- *VAR-based tests*. When $\mathbf{x}_t = (p_t, d_t)'$ follows a vector autoregressive process, then the restriction $E_t(r_{t+i}) = r$, or equivalently $E_t(p_t^*) = p_t$, can be written as a restriction on the coefficient matrices. Specifically, if $\mathbf{x}_t$ follows a VAR(1) with coefficient matrix $\mathbf{A}$, then this restriction is $\mathbf{e}_1' = (1-\rho)\mathbf{e}_2'\mathbf{A}(\mathbf{I}-\rho\mathbf{A})^{-1}$, or equivalently $\mathbf{e}_1'(\mathbf{I}-\rho\mathbf{A}) = (1-\rho)\mathbf{e}_2'\mathbf{A}$, see the lecture notes for a full derivation. Given a value of ρ, this may be tested with a Wald or likelihood ratio test in a VAR. Again, the typical result is that this restriction is rejected.

4. (a) Let $\mathbf{Z}_t = (Z_{it}, Z_{mt})$, with mean $\boldsymbol{\mu}$ and variance matrix $\boldsymbol{\Sigma}$. Mean variance optimization involves minimizing $\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}$ subject to

$$E(\mathbf{w}'\mathbf{R}_t + (1-\mathbf{w}'\boldsymbol{\iota})R_f) = E(R_{pt}) \qquad \Leftrightarrow \qquad \mathbf{w}'E(\mathbf{R}_t - \boldsymbol{\iota}R_f) = E(R_{pt} - R_f),$$

or equivalently $\mathbf{w}'\boldsymbol{\mu} = \mu_p$, for fixed expected excess portfolio return μ_p. The Lagrangean is $\frac{1}{2}\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w} - \lambda(\mathbf{w}'\boldsymbol{\mu} - \mu_p)$, and the first-order condition is $\boldsymbol{\Sigma}\mathbf{w} = \lambda\boldsymbol{\mu}$. When the mean-variance frontier is spanned by Z_{mt} only, then for all μ_p, the optimal portfolio satisfies $\mathbf{w} = (0, w_m)'$, so that the first-order condition is

$$\begin{aligned} \sigma_{im} w_m &= \lambda\mu_i, \\ \sigma_m^2 w_m &= \lambda\mu_m, \end{aligned}$$

which implies $\mu_i = (\sigma_{im}/\sigma_m^2)\mu_m$. From the properties of the bivariate normal distribution we know that $E(Z_{it}|Z_{mt}) = \alpha + \beta Z_{mt}$, with $\beta = \sigma_{im}/\sigma_m^2$ and $\alpha = \mu_i - \beta\mu_m$. Therefore, the spanning condition corresponds to $H_0 : \alpha = 0$.

(b) It is convenient to use matrix notation. Let $\mathbf{X}_t = (1, Z_{mt})'$, and $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_T)'$. Next, let $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{iT})'$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_T)'$ and $\boldsymbol{\gamma} = (\alpha, \beta)'$. Then

$$\hat{\boldsymbol{\gamma}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}_i = \boldsymbol{\gamma} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon},$$

so that $\hat{\boldsymbol{\gamma}}|\mathbf{X} \sim N(\boldsymbol{\gamma}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$. Now

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} T & \sum_{t=1}^T Z_{mt} \\ \sum_{t=1}^T Z_{mt} & \sum_{t=1}^T Z_{mt}^2 \end{bmatrix}^{-1} = \frac{1}{T}\begin{bmatrix} 1 & \hat{\mu}_m \\ \hat{\mu}_m & \hat{\mu}_m^2 + \hat{\sigma}_m^2 \end{bmatrix}^{-1},$$

where $\hat{\mu}_m$ and $\hat{\sigma}_m^2$ are the sample mean and variance (not degrees-of-freedom-corrected) of Z_{mt}. The first diagonal element of this matrix is

$$\frac{1}{T}\left(1 - \frac{\hat{\mu}_m^2}{\hat{\mu}_m^2 + \hat{\sigma}_m^2}\right)^{-1} = \frac{1}{T}\frac{\hat{\mu}_m^2 + \hat{\sigma}_m^2}{\hat{\sigma}_m^2} = \frac{1}{T}(1 + sr_m^2),$$

where $sr_m = \hat{\mu}_m/\hat{\sigma}_m$. This yields the required result $\hat{\alpha}|\mathbf{X} \sim N\left(\alpha, \frac{1}{T}\sigma^2(1+sr_m^2)\right)$. The Gibbons-Ross-Shanken test statistic is essentially the squared t-statistic for $H_0 : \alpha = 0$. The conditional normality of $\hat{\alpha}$ implies that under H_0,

$$T\frac{\hat{\alpha}^2}{\sigma^2(1+sr_m^2)}\bigg|\mathbf{X} \sim \chi^2(1).$$

Moreover, the usual least-squares variance estimator $s^2 = \mathbf{Z}_i'(\mathbf{I}_T - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{Z}_i/(T-2)$ satisfies $(T-2)s^2|\mathbf{X} \sim \chi^2(T-2)$, independent of $\hat{\alpha}$. This in turn implies that, still under H_0,

$$J_1 = T\frac{\hat{\alpha}^2}{s^2(1+sr_m^2)} \sim F(1, T-2),$$

conditionally on $\mathbf{X}$, but because this distribution does not depend on $\mathbf{X}$, it also holds marginally.

5. (a) The BEKK model, in its simplest form, is

$$H_t = C'C + A'y_{t-1}y_{t-1}'A + B'H_{t-1}B,$$

where C is upper triangular, and A and B are general $N \times N$ matrices.
The DCC model takes the form

$$H_t = D_t^{1/2} R_t D_t^{1/2},$$

here $D_t = \text{diag}(h_{1t}, \ldots, h_{Nt})$ consists of univariate GARCH specifications for $h_{it} = \text{Var}(y_{it}|\Omega_{t-1})$, and R_t is a conditional correlation matrix, with typical elements $\rho_{ij,t} = q_{ij,t}/\sqrt{q_{ii,t}q_{jj,t}}$, where

$$Q_t = \bar{Q}(1-\alpha-\beta) + \alpha u_{t-1}u_{t-1}' + \beta Q_{t-1},$$

with $u_t = D_t^{-1/2} y_t$ and $\bar{Q} = E(u_t u_t')$.

- *Positive definiteness*:
  - The BEKK model is constructed to deliver a p.d. H_t, provided that the starting value H_0 is chosen p.d. This follows directly from $w'H_tw = w'C'Cw + w'A'y_{t-1}y_{t-1}'Aw + w'B'H_{t-1}Bw > 0$ for all $w \neq 0$.
  - In the DCC model, the matrix Q_t is p.d. (again, provided that Q_0 is p.d.), which in turn implies that R_t and hence H_t is p.d.

  This holds for all parameter values in both the BEKK and DCC model (except, of course, for special cases such as $C = A = B = 0$).
- *Computational aspects*:
  - The simplest BEKK model has $\frac{1}{2}N(N+1)+2N^2$ unknown parameters in A, B and C. This is already a lot less than the unrestricted VEC model, but still considerable for moderate to large N (for example, $N = 10$ implies 230 unknown parameters). Likelihood maximization on such high-dimensional parameter spaces is very hard, one is bound to end up in local minima, maxima or saddle points instead of the global maximum (if the Newton-type iterations converge at all). Therefore, application of this model is feasible only in small systems.
  - The DCC model can be applied to any dimension N. Two-step estimation of (the simplest version of) the model involves estimating N univariate GARCH models, and subsequently maximizing a likelihood function over only 2 parameters α and β (estimating $\bar{Q}$ by the sample variance matrix of $\hat{u}_t$).
- *Flexibility / restrictiveness*: Both the BEKK model and the DCC model allow for time-varying variances, covariances and correlations. However, the DCC model imposes much more restrictions on the type of dynamic effects that are allowed. In particular, the conditional variance h_{it} of y_{it} only depends on the past squared returns y_{it-j}^2, such that "volatility spillovers" are excluded. Similarly, feedback from past volatilities or squared returns on correlations is severly limited in the DCC model.
- *Other criteria*: One aspect that may be mentioned is that the BEKK model is closed under non-singular transformations (Qy_t also follows a BEKK model if y_t does and $|Q| \neq 0$), whereas the DCC model is not.

(b) The assumptions imply that $w'y_t|\Omega_{t-1} \sim N(0, w'H_tw)$. The 5% Value at Risk $\text{VaR}_t^{0.05}$ is defined by the property $\mathbb{P}\left(w'y_t \leq -\text{VaR}_t^{0.05}|\Omega_{t-1}\right) = 0.05$. Standard normality of $(w'H_tw)^{-1/2}w'y_t$ implies that

$$\widehat{\text{VaR}}_t^{0.05} = 1.65\sqrt{w'\hat{H}_tw}.$$

(c) Defining $I_t = \mathbf{1}_{\{w'y_t < -\mathrm{VaR}_t^{0.05}\}}$, correct conditional coverage requires $E(I_t|\Omega_{t-1}) = 0.05$, whereas unconditional coverage relates to $E(I_t) = 0.05$. The latter can be checked by testing $H_0 : p = 0.05$ in a binomial model, based on $\hat{p} = T^{-1}\sum_{t=1}^{T} I_t$; an asymptotically valid version is obtained by regressing $I_t - 0.05$ on a constant and testing whether the constant term is zero with a t-test. Conditional coverage can be checked by testing whether $I_t - 0.05$ can be significantly explained by lagged I_{t-j}'s or other predetermined variables ($\widehat{\mathrm{VaR}}_t$, lagged returns, etc.). Testing for first-order dependence can be carried out in a Markov chain, via a likelihood ratio test for the hypothesis that $\mathbb{P}(I_t = 1|I_{t-1} = 1) = \mathbb{P}(I_t = 1|I_{t-1} = 0) = 0.05$.