Financial Econometrics

- Week 5, Lecture 2 -

Indirect Inference

VU Econometrics and Data Science
2024-2025

Paolo Gorgi

VU Vrije Universiteit Amsterdam

Today's class

1. Indirect inference
   - Indirect inference estimation
   - Example: estimation of MA model
2. Estimation of the SV model
   - SV model by Indirect Inference
   - Filtering paths for the SV model

Indirect inference

Indirect Inference

Types of estimation problems:

- Simple models have tractable likelihoods and estimators (linear regression model).
  - parameter estimates can be directly obtained by using analytic expression of the estimator.
- Complicated models have tractable likelihoods but intractable estimators (GARCH model).
  - parameter estimates can be obtained by optimizing the likelihood function numerically.
- Very complex models have intractable likelihoods and estimators (SV model).
  - advanced simulation-based estimation techniques are needed!

Indirect Inference: Discovery

Problem: How can we estimate models were we cannot write down the log-likelihood function?

A solution: Indirect Inference

- Introduced by Smith (1993) and independently by Gourieroux, Monfort and Renault (1993)
- Encompasses SMM, SML, EMM, IRF-Matching, etc

Applied to many problems in economics and finance!

Indirect Inference: Examples

Application of indirect inference:

- Bias correction (near unit-root);
- Parameter-driven models (SV);
- Regression models with time-varying parameters;
- Nonlinear dynamic models with latent variables;
- Structural models in economics and finance (dynamic stochastic general equilibrium models, real business cycle models).

Indirect Inference: how it works (i)

Suppose we have a sample of observed data $y_1, ..., y_T$ and we wish to estimate the true parameter vector θ_0 of a *parameter-driven model*.

- **First:** describe the properties of the observed data $y_1, ..., y_T$ using a vector of *auxiliary statistics* $\hat{B}_T$.
- **Second:** simulate a sample of length H from our parameter-driven model $\tilde{y}_1(\theta), ..., \tilde{y}_H(\theta)$ for a given value of θ.
- **Third:** obtain the vector of *auxiliary statistics* for the simulated data $\tilde{B}_H(\theta)$.
- **Fourth:** find the value of θ that makes $\tilde{B}_H(\theta)$ as close as possible to $\hat{B}_T$.

Note: the observed data are assumed to be generated by the parameter-driven model under θ_0.

Indirect Inference: how it works (ii)

Note:

- $\hat{B}_T$ and $\tilde{B}_H(\theta)$ may contain moments (like the mean, variance and covariances), or parameter estimates of models that are simple to estimate (regression, ARMA, etc.).
- The *auxiliary statistics* $\hat{B}_T$ and $\tilde{B}_H(\theta)$ must provide an appropriate description of the dynamic properties of the data!
- Notice that we can simulate a very long path $\tilde{y}_1(\theta), ..., \tilde{y}_H(\theta)$ from our model and therefore obtain a very accurate auxiliary statistics $\tilde{B}_H(\theta)$.
- The *indirect inference estimate* is the value of θ that makes simulated data as similar as possible to observed data!

Indirect Inference: definition (i)

Formally: the **indirect inference estimator** $\hat{\theta}_{TH}$ is defined as

$$\hat{\theta}_{TH} = \arg\min_{\theta \in \Theta} d\big(\hat{B}_T, \tilde{B}_H(\theta)\big).$$

where $d\big(\hat{B}_T, \tilde{B}_H(\theta)\big)$ denotes the quadratic distance

$$d\big(\hat{B}_T \ , \ \tilde{B}_H(\theta)\big) = \Big(\hat{B}_T - \tilde{B}_H(\theta)\Big)^\top W \Big(\hat{B}_T - \tilde{B}_H(\theta)\Big).$$

Note: W is a weighting matrix that can give different weights to each auxiliary statistic.

In practice: W can be set equal to the identity matrix.

Indirect Inference: definition (ii)

Note: the quadratic distance $d\big(\hat{B}_T, \tilde{B}_H(\theta)\big)$ can be minimized using numerical methods as the Newton-Raphson algorithm (see Week 2).

Important: The simulations must be carried out using the same random *seed* value for any θ

> Otherwise estimation error renders the criterion function $d\big(\hat{B}_T, \tilde{B}_H(\theta)\big)$ non smooth and difficult to optimize!

Indirect Inference: a diagram

$$\begin{array}{ccc}
\text{Observed data} & & \text{Simulated data} \\
y_1, ..., y_T & & \tilde{y}_1(\theta), ..., \tilde{y}_H(\theta) \\
\downarrow & & \downarrow \\
\hat{\beta}_T & & \tilde{\beta}_H(\theta) \\
\downarrow & & \downarrow
\end{array}$$

$$\hat{\theta}_{TH} = \arg\min_{\theta \in \Theta} d\big(\hat{\beta}_T \ , \ \tilde{\beta}_H(\theta)\big)$$

Indirect Inference: Asymptotic properties

Lemma (consistency and asymptotic normality)

Under appropriate conditions, the indirect inference estimator is consistent as both H and T go to infinity,

$$\hat{\theta}_{TH} \xrightarrow{p} \theta_0 \quad \text{as} \quad T \to \infty \quad \text{and} \quad H \to \infty.$$

Moreover, the indirect inference estimator is asymptotically Gaussian (here H is a multiple of T given by $H = \Delta T$)

$$\sqrt{T}(\hat{\theta}_{TH} - \theta_0) \xrightarrow{d} N\left(\mathbf{0}, \left(1 + \frac{1}{\Delta}\right)\Sigma\right).$$

Note: The asymptotic covariance matrix of the indirect inference estimator depends on H. The uncertainty and inefficiency introduced by simulations vanishes as $H \to \infty$.

Indirect Inference: identifiability and consistency (i)

Important:

- As sample size increases ($T \to \infty$ and $H \to \infty$), the auxiliary statistics converge to limit values that depend only on θ

$$\hat{B}_T \overset{p}{\to} B(\theta_0) \quad \text{and} \quad \tilde{B}_H(\theta) \overset{p}{\to} B(\theta),$$

  where the limit function $B(\theta)$ is called the *binding function*.

- Therefore, we have that

$$\hat{\theta}_{TH} \overset{p}{\to} \arg\min_{\theta \in \Theta} d\big(B(\theta_0), B(\theta)\big).$$

Indirect Inference: identifiability and consistency (ii)

Important:

- The II estimator is consistent ($\hat{\theta}_{TH} \overset{p}{\to} \theta_0$) if $\theta = \theta_0$ is the unique minimizer of $d\big(B(\theta_0), B(\theta)\big)$.
- When $\theta = \theta_0$ is the unique minimizer, we say that the parameter vector θ is identified by the *auxiliary statistics*.
- **In practice:** It is important to choose auxiliary statistics $\hat{B}_T$ that identify all parameters in θ.
- **In practice:** the number of auxiliary statistics has always to be equal or larger than the number of parameters in the vector θ

Important: If $\dim(\theta) > \dim(\hat{B}_T)$ then θ is not identified!

Example: Estimating an MA model (i)

Example: Consider the following MA(1) model

$$y_t = \epsilon_t + \phi\epsilon_{t-1} \ , \quad \epsilon_t \sim N(0, \sigma^2)$$

Recall: ϕ and σ^2 determine a number of properties of $\{y_t\}_{t\in\mathbb{Z}}$ including its variance and autocorrelation structure

Therefore: the *sample variance* and *first-order autocovariance* are natural choices as auxiliary statistics.

Note: we have 2 parameters to estimate $\theta = (\phi, \sigma^2)^\top$ and 2 auxiliary statistics (sample variance, and first-order autocovariance).

Example: Estimating an MA model (ii)

Hence: we have the following auxiliary statistics

$$\hat{B}_T = \begin{bmatrix} \hat{\gamma}_0 \\ \hat{\gamma}_1 \end{bmatrix} = \begin{bmatrix} (1/T)\sum_{t=1}^{T} y_t^2 \\ (1/T)\sum_{t=2}^{T} y_t y_{t-1} \end{bmatrix}, \quad \text{and}$$

$$\tilde{B}_H(\theta) = \begin{bmatrix} \tilde{\gamma}_0(\theta) \\ \tilde{\gamma}_1(\theta) \end{bmatrix} = \begin{bmatrix} (1/H)\sum_{t=1}^{H} \tilde{y}_t^2(\theta) \\ (1/H)\sum_{t=2}^{H} \tilde{y}_t(\theta)\tilde{y}_{t-1}(\theta) \end{bmatrix}.$$

Note: the criterion function can be minimized with a standard algorithm.

Let us take a look at the precision of the II estimator...

Example: Estimating an MA model (iii)

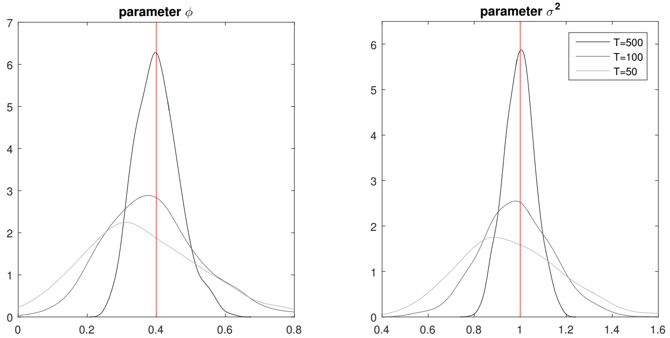

Figure: Distribution of the indirect inference estimator for different sample sizes T. The length of the simulations is set $H = 20T$.

Example: Estimating an MA model (iv)

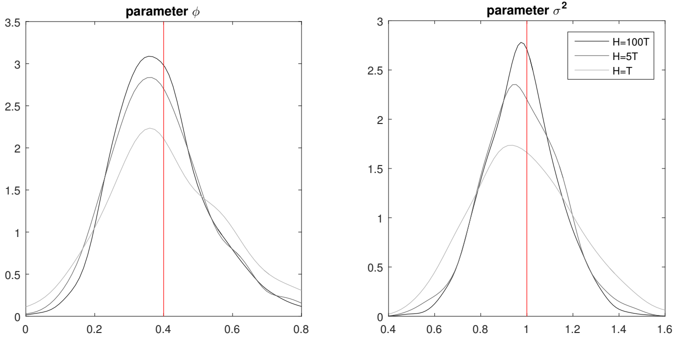

Figure: Distribution of the indirect inference estimator for different length of the simulated series H. The sample size of the series is set $T = 100$.

Estimation of the SV model

Estimation of the SV model (i)

Stochastic volatility model:

$$y_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \exp(f_t),$$
$$f_t = \omega + \beta f_{t-1} + \eta_t.$$

Recall: ω, β and σ_η^2 determine certain moments of y_t

- Unconditional variance of y_t;
- Unconditional kurtosis;
- Autocovariance in squared log-returns y_t^2.

Hence: we can use these moments as *auxiliary statistics*!

Estimation of the SV model (ii)

Auxiliary statistics for SV model:

$$\hat{B}_T = \begin{bmatrix} \hat{s}^2 \\ \hat{k}^2 \\ \hat{\gamma}_1^2 \\ \hat{\gamma}_2^2 \end{bmatrix} = \begin{bmatrix} (1/T)\sum_{t=1}^{T} y_t^2 \\ (1/T)\sum_{t=1}^{T} y_t^4 \\ (1/T)\sum_{t=2}^{T}(y_t^2 - \hat{s}^2)(y_{t-1}^2 - \hat{s}^2) \\ (1/T)\sum_{t=3}^{T}(y_t^2 - \hat{s}^2)(y_{t-2}^2 - \hat{s}^2) \end{bmatrix} \quad \text{and}$$

$$\tilde{B}_H(\theta) = \begin{bmatrix} \tilde{s}^2(\theta) \\ \tilde{k}^2(\theta) \\ \tilde{\gamma}_1^2(\theta) \\ \tilde{\gamma}_2^2(\theta) \end{bmatrix} = \begin{bmatrix} (1/H)\sum_{t=1}^{H} \tilde{y}_t^2(\theta) \\ (1/H)\sum_{t=1}^{H} \tilde{y}_t^4(\theta) \\ (1/H)\sum_{t=2}^{H}\left(\tilde{y}_t^2(\theta) - \tilde{s}^2(\theta)\right)\left(y_{t-1}^2(\theta) - \tilde{s}^2(\theta)\right) \\ (1/H)\sum_{t=3}^{H}\left(\tilde{y}_t^2(\theta) - \tilde{s}^2(\theta)\right)\left(y_{t-2}^2(\theta) - \tilde{s}^2(\theta)\right) \end{bmatrix}.$$

Estimation of the SV model (iii)

Note: instead of squared log-returns other suitable transformations could be used as for instance absolute log-returns $|y_t|$.

Alternative auxiliary statistics: the parameters of an AR(p) model for squared log-returns y_t^2 constitute a natural alternative to these raw moments of the data.

AR(p) model:

$$y_t^2 = b_0 + \sum_{i=1}^{p} b_i y_{t-i}^2 + \epsilon_t \ , \quad \epsilon_t \sim N(0, c^2).$$

Estimation of the SV model (iv)

Important: the parameters of the AR(p) model describe the autocovariance structure of the squared log-returns, and this determines precisely the moments of y_t.

Auxiliary statistics: correspond to the estimated AR(p) parameters $\hat{b}_0, \ldots, \hat{b}_p$ and $\hat{c}^2$.

$$\hat{B}_T = \begin{bmatrix} \hat{b}_0 \\ \vdots \\ \hat{b}_p \\ \hat{c}^2 \end{bmatrix} \qquad \text{and} \qquad \tilde{B}_H(\theta) = \begin{bmatrix} \tilde{b}_0(\theta) \\ \vdots \\ \tilde{b}_p(\theta) \\ \tilde{c}^2(\theta) \end{bmatrix}.$$

Estimation of the SV model with R (i)

Estimation with R: `estimate_SV_II.R` and `sim_m_SV.R`

Step 1: Write an R function that generates the *auxiliary statistics* of the SV model from simulated data $\tilde{B}_H(\theta)$ for any given parameter value θ.

Function `sim_m_SV()`: uses as auxiliary statistics:

- The sample variance of log-returns y_t.
- The sample kurtosis of log-returns y_t.
- The first-order autocorrelation of absolute-log-returns $|y_t|$.

Estimation of the SV model with R (ii)

```
sim_m_SV <- function(e,par){
```

Input of `sim_m_SV()`:

- Parameter vector `par`.
- Simulated errors `e` of length `H` for both the observation equation and the transition equation.

Output of `sim_m_SV()`:

- The function returns the vector `output` that contains *auxiliary statistics* of data simulated from SV model, as mentioned in the previous slide.

Estimation of the SV model with R (iii)

First: Use inputs to define SV parameters, simulation length H, and innovations $\{\epsilon_t\}_{t=1}^{H}$ and $\{\eta_t\}_{t=1}^{H}$

```
omega <- par[1]
beta <- exp(par[2])/(1+exp(par[2]))
sig2f <- exp(par[3])
H <- length(e[,1])


epsilon <- e[,1]
eta <- sqrt(sig2f)*e[,2]
```

Define data vectors:

```
x <- rep(0,H)
f <- rep(0,H)
```

Estimation of the SV model with R (iv)

Simulate from the SV model:

```
f[1] <- omega/(1-beta)
x[1] = exp(f[1]/2) * epsilon[1]
for(t in 2:H){
f[t] <- omega + beta * f[t-1] + eta[t]
x[t] <- exp(f[t]/2) * epsilon[t]
}
```

Finally: Return output vector

```
xa <- abs(x)
output <- c(var(x),kurtosis(x),cor(xa[2:H],xa[1:(H-1)]))
return(output)
```

Estimation of the SV model with R (v)

Step 2: Minimize the quadratic distance between the *auxiliary moments* $\tilde{B}_H(\theta)$ and $\hat{B}_T$.

R file: `estimate_SV_II.R`.

First: Estimate moments of observed data `x`

```
n <- length(x)
xa <- abs(x)
sample_m <- c(var(x), kurtosis(x), cor(xa[2:n],xa[1:(n-1)]))
```

Second: Choose simulation length `H` and simulate N(0,1) innovations

```
set.seed(123)
H <- 50*n
epsilon <- rnorm(H)
eta <- rnorm(H)
e <- cbind(epsilon,eta)
```

Estimation of the SV model with R (vi)

Third: set the starting values for optimization

```
b <- 0.90
sig2f <- 0.1
omega <- log(var(x))*(1-b)
par_ini <- c(omega,log(b/(1-b)),log(sig2f))
```

Finally: minimize $d(\tilde{B}_H(\theta), \hat{B}_T)$ with respect to theta

```
est <- optim(par=par_ini,
        fn=function(par) mean((sim_m_SV(e,par)-sample_m)^2),method = "B

omega_hat <- est$par[1]
beta_hat <- exp(est$par[2])/(1+exp(est$par[2]))
sig2f_hat <- exp(est$par[3])
theta_hat <- c(omega_hat,beta_hat,sig2f_hat)
```

Filtering paths for the SV model (i)

So far, we have seen how to estimate the parameter vector θ of the SV model by Indirect Inference.

However: we also wish to obtain a *filtered* path for the unobserved time-varying volatility $\{\sigma_t^2\}_{t=1}^T$!

Problem: how can we filter $\{\sigma_t^2\}_{t=1}^T$ if the updating equation depends on the unobserved innovations?

Solution 1: *Kalman filter* and *Particle filter* (Master)

Solution 2: Approximate ML filter!

Filtering paths for the SV model (ii)

Filtering paths of $\{\sigma_t^2\}_{t=1}^T$ for the SV model:

We want the path $\{\sigma_t^2\}_{t=1}^T$ that maximizes the likelihood; i.e. the joint density of $(y_1, \ldots, y_T, \sigma_1^2, \ldots, \sigma_T^2)$

$$p(y_1, \ldots, y_T, \sigma_1^2, \ldots, \sigma_T^2; \theta).$$

Problem: High-dimensional problem... maximization is difficult!

Solution: We consider a sequential maximization procedure.

Filtering paths for the SV model (iii)

Solution: we consider a sequential maximization procedure.

1. Start with a given value of σ_1^2.
2. Obtain the joint distribution of y_2 and σ_2^2 given σ_1^2.

$$p(y_2, \sigma_2^2|\sigma_1^2) = p(y_2|\sigma_2^2)p(\sigma_2^2|\sigma_1^2).$$

3. Find the value σ_2^2 that maximizes $p(y_2, \sigma_2^2|\sigma_1^2)$ (σ_1^2 is given).
4. Next, take σ_2^2 as given and optimize w.r.t. σ_3^2.
5. Repeat this procedure for any σ_t^2

$$p(y_t, \sigma_t^2|\sigma_{t-1}^2) = p(y_t|\sigma_t^2)p(\sigma_t^2|\sigma_{t-1}^2).$$

Filtering paths for the SV model (iv)

Note: we can maximize log-likelihoods instead of likelihoods

$$\log p(y_t, \sigma_t^2 | \sigma_{t-1}^2) = \log p(y_t | \sigma_t^2) + \log p(\sigma_t^2 | \sigma_{t-1}^2).$$

As a result, we can obtain the filtered path by solving for every $t = 1, \ldots, T$

$$\sigma_t^2 = \arg\max \left\{ \log p(y_t | \sigma_t^2) + \log p(\sigma_t^2 | \sigma_{t-1}^2) \right\},$$

where the initial value σ_1^2 is fixed.

Filtering paths for the SV model (v)

Important: The conditional densities $p(y_t|\sigma_t^2)$ and $p(\sigma_t^2|\sigma_{t-1}^2)$ are simple and analytically tractable!!!

$p(y_t|\sigma_t^2)$ **is determined** by the observation equation

$$y_t = \sigma_t \varepsilon_t$$

Hence: $y_t|\sigma_t^2 \sim N(0, \sigma_t^2)$

$$p(y_t|\sigma_t^2) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{y_t^2}{2\sigma_t^2}\right).$$

Recall: it is only the conditional density $p(y_t|Y^{t-1})$ that is intractable!!

Filtering paths for the SV model (vi)

$p(\sigma_t^2|\sigma_{t-1}^2)$ **is determined** by the transition equation

$$\sigma_t^2 = \exp(f_t), \quad f_t = \omega + \beta f_{t-1} + \eta_t$$

Hence: $\sigma_t^2|\sigma_{t-1}^2 \sim \log\text{-}N(\omega + \beta \log \sigma_{t-1}^2, \sigma_\eta^2)$

$$p(\sigma_t^2|\sigma_{t-1}^2) = \frac{1}{\sigma_t^2\sqrt{2\pi\sigma_\eta^2}} \exp\left(-\frac{(\log \sigma_t^2 - \omega - \beta \log \sigma_{t-1}^2)^2}{2\sigma_\eta^2}\right).$$

Filtering paths for the SV model (vii)

Note: maximizing with respect to σ_t is the same as maximizing with respect to f_t since $f_t = \log \sigma_t^2$.

Hence: we can restate the problem as follows

$$\hat{f}_t = \arg\min \left\{ y_t^2 \exp(-f_t) + 3 f_t + \frac{(f_t - \omega - \beta f_{t-1})^2}{\sigma_\eta^2} \right\},$$

for $t = 1, ..., T$ and a given initial value of f_1.

Filtering paths for the SV model (vii)

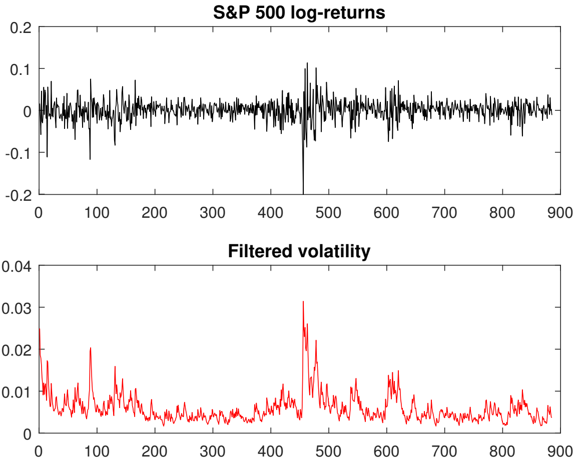

Figure: Weekly log-returns of S&P 500 and estimated volatility.

Filtering paths for SV model with R (i)

Filtering with R: `filter_SV.R` and `estimate_SV_II.R`.

Step 1: Write R function (`filter_SV()`) that contains *filtering criterion* that needs to be optimized wrt f_t (see slide 36).

Input `filter_SV()`: y_t, f_t, f_{t-1} and θ which are labeled `yt`, `ft`, `ft1` and `theta`.

Output `filter_SV()`: the value of the filtering criterion.

The function `filter_SV()` is given by the following code

```
filter_SV <- function(yt,ft,ft1,theta){
omega <- theta[1]
beta <- theta[2]
sig2f <- theta[3]
output <- yt^2*exp(-ft)+3*ft+(ft-omega-beta*ft1)^2/sig2f
return(output) }
```

Filtering paths for SV model with R (ii)

Step 2: optimize the function `filter_SV()` and obtain the filtered f_t for each time period $t = 1, \dots, T$.

```
f <- rep(0,n)
f[1] <- log(var(x))

for(t in 2:n){
  ft_ini <- f[t-1]
  f_est <- optim(par=ft_ini,
      fn= function(ft) filter_SV(x[t],ft,f[t-1],theta_hat),
        method = "BFGS")
  f[t] <- f_est$par
}
```