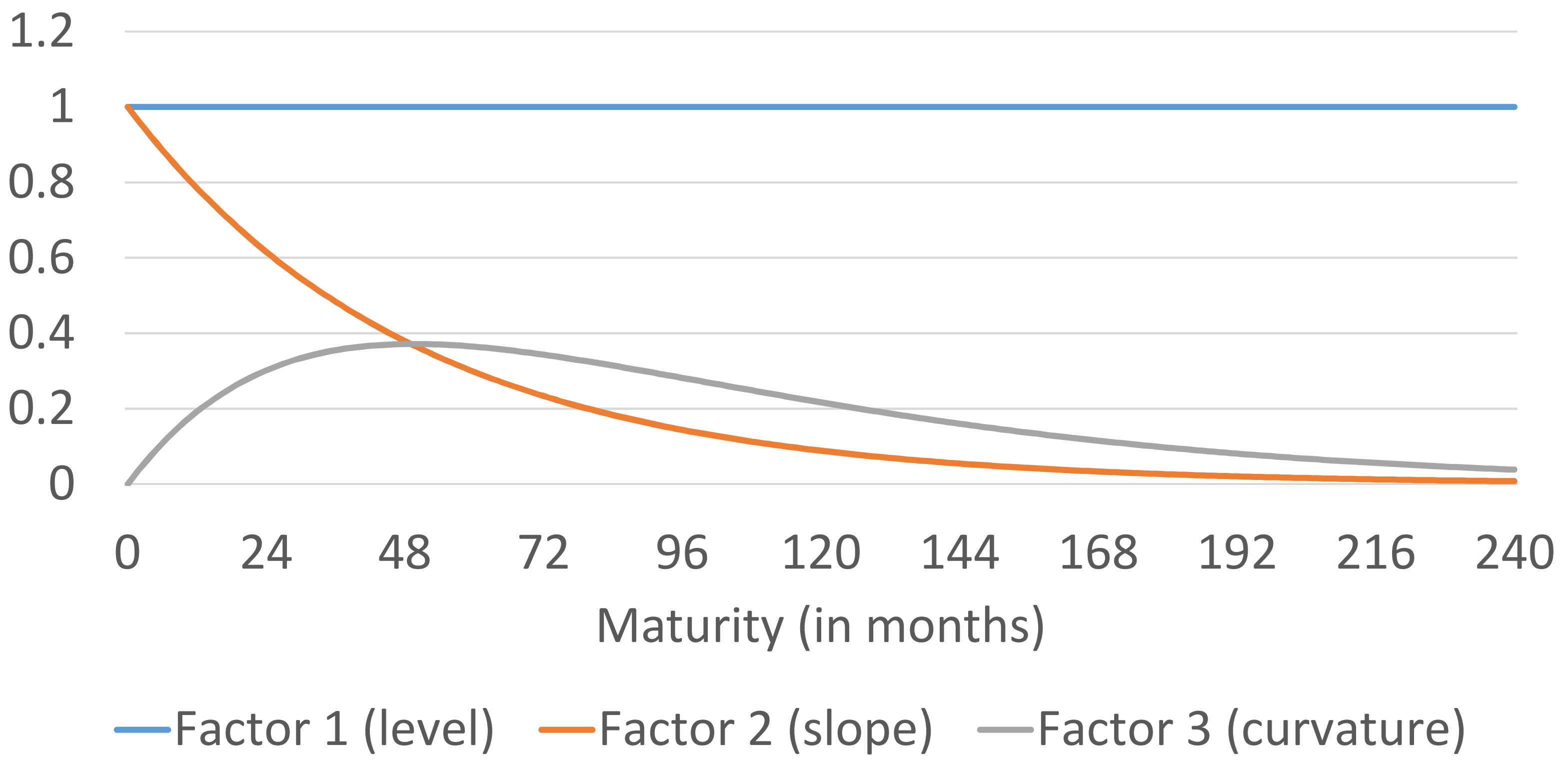
So forward rate f_{nt} loads on

factor 1 with weight 1

factor 2 with weight $\gamma^n \xi_{2t}$

factor 3 with weight $n\gamma^{n-1}(1-\gamma)$

Factor loadings for yields of different maturity (gamma = 0.98)



Dynamic Nelson-Siegel: can then estimate P-measure dynamics for state vector as

$$\boldsymbol{\xi}_t = \mathbf{c} + \boldsymbol{\rho}\boldsymbol{\xi}_{t-1} + \mathbf{v}_t$$

which gives complete dynamic description of process for all yields.

E. Small-sample bias (Bauer, Rudebusch, Wu)

$$\boldsymbol{\xi}_t = \mathbf{c} + \boldsymbol{\rho}\boldsymbol{\xi}_{t-1} + \mathbf{v}_t$$

Caution: for unrestricted OLS estimation of $\boldsymbol{\rho}$, eigenvalues biased downard.

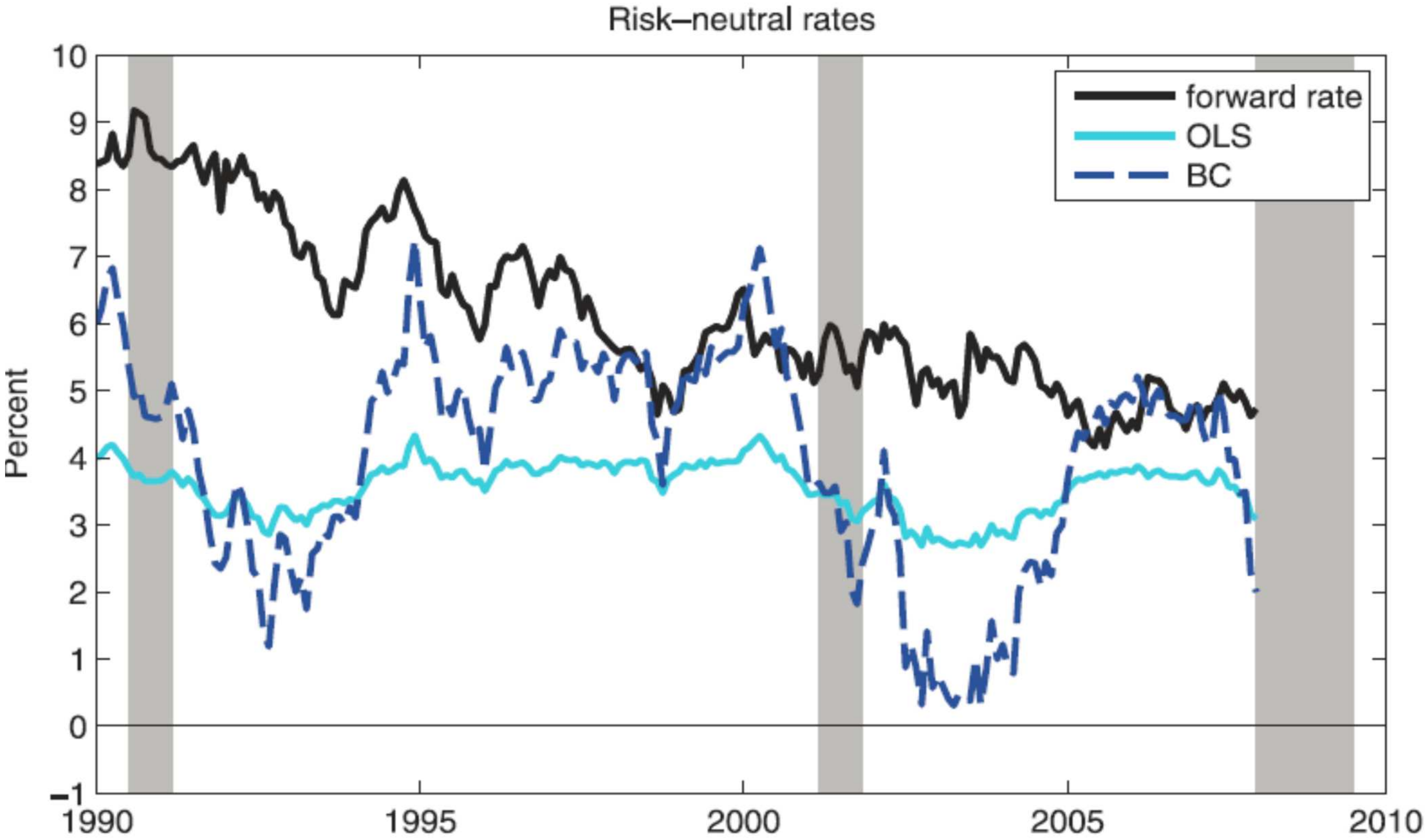
Implication: if $\boldsymbol{\Lambda} = \mathbf{0}$, $\hat{\boldsymbol{\rho}}^n \to \mathbf{0}$ for large n

Empirical models want to attribute most of fluctuation in y_{nt} for large n to $\mathbf{\Lambda}$ (changes in risk premium) not $\boldsymbol{\rho}^n$ (expectations component).

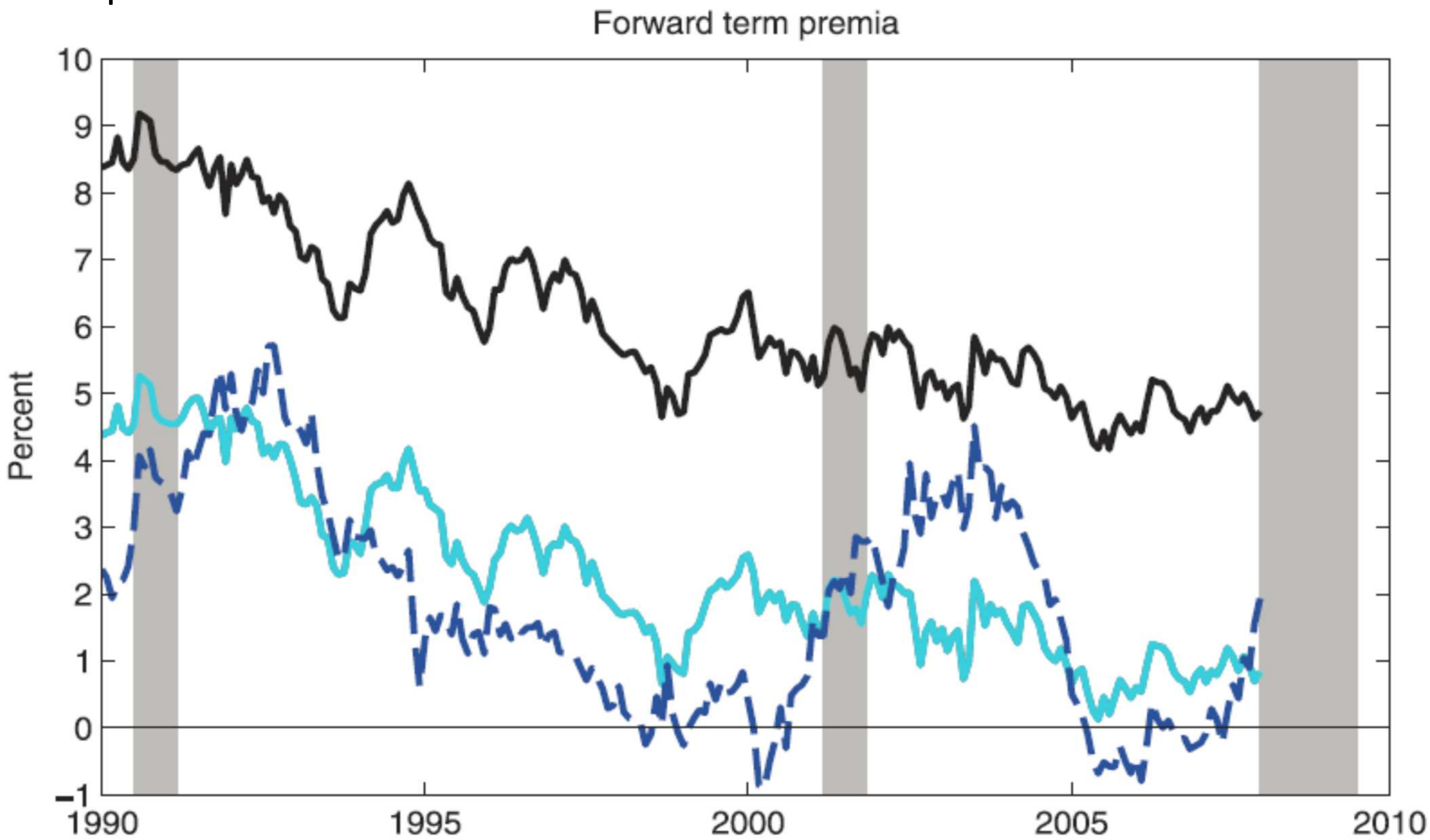
Bauer, Rudebusch and Wu (JBES, 2012).

Correct estimate of $\hat{\boldsymbol{\rho}}$ for small-sample bias.

4-year forward rates (black) and expected 4-year-ahead short rates with and without bias correction



4-year forward rates (black) and component attributed to risk premium with and without bias correction



Big picture:

(1) Methods exist to decompose long yield into expectations component and risk premium.

(2) Identification comes from fact that predictable excess returns attributed to risk premium.

(3) Specific answer sensitive to assumed underlying forecasting model.