F500 Revision Session - Solutions

Dr Tom Auld*

March 25, 2019

2014 A4

Lets start by trying to understand the objects in the factor equation:

$$Z_{it} = \mu_i + \sum_{j=1}^{K} b_{ij} f_{jt} + \epsilon_{it} \tag{1}$$

This describes a factor model with k factors. This models attempts to explain the large covariance matrix of the observed returns using a much smaller number of factors that dominate the variation. Note that the market (β) model we studied in earlier problems is in fact a one factor model. However, the market model not explain observed returns well at all whereas including a few more factors provides a much better better fit to data. Equation 1 is perhaps easier to understand if we write it out in vector notation. We are given that there are K factors. Suppose there are N stocks;

$$\begin{pmatrix} Z_{1t} \\ \vdots \\ ZNt \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_N \end{pmatrix} + f_{1t} \times \begin{pmatrix} b_{11} \\ \vdots \\ b_{N1} \end{pmatrix} + \cdots + f_{1t} \times \begin{pmatrix} b_{1K} \\ \vdots \\ b_{NK} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \vdots \\ \epsilon_{Nt} \end{pmatrix}$$

Note that $f_{1t}, \ldots, f_{Kt}$ are K univariate random variables that vary with time. b_{ij} is a $N \times K$ constant matrix describing the loading of the factors within the space of the N stocks. ϵ is a vector of shocks, usually taken to be i.i.d. (in some versions contemporaneous correlations are allowed between stocks, as long as the idiosyncratic risk associated with ϵ vanishes as the universe of stocks N gets large). Now lets answer the question:

a) i)

When T is large and we do not observe factors or b, we treat b as fixed and f as stochastic. Estimates for b and μ proceed by MLE. We impose ϵ as i.i.d. and define the covariances as follows:

$$\begin{aligned} D &= VAR(\epsilon) \\ \Sigma_f &= VAR(f) \\ \Omega &= VAR(Z) = b\Sigma_f b^T + D \end{aligned}$$

Ω is the covariance of observed excess returns. We do not know Σ_f or b and so there is an identification issue in the model; we cannot untangle b from Σ_f - this is not surprising as the purpose of fitting this model is to separate out the variance of returns into that caused by a small number of factors ($b\Sigma_f b^T$)

*Comments and corrections to tja20@cam.ac.uk

and an idiosyncratic part (D). Thus we set $\Sigma_f = I_K$ which allows unique identification of b (up to sign). The LogLikelihood is:

$$\ell(\mu, b, D) = \text{const} - \frac{T}{2} log\left[det(bb^T + D)\right] - \frac{1}{2}\sum_{t=1}^{T}(z_t - \mu)^T(bb^T + D)^{-1}(z_t - \mu)$$

This is a nicely behaved function for μ and the M.L.E. yields the same expressions as OLS (because to maximise the likelihood we minimise the ordinary square error term $\sum_{t=1}^{T}(z_t - \mu)^T\Omega(z_t - \mu)$);

$$\hat{\mu} = \arg\max_{\mu} \ell(\mu, b, D) = 0 \quad \Rightarrow \quad \hat{\mu} = \overline{z}$$

As in OLS the MLE for μ is the sample average of the means. Write also:

$$\widetilde{\Omega} = \frac{1}{T}\sum_{t=1}^{T}(z - \hat{\mu})(z - \hat{\mu})^T$$

We can substitute $\hat{\mu}$ (and the expression $\widetilde{\Omega}$) back into the LogLikelihood to form a likelihood for b and D (this process is called concentrating out μ):

$$\ell(\hat{\mu}, b, D) = \text{const} - \frac{T}{2} log\left[det(bb^T + D)\right] - \frac{1}{2} tr(\widetilde{\Omega}(bb^T + D)^{-1}) \qquad \text{using } tr(ABC) = tr(CAB)$$

This is a horrendous non-linear equation in b and D and the maximum has to be found using a numerical procedure. Once $\hat{b}$ and $\hat{D}$ have been obtained estimates for $\hat{f}_t$ can be found by cross-sectional regression. Note that all factors must be pervasive ($b^Tb \to \infty$) for these estimates to be consistent. Due to the errors-in-variables issue caused by the use of an estimate for b, we are not able to find correct standard errors.

a) ii)

When T is small compared with N estimation is conducted by Asymptotic Principle Components. We now assume that the realisation of the factors f_{jt} are fixed and consider the TxK matrix F where $F_{tj} = f_{jt}$. b is treated as the stochastic variable having a realisation for each stock i. b_i is a vector of length K (ie the loadings of the factors for stock i) and has variance Σ_b. The model is now written for each stock i as

$$Z_i = \mu_i i_T + Fb_i + \epsilon_i \qquad i = 1, \ldots, n$$

Written explicitly as vectors:

$$\begin{pmatrix} Z_{i1} \\ \vdots \\ Z_{iT} \end{pmatrix} = \mu_i \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + \begin{pmatrix} f_{11} & \cdots & f_{1K} \\ \vdots & \ddots & \vdots \\ f_{T1} & \cdots & f_{TK} \end{pmatrix} \begin{pmatrix} b_{i1} \\ \vdots \\ b_{iK} \end{pmatrix} + \begin{pmatrix} \epsilon_{i1} \\ \vdots \\ \epsilon_{iT} \end{pmatrix} \qquad i = 1, \ldots, n$$

In contrast to the method above when T is large, the procedure aims to separate the excess return variance across *stocks* as opposed to across *time*. Again there is a separation in variance caused by the factors (Fb_i) as well as a idiosyncratic part (ϵ_i). The identification arises and this is solved here by making Σ_b diagonal, with the K diagonal elements ordered by decreasing size. The procedure for fitting is outlined in lectures and the steps are summarised as:

1. Set the columns of F equal to the K largest eigenvectors of the estimate of the sample covariance matrix of excess returns across stocks; $\hat{\Psi}_{ts} = \frac{1}{N}\sum_{i=1}^{n}(Z_{it} - \overline{Z}_t)(Z_{is} - \overline{Z}_s)$
2. Given $\widehat{F}$, estimate the factor loadings $\hat{b}_i$ by times-series regression

The above approach assumes ϵ is homoskedastic in the time dimension. This can be overcome by estimating $\hat{\sigma}_1^2, \ldots, \hat{\sigma}_T^2$ iteratively ($\sigma_t = var(\epsilon_{it})$). Estimates of the heteroskedasticity are used in subsequent time-series regressions which themselves can be then used to update $\hat{\sigma}_1^2, \ldots, \hat{\sigma}_T^2$. You should have seen a similar approach when performing FGLS in the Econometrics course.

b)

When factor returns f are observed, the test of APT is analogous to the test of the single factor CAPM (which we discussed in problem set 2). In both cases theory implies that the intercept $\mu = 0$. The MLE is the same as the that provided by equation by equation OLS. As in the CAPM test the (large T asymptotic) test is conducted by a Chi Squared test on the restriction $\mu = 0$.

2014 B1

As usual using lower case variables to represent log quantities:

$$
\begin{aligned}
\text{"Return"} &= \text{"Captial Gains"} + \text{"Dividends"} \\
R_{t+1} &= \frac{P_{t+1} - P_t}{P_t} + \frac{D_{t+1}}{P_t} \\
r_{t+1} &= log(1 + R_{t+1}) \\
&= log\left(\frac{P_t}{P_t} + \frac{P_{t+1} - P_t + D_{t+1}}{P_t}\right) \\
&= log(P_{t+1} + D_{t+1}) - log(P_t) \qquad \because log(A/B) = log(A) \times log(B^{-1}) = log(A) - log(B) \\
&= log\left(P_{t+1}\left[1 + \frac{D_{t+1}}{P_{t+1}}\right]\right) - log(P_t) \\
&= log(P_{t+1}) - log(P_t) + log\left(1 + \frac{D_{t+1}}{P_{t+1}}\right) \\
&= log(P_{t+1}) - log(P_t) + log\left(1 + exp\left[log\left(\frac{D_{t+1}}{P_{t+1}}\right)\right]\right) \qquad \because exp(log(x)) = x \;\forall x \\
&= log(P_{t+1}) - log(P_t) + log(1 + exp(log(D_{t+1}) - log(P_{t+1}))) \\
&= p_{t+1} - p_t + log(1 + exp(d_{t+1} - p_{t+1})) \qquad (2)
\end{aligned}
$$

The approximation in Campbell's model comes from using a first order approximation to

$$f(\cdot) = log(1 + exp(\cdot))$$

and expanding about the average values of the dividend and price ratios. Note that:

$$\frac{d}{dx}f(x) = \frac{exp(x)}{1 + exp(x)} = \frac{1}{1 + exp(-x)}$$

So (using Taylor's approximation)

$$f(x) = f(\bar{x}) + f'(\bar{x})(x - \bar{x}) + o(x - \bar{x}).(x - \bar{x})$$

Substituting $x = d_{t+1} - p_{t+1}$ and $\bar{x} = \bar{d} - \bar{p}$ into above

$$\begin{aligned} log(1 + exp(d_{t+1} - p_{t+1})) &\approx log(1 + exp(\bar{d} - \bar{p})) + \frac{d_{t+1} - p_{t+1} - \bar{d} + \bar{p}}{1 + exp(\bar{p} - \bar{d})} \\ &= log(1 + exp(\bar{d} - \bar{p})) + \frac{\bar{p} - \bar{d}}{1 + exp(\bar{p} - \bar{d})} + \frac{d_{t+1} - p_{t+1}}{1 + exp(\bar{p} - \bar{d})} \\ &= k + (1 - \rho).\, (d_{t+1} - p_{t+1}) \end{aligned}$$

where

$$\begin{aligned} (1 - \rho) &= \frac{1}{1 + exp(\bar{p} - \bar{d})} \qquad (3) \\ \rho &= 1 - \frac{1}{1 + exp(\bar{p} - \bar{d})} \\ &= \frac{exp(\bar{p} - \bar{d})}{1 + exp(\bar{p} - \bar{d})} \\ &= \frac{1}{1 + exp(\bar{d} - \bar{p})} \end{aligned}$$

and

$$\begin{aligned} k &= log(1 + exp(\bar{d} - \bar{p})) + \frac{\bar{p} - \bar{d}}{1 + exp(\bar{p} - \bar{d})} \\ &= log(\rho^{-1}) + (1 - \rho).\, (\bar{p} - \bar{d}) \\ &= -log(\rho) + (1 - \rho).\, (\bar{p} - \bar{d}) \end{aligned}$$

note from equation 3

$$\begin{aligned} 1 + exp(\bar{p} - \bar{d}) &= \frac{1}{1 - \rho} \\ exp(\bar{p} - \bar{d}) &= \frac{\rho}{1 - \rho} \\ \bar{p} - \bar{d} &= log\left(\frac{\rho}{1 - \rho}\right) \end{aligned}$$

so

$$k = -log(\rho) + (1 - \rho).\, log\left(\frac{\rho}{1 - \rho}\right)$$

Substituting these expressions into equation 2 above yields the result from lectures:

$$\begin{aligned} r_{t+1} &\approx k + p_{t+1} - p_t + (1 - \rho).\, (d_{t+1} - p_{t+1}) \\ &= k + (1 - \rho)d_{t+1} + \rho.\, p_{t+1} - p_t \end{aligned}$$

Rearranging, taking expectations at time t, and iterating forward:

$$
\begin{aligned}
p_t &= k + (1-\rho)E_t d_{t+1} - E_t r_{t+1} + \rho E_t p_{t+1} \\
&= k + (1-\rho)E_t d_{t+1} - E_t r_{t+1} + \rho\left(k + (1-\rho)E_t d_{t+2} - E_t r_{t+2} + \rho E_t p_{t+2}\right) \\
&= k(1+\rho) + (1-\rho)E_t(d_{t+1} + \rho d_{t+2}) - E_t(r_{t+1} + \rho r_{t+2}) + \rho^2 E_t p_{t+2} \\
&\vdots \qquad\qquad \text{iterating } s \text{ times} \\
&= k\sum_{j=0}^{n}\rho^j + (1-\rho)\sum_{j=0}^{n}\rho^j E_t d_{t+j+1} - \sum_{j=0}^{n}\rho^j E_t r_{t+j+1} + \rho^n E_t p_{t+n} \\
&= \frac{k}{1-\rho} + (1-\rho)\sum_{j=0}^{\infty}\rho^j E_t d_{t+j+1} - \sum_{j=0}^{\infty}\rho^j E_t r_{t+j+1} + lim_{n\to\infty}\rho^n E_t p_{t+n} \\
&= \frac{k}{1-\rho} + (1-\rho)\sum_{j=0}^{\infty}\rho^j E_t d_{t+j+1} - \sum_{j=0}^{\infty}\rho^j E_t r_{t+j+1} \qquad \text{assuming the usual no-bubble condition}
\end{aligned}
$$

As required. Noting also

$$
\rho = \frac{1}{1 + exp(\bar{d} - \bar{p})} \approx \frac{1}{1 + \frac{\bar{D}}{\bar{P}}} \qquad \because\ exp(\overline{log(D)}) \approx \overline{exp(log(D))} = \bar{D}, \text{ and similarly for } \overline{P}
$$

shows ρ is approximately the discount factor associated with a discount rate of approximately the average dividend yield ratio. Campbell's model can also be explained as

$$
p_t = \text{constant} + (1-\rho) \times D - R
$$

where

$$
\begin{aligned}
D &= \sum_{j=0}^{\infty}\rho^j E_t d_{t+j+1} = \text{"Expected future dividends discounted at rate } \rho\text{"} \\
R &= \sum_{j=0}^{\infty}\rho^j E_t r_{t+j+1} = \text{"Expected future returns discounted at rate } \rho\text{"}
\end{aligned}
$$

Whereas with, say, bond pricing, future cashflows are discounted by the bond yield, Campbell's model has expected future dividends and returns being approximately discounted at the average dividend yield rate ρ. This makes some appealing intuitive sense.

Campbell's model is an improvement on earlier fundamental models of stock prices, such as the Gordon Growth model, that use constant expected returns. The preceding class of models predict much lower variability of prices than that observed. In those models prices only vary with changes in expected dividends. These simply do not vary enough. However in Campbell's model, expectations of returns may vary which can explain much higher volatility of prices. In particular when returns are slowly varying, for example with a constant plus an AR(1) process with a coefficient close to unity, prices become very sensitive to changes in expected returns.

VAR modeling can be performed when fundamental data is available as a time series. For example, If $x_1, \ldots, x_k$ are available fundamental data then the following VAR(1) model can be estimated consistently by OLS:

$$\begin{pmatrix} r_t \\ x_{1t} \\ \vdots \\ x_{kt} \end{pmatrix} = \Phi \begin{pmatrix} r_{t-1} \\ x_{1,t-1} \\ \vdots \\ x_{k,t-1} \end{pmatrix} + \epsilon$$

Given the estimate of the $k+1$ squared matrix $\hat{\Phi}$, expectations for future returns at time t can be found via:

$$\widehat{E_t(r_{t+k+1})} = \widehat{\Phi}^k \begin{pmatrix} r_t \\ x_{1t} \\ \vdots \\ x_{kt} \end{pmatrix}$$

2015 A2

The model in this question is a particular type of ARCH(5) model (Note GARCH includes lags of the conditional variance in the dynamic equation whereas ARCH does not). In the below model, today, conditional variance depends on the squared return from 5 days ago:

$$\begin{aligned} r_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= \omega + \gamma r_{t-5}^2 \\ \epsilon_t &\sim NID(0,1) \end{aligned}$$

Not given in the question are the stationarity conditions $\omega \geq 0,\ 0 \leq \gamma < 1$.

a)

Simple answer: Yes. Returns do not depend on any public information in this model and knowing it does not help you make money (price-information is also fully reflected - see the next part).
My preferred answer: This model does not say anything about Semi-Strong Market Efficiency. Fundamental information may or may not be transmitted to the price via the shocks ϵ_t.
Note, the “simple answer” is probably the answer the lecturer was expecting. The question is not asking “is this a good model for financial assets” it is stating that if financial assets abide by this model does semi-strong market efficiency hold. Neither answer is incorrect.

b)

This model is consistent with weak form efficiency as returns are a martingale difference (ie it obeys RW 2.5). Specifically $E(r_t|r_{t-1}, r_{t-2}, \ldots) = 0 \because r_t|r_{t-1}, r_{t-2}, \ldots \sim N(0, \omega + \gamma r_{t-5}^2)$. Past information may effect higher moments of the distribution of returns (specifically the variance) but the conditional expectation is always zero no matter what. An investor cannot make abnormal returns systematically in this model.

c)

This model is not consistent with the stylized empirical fact of strictly positive serial correlation of squared returns. As in lectures, the most straightforward way of finding the ACF of r_t^2 is to eliminate σ_t^2 from the dynamic equation using a Martingale Difference (MD) innovation that depends on r_t^2:

$$\begin{aligned} r_t^2 &= \sigma_t^2 + \sigma_t^2(\epsilon_t^2 - 1) \\ &= \omega + \gamma r_{t-5}^2 + \eta_t \\ \eta_t &= \sigma_t^2(\epsilon_t^2 - 1) \end{aligned} \tag{4}$$

η_t is the MD as $E(\eta_t^2|\text{"the world at t-1"}) = 0 \quad \because \epsilon_t$ is i.i.d. You should notice that r_t^2 follows a particular type of AR(5) process which would produce an ACF at lags of multiples of 5 only. I will use a slightly more sophisticated method of proving this:

$$\begin{aligned} (1-\gamma L^5)r_t^2 &= \omega + \eta_t \\ r_t^2 &= \frac{\omega}{1-\gamma} + \frac{\eta_t}{(1-\gamma L^5)} && \because L \text{ has no effect on the constant } \omega \\ &= \frac{\omega}{1-\gamma} + \sum_{s\geq 0}(\gamma L^5)^s \eta_t && \because |\gamma| < 1,\ \frac{1}{1-x} = \sum_{s\geq 0} x^s \ \ \forall\, |x| < 1 \\ &= \frac{\omega}{1-\gamma} + \sum_{s\geq 0}\gamma^s \eta_{t-5s} \end{aligned}$$

So r_t^2 is equal to the unconditional mean ($\frac{\omega}{1-\gamma}$) plus a linear combination of *past* shocks of the MD η_t separated in time by multiples of 5. This shows that if k is not a multiple of 5 then r_t^2 and r_{t-k}^2 will share no η_t terms and thus have zero covariance. This ACF does not describe the expected shape of a financial asset which would be positive at all lags and decaying. This model also does not makes much intuitive sense; why would the volatility depend of the market move from one week ago only, rather than yesterday?

d)

This is the leverage assumption written in a different way. GARCH models (of which ARCH is a special case) do ***not*** have this property. Proving this follows from the fact that there is a r_{t-k} linear term in the covariance which has zero expectation:

$$\begin{aligned} Cov(r_t^2, r_{t-k}) &= E(r_t^2 r_{t-k}) && \text{since } E(r_{t-k}) = 0 \\ &= E(\sigma_t^2 \sigma_{t-k}\epsilon_t^2 \epsilon_{t-k}) \\ &= E_{\sigma_t,\sigma_{t-k}}\left[E(\sigma_t^2\sigma_{t-k}\epsilon_t^2\epsilon_{t-k}|\sigma_t,\sigma_{t-k})\right] && \text{Using L.I.E.} \\ &= E_{\sigma_t,\sigma_{t-k}}\left[\sigma_t^2\sigma_{t-k}E(\epsilon_t^2\epsilon_{t-k}|\sigma_t,\sigma_{t-k})\right] \\ &= E_{\sigma_t,\sigma_{t-k}}\left[\sigma_t^2\sigma_{t-k}E(\epsilon_t^2)E(\epsilon_{t-k})\right] && \because \epsilon_t \text{ is i.i.d.} \\ &= E_{\sigma_t,\sigma_{t-k}}\left[\sigma_t^2\sigma_{t-k}\times 1 \times 0\right] \\ &= 0 \end{aligned}$$

2016 A3

Returns depend on fundamentals x_t which are themselves are serialling correlated. Essentially the question is asking us to show that the ACF of returns is small when β is significant despite the fact that the fundamentals have persistence. The key is the fact that error terms in the price equation and the dynamic AR(1) equation are contemporaneously correlated (but not serially correlated). First note that x_t is an AR(1) process and is thus a linear combination of past shocks of η_t. This means that x_t is independent of future shocks of both ϵ_{t+k} and η_{t+k}. Start by writing down some properties of this AR(1) process:

$$\begin{aligned}
E(x_t^2) &= \frac{\sigma_{\eta\eta}}{1-\rho^2} && \text{The usual AR(1) unconditional variance covered previously} \\
E(x_t x_{t-k}) &= \rho^k E(x_t^2) && \text{The usual covariance} \qquad (5) \\
(1-\rho L)x_t &= \eta_{t+1} \\
x_t &= \frac{1}{1-\rho L}\eta_{t+1} \\
&= \sum_{s\geq 0} (\rho L)^s \eta_{t+1} && \text{Expanding as a geometric sum as in previous questions} \\
&= \sum_{s\geq 0} \rho^s \eta_{t+1-s} && (6)
\end{aligned}$$

Consider the covariance of returns at lag 1:

$$\begin{aligned}
COV(r_{t+1}, r_t) &= E(r_{t+1} r_t) && \because E(r_t) = E(x_{t-1}) = 0 \\
&= E\left[(\beta x_t + \epsilon_{t+1}) \times (\beta x_{t-1} + \epsilon_t)\right] \\
&= E\left[(\beta\rho x_{t-1} + \eta_t + \epsilon_{t+1}) \times (\beta x_{t-1} + \epsilon_t)\right] \\
&= E\left[\beta^2 \rho x_{t-1}^2 + \beta \eta_t \epsilon_t\right]
\end{aligned}$$

because there is no serial correlation of shocks and x_{t-1} is independent of η_t, ϵ_t and ϵ_{t+1}. So

$$\begin{aligned}
COV(r_{t+1}, r_t) &= \beta^2 \rho VAR(x_t) + \beta \sigma_{\eta\epsilon} \\
&= \frac{\beta^2 \rho \sigma_{\eta\eta}}{1-\rho^2} + \beta \sigma_{\eta\epsilon}
\end{aligned}$$

Calculate the covariance for higher lags by forming an iterative relationship:

$$\begin{aligned}
COV(r_t, r_{t-k}) &= E\left[(\beta x_{t-1} + \epsilon_t) \times (\beta x_{t-1-k} + \epsilon_{t-k})\right] \\
&= E\left[(\beta\rho x_{t-2} + \beta\eta_{t-1} + \epsilon_t) \times (\beta x_{t-1-k} + \epsilon_{t-k})\right] \\
&= E\left[(\rho(\beta x_{t-2} + \epsilon_{t-1}) + \beta\eta_{t-1} + \epsilon_t - \rho\epsilon_{t-1}) \times (\beta x_{t-1-k} + \epsilon_{t-k})\right] \\
&= E\left[\rho r_{t-1} r_{t-k}\right] + E\left[(\beta\eta_{t-1} + \epsilon_t - \rho\epsilon_{t-1}) \times (\beta x_{t-1-k} + \epsilon_{t-k})\right] \\
&= \rho . COV(r_t, r_{t-k-1}) \qquad \forall\, k > 1
\end{aligned}$$

since everything in the second expectation term vanishes if there are no shared shocks and if the shocks $\eta_{t-1} + \epsilon_t - \epsilon_{t-1}$ occur "after" x_{t-1-k}, which they do for $k > 1$. Thus

$$\begin{aligned}
COV(r_t, r_{t-k}) &= \rho^{k-1} COV(r_t, r_{t-1}) \\
&= \rho^{k-1} \beta \left[\frac{\beta\rho\sigma_{\eta\eta}}{1-\rho^2} + \sigma_{\eta\epsilon}\right]
\end{aligned}$$

Thus

$$\frac{\sigma_{\eta\epsilon}}{\sigma_{\eta\eta}} \approx -\frac{\beta\rho\sigma_{\eta\eta}}{1-\rho^2} \qquad \Rightarrow \qquad COV(r_t, r_{t-k}) \approx 0$$

Thus if dividend price ratios are negatively correlated to shocks in returns, which makes sense, and the correlation is just the right size, large βs are indeed possible without the persistence in the fundamentals causing serial correlation in observed returns. (Stambaugh 1999).