Online Appendix – Relative Value in Rates: presenting and challenging the Gauss+ model

This appendix explicitly derives the model solution sticking to the notation in Tuckman and Serrat (2022). All expectation operators and brownian motions are defined under the $\mathbb{Q}$ measure unless otherwise specified.

Solving the model

We start from the three equations, where $\alpha_r, \alpha_m, \alpha_l > 0$:

$$dr_t = -\alpha_r(r_t - m_t)dt$$

$$dm_t = -\alpha_m(m_t - l_t)dt + \sigma_m\left(\rho dW_t^1 + \sqrt{1-\rho^2}dW_t^2\right)$$

$$dl_t = -\alpha_l(l_t - \mu)dt + \sigma_l dW_t^1$$

Where the brownian motions are specified under the risk-neutral measure. Expanding the drift terms:

$$dr_t = (-\alpha_r r_t + \alpha_r m_t)dt$$

$$dm_t = (-\alpha_m m_t + \alpha_m l_t)dt + \sigma_m\left(\rho dW_t^1 + \sqrt{1-\rho^2}dW_t^2\right)$$

$$dl_t = (-\alpha_l l_t + \alpha_l \mu)dt + \sigma_l dW_t^1$$

Now let $x_t = (r_t, m_t, l_t)$ and note how (signs chosen to follow the book):

$$\underbrace{\begin{pmatrix} dr_t \\ dm_t \\ dl_t \end{pmatrix}}_{dx_t} = \underbrace{\begin{bmatrix} -\alpha_r & +\alpha_r & 0 \\ 0 & -\alpha_m & +\alpha_m \\ 0 & 0 & -\alpha_l \end{bmatrix}}_{R(\alpha)} \underbrace{\begin{pmatrix} r_t \\ m_t \\ l_t \end{pmatrix}}_{x_t} dt + \cdots$$

Where $\alpha = (\alpha_r, \alpha_m, \alpha_l)$. If we further define $\bar{\mu} = (\mu, \mu, \mu)$:

$$dx_t = R(\alpha)x_t dt + \underbrace{\begin{pmatrix} 0 \\ 0 \\ +\alpha_l \mu \end{pmatrix}}_{-R(\alpha)\bar{\mu}} dt + \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ \rho\sigma_m & \sqrt{1-\rho^2}\sigma_m & 0 \\ \sigma_l & 0 & 0 \end{bmatrix}}_{\Omega(\sigma)} \underbrace{\begin{pmatrix} 0 \\ dW_t^1 \\ dW_t^2 \end{pmatrix}}_{dW_t}$$

Where $\sigma = (\sigma_m, \sigma_l, \rho)$. Which brings us to a clean matrix formulation:

$$dx_t = R(\alpha)[x_t - \bar{\mu}]dt + \Omega(\sigma)dW_t$$

Now the problem is that $R(\alpha)$ is not diagonal, which means the model is still in cascade form. We want to transform it so as to derive a form with three factors that move independently of each other. Assume we want new factors X_t that are related to x_t in this way, for some A:

$$x_t - \bar{\mu} = AX_t$$

If this is true, then also:

$$dx_t = AdX_t$$

And substituting into the cascade form equation:

$$AdX_t = R(\alpha)AX_t dt + \Omega(\sigma)dW_t$$

$$dX_t = [A^{-1}R(\alpha)A]X_t dt + A^{-1}\Omega(\sigma)dW_t$$

A convenient choice for A would be that of the eigenvectors of $R(\alpha)$. This makes the matrix multiplying X_t diagonal. Recall that the eigenvalues of $R(\alpha)$ are the elements on its diagonal since it's triangular. Hence $A = A(\alpha)$ satisfies:

$$R(\alpha)A(\alpha) = A(\alpha)\begin{bmatrix} -\alpha_r & 0 & 0 \\ 0 & -\alpha_m & 0 \\ 0 & 0 & -\alpha_l \end{bmatrix}$$

Which leads us to the reduced form:

$$dX_t = -\text{diag}(\alpha)X_t dt + A^{-1}(\alpha)\Omega(\sigma)dW_t$$

Rewriting the link between x_t and X_t gives:

$$x_t = A(\alpha)X_t + \bar{\mu}$$

It can be shown that $A(\alpha)$ takes the form:

$$A(\alpha) = \begin{bmatrix} 1 & 1 & 1 \\ 0 & \dfrac{\alpha_r - \alpha_m}{\alpha_r} & \dfrac{\alpha_r - \alpha_l}{\alpha_r} \\ 0 & 0 & \dfrac{(\alpha_r - \alpha_l)(\alpha_m - \alpha_l)}{\alpha_r \alpha_m} \end{bmatrix}$$

Reading this statement only in the first row:

$$r_t = 1'X_t + \mu$$

By no-arbitrage, the price of a zero-coupon bond with maturity τ at time t is given by:

$$P_t(\tau) = E_t^Q\left[e^{-\int_t^{t+\tau} r_s ds}\right]$$

Which, incorporating the relationship with X_t and μ, becomes:

$$P_t(\tau) = E_t\left[\exp\left\{-\int_t^{t+\tau}(1'X_s + \mu)\,ds\right\}\right]$$

Where the expectation is taken with respect to the risk-neutral measure. We need to figure out a way to rewrite this in terms of the parameters. The reasoning is: if we can prove that X_t is a Gaussian process, then also r_s is Gaussian as linear functional of a Gaussian process; ultimately, the random variable inside the expectation is lognormally distributed. If we know the mean and variance characterizing the distribution of X_t then we can solve for the expectation of $e^{-\int_t^{t+\tau} r_s ds}$.

First of all, we solve the SDE governing X_t, so we can show that it's normally distributed and compute its mean and variance.

$$dX_t = -\text{diag}(\alpha)X_t dt + A^{-1}(\alpha)\Omega(\sigma)dW_t$$

With the integrating factor $e^{\text{diag}(\alpha)t} = \text{diag}(e^{\alpha_r t}, e^{\alpha_m t}, e^{\alpha_l t})$:

$$e^{\text{diag}(\alpha)t} \cdot dX_t + e^{\text{diag}(\alpha)t}\text{diag}(\alpha)X_t dt = e^{\text{diag}(\alpha)t} \cdot A^{-1}(\alpha)\Omega(\sigma)dW_t$$

Notice how the left hand side is the differential of $e^{\text{diag}(\alpha)t}X_t$:

$$d\left(e^{\text{diag}(\alpha)t}X_t\right) = e^{\text{diag}(\alpha)t} \cdot dX_t + d\left(e^{\text{diag}(\alpha)t}\right)X_t dt$$

$$d\left(e^{\text{diag}(\alpha)t}X_t\right) = e^{\text{diag}(\alpha)t} \cdot dX_t + e^{\text{diag}(\alpha)t}\text{diag}(\alpha)X_t dt$$

So we can rewrite the expression as:

$$d(e^{\text{diag}(\alpha)t} \cdot X_t) = e^{\text{diag}(\alpha)t} \cdot A^{-1}(\alpha)\Omega(\sigma)dW_t$$

Integrating between t and $t+\tau$ on both sides we get:

$$e^{\text{diag}(\alpha)(t+\tau)}X_{t+\tau} - e^{\text{diag}(\alpha)t}X_t = \int_t^{t+\tau} e^{\text{diag}(\alpha)s} \cdot A^{-1}(\alpha)\Omega(\sigma)dW_s$$

And solving for $X_{t+\tau}$ we get:

$$X_{t+\tau} = e^{-\text{diag}(\alpha)\tau}X_t + \int_t^{t+\tau} e^{-\text{diag}(\alpha)(t+\tau-s)} \cdot A^{-1}(\alpha)\Omega(\sigma)dW_s$$

If we have a starting value X_t that is a constant, then all the randomness in $X_{t+\tau}$ comes from the stochastic integral. Since the stochastic integral is a linear function of the brownian increment vectors dW_s then the integral follows a normal distribution. And thus also $X_{t+\tau}$ follows a normal. Since the stochastic integral has zero mean, we easily compute the expectation of $X_{t+\tau}$:

$$E_t[X_{t+\tau}] = e^{-\text{diag}(\alpha)\tau}X_t$$

By ito isometry $V\left(\int H_s dW_s\right) = \int H_s H_s' ds$, we can compute its variance:

$$V_t(X_{t+\tau}) = V_t\left(\int_t^{t+\tau} e^{-\text{diag}(\alpha)(t+\tau-s)} \cdot A^{-1}(\alpha)\Omega(\sigma)dW_s\right)$$

$$V_t(X_{t+\tau}) = \int_t^{t+\tau} \left[e^{-\text{diag}(\alpha)(t+\tau-s)} \cdot A^{-1}(\alpha)\Omega(\sigma)\right]\left[e^{-\text{diag}(\alpha)(t+\tau-s)} \cdot A^{-1}(\alpha)\Omega(\sigma)\right]' ds$$

$$V_t(X_{t+\tau}) = \int_t^{t+\tau} e^{-\text{diag}(\alpha)(t+\tau-s)}[A^{-1}(\alpha)\Omega(\sigma)][\Omega(\sigma)'A^{-1}(\alpha)']e^{-\text{diag}(\alpha)(t+\tau-s)} ds$$

Taking out terms that do not depend on the variable with respect to which we're integrating:

$$V_t(X_{t+\tau}) = [A^{-1}(\alpha)\Omega(\sigma)][\Omega(\sigma)'A^{-1}(\alpha)'] \int_t^{t+\tau} e^{-\text{diag}(\alpha)(t+\tau-s)}e^{-\text{diag}(\alpha)(t+\tau-s)} ds$$

Evaluating $V_t(X_{t+\tau})$ entrywise:

$$[V_t(X_{t+\tau})]_{ij} = \{[A^{-1}(\alpha)\Omega(\sigma)][\Omega(\sigma)'A^{-1}(\alpha)']\}_{ij} \int_t^{t+\tau} e^{-(\alpha_i+\alpha_j)(t+\tau-s)} ds$$

The integral can be solved easily with a change of variable $u = t + \tau - s$

$$\int_t^{t+\tau} e^{-(\alpha_i+\alpha_j)(t+\tau-s)} ds = \int_0^{\tau} e^{-(\alpha_i+\alpha_j)u}\, du = \left[\frac{e^{-(\alpha_i+\alpha_j)u}}{-(\alpha_i+\alpha_j)}\right]_0^{\tau} = \frac{e^{-(\alpha_i+\alpha_j)\tau}}{-(\alpha_i+\alpha_j)} + \frac{1}{-(\alpha_i+\alpha_j)} = \frac{1-e^{-(\alpha_i+\alpha_j)\tau}}{\alpha_i+\alpha_j}$$

Which leads to the formulation:

$$[V_t(X_{t+\tau})]_{ij} = \{[A^{-1}(\alpha)\Omega(\sigma)][\Omega(\sigma)'A^{-1}(\alpha)']\}_{ij} \cdot \frac{1-e^{-(\alpha_i+\alpha_j)\tau}}{\alpha_i+\alpha_j}$$

Ok. So let's go back to the formulation of $P_t(\tau)$ in terms of X_t. We simplify it a bit:

$$P_t(\tau) = E_t\left[\exp\left\{\left(-\int_t^{t+\tau} 1'X_s\, ds\right) + \mu\tau\right\}\right]$$

The integral $\int_t^{t+\tau} 1'X_s\, ds$ is a linear function of X_s, which as we said follows a normal conditional to X_t for all $s > t$. Hence the integral follows a univariate normal, with a certain mean m and variance v^2. If we can compute m and v, we then know that:

$$P_t(\tau) = E_t\left[\exp\left\{\left(-\int_t^{t+\tau} 1'X_s\, ds\right)\right\}\right]e^{-\mu\tau} = e^{-\mu\tau-m+\frac{1}{2}v^2}$$

Ok. Now let's compute m and v^2. We start with m:

$$m = E_t\left[\left(\int_t^{t+\tau} 1' X_s\, ds\right)\right] = \int_t^{t+\tau} 1' E_t[X_s]\, ds$$

Since $|1'X_s|$ has finite expectation, we can swap integration and summation by Fubini's theorem Then by substituting in the formula we derived for $E_t[X_s]$ and with a change of variable $u = s - t$ we get:

$$m = \int_t^{t+\tau} 1' e^{-\mathrm{diag}(\alpha)(s-t)} X_t\, ds = X_t \int_0^{\tau} 1' e^{-\mathrm{diag}(\alpha)u}\, du$$

$$m = \sum_{i=1}^{3}\left(X_t^i \int_0^t e^{-\alpha_i u} du\right) = \sum_{i=1}^{3}\left(X_t^i \frac{1 - e^{-\alpha_i \tau}}{\alpha_i}\right)$$

Now define a vector $\tilde{B}(\tau, \alpha)$ of the form:

$$\tilde{B}(\tau, \alpha) \coloneqq \begin{pmatrix} \frac{1 - e^{-\alpha_r \tau}}{\alpha_r} & \frac{1 - e^{-\alpha_m \tau}}{\alpha_m} & \frac{1 - e^{-\alpha_l \tau}}{\alpha_l} \end{pmatrix}'$$

Then m can be conveniently rewritten as:

$$m(\tau, \alpha) = \tilde{B}(\tau, \alpha) X_t$$

Now, to compute the variance of the integral, let's rewrite it with the solution of the SDE for X_t:

$$1' X_{t+\tau} = 1'\left(e^{-\mathrm{diag}(\alpha)\tau} X_t + \int_t^{t+\tau} e^{-\mathrm{diag}(\alpha)(t+\tau-s)} \cdot A^{-1}(\alpha)\Omega(\sigma) dW_s\right)$$

$$1' X_{t+\tau} = 1' e^{-\mathrm{diag}(\alpha)\tau} X_t + 1' \int_t^{t+\tau} e^{-\mathrm{diag}(\alpha)(t+\tau-s)} \cdot A^{-1}(\alpha)\Omega(\sigma) dW_s$$

The variance of the integral of $1'X_s$ only comes from the integral of the stochastic integral:

$$v^2 = V_t\left(\int_t^{t+\tau} 1' X_s\, ds\right) = V_t\left(\int_t^{t+\tau}\left(\int_t^{s} 1' e^{-\mathrm{diag}(\alpha)(s-u)} \cdot A^{-1}(\alpha)\Omega(\sigma) dW_u\right) ds\right)$$

Since the term $H(s, u) = 1' e^{-\mathrm{diag}(\alpha)(s-u)} \cdot A^{-1}(\alpha)\Omega(\sigma)$ is square integrable, by Fubini we can swap the orders of integration:

$$v^2 = V_t\left(\int_t^{t+\tau}\left(\int_u^{t+\tau} 1' e^{-\mathrm{diag}(\alpha)(s-u)} ds\right) \cdot A^{-1}(\alpha)\Omega(\sigma) dW_u\right)$$

Setting $z = s - u$ in the inner integral we get

$$\int_u^{t+\tau} 1' e^{-\mathrm{diag}(\alpha)(s-u)} ds = \int_0^{t+\tau-u} 1' e^{-\mathrm{diag}(\alpha) z} dz = \left(\frac{1-e^{\alpha_r(t+\tau-u)}}{\alpha_r} \quad \frac{1-e^{\alpha_m(t+\tau-u)}}{\alpha_m} \quad \frac{1-e^{\alpha_l(t+\tau-u)}}{\alpha_l} \right)' : \\ = g(t+\tau-u)$$

Plugging this new term into the variance of the other integral we get:

$$v^2 = V_t\left(\int_t^{t+\tau} g(t+\tau-u)' A^{-1}(\alpha)\Omega(\sigma)\, dW_u\right)$$

We use Ito isometry and we obtain:

$$v^2 = \int_t^{t+\tau} [g(t+\tau-u)' A^{-1}(\alpha)\Omega(\sigma)][g(t+\tau-u)' A^{-1}(\alpha)\Omega(\sigma)]'\, du$$

Expanding componentwise the quadratic form inside the integral:

$$v^2 = \sum_{i,j} [A^{-1}(\alpha)\Omega(\sigma)\Omega(\sigma)' A^{-1}(\alpha)']_{ij} \int_t^{t+\tau} g(t+\tau-u)'_i g(t+\tau-u)_j\, du$$

The TS appendix defines $\sigma_{ij} = [A^{-1}(\alpha)\Omega(\sigma)\Omega(\sigma)' A^{-1}(\alpha)']_{ij}$ for ease of notation.

$$v^2 = \sum_{i,j} \left\{ \sigma_{ij} \int_t^{t+\tau} \left[\frac{1-e^{\alpha_i(t+\tau-u)}}{\alpha_i}\right]\left[\frac{1-e^{\alpha_j(t+\tau-u)}}{\alpha_j}\right] du \right\}$$

By the change of variable $h = t+\tau-u$, and expanding the product of the two exponentials:

$$v^2 = \sum_{i,j} \left\{ \frac{\sigma_{ij}}{\alpha_i\alpha_j} \int_0^{\tau} (1-e^{\alpha_i h})(1-e^{\alpha_j h})\, dh \right\}$$

$$v^2 = \sum_{i,j} \left\{ \frac{\sigma_{ij}}{\alpha_i\alpha_j} \int_0^{\tau} 1 - e^{-\alpha_i h} - e^{-\alpha_j h} + e^{-(\alpha_i+\alpha_j)h}\, dh \right\}$$

$$v^2 = \sum_{i,j} \left\{ \frac{\sigma_{ij}}{\alpha_i\alpha_j} \left[\tau - \frac{1-e^{-\alpha_i \tau}}{\alpha_i} - \frac{1-e^{-\alpha_j \tau}}{\alpha_j} + \frac{1-e^{-(\alpha_i+\alpha_j)\tau}}{\alpha_i+\alpha_j} \right] \right\}$$

Ok! So now we can plug m and v into the expression for the price of a ZCB. More cleanly, let's define the price of a ZCB in terms of its ZCB yield:

$$P_t(\tau) = \exp(-y_t(\tau)\tau)$$

$$y_t(\tau) = -\frac{1}{\tau} \log P_t(\tau)$$

Recall that $P_t(\tau) = e^{-\mu\tau - m + \frac{1}{2}v^2}$ so then:

$$y_t(\tau) = -\frac{1}{\tau}\left\{-\mu\tau - m + \frac{1}{2}v^2\right\}$$

$$y_t(\tau) = \mu + \frac{m}{\tau} - \frac{1}{2\tau}v^2$$

Good. Now, in order to stick to the book's notation, define a new vector:

$$B(\tau,\alpha) = \frac{1}{\tau}\tilde{B}(\tau,\alpha) = \begin{pmatrix}\frac{1-e^{-\alpha_r\tau}}{\alpha_r\tau} & \frac{1-e^{-\alpha_m\tau}}{\alpha_m\tau} & \frac{1-e^{-\alpha_l\tau}}{\alpha_l\tau}\end{pmatrix}'$$

And also a new quantity $C(\tau,\alpha,\sigma)$ defined as:

$$C(\tau,\alpha,\sigma) = \sum_{i=1}^{3}\sum_{j=1}^{3}\frac{\sigma_{ij}}{2\alpha_i\alpha_j}\left[1 - \frac{1-e^{-\alpha_i\tau}}{\alpha_i\tau} - \frac{1-e^{-\alpha_j\tau}}{\alpha_j\tau} + \frac{1-e^{-(\alpha_i+\alpha_j)\tau}}{(\alpha_i+\alpha_j)\tau}\right]$$
$$= \sum_{i=1}^{3}\sum_{j=1}^{3}\frac{\sigma_{ij}}{2\alpha_i\alpha_j}\left[1 - B(\tau,\alpha)_i - B(\tau,\alpha)_j + \frac{1-e^{-(\alpha_i+\alpha_j)\tau}}{(\alpha_i+\alpha_j)\tau}\right] = \frac{1}{2\tau}v^2$$

Since $m = \tilde{B}(t,\alpha)X_t$ we finally get to the formula from the book:

$$y_t(\tau) = \mu - C(\tau,\alpha,\sigma) + B(\tau,\alpha)X_t$$

Now, this gives us the link between a yield of a ZCB with maturity τ at time t (so essentially the whole yield curve) and the reduced form factors X_t. How do we link yields $y_t(\tau)$ to the cascade form factors x_t that we began with?

Recall that $x_t = A(\alpha)X_t + \bar{\mu}$, which equivalently can be rewritten as $X_t = A^{-1}(\alpha)(x_t - \bar{\mu})$. We substitute this into the formula:

$$y_t(\tau) = \mu - C(\tau,\alpha,\sigma) + B(\tau,\alpha)[A^{-1}(\alpha)(x_t - \bar{\mu})]$$

$$y_t(\tau) = \mu\big(1 - \mathbf{1}'B(\tau,\alpha)A^{-1}(\alpha)\big) - C(\tau,\alpha,\sigma) + B(\tau,\alpha)A^{-1}(\alpha)x_t$$

$$y_t(\tau) = \mu\big(1 - \mathbf{1}'\gamma(\tau,\alpha)\big) - C(\tau,\alpha,\sigma) + \gamma(\tau,\alpha)x_t$$

The book calls $\gamma(\tau,\alpha) := B(\tau,\alpha)A^{-1}(\alpha)$ the vector of factor loadings. In other words, they express the partial derivative of $y_t(\tau)$ to each of the cascade form factors.

We can also obtain an expression for the forward rate at time t for a forward between times τ and $\tau + \tau'$:

$$f_t(\tau) = \mu\big(1 - \mathbf{1}'\gamma^*(\tau,\alpha,\tau')\big) + \gamma^*(\tau,\alpha,\tau')x_t - C^*(\tau,\alpha,\sigma,\tau')$$

Where we have defined factor loadings and a convexity term for forwards as:

$$\gamma^*(\tau,\alpha,\tau') = \frac{1}{\tau'}\cdot[B(\tau+\tau',\alpha) - B(\tau,\alpha)]A(\alpha)^{-1}$$

$$C^*(\tau,\alpha,\sigma,\tau') = \frac{1}{\tau'}\cdot[C(\tau+\tau',\alpha,\sigma) - C(\tau,\alpha,\sigma)]$$

Calibration of the model

Tuckman assumes that r_t is the observed fed funds rate at t. It can be shown that the first factor loading (the loading to the short rate) depends only on α_r.

It can be shown that $A^{-1}(\alpha)$ takes the form:

$$A^{-1}(\alpha) = \begin{pmatrix} 1 & -\frac{\alpha_r}{\alpha_r-\alpha_m} & \frac{\alpha_r\alpha_m}{(\alpha_r-\alpha_m)(\alpha_r-\alpha_l)} \\ 0 & \frac{\alpha_r}{\alpha_r-\alpha_m} & \frac{\alpha_r\alpha_m}{(\alpha_r-\alpha_m)(\alpha_m-\alpha_l)} \\ 0 & 0 & \frac{\alpha_r\alpha_m}{(\alpha_r-\alpha_l)(\alpha_m-\alpha_l)} \end{pmatrix}$$

Thus the product $\gamma(\tau,\alpha) = B(\tau,\alpha)A^{-1}(\alpha)$ has first entry equal to $B(\tau,\alpha)_1 = \frac{1-e^{-\alpha_r\tau}}{\alpha_r\tau}$. The procedure in Tuckman works as follows:

1. Take a candidate value for α_r
2. Subtract $\gamma(\tau,\alpha)_1\cdot r_t$ from both sides of the relationship between yields and factors

$$y_t(\tau) - \gamma(\tau,\alpha)_1 r_t = \mu\big(1-\mathbf{1}'\gamma(\tau,\alpha)\big) - C(\tau,\alpha,\sigma) + \gamma(\tau,\alpha)x_t - \gamma(\tau,\alpha)_1 r_t$$

The equation now does not involve r_t any longer on its RHS, since

$$\gamma(\tau,\alpha)x_t - \gamma(\tau,\alpha)_1 r_t = \gamma(\tau,\alpha)_2 m_t + \gamma(\tau,\alpha)_3 l_t$$

Now define $\tilde{y}_t(\tau) = y_t(\tau) - \gamma(\tau,\alpha)_1 r_t$ and $\tilde{x}_t = (m_t, l_t)$. Also, let's write $\tilde{\gamma}(\tau,\alpha) = (\gamma(\tau,\alpha)_2, \gamma(\tau,\alpha)_3)$. We have:

$$\tilde{y}_t(\tau) = \mu\big(1-\mathbf{1}'\gamma(\tau,\alpha)\big) - C(\tau,\alpha,\sigma) + \tilde{\gamma}(\tau,\alpha)\tilde{x}_t$$

3. We now assume that two points on the yield curve are priced exactly by the model. For example, the 2y and 10y yields of the curve. We stack them into a vector $yb_t = \big(y_t(2), y_t(10)\big)'$. With the usual transformation, we also define:

$$\widetilde{yb}_t = \begin{bmatrix} \tilde{y}(2) \\ \tilde{y}(10) \end{bmatrix} = \begin{bmatrix} y(2) - \gamma(\tau,\alpha)_1 r_t \\ y(10) - \gamma(\tau,\alpha)_1 r_t \end{bmatrix}$$

This means we can invert the equation linking a generic yield $\tilde{y}_t(\tau)$ to $\tilde{x}_t$ to solve for $\tilde{x}_t$ based on $\widetilde{yb}_t$:

$$\begin{bmatrix} \tilde{y}_t(2) \\ \tilde{y}_t(10) \end{bmatrix} = \begin{bmatrix} \mu(1 - \mathbf{1}'\gamma(2,\alpha) \\ \mu(1 - \mathbf{1}'\gamma(10,\alpha) \end{bmatrix} - \begin{bmatrix} C(2,\alpha,\sigma) \\ C(10,\alpha,\sigma) \end{bmatrix} + \begin{pmatrix} \gamma(2,\alpha)_2 & \gamma(2,\alpha)_3 \\ \gamma(10,\alpha)_2 & \gamma(10,\alpha)_3 \end{pmatrix} \begin{bmatrix} m_t \\ l_t \end{bmatrix}$$

If we take the first difference of both sides of the equation, we get:

$$\begin{bmatrix} \Delta\tilde{y}_t(2) \\ \Delta\tilde{y}_t(10) \end{bmatrix} = \underbrace{\begin{pmatrix} \gamma(2,\alpha)_2 & \gamma(2,\alpha)_3 \\ \gamma(10,\alpha)_2 & \gamma(10,\alpha)_3 \end{pmatrix}}_{Y_b(\alpha)} \begin{bmatrix} \Delta m_t \\ \Delta l_t \end{bmatrix}$$

Which then gives us $\Delta\tilde{x}_t = Y_b(\alpha)^{-1}\Delta\widetilde{yb}_t$. We use this to relate the first difference of any yields in terms of the first difference of the 2y and 10y.

$$\Delta\tilde{y}_t(\tau) = \tilde{\gamma}(\tau,\alpha)\Delta\tilde{x}_t$$

$$\Delta\tilde{y}_t(\tau) = [\tilde{\gamma}(\tau,\alpha)Y_b(\alpha)^{-1}]\Delta\widetilde{yb}_t$$

In the first stage, we estimate $\beta(\tau) = \tilde{\gamma}(\tau,\alpha)Y_b(\alpha)^{-1}$ with an OLS regression for each maturity τ.

$$\hat{\beta}(\tau) = \left(\Delta\widetilde{yb}_t'\Delta\widetilde{yb}_t\right)^{-1}\Delta\widetilde{yb}_t'\Delta y$$

Then, as $\tilde{\gamma}(\tau,\alpha)Y_b(\alpha)^{-1}$ is a nonlinear function of α, we pick α solving:

$$\min_{\alpha} \sum_{\tau} \left|\left|\tilde{\gamma}(\tau,\alpha)Y_b(\alpha)^{-1} - \hat{\beta}(\tau)\right|\right|^2$$

4. Once α is estimated, we use the estimate $\hat{\alpha}$ to estimate σ minimizing the distance between model implied yield volatilities and realized volatilities of constant-maturity yields.

   According to the model, the benchmark yields $\Delta\widetilde{yb}_t$ should have varcov equal to:

$$V_t\left(\Delta\widetilde{yb}_t\right) = Y_b(\alpha)V_t(\Delta\tilde{x}_t)Y_b(\alpha)'$$

   We call $\widetilde{\Omega}(\sigma)$ the diffusion matrix of (m_t, l_t), taking the same form as $\Omega(\alpha)$ but without the zeros in the entries relating to r_t. This gives:

$$V_t(\Delta\tilde{x}_t) = \widetilde{\Omega}(\sigma)\widetilde{\Omega}(\sigma)'\Delta t = \begin{pmatrix} \sigma_m^2 & \rho\sigma_m\sigma_l \\ \rho\sigma_m\sigma_l & \sigma_l^2 \end{pmatrix}\Delta t$$

   So the model-implied variance of changes in benchmark yields shall be:

$$V_t\left(\Delta\widetilde{yb}_t\right) = \left[Y_b(\alpha)\widetilde{\Omega}(\sigma)\widetilde{\Omega}(\sigma)'Y_b(\alpha)'\right]\Delta t$$

   Using this, we can compute the model-implied variance of $\Delta\tilde{y}_t(\tau)$ for a generic τ:

$$V_t\left(\Delta\tilde{y}_t(\tau)\right) = [\tilde{\gamma}(\tau,\alpha)Y_b(\alpha)^{-1}]V_t\left(\Delta\widetilde{yb}_t\right)[\tilde{\gamma}(\tau,\alpha)Y_b(\alpha)^{-1}]'$$

$$V_t\big(\Delta\tilde{y}_t(\tau)\big) = \tilde{\gamma}(\tau,\alpha)\underbrace{Y_b(\alpha)^{-1}Y_b(\alpha)}_{I}\cdot\tilde{\Omega}(\sigma)\tilde{\Omega}(\sigma)'\cdot\underbrace{Y_b(\alpha)'(Y_b(\alpha)^{-1})'}_{I}\tilde{\gamma}(\tau,\alpha)'$$

$$V_t\big(\Delta\tilde{y}_t(\tau)\big) = \tilde{\gamma}(\tau,\alpha)\tilde{\Omega}(\sigma)\tilde{\Omega}(\sigma)'\tilde{\gamma}(\tau,\alpha)'$$

We shall compute the difference between the varcov of all $\Delta\tilde{y}_t(\tau)$ for all τ, and the sample varcov of all yields. Note that the model implied varcov is calculated with our previous estimate of α.

$$\hat{\sigma} = \arg\min_{\sigma}\sum_{\tau}\left\|\tilde{\gamma}(\tau,\hat{\alpha})\tilde{\Omega}(\sigma)\tilde{\Omega}(\sigma)'\tilde{\gamma}(\tau,\hat{\alpha})' - \widehat{\text{Var}}\big(\Delta y(\tau)\big)\right\|^2$$

5. Lastly, we pick μ to minimize squared yield errors in time across all maturities. How? Given our estimates of all other parameters, and a candidate value for μ, we can obtain the latent factors (m_t, l_t) inverting the equation that links them to the 2y and 10y rates:

$$\begin{pmatrix}\gamma(2,\alpha)_2 & \gamma(2,\alpha)_3\\ \gamma(10,\alpha)_2 & \gamma(10,\alpha)_3\end{pmatrix}^{-1}\left(\begin{bmatrix}\tilde{y}_t(2)\\ \tilde{y}_t(10)\end{bmatrix} - \begin{bmatrix}\mu(1-\mathbf{1}'\gamma(2,\alpha)\\ \mu(1-\mathbf{1}'\gamma(10,\alpha)\end{bmatrix} + \begin{bmatrix}C(2,\alpha,\sigma)\\ C(10,\alpha,\sigma)\end{bmatrix}\right) = \begin{bmatrix}m_t\\ l_t\end{bmatrix}$$

Or otherwise by inverting the equation linking the latent factors to two forward rates, say the 2-year and 10-year forward 1y rates. Recall that the 2-year and 10-year forward 1-year rates are linked to x_t via:

$$f_t(2) = \mu\big(1-\mathbf{1}'\gamma^*(2,\alpha,1)\big) + \gamma^*(2,\alpha,1)x_t - C^*(2,\alpha,\sigma,1)$$

$$f_t(10) = \mu\big(1-\mathbf{1}'\gamma^*(10,\alpha,1)\big) + \gamma^*(10,\alpha,1)x_t - C^*(10,\alpha,\sigma,1)$$

Which means that we can solve for x_t with:

$$\begin{bmatrix}f_t(2)\\ f_t(10)\end{bmatrix} - \begin{bmatrix}\gamma^*(2,\alpha,1)_1 r_t\\ \gamma^*(2,\alpha,1)_1 r_t\end{bmatrix} = \begin{bmatrix}\mu\big(1-\mathbf{1}'\gamma^*(2,\alpha,1)\big)\\ \mu\big(1-\mathbf{1}'\gamma^*(10,\alpha,1)\big)\end{bmatrix} + \begin{bmatrix}\gamma^*(2,\alpha,1)_2 & \gamma^*(2,\alpha,1)_3\\ \gamma^*(10,\alpha,1)_2 & \gamma^*(10,\alpha,1)_3\end{bmatrix}\begin{bmatrix}m_t\\ l_t\end{bmatrix} - \begin{bmatrix}C^*(2,\alpha,\sigma,1)\\ C^*(10,\alpha,\sigma,1)\end{bmatrix}$$

Adopting the same notation that we used when subtracting the contribution of the short rate from zero coupon rates, so that $\tilde{f}_t(\tau) = f_t(\tau) - \gamma^*(\tau,\alpha,\tau')_1 r_t$, we have:

$$\begin{bmatrix}\gamma^*(2,\alpha,1)_2 & \gamma^*(2,\alpha,1)_3\\ \gamma^*(10,\alpha,1)_2 & \gamma^*(10,\alpha,1)_3\end{bmatrix}^{-1}\left(\begin{bmatrix}\tilde{f}_t(2)\\ \tilde{f}_t(10)\end{bmatrix} - \begin{bmatrix}\mu\big(1-\mathbf{1}'\gamma^*(2,\alpha,1)\big)\\ \mu\big(1-\mathbf{1}'\gamma^*(10,\alpha,1)\big)\end{bmatrix} + \begin{bmatrix}C^*(2,\alpha,\sigma,1)\\ C^*(10,\alpha,\sigma,1)\end{bmatrix}\right) = \begin{bmatrix}m_t\\ l_t\end{bmatrix}$$

The difference between the two methods is that inverting the equation with forward rates gives a clearer meaning to m_t and l_t. They're factors tracking the risk-neutral expected path of the short rate.

Once we've done that, we use the fitted factors to price all yields in conjunction with the candidate value for μ and the other six estimated parameters. This allows us to compute a difference between model-implied and actual yields, which we use to pick the best $\hat{\mu}$:

$$\hat{\mu} = \arg\min_{\mu} \sum_{t=1}^{T} ||Y_t - y_t||$$

Where Y_t is the vector of model implied yields across all maturities, and y_t is the vector of actual yields across all maturities, both at a given time t.

Deriving risk premia

We assume only the long factor earns a risk premium. Define $P(t,T)$ to be the price of a bond with maturity T at time t. We're going to derive its dynamics under Q. We know that under Q, its expected return over a very small time horizon dt shall be $r_t dt$, because every discounted price is a Q-martingale. For what concerns its diffusion, we will derive it starting from the very first source of risk: the yield. Recall the equation:

$$y(t,T) = \mu\big(1 - \mathbf{1}'\gamma(T-t)\big) - C(T-t) + \gamma(T-t)x_t$$

We omit dependence on model parameters for ease of writing. Now, since $y(t,T)$ is affine in x_t:

$$dy(t,T) = [\mu_y]dt + \gamma(T-t)\Omega dW_t$$

We denote by μ_y the drift term for $y(t,T)$, which we're not interested in calculating for now. The log price is an affine function of the yield, since $p(t,T) \coloneqq \log P(t,T) = -(T-t)y(t,T)$. Hence:

$$dp(t,T) = [\mu_p]dt + [-(T-t)\gamma(T-t)\Omega]dW_t$$

Once again, we don't derive μ_p for now since we're only interested in diffusions for now. Now, the price of a bond $P(t,T)$ is an exponential function of $p(t,T)$, so:

$$dP(t,T) = \left[\frac{\partial P(t,T)}{\partial t} + \frac{\partial P(t,T)}{\partial p}\mu_p + \frac{1}{2}\frac{\partial^2 P(t,T)}{\partial p^2}(-(T-t)\gamma(T-t)\Omega)\cdot(-(T-t)\gamma(T-t)\Omega)'\right]dt + \left[\frac{\partial P(t,T)}{\partial p}(-(T-t)\gamma(T-t)\Omega)\right]dW_t$$

And since:

$$\frac{\partial P(t,T)}{\partial t} = 0 \qquad \frac{\partial P(t,T)}{\partial p} = \frac{\partial^2 P(t,T)}{\partial p^2} = P$$

$$\frac{dP(t,T)}{P} = \left[\mu_p + \frac{1}{2}(T-t)^2\gamma(T-t)\Omega\Omega'\gamma(T-t)'\right]dt + [-(T-t)\gamma(T-t)\Omega]dW_t$$

But as we mentioned in the beginning we must have $E_t\left[\frac{dP(t,T)}{P}\right] = r_t dt$. So then:

$$\frac{dP(t,T)}{P} = r_t dt - (T-t)\gamma(T-t)\Omega dW_t$$

By Girsanov, if we switch to the P measure, we would define a brownian motion W_t^P with:

$$dW_t^P = dW_t + \boldsymbol{\lambda}_t dt$$

Where $\boldsymbol{\lambda}_t$ is a 3-item vector of prices of risk, one per factor. We assume that only the long factor earns a risk premium. So that:

$$\boldsymbol{\lambda}_t = (0 \quad 0 \quad \lambda_t)$$

This means that under the P measure, the differential of P should look like this:

$$\frac{dP(t,T)}{P(t,T)} = \left[r_t + \underbrace{\lambda_t (T-t)\gamma(T-t,\alpha)_3 \sigma_l}_{\text{compensation for risk}}\right] dt + [-(T-t)\gamma(T-t,\alpha)\Omega] dW_t^P$$

The compensation for risk is the product of the bond's duration $(T-t)$, the loading to the long factor $\gamma(T-t,\alpha)_3$, the volatility of the long factor, and a price of risk λ_t. Now define:

$$\mu_P = P[r_t + \lambda_t (T-t)\gamma(T-t,\alpha)_3 \sigma_l]$$

$$\sigma_P = P[-(T-t)\gamma(T-t,\alpha)\Omega]$$

Now consider the log price of a bond, $p(t,T) := \log P(t,T)$. By Ito's lemma:

$$dp(t,T) = \left[\frac{1}{P}\mu_P - \frac{1}{2}\frac{1}{P^2}\sigma_P^2\right] dt + \left[\frac{1}{P}\sigma_P\right] dW_t^P$$

$$dp(t,T) = \left[r_t + \lambda_t (T-t)\gamma(T-t,\alpha)_3 \sigma_l - \frac{1}{2}(T-t)^2 \gamma(T-t,\alpha)\Omega\Omega'\gamma(T-t,\alpha)'\right] dt + [-(T-t)\gamma(T-t,\alpha)\Omega] dW_t^P$$

This will help us in a second. Consider a strategy designed as follows:

1. Buy a ZCB with maturity $\tau + \Delta\tau$ at time t

1. Sell a ZCB with maturity τ at time t

The returns of the strategy should be equal to the forward rate with tenor $\Delta\tau$ locked in at time t, minus the $\Delta\tau$ maturity yield prevailing at time τ. The strategy's returns can be written as:

$$R_t^\tau = [p(\tau,\tau+\Delta\tau) - p(t,\tau+\Delta\tau)] - [p(\tau,\tau) - p(t,\tau)]$$

Recall the definition of forward rate at time t with maturity τ and tenor $\Delta\tau$:

$$f_t(\tau) = \frac{p(t,\tau) - p(t,\tau+\Delta\tau)}{\Delta\tau}$$

We rearrange the terms in R_t^τ and get:

$$R_t^\tau = [p(t,\tau) - p(t,\tau+\Delta\tau)] - [p(\tau,\tau) - p(\tau,\tau+\Delta\tau)]$$

$$R_t^\tau = [f_t(\tau) - f_\tau(\tau)]\Delta\tau$$

What $E^P[R_t^\tau]$? Well, we have a form for dp and we can integrate over time.

$$E_t^P[R_t^\tau] = E_t\left[\int_t^\tau dp(u,\tau+\Delta\tau) - \int_t^\tau dp(u,\tau)\right]$$

$$E_t^P[R_t^\tau] = E_t\left[\int_t^\tau dp(u,\tau+\Delta\tau)\right] - E_t\left[\int_t^\tau dp(u,\tau)\right]$$

To evaluate the two expectations we only care about the drift summand of each of the two differentials, as the diffusions have expectation equal to zero.

$$\begin{aligned} E_t^P[R_t^\tau] = \int_t^\tau & r_u + \lambda_u(\tau+\Delta\tau-u)\gamma(\tau+\Delta\tau-u,\alpha)_3\sigma_l \\ & - \frac{1}{2}(\tau+\Delta\tau-u)^2\gamma(\tau+\Delta\tau-u,\alpha)\Omega\Omega'\gamma(\tau+\Delta\tau-u,\alpha)'du \\ & - \int_t^\tau r_u + \lambda_u(\tau-u)\gamma(\tau-u,\alpha)_3\sigma_l - \frac{1}{2}(\tau-u)^2\gamma(\tau-u,\alpha)\Omega\Omega'\gamma(\tau-u,\alpha)'du \end{aligned}$$

Tuckman assumes $E_t(\lambda_u) \approx \lambda_t$. So we treat it as constant. To simplify this integral, define:

$$h(z) = \lambda_t z\gamma(z,\alpha)_3\sigma_l - \frac{1}{2}z^2\gamma(z,\alpha)\Omega\Omega'\gamma(z,\alpha)'$$

And then

$$E_t^P[R_t^\tau] = \int_t^\tau r_u + h(\tau+\Delta\tau-u)du - \int_t^\tau r_u + h(\tau-u)du$$

$$E_t^P[R_t^\tau] = \int_t^\tau [h(\tau+\Delta\tau-u) - h(\tau-u)]\,du$$

Now perform a change of variable $s = \tau - u$:

$$E_t^P[R_t^\tau] = \int_0^{\tau-t} [h(s+\Delta\tau) - h(s)]\,ds$$

We split the integral again:

$$E_t^P[R_t^\tau] = \int_0^{\tau-t} h(s+\Delta\tau)\,ds - \int_0^{\tau-t} h(s)\,ds$$

$$E_t^P[R_t^\tau] = \int_{\Delta\tau}^{\tau+\Delta\tau-t} h(u)\,du - \int_0^{\tau-t} h(u)\,du$$

Which one may rewrite as:

$$E_t^P[R_t^\tau] = \int_{\tau-t}^{\tau+\Delta\tau-t} h(u)\,du - \int_0^{\Delta\tau} h(u)\,du$$

Tuckman now assumes that the contribution of the second term is approximately zero for very short u. Hence the expected return is entirely given by:

$$E_t^P[R_t^\tau] = \int_{\tau-t}^{\tau+\Delta\tau-t} h(u)\,du$$

And with a last change of variable $u = t + \Delta\tau - t - s$ for $s \in [0, \Delta\tau]$:

$$E_t^P[R_t^\tau] = \int_0^{\Delta\tau} h(\tau + \Delta\tau - t - s)\,ds$$

$$E_t^P[R_t^\tau] = \lambda_t \overbrace{\int_0^{\Delta\tau} [(\tau + \Delta\tau - t - s)\gamma(\tau + \Delta\tau - t - s, \alpha)_3 \sigma_l\,ds}^{RP_d(t,\tau,\Delta\tau)} \\ \underbrace{- \int_0^{\Delta\tau} \frac{1}{2}(\tau + \Delta\tau - t - s)^2 \gamma(\tau + \Delta\tau - t - s, \alpha)\Omega\Omega'\gamma(\tau + \Delta\tau - t - s, \alpha)']\,ds}_{RP_c(t,\tau,\Delta\tau)}$$

Using these two amount of risk terms we get:

$$E_t^P[R_t^\tau] = \lambda_t RP_d(t, \tau, \Delta\tau) + RP_c(t, \tau, \Delta\tau)$$

So we can tie it all together with the forward rates:

$$E_t^P[R_t^\tau] = E_t^P[f_t(\tau) - f_\tau(\tau)]\Delta\tau$$

$$E_t^P[R_t^\tau] = (f_t(\tau) - E_t^P[f_\tau(\tau)])\Delta\tau$$

The forward rate at τ for a very short horizon $\Delta\tau \to 0$ can be approximated with the short rate r_τ. Why? Here is a simple argument:

$$f_t(\tau) = \frac{p(t, \tau) - p(t, \tau + \Delta\tau)}{\Delta\tau} = -\frac{p(\tau, \tau + \Delta\tau)}{\Delta\tau}$$

$$f_t(\tau) = -\frac{\log E_\tau\left[e^{-\int_\tau^{\tau+\Delta\tau} r_s ds}\right]}{\Delta\tau}$$

If we expand the integral to first order:

$$e^{-\int_{\tau}^{\tau+\Delta\tau} r_s ds} = 1 - r_\tau \Delta\tau + o(\Delta\tau)$$

So then

$$f_t(\tau) = -\frac{1}{\Delta\tau}\log E_\tau[1 - r_\tau\Delta\tau + o(\Delta\tau)]$$

Since everything in the expectation is now τ-measurable:

$$f_t(\tau) = -\frac{1}{\Delta\tau}\log[1 - r_\tau\Delta\tau + o(\Delta\tau)]$$

And by a first-order expansion of $\log(1+x)$:

$$f_t(\tau) = \frac{r_\tau\Delta\tau + o(\Delta\tau)}{\Delta\tau} = r_\tau + o(1)$$

Which explains why $f_t(\tau) \to r_\tau$ as $\Delta\tau \to 0$. We now use this fact to write:

$$\lambda_t RP_d(t,\tau,\Delta\tau) + RP_c(t,\tau,\Delta\tau) = [f_t(\tau) - E_t^P[r_\tau]]\Delta\tau$$

In order to identify λ_t, we need to choose two maturities τ, τ' that fulfill this condition.

$$\frac{\lambda_t RP_d(t,\tau,\Delta\tau) + RP_c(t,\tau,\Delta\tau)}{\Delta\tau} = [f_t(\tau) - E_t^P[r_\tau]]$$

$$\frac{\lambda_t RP_d(t,\tau',\Delta\tau) + RP_c(t,\tau',\Delta\tau)}{\Delta\tau} = [f_t(\tau') - E_t^P[r_{\tau'}]]$$

$$\frac{\lambda_t[RP(t,\tau',\Delta\tau) - RP(t,\tau,\Delta\tau)] + [RP_c(t,\tau',\Delta\tau) - RP_c(t,\tau,\Delta\tau)]}{\Delta\tau} = [f_t(\tau') - f_t(\tau)] - [E_t^P(r_{\tau'}) - E_t^P(r_\tau)]$$

For very large τ, τ', Tuckman simplifies $E_t^P(r_\tau) - E_t^P(r_{\tau'})$ by the properties of mean reversion of the short rate. Hence:

$$\lambda_t = \frac{[f_t(\tau') - f_t(\tau)] + [RP_c(t,\tau',\Delta\tau) - RP_c(t,\tau,\Delta\tau)]}{\frac{1}{\Delta\tau}[RP_d(t,\tau',\Delta\tau) - RP_d(t,\tau,\Delta\tau)]}$$

And if we can make than estimate for $\hat{\lambda}_t$, then we can compute the real expectation of $E_t^P(r_\tau)$ as:

$$E_t^P(r_\tau) = f_t(\tau) - \frac{\hat{\lambda}_t RP_d(t,\tau,\Delta\tau) + RP_c(t,\tau,\Delta\tau)}{\Delta\tau}$$