Heuristic Methods for Financial Modelling
Concepts and Applications

Dietmar Maringer

CCFEA Summer School, University of Essex

September 2007
The Importance of Financial Modeling in (Computational) Finance

- discover relationships
- understand processes
- test models
- modeling
How it’s usually done

1. state a model
2. apply to data
3. test and evaluate the outcome

Example

1. \( r_t = f(x_t, \psi) \)
2. Estimate parameters \( \psi \)
3. test \( \psi \) for statistical significance
A Short Primer in Risk Modelling


$$r_t = \mu + e_t$$
$$e_t \sim N\left(0; \sigma_t^2\right)$$

$$\hat{e}_t^2 = \sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i e_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2$$

Finding the parameters $\psi = [\mu \quad \alpha_0 \quad \alpha_1 \quad \beta_1]$ for GARCH(1,1):

$$\max_\psi \mathcal{L} = -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^T \left( \ln(\sigma_t^2) + \frac{e_t^2}{\sigma_t^2} \right)$$
Estimation Issues

McCullough and Vinod (1999, p. 635)

“Many textbooks convey the impression that all one has to do is use a computer to solve the problem, the implicit and unwarranted assumptions being that the computer’s solution is accurate and that one software package is as good as any other.”
Estimation at Work

GARCH(1,1) for DEM / GPB exchange rate

- 1974 daily observations
- actual result:
  - Fiorentini, Calzolari, Panattoni (1996) provide analytical derivatives for GARCH(1,1)
  - benchmark results in Bollerslev and Ghysels (1996)
- results from software packages:
  - Brooks et al. (2001)
  - estimate the parameters with nine standard software packages
**GARCH Estimation at Work**


<table>
<thead>
<tr>
<th>Method</th>
<th>$\mu$</th>
<th>$\alpha_0$</th>
<th>$\alpha_1$</th>
<th>$\beta_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benchmark</td>
<td>–0.00619041</td>
<td>0.0107613</td>
<td>0.153134</td>
<td>0.805974</td>
</tr>
<tr>
<td>E-Views</td>
<td>–0.00540</td>
<td>0.0096</td>
<td>0.143</td>
<td>0.821</td>
</tr>
<tr>
<td>Gauss-Fanpac</td>
<td>–0.00600</td>
<td>0.0110</td>
<td>0.153</td>
<td>0.806</td>
</tr>
<tr>
<td>Limdep</td>
<td>–0.00619</td>
<td>0.0108</td>
<td>0.153</td>
<td>0.806</td>
</tr>
<tr>
<td>Matlab</td>
<td>–0.00619</td>
<td>0.0108</td>
<td>0.153</td>
<td>0.806</td>
</tr>
<tr>
<td>Microfit</td>
<td>–0.00621</td>
<td>0.0108</td>
<td>0.153</td>
<td>0.806</td>
</tr>
<tr>
<td>SAS</td>
<td>–0.00619</td>
<td>0.0108</td>
<td>0.153</td>
<td>0.806</td>
</tr>
<tr>
<td>Shazam</td>
<td>–0.00613</td>
<td>0.0107</td>
<td>0.154</td>
<td>0.806</td>
</tr>
<tr>
<td>Rats</td>
<td>–0.00625</td>
<td>0.0108</td>
<td>0.153</td>
<td>0.806</td>
</tr>
<tr>
<td>TSP</td>
<td>–0.00619</td>
<td>0.0108</td>
<td>0.153</td>
<td>0.806</td>
</tr>
</tbody>
</table>
Estimation Issues

Uphill search and Starting Point

Objective: find parameter(s) $\psi^*$ s.t. $\mathcal{L}(\psi) \rightarrow \text{max}$
Traditional vs. Heuristic Methods

Objective: find $\psi^*$ that maximizes $\mathcal{L}(\psi)$

Uphill search

Initialize $\psi$ with “good” guess

REPEAT
make sophisticated guess for $\Delta \psi$

$\psi^n := \psi + \Delta \psi$
if $\mathcal{L}(\psi^n) > \mathcal{L}(\psi)$
then $\psi := \psi^n$

UNTIL converged

report last $\psi$

Threshold Accepting

Initialize $\psi$ with “good” random guess

REPEAT
make sophisticated random guess for $\Delta \psi$

$\psi^n := \psi + \Delta \psi$
if $\mathcal{L}(\psi^n) > \mathcal{L}(\psi)$ – Threshold
then $\psi := \psi^n$
lower Threshold
UNTIL converged halting criterion met

report last best $\psi$
GARCH Estimation with Threshold Accepting

First Results (Maringer, 2005)

<table>
<thead>
<tr>
<th>Method</th>
<th>$\mu$</th>
<th>$\alpha_0$</th>
<th>$\alpha_1$</th>
<th>$\beta_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benchmark</td>
<td>-0.00619041</td>
<td>0.0107613</td>
<td>0.153134</td>
<td>0.805974</td>
</tr>
<tr>
<td>E-Views</td>
<td>-0.00540</td>
<td>0.0096</td>
<td>0.143</td>
<td>0.821</td>
</tr>
<tr>
<td>Gauss-Fanpac</td>
<td>-0.00600</td>
<td>0.0110</td>
<td>0.153</td>
<td>0.806</td>
</tr>
<tr>
<td>Limdep</td>
<td>-0.00619</td>
<td>0.0108</td>
<td>0.153</td>
<td>0.806</td>
</tr>
<tr>
<td>Matlab</td>
<td>-0.00619</td>
<td>0.0108</td>
<td>0.153</td>
<td>0.806</td>
</tr>
<tr>
<td>Microfit</td>
<td>-0.00621</td>
<td>0.0108</td>
<td>0.153</td>
<td>0.806</td>
</tr>
<tr>
<td>SAS</td>
<td>-0.00619</td>
<td>0.0108</td>
<td>0.153</td>
<td>0.806</td>
</tr>
<tr>
<td>Shazam</td>
<td>-0.00613</td>
<td>0.0107</td>
<td>0.154</td>
<td>0.806</td>
</tr>
<tr>
<td>Rats</td>
<td>-0.00625</td>
<td>0.0108</td>
<td>0.153</td>
<td>0.806</td>
</tr>
<tr>
<td>TSP</td>
<td>-0.00619</td>
<td>0.0108</td>
<td>0.153</td>
<td>0.806</td>
</tr>
<tr>
<td>Heuristic optimization</td>
<td>-0.00619034</td>
<td>0.0107614</td>
<td>0.153134</td>
<td>0.805973</td>
</tr>
</tbody>
</table>
Estimation Issues

Which parameters?

underlying model: $\psi^{TR}$

“best” estimator from observed data: $\psi^{ML}$

result from optimizer: $\psi$

Main Problems with Parameter Estimation

1. identifying the Maximum Likelihood estimator
2. differences between Maximum Likelihood estimator and “True” estimator
3. differences between reported and actual Maximum Likelihood estimator
Computational Study (Winker and Maringer (2006))

Data Driven Monte Carlo

- Data Generating Process: GARCH(1,1) model with $\psi^{TR}$ from Bollerslev and Ghysels (1996)
- 100 data series with 2000 (+100) observations each
- estimate parameters for first $T = [50, 100, 200, 400, 1000, 2000]$ observations

⇒ 600 cases

Testing the Heuristic

- allow for $I = [1, 5, 10, 25, 50, 100] \cdot 1000$ iterations
- for each instance: approx. 1700 runs with TA
Computational Study (Winker and Maringer (2006))

“Reliability” of Threshold Acceptance

- Benchmark: Matlab GARCH toolbox
- Let $\Delta = L^{*,TA} - L^{Matlab}$

<table>
<thead>
<tr>
<th>Case</th>
<th>Condition</th>
<th>Count</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>88 cases</td>
<td>$\Delta &lt; -10^{-10}$</td>
<td></td>
<td>(extreme: $-0.0000001$)</td>
</tr>
<tr>
<td>259 cases</td>
<td>$</td>
<td>\Delta</td>
<td>&lt; 10^{-10}$</td>
</tr>
<tr>
<td>253 cases</td>
<td>$\Delta &gt; +10^{-10}$</td>
<td></td>
<td>(extreme: $+9.45$)</td>
</tr>
<tr>
<td>600 cases</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Convergence of Actual ML Estimator

Drawing from a sample
- true parameters of DGP, $\psi^{TR}$
- $T$ observations
  $\rightarrow \psi^{ML,T}$

Herwartz (2004, p. 202)
If $\psi^{ML,T}$ is a consistent estimator
- $\psi^{ML,T} \xrightarrow{\text{asym}} N$
- converges with standard rate $\sqrt{T}$ to $\psi^{TR}$
- $\rightarrow$ choose $T$ proportional to $1/\varepsilon^2$ to obtain
  $P\left(\left|\psi^{ML,T} - \psi^{TR}\right| < \varepsilon\right) > 1 - \delta$
Convergence of Reported Estimator

Threshold Acceptance
- heuristic search & optimization method
- stochastic elements
- given $T$: $\psi^{T,I,r}$ from run $r$ with $I$ iterations
- ideally: $\psi^{T,I,r} = \psi^{ML,T}$

Based on convergence result by Althöfer & Koschnik (1991):
number of iterations, $I(\delta, \epsilon)$, such that

$$P\left(|\psi^{T,I,r} - \psi^{ML,T}| < \epsilon\right) > 1 - \delta$$
Joint Convergence

Achieving a precision $\varepsilon$ with probability $\delta$

reported and ML $\text{prob}(\left|\psi_{ML,T} - \psi_{T,I,r}^{T}\right| < \varepsilon) > 1 - \delta$ $I \geq I(\varepsilon, \delta)$

ML and true $\text{prob}(\left|\psi_{ML,T} - \psi_{TR}^{T}\right| < \varepsilon) > 1 - \delta$ $T \geq T(\varepsilon, \delta)$

reported and true $\text{prob}(\left|\psi_{T,I,r}^{T} - \psi_{TR}^{T}\right| < \varepsilon) > 1 - \delta$ $I \geq I(T(\varepsilon, \delta))$
Convergence of the ML estimators

Distribution of $ψ_{ML,T}^T$ for 100 data series
Convergence of the reported and ML estimators

Distribution of reported results for one specimen case

1 000 Iterations

\[ \cdots \psi^{TR} \]
\[ * \psi^{ML,T} \]
\[ \rightarrow \text{median} \]
\[ - - 10\% (90\%) \text{ quantile} \]
Convergence of the reported and ML estimators

Distribution of reported results for one specimen case

5 000 Iterations

\[ \text{\psi}^{\text{TR}} \]
\[ * \ \psi^{\text{ML},T} \]

$\rightarrow$ median

- - 10% (90%) quantile
Convergence of the reported and ML estimators

Distribution of reported results for one specimen case

25,000 Iterations

\[ \psi^{TR} \]

\[ * \psi^{ML,T} \]

\[ \Rightarrow \text{median} \]

\[ - \sim \text{10\% (90\%) quantile} \]
Evaluation

Measures of Deviation I: Reported and Actual ML Parameters

\[
MSD_{p}^{\text{ML}} = \frac{1}{R} \cdot \sum_{r} \left( \psi_{p}^{d,T,I,r} - \psi_{p}^{\text{ML},d,T} \right)^{2}
\]

\[
\ln \left( MSD_{p}^{\text{ML},d,T} \right) = a + b \cdot \ln(I)
\]

\[
\begin{array}{c|cccc}
 p & \mu & \alpha_0 & \alpha_1 & \beta_1 \\
 b & -1.527 & -2.561 & -2.207 & -2.520 \\
 R^2 & .913 & .901 & .891 & .897 \\
\end{array}
\]
averaged over data series \(d; T = 400\)
Evaluation

Measures of Deviation II: Reported and “True” Parameters

\[
MSD_{p}^{TR,d,T,I} = \frac{1}{R} \cdot \sum_{r} \left( \psi_{p}^{d,T,I,r} - \psi_{p}^{TR} \right)^{2}
\]

\[
\ln \left( MSD_{p}^{TR,d,I} \right) = a + b \cdot \ln(T)
\]

<table>
<thead>
<tr>
<th>( p )</th>
<th>( \mu )</th>
<th>( \alpha_0 )</th>
<th>( \alpha_1 )</th>
<th>( \beta_1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( b )</td>
<td>-0.962</td>
<td>-1.890</td>
<td>-1.159</td>
<td>-1.643</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>.428</td>
<td>.615</td>
<td>.567</td>
<td>.639</td>
</tr>
</tbody>
</table>

averaged over data series \( d; I = 100000 \)
Linear Factor Models

A Simple Linear Factor Model

\[ r_{i,t} = \alpha_0 + \beta_1 f_{1,t} + \beta_2 f_{2,t} + \cdots + \beta_n f_{n,t} + \varepsilon_{i,t} \]

Ideally:

- data: “reasonable” frequency
- individually for each stock \( i \)
- parsimonious, e.g., at most \( k \) out of \( n \) different factors

The Modelling Problem:

+ estimation of parameters \( \beta \)
- selection of factors the \( k \) factors
Genetic Algorithms for Linear Factor Models

Basic Idea

binary strings represent inclusion / exclusion of factors:

\[
\begin{array}{cccccccc}
  f_1 & f_2 & f_3 & f_4 & f_5 & f_6 & f_7 \\
  1 & 1 & 0 & 0 & 1 & 0 & 0 \\
\end{array}
\]

✓ ✓ − − ✓ − −

Ingredients

1. population of candidate solutions
2. cross-over:

<table>
<thead>
<tr>
<th>parents</th>
<th>offspring</th>
</tr>
</thead>
</table>
| \begin{array}{cccc}
  1 & 0 & 0 & 1 \\
  0 & 1 & 1 & 1 \\
\end{array} | \begin{array}{cccc}
  1 & 0 & 1 & 1 \\
  0 & 1 & 0 & 1 \\
\end{array} |

3. mutation:

<table>
<thead>
<tr>
<th>before</th>
<th>after</th>
</tr>
</thead>
</table>
| \begin{array}{cccc}
  \ldots & 0 & 1 & 1 \\
\end{array} | \begin{array}{cccc}
  \ldots & 0 & 0 & 1 \\
\end{array} |

4. survival of the fittest
Pseudo-Algorithm

randomly initialize population of $P$ binary strings
REPEAT

%% generate new solutions:
    FOR offspring = 1...$P$
    pick 2 current solutions = parents;
    cross-over parents;
    with probability $\pi$, mutate offspring;
    compute offspring’s fitness;
END

%% select new population
    pool = current population + offspring;
    pick $P$ solutions from pool based on fitness;
UNTIL converged;
Heuristic Factor Selection (Maringer (2004))

Computational Study

- S & P 100 stocks, 1995–2000
- 103 MSCI indices:
  - 23 country indices
  - 42 regional indices
  - 38 industries
- combinatorial problem: $2^n$ combinations to (not) include $n$ indices

<table>
<thead>
<tr>
<th>factors:</th>
<th>23</th>
<th>42</th>
<th>38</th>
<th>103</th>
</tr>
</thead>
<tbody>
<tr>
<td>combinations:</td>
<td>$8.39E+06$</td>
<td>$4.40E+12$</td>
<td>$2.75E+11$</td>
<td>$1.01E+31$</td>
</tr>
</tbody>
</table>

- heuristic: Memetic Algorithm
Parsimonious Linear Regression Models

Computational Study: $R^2$ with $k = 5$ factors
Parsimonious Linear Regression Models

Computational Study: Picked factors when $k = 5$
VAR– and VEC–Models

Basic Idea

\[ Y_t = \sum_{i=1}^{k+1} \Pi_i Y_{t-i} + \varepsilon_t \]

with initial values \([ Y_{-k} \ldots Y_0 ]\) and \( Y_t = \begin{bmatrix} Y_{1,t} \\ \vdots \\ Y_{d,t} \end{bmatrix} \)

Expressed in error-correction notation (Ahn and Reinsel (1990))

\[ \Delta Y_t = \sum_{i=1}^k \Gamma_i \Delta Y_{t-i} + \Pi Y_{t-k-1} + \varepsilon_t \]
VAR– and VEC–Models (cont.’d)

Parsimonious Models and Rank Identification (Winker and Maringer (2005))

\[
\begin{align*}
\Pi &= \begin{bmatrix} 0 & 0 \\ 0.20 & -0.20 \\ -0.07 & 0.17 \end{bmatrix}, \\
\Omega_\epsilon &= \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix}, \\
\Gamma_1 &= \begin{bmatrix} \cdot & \spadesuit & -\spadesuit \\ \cdot & \cdot & -\spadesuit \\ \cdot & \cdot & \spadesuit \end{bmatrix}, \\
\Gamma_2 &= \begin{bmatrix} -\spadesuit & \cdot & -\spadesuit \\ -\spadesuit & \cdot & -\spadesuit \\ -\spadesuit & -\spadesuit & \cdot \end{bmatrix}, \\
\Gamma_3 &= \begin{bmatrix} -\spadesuit & \cdot & -\spadesuit \\ -\spadesuit & \cdot & -\spadesuit \\ -\spadesuit & -\spadesuit & \cdot \end{bmatrix}, \\
\Gamma_4 &= \begin{bmatrix} -\spadesuit & \cdot & -\spadesuit \\ -\spadesuit & \cdot & -\spadesuit \\ -\spadesuit & -\spadesuit & \cdot \end{bmatrix}, \\
\Gamma_5 &= \begin{bmatrix} \cdot & \cdot & -\spadesuit \\ \cdot & -\spadesuit & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix}, \\
\Gamma_{i \geq 6} &= 0
\end{align*}
\]

Parameter estimation (Ahn and Reinsel (1990)):

\[
\begin{align*}
\hat{b}_{t+1} &= \hat{b}_{t} + \left( \sum_{t=1}^{T} U_t^{*} \hat{\Omega}_{\epsilon}^{-1} \epsilon_t \right)^{(-1)} \left( \sum_{t=1}^{T} U_t^{*} \hat{\Omega}_{\epsilon}^{-1} \epsilon_t \right) \\
U_t^{*} &= \left( \alpha' \otimes [0, I_{d-r}] Y_{t-1}, I_d \otimes [(\beta Y_{t-1})', \Delta Y_{t-1}, \ldots, \Delta Y_{t-k}]' \right)'
\end{align*}
\]
Smooth Transition Autoregressive (STAR) Models

Basic Idea of Smooth Transition Autoregressive (STAR) Models

- two regimes
- transition variable, $z_t$
- transition function $G(z_t) \rightarrow [0, 1]$

The Model

$$x_t = (1 - G(z_t)) \cdot \sum_{\ell \in \mathcal{L}_1} \alpha_{\ell} x_{t-\ell}$$

$$+ G(z_t) \cdot \sum_{\ell \in \mathcal{L}_2} \phi_{\ell} x_{t-\ell} + \varepsilon_t$$

Problem (Maringer and Meyer (2007); Chen and Maringer (wip))

- parameter estimation: $\gamma, c$
- lag structures $\mathcal{L}_1, \mathcal{L}_2$
- factor selection
Summary & Conclusions for Parameter Estimation

Main results

- heuristic methods are capable, appropriate and useful
- more advanced financial modelling possible
- more CPU time (iterations, runs) = higher precision
- analysis of convergence

Further research

- alternative models
- choice of heuristic
- distribution of estimations


