MFE Financial Econometrics
2018 Final Exam Model Solutions

Tuesday 12th March, 2019

1. **If $(X, \epsilon) \sim N(0, I_2)$ what is the distribution of $Y = \mu + \beta X + \epsilon$?**

$$Y \sim N\left(\mu, \beta^2 + 1\right)$$

2. **What is the Cramer-Rao lower bound and why is it useful?**
The Cramer-Rao lower bound is the smallest variance that an estimator can achieve within a large family of consistent estimators. MLE usually acheie the CRLB which provides a strong justification for using estimators from this class.

3. **Derive the OLS estimator for the model $y_i = \beta x_i + \epsilon_i$.**

$$\min_{\beta} \sum_{i=1}^{n} (y_i - \beta x_i)^2 \Rightarrow -\left(\sum_{i=1}^{n} x_i y_i - \hat{\beta} x_i^2\right) = 0$$
$$\Rightarrow \sum_{i=1}^{n} x_i y_i = \hat{\beta} \sum_{i=1}^{n} x_i^2$$
$$\Rightarrow \hat{\beta} = \sum_{i=1}^{n} x_i y_i / \sum_{i=1}^{n} x_i^2$$

4. **Describe the steps to implment k-fold cross validation in a regrssion to select a model.**
For each model:

   (a) Randomly divide observations into k-equally sized blocks, $\mathcal{S}_j$, $j = 1, \ldots, k$
   (b) For $j = 1, \ldots, k$ estimate $\hat{\boldsymbol{\beta}}_j$ by excluding the observations in block i
   (c) Compute cross-validated SSE using observations in block j and $\hat{\boldsymbol{\beta}}_j$

$$SSE_{xv} = \sum_{j=1}^{k} \sum_{i \in \mathcal{S}_j} \left(y_i - \mathbf{x}_i \hat{\boldsymbol{\beta}}_j\right)^2$$

   (d) Select model with lowest cross-validated SSE

5. **Under what conditions on ϕ_0, ϕ_1 and θ_1 is the process $\{y_t\}$ stationary where $y_t = \phi_0 + \phi_1 y_{t-1} + \theta_1 \epsilon_{t-1} + \epsilon_t$ and $\{\epsilon_t\}$ is a white noise process.**

$$|\phi_1| < 1$$

with no other restirctions other than coefficients are finite numbers.

6. **What properties must a covariance stationary time series satisfy?**
A stochastic process $\{y_t\}$ is covariance stationary if

$$\begin{aligned} \mathrm{E}[y_t] &= \mu \quad \text{for } t = 1, 2, \ldots \\ \mathrm{V}[y_t] &= \sigma^2 < \infty \quad \text{for } t = 1, 2, \ldots \\ \mathrm{E}[(y_t - \mu)(y_{t-s} - \mu)] &= \gamma_s \quad \text{for } t = 1, 2, \ldots, s = 1, 2, \ldots, t-1. \end{aligned}$$

7. **Outline the steps to objectively evaluate a sequence of variance forcases $\left\{\hat{\sigma}^2_{t+1|t}\right\}$ using a set of returns in the Mincer-Zarnowitch framework.**
Generalized Mincer-Zarnowitz regressions can be used to assess forecast optimality,

$$r^2_{t+h} - \hat{\sigma}^2_{t+h|t} = \gamma_0 + \gamma_1 \hat{\sigma}^2_{t+h|t} + \gamma_2 z_{1t} + \ldots + \gamma_{K+1} z_{Kt} + \eta_t$$

where z_{jt} are any instruments known at time t. Common choices for z_{jt} include r_t^2, $|r_t|$, r_t or indicator variables for the sign of the lagged return. The GMZ regression has a heteroskedastic variance and that a better estimator, GMZ-GLS, can be constructed as

$$\begin{aligned} \frac{r^2_{t+h} - \hat{\sigma}^2_{t+h|t}}{\hat{\sigma}^2_{t+h|t}} &= \gamma_0 \frac{1}{\hat{\sigma}^2_{t+h|t}} + \gamma_1 1 + \gamma_2 \frac{z_{1t}}{\hat{\sigma}^2_{t+h|t}} + \ldots + \gamma_{K+1} \frac{z_{Kt}}{\hat{\sigma}^2_{t+h|t}} + \nu_t \\ \frac{r^2_{t+h}}{\hat{\sigma}^2_{t+h|t}} - 1 &= \gamma_0 \frac{1}{\hat{\sigma}^2_{t+h|t}} + \gamma_1 1 + \gamma_2 \frac{z_{1t}}{\hat{\sigma}^2_{t+h|t}} + \ldots + \gamma_{K+1} \frac{z_{Kt}}{\hat{\sigma}^2_{t+h|t}} + \nu_t \end{aligned}$$

by dividing both sized by the time t forecast, $\hat{\sigma}^2_{t+h|t}$ where $\nu_t = \eta_t / \hat{\sigma}^2_{t+h|t}$.
These models are estimated by OLS (or GLS) and the coefficients are tested under the null $H_0 : \boldsymbol{\gamma} = 0$ against an alternative that one or more is non-zero. The test can be implemented as a Wald, LM or LR test.

8. **What is Principal Component Analysis and how is PCA useful in covariance modeling.**
PCA uses a panel of data to extract the k components which extract the most variance in the panel. These components are uncorrelated by construction. These components can then be used to estimate a k-factor covariance model where each return series is regressed on the k factors and the idiosyncratic variance is used to complete the model. The final covariance is

$$\beta \Sigma_f \beta' + \Omega$$

where β is the m by k matrix of factor loadings for the m assets, Σ_f is the diagonal covariance of the factors and Ω is diagonal matrix with the idiosyncratic variance of series i in position (i, i).

1. **Consider the APT regression**

$$r_t^e = \alpha + \beta_m r_{m,t}^e + \beta_s r_{smb,t} + \beta_v r_{hml,t} + \epsilon_t$$

where $r_{m,t}^e$ is the excess return on the market, $r_{smb,t}$ is the return on the size factor, $r_{hml,t}$ is the return on value factor and r_t^e is an excess return on a portfolio of assets. Using the information provided in the tables below below, answer the following questions:

(a) **Is there evidence that this portfolio is market neutral?**
Using a t-test, the test statistics is

$$\sqrt{n}\frac{\hat{\beta}_m}{\sqrt{s.e.\left(\hat{\beta}_m\right)}}.$$

The null is $H_0 : \beta_m = 0$. Using the two models and two covariances, these values are

	Homosk.	Heterosk.
CAP-M	8.27	6.77
FF3	5.14	5.64

All are larger than 1.96 and so we reject the null of market neutrality at the 5% level.

(b) **Are the size and value factors needed to adequately capture the cross-sectional dynamics in this portfolio?**
Here the null is $\beta_{smb} = \beta_{hml} = 0$. The test has 2 restrictions and so can be implemented as a Wald test using the test statistic

$$nR\hat{\beta}\left(RCR'\right)^{-1}\hat{\beta}R$$

where C is a covariance estimator and

$$R = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The value of the test statistics are 25.46 (Homosk.) and 26.47 (Heterosk.). There are 2 resitrictions and the asymptotic distribution is a χ_2^2. Both are well above the CV of 5.99.

(c) **Is there evidence of conditional heteroskedasticity in this model?**
We can use White's test based on nR^2. Using Model 3 which corresponds to the CAP-M, the test statistic is 13.4. There are 2 restrictions and so the null of homoskedasticity is rejected. In the APT, White's test corresponds to Model 4, and the test statistic is 12.6. The distribution here is a χ_9^2 and to the critical value is 16.91, and the null cannot be rejected. This is mixed evidence.

(d) **What are the trade-offs made when choosing a covariance estimator to use when making inference on this model?**
When the data are homoskedastic, both covariance estimators are consistent. When the data are not homoskedastic, only White's is consistent. This would favor choosing White's covariance estiamtor. However, when the data are homoskedastic White's estimator is noisier than the classic covariance estimator, and so test statistics will have worse finite sample properties. This suggests using the classic covariance estimator unless there is evidence that the data are heteroskedastic.

(e) **Define the size and power of a statistical test.**
The size is the probability of a Type I error – that is, the chance that a true null is rejected. The power is 1 minus the probability of a Type II error, or the chance that the null is not rejected when teh alternative is true.

(f) **What factors affect the power of a statistical test?**

- Sample size. Larger samples increase power since they decrease the estimation error.
- Estimator efficiency. More efficient estimators increase power by reducing estimation error.
- Distance between null and true value. Larer differences are easier to detect.

(g) **Outline the steps to implement the correct bootstrap covariance estimator for these parameters. Justify the method you chose using the information provided.**
Assuming the data is heteroskedastic,

i. Generate a sets of n uniform integers $\{u_i\}_{i=1}^{n}$ on $[1, 2, \ldots, n]$.
ii. Construct a simulated sample $\{y_{u_i}, \mathbf{x}_{u_i}\}$.
iii. Estimate the parameters of interest using $y_{u_i} = \mathbf{x}_{u_i}\boldsymbol{\beta} + \epsilon_{u_i}$, and denote the estimate $\tilde{\boldsymbol{\beta}}_b$.
iv. Repeat steps 1 through 3 a total of B times.
v. Estimate the variance of $\hat{\boldsymbol{\beta}}$ using

$$\begin{aligned}\widehat{\mathrm{V}}\left[\hat{\boldsymbol{\beta}}\right] &= \mathrm{B}^{-1}\sum_{b=1}^{B}\left(\tilde{\boldsymbol{\beta}}_j - \hat{\boldsymbol{\beta}}\right)\left(\tilde{\boldsymbol{\beta}}_j - \hat{\boldsymbol{\beta}}\right)' \text{ or} \\ \widehat{\mathrm{V}}\left[\hat{\boldsymbol{\beta}}\right] &= \mathrm{B}^{-1}\sum_{b=1}^{B}\left(\tilde{\boldsymbol{\beta}}_j - \overline{\tilde{\boldsymbol{\beta}}}\right)\left(\tilde{\boldsymbol{\beta}}_j - \overline{\tilde{\boldsymbol{\beta}}}\right)'\end{aligned}$$

Notes: All models were estimated on $n = 100$ data points. Models 1 and 2 correspond to the specification above. In model 1 r_{smb} and r_{hml} have been excluded. Model 3, 4 and 5 are all version of

$$\hat{\epsilon}_t^2 = \gamma_0 + \gamma_1 r_{m,t}^e + \gamma_2 r_{smb,t} + \gamma_3 r_{hml,t} + \gamma_4 \left(r_{m,t}^e\right)^2 + \gamma_5 r_{m,t}^e r_{smb,t} + \gamma_6 r_{m,t}^e r_{hml,t} + \gamma_7 r_{smb,t}^2 + \gamma_8 r_{smb,t} r_{hml,t} + \gamma_9 r_{hml,t}^2 + \eta_t$$

$\hat{\epsilon}_t$ was computed using Model 1 for the results under Model 3, and using model 2 for the results under Models 4 and 5. R^2 is the R-squared and n is the number of observations.

Parameter Estimates

	Model 1	Model 2		Model 3	Model 4	Model 5
α	0.128	0.089	γ_0	0.984	0.957	0.931
β_m	1.123	0.852	γ_1	-0.779	-0.498	
β_{smb}		0.600	γ_2		-0.046	
β_{hml}		-0.224	γ_3		0.124	
			γ_4	0.497	0.042	0.295
			γ_5		0.049	
			γ_6		0.684	
			γ_7		0.036	-0.149
			γ_8		-0.362	
			γ_9		-0.005	0.128
R^2	0.406	0.527		0.134	0.126	0.037

Parameter Covariance Estimates

The estimated covariance matrices from the asymptotic distribution

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_0\right) \xrightarrow{d} N(0, C)$$

are below where C is either $\hat{\sigma}^2\hat{\Sigma}_{XX}^{-1}$ or $\hat{\Sigma}_{XX}^{-1}\hat{S}\hat{\Sigma}_{XX}^{-1}$.

CAP-M

$\hat{\sigma}^2\hat{\Sigma}_{XX}^{-1}$

	α	β_m
α	1.365475	0.030483
β_m	0.030483	1.843262

$\hat{\Sigma}_{XX}^{-1}\hat{S}\hat{\Sigma}_{XX}^{-1}$

	α	β_m
α	1.341225	-0.695235
β_m	-0.695235	2.747142

Fama-French Model

$$\hat{\sigma}^2 \hat{\Sigma}_{XX}^{-1}$$

	α	β_m	β_{smb}	β_{hml}
α	1.100680	0.103611	-0.088259	-0.063529
β_m	0.103611	1.982761	-0.619139	-0.341118
β_{smb}	-0.088259	-0.619139	1.417318	-0.578388
β_{hml}	-0.063529	-0.341118	-0.578388	1.686200

$$\hat{\Sigma}_{XX}^{-1} \hat{S} \hat{\Sigma}_{XX}^{-1}$$

	α	β_m	β_{smb}	β_{hml}
α	1.073227	-0.361618	-0.072784	0.045732
β_m	-0.361618	2.276080	-0.684809	0.187441
β_{smb}	-0.072784	-0.684809	1.544745	-1.074895
β_{hml}	0.045732	0.187441	-1.074895	1.947117

χ^2_m critical values

Critical value for a 5% test when the test statistic has a χ^2_m distribution.

m	1	2	3	4	8	9	10
Crit Val.	3.84	5.99	7.81	9.48	15.50	16.91	18.30

m	90	91	98	99	100
Crit Val.	113.14	114.26	122.10	123.22	124.34

Matrix Inverse

The inverse of a 2 by 2 matrix

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

(h)

2. Consider the MA(2)-GARCH(1,1) model

$$
\begin{aligned}
y_t &= \phi_0 + \theta_1\epsilon_{t-1} + \phi_2\epsilon_{t-2} + \epsilon_t \\
\epsilon_t &= \sigma_t e_t \\
\sigma_t^2 &= \omega + \alpha_1\epsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2 \\
e_t &\overset{\text{i.i.d.}}{\sim} N(0,1)
\end{aligned}
$$

(a) **What conditions are required for ϕ_0, θ_1 and θ_2 for the model to be covariance stationary?**
The mean is an MA(2) and so there are no restrictions on these parameters (other than they are finite numbres) for stationarity.

(b) **What conditions are required for ω,α_1, β_1 for the model to be covariance stationary?**
$\omega > 0$, $\alpha \geq 0$,$\beta \geq 0$,$\alpha + \beta < 1$.

(c) **Show that $\{\epsilon_t\}$ is a white noise process.**

$$\mathrm{E}[\epsilon_t] = \mathrm{E}[e_t\sigma_t] = \mathrm{E}[\mathrm{E}_{t-1}[e_t\sigma_t]] = \mathrm{E}[\sigma_t\mathrm{E}_{t-1}[e_t]] = \mathrm{E}[\sigma_t\mathrm{E}_{t-1}[e_t]] = 0$$

$$
\begin{aligned}
\mathrm{Cov}[\epsilon_t, \epsilon_{t-s}] &= \mathrm{Cov}[e_t\sigma_t, e_{t-s}\sigma_{t-s}] \\
&= \mathrm{E}[e_t\sigma_t e_{t-s}\sigma_{t-s}] \\
&= \mathrm{E}[\mathrm{E}_{t-1}[e_t\sigma_t e_{t-s}\sigma_{t-s}]] \\
&= \mathrm{E}[\sigma_t e_{t-s}\sigma_{t-s}\mathrm{E}_{t-1}[e_t]] \\
&\mathrm{E}[\sigma_t e_{t-s}\sigma_{t-s} \times 0] = 0
\end{aligned}
$$

(d) **Are ϵ_t and ϵ_{t-1} independent?**
The previous problem shoed they are uncorrelation. They are not independent since the magnitute of the shock to ϵ_{t-1} affects the variance of ϵ_t. Moreover, since this model can be written as an ARMA(1,1), the squared shocks ϵ_t^2 and ϵ_{t-1}^2 are correlated.

(e) **What are the values of the following quantities:**

i. $\mathrm{E}[y_t] = \mathrm{E}[\phi_0 + \theta_1\epsilon_{t-1} + \phi_2\epsilon_{t-2} + \epsilon_t] = \phi_0 + \theta_1\mathrm{E}[\epsilon_{t-1}] + \phi_2\mathrm{E}[\epsilon_{t-2}] + \mathrm{E}[\epsilon_t] = \phi_0$

ii. $\mathrm{E}_t[y_{t+1}] = \mathrm{E}_t[\phi_0 + \theta_1\epsilon_t + \phi_2\epsilon_{t-1} + \epsilon_{t+1}] = \phi_0 + \theta_1\mathrm{E}_t[\epsilon_t] + \phi_2\mathrm{E}_t[\epsilon_{t-1}] + \mathrm{E}_t[\epsilon_{t+1}] = \phi_0 + \theta_1\epsilon_t + \phi_2\epsilon_{t-1}$

iii. $\mathrm{E}_t[y_{t+2}] = \mathrm{E}_t[\phi_0 + \theta_1\epsilon_{t+1} + \phi_2\epsilon_t + \epsilon_{t+2}] = \phi_0 + \theta_1\mathrm{E}_t[\epsilon_{t+1}] + \phi_2\mathrm{E}_t[\epsilon_t] + \mathrm{E}_t[\epsilon_{t+2}] = \phi_0 + \phi_2\epsilon_t$

iv. $\lim_{h\to\infty}\mathrm{E}_t[y_{t+h}] = \phi_0$
since the mean is an MA and all forecasts for $h > 2$ have no dynamics.

v. $\mathrm{V}_t[y_{t+1}] = \mathrm{V}_t[\epsilon_{t+1}] = \omega + \alpha_1\epsilon_t^2 + \beta_1\sigma_t^2$

vi. $\mathrm{V}_t[y_{t+2}] = \mathrm{V}_t[\epsilon_{t+2} + \theta_1\epsilon_{t+1}] = \mathrm{V}_t[\epsilon_{t+2}] + \theta_1^2\mathrm{V}_t[\epsilon_{t+1}]$
since we know form above that ϵ is a white noise process so that the covariance is 0. Finally

$$
\begin{aligned}
\mathrm{V}_t[\epsilon_{t+2}] &= \mathrm{E}_t\left[\omega + \alpha_1\epsilon_{t+1}^2 + \beta_1\sigma_{t+1}^2\right] \\
&= \omega + \alpha_1\mathrm{E}_t\left[\epsilon_{t+1}^2\right] + \beta_1\mathrm{E}\left[\sigma_{t+1}^2\right] \\
&= \omega + (\alpha_1 + \beta_1)\mathrm{V}_t\left[\epsilon_{t+1}^2\right] \\
&= \omega + (\alpha_1 + \beta_1)\left(\omega + \alpha_1\epsilon_t^2 + \beta_1\sigma_t^2\right)
\end{aligned}
$$

3. **Consider the VAR(P)**

$$\mathbf{y_t} = \mathbf{\Phi_1 y_{t-1}} + \mathbf{\Phi_2 y_{t-2}} + \epsilon_\mathbf{t}.$$

(a) **Write this in companion form. Under what conditions is the VAR(P) stationary?**
The companion form of this is

$$\begin{bmatrix} \mathbf{y_t} \\ \mathbf{y}_{t-1} \end{bmatrix} = \begin{bmatrix} \mathbf{\Phi}_1 & \mathbf{\Phi}_2 \\ \mathbf{I_2} & \mathbf{0}_{2\times 2} \end{bmatrix} \begin{bmatrix} \mathbf{y_{t-1}} \\ \mathbf{y_{t-2}} \end{bmatrix} + \begin{bmatrix} \epsilon_\mathbf{t} \\ \mathbf{0}_{2\times 1} \end{bmatrix}.$$

(b) **Consider the 2-dimentional VAR(1)**

$$\mathbf{y_t} = \mathbf{\Phi_1 y_{t-1}} + \epsilon_\mathbf{t}.$$

i. **What conditions on Φ_1 are required for the VAR(1) to have cointegration?**
The system is cointegrated if ˘$_\mathbf{1}$ has one eigenvalue equal to one and the other eigenvalue less than one in complex modulus.

ii. **Describe how to test for conintegration using the Engle-Granger method.**
First, test that each time series is nonstationary with an augmented Dickey-Fuller test. If you cannot reject nonstationarity, you can proceed. If you reject nonstationarity, then there is no cointegration. Second, estimate an OLS regression of $\mathbf{y_{t,1}}$ on $\mathbf{y_{t,2}}$ and collect the estimated residuals. Then test if the estimated residuals are stationary. If they are, cointegration is present.

(c) **Define conditional Value-at-Risk. Describe two methods for estimating this and compare their strengths and weaknesses.**
Conditional Value-at-Risk is defined as the value $VaR_{t+1|t}$ such that, given the information at period t (written $\mathcal{F}_t$), next period's asset return r_{t+1} satisfies the following for a given $0 < \alpha < 1$:

$$Pr(r_{t+1} < -VaR_{t+1|t}|\mathcal{F}_t) = \alpha$$

There are many answers to the second part of the question. Conditional Value-at-Risk can be estimated with RiskMetrics, a GARCH model assuming conditional normality, A GARCH model assuming no distribution on the shocks or a CaViaR model. If we are prepared to argue that the conditonal aspect of the model is irrelevant, then Value-at-Risk can be estimated with an unconditional model. This includes parametric and nonparametric estimation. Each of these models requires different assumptions and estimation methods. The basic trade-off is between complexity of the model and difficulty in estimating it accurately given the available data.

(d) **Define conditional expected shortfall. Is this a more or less difficult object to estimate than Value-at-Risk? Why?**
Expected shortfall is defined as

$$ES = E_t(r_{t+1}|r_{t+1} < -VaR_{t+1|t}).$$

This is a more difficult object to estimate than Value-at-Risk because it requires determining the Value-at-Risk to compute it. Then, you must compute the exptected value of returns conditional on a Value-at-Risk exceedance. This requires knowledge of the entire left tail of the conditional return distribution.

(e) **Give the formula for the original 1996 RiskMetrics model. How does this differ from the updated 2006 RiskMetrics model? How is this 1996 model estimated?**
The 1996 RiskMetrics formula is an exponentially weighted moving average:

$$\Sigma_t = (1-\lambda)\sum_{i=1}^{\infty} \lambda^{i-1} \epsilon_{\mathrm{t-i}} \epsilon'_{\mathrm{t-i}}.$$

The 2006 RiskMetrics formula is a similar weighted average of past values $\epsilon_{t-i}\epsilon'_{t-i}$, but with more weight on recent and very distant observations and less on intermediate times. The 1996 RiskMetrics formula is not estimated. It uses $\lambda = 0.94$ for daily data and $\lambda = 0.97$ for monthly data.