ECON 515 Time Series Analysis

Volatility Models, Vector Autoregression, and Forecasting

Zhan Gao

12 November 2024

Acknowledgements: This set of slides is heavily dependent on Professor Constantin Colonescu's book (Principles of Econometrics with R), Professor Zhentao Shi's lecture notes at Goergia Tech (ECON 4160), and Professor Hashem Pesaran's textbook *Time series and panel data econometrics*.

Overview

1. Volatility Models
2. Vector Autoregression (VAR)
3. Forecasting

Volatility Models

Volatility Models

- Consider a linear regression model,

$$r_t = \boldsymbol{\beta}'\mathbf{x}_{t-1} + \varepsilon_t,$$

where r_t can be stock return, inflatin rate, or output growth rate.

- Let $\mathrm{E}\left(\varepsilon_t | \Omega_{t-1}\right) = \mathbf{0}$, and

$$\mathbf{var}\left(\varepsilon_t | \Omega_{t-1}\right) = h_t^2,$$

where the conditional variance is time varying.

- This need not to hold *unconditionally*.
- The goal is to model h_t^2.

ARCH Model

Autoregressive conditional heteroskedasticity (ARCH) model

- The ARCH(1) model is defined as

$$h_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2, \quad \alpha_0 > 0.$$

- *Unconditionally*, when $|\alpha_1| < 1$,

$$\mathbf{var}\left(\varepsilon_t\right) = \sigma^2 = \mathrm{E}\left(h_t^2\right) = \frac{\alpha_0}{1-\alpha_1} > 0,$$

  i.e. unconditionally the ARCH(1) model is stationary if $|\alpha_1| < 1$.

- ARCH(p) model

$$\mathbf{var}\left(\varepsilon_t \mid \Omega_{t-1}\right) = h_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_p \varepsilon_{t-p}^2,$$

  where $\mathbf{var}\left(\varepsilon_t\right) = \sigma^2 = \alpha_0 / \left(1 - \alpha_1 - \ldots - \alpha_p\right)$ if roots of $1 - \sum_{i=1}^{p} \alpha_i \lambda^i = 0$ lie out of unit circle.

ARCH Model

Testing an ARCH model

- Step 1: regress r_t on x_{t-1}, and the estimated residuals $\hat{\varepsilon}_t = r_t - x'_{t-1}\beta$
- Step 2: Regress $\hat{\varepsilon}_t$ on a constant and its lagged terms,

$$\hat{\varepsilon}_t^2 = \alpha_0 + \alpha_1 \hat{\varepsilon}_{t-1}^2 + \cdots + \alpha_q \hat{\varepsilon}_{t-q}^2 + \text{ Error.}$$

Test the null hypothesis $\mathbb{H}_0: \ \alpha_1 = \alpha_2 = \cdots = \alpha_q = 0$ (by Lagrangian Multiplier (LM) Test).

```
data("byd", package="PoEdata")
FinTS::ArchTest(byd$r, lags=1, demean=TRUE)
```

```
##
##      ARCH LM-test; Null hypothesis: no ARCH effects
##
## data:  byd$r
## Chi-squared = 62.16, df = 1, p-value = 3.167e-15
```

```r
byd_arch <- tseries::garch(ts(byd$r), c(0,1))
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo

##
##  ***** ESTIMATION WITH ANALYTICAL GRADIENT *****
##
##
##      I     INITIAL X(I)        D(I)
##
##      1     1.334069e+00     1.000e+00
##      2     5.000000e-02     1.000e+00
##
##     IT   NF      F         RELDF    PRELDF    RELDX   STPPAR   D*STEP   NPRELDF
##      0    1  5.255e+02
##      1    2  5.087e+02  3.20e-02  7.13e-01  3.1e-01  3.8e+02  1.0e+00  1.34e+02
```

GARCH Model

Genaralized Autoregressive conditional heteroskedasticity (GARCH) model

- A GARCH(1,1) model

$$h_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \phi_1 h_{t-1}^2, \quad \alpha_0 > 0$$

  which can also be viewed as restricted form of an ARCH(∞) model.

- The unconditional variance exists and is fixed if $|\alpha_1 + \phi_1| < 1$.

- Higher-order GARCH

$$h_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \phi_i h_{t-i}^2.$$

- By invertibility, GARCH(1,1) models can be approximated by a ARCH model, which provides a way to test $\mathbb{H}_0: \ \alpha_1 = 0$.

In additional to `tseries::garch( ... )`, packages `rugarch` and `ugarch` are also powerful to

Some references

- The Risk Lab https://dachxiu.chicagobooth.edu/

  > We provide up-to-date daily annualized realized volatilities for individual stocks, ETFs, and future contracts, which are estimated from high-frequency data. We are in the process of incorporating equities from global markets.

- Textbook Treatment: Pesaran, M. H. (2015, Chapter 18). Time series and panel data econometrics. Oxford University Press.
  - or the textbooks listed in the syllabus.
- It is well-documented that there is stonger predictability in volatility than stock returns, for example see the volatility counterpart of Welch and Goyal (2008): Christiansen, C., Schmeling, M., & Schrimpf, A. (2012). A comprehensive look at financial volatility prediction by economic variables. Journal of Applied Econometrics, 27(6), 956-977.

Vector Autoregression

Vector Autoregression (VAR)

- Extension of AR to random vectors, which is a simple multivariate regression device
- Christopher A. Sims, 2011 Noble Prize

> How is the economy affected by unexpected events and changes in economic policy? What effects do interest rate hikes and tax reductions have on the production of good and services, unemployment, inflation and investment?

- A VAR(p) system,

$$\mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\Phi}_1 \mathbf{y}_{t-1} + \boldsymbol{\Phi}_2 \mathbf{y}_{t-2} + \ldots + \boldsymbol{\Phi}_p \mathbf{y}_{t-p} + \mathbf{u}_t$$

where $\mathbf{y}$ and μ are K-dimensional vectors, and $\boldsymbol{\Phi}_j$ is a $K \times K$ parameter matrix.

- Example: bivaraite VAR(1)

$$\begin{aligned} y_{1t} &= \mu_1 + \varphi_{11} y_{1,t-1} + \varphi_{12} y_{2,t-1} + u_{1t} \\ y_{2t} &= \mu_2 + \varphi_{21} y_{1,t-1} + \varphi_{22} y_{2,t-1} + u_{2t} \end{aligned}$$

Stationarity

Write VAR(p) to VAR(1)

$$\begin{pmatrix} \mathbf{y}_t \\ \mathbf{y}_{t-1} \\ \vdots \\ \mathbf{y}_{t-p+2} \\ \mathbf{y}_{t-p+1} \end{pmatrix} = \begin{pmatrix} \mathbf{\Phi}_1 & \mathbf{\Phi}_2 & \dots & \mathbf{\Phi}_{p-1} & \mathbf{\Phi}_p \\ \mathbf{I}_m & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_m & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{I}_m & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{y}_{t-1} \\ \mathbf{y}_{t-2} \\ \vdots \\ \mathbf{y}_{t-p+1} \\ \mathbf{y}_{t-p} \end{pmatrix} + \begin{pmatrix} \mathbf{u}_t \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix},$$

namely,

$$\mathbf{Y}_t = \mathbf{\Phi} \mathbf{Y}_{t-1} + \mathbf{U}_t.$$

Iteratively,

$$\mathbf{Y}_t = \mathbf{\Phi}^{t+M-p} \mathbf{Y}_{-M+p} + \sum_{j=0}^{t+M-p-1} \mathbf{\Phi}^j \mathbf{U}_{t-j}$$

. Need eigenvalues of $\mathbf{\Phi}$ to lie inside unit circle.

Estimation

- OLS estimation to obtain $\hat{\mathbf{\Phi}}_j$, $\boldsymbol{j = 1, 2, \cdots, p}$
- Compute residuals $\mathbf{U}_t$
- Estimated variance structure $\hat{\mathbf{\Omega}} = T^{-1}\sum_{t=1}^{T}\mathbf{U}_t\mathbf{U}_t'$.

Estimation

```
data("fred", package="PoEdata")
varmat <- as.matrix(cbind(dc = diff(fred[,"c"]), dy = diff(fred[,"y"])))
varfit <- vars::VAR(varmat)
summary(varfit)
```

```
##
## VAR Estimation Results:
## =========================
## Endogenous variables: dc, dy
## Deterministic variables: const
## Sample size: 198
## Log Likelihood: 1400.444
## Roots of the characteristic polynomial:
## 0.3441 0.3425
## Call:
## vars::VAR(y = varmat)
##
##
```

Impulse response

Consider the example bivaraite VAR(1)

$$
\begin{aligned}
y_{1t} &= \phi_{11} y_{1,t-1} + \phi_{12} y_{2,t-1} + u_{1t} \\
y_{2t} &= \phi_{21} y_{1,t-1} + \phi_{22} y_{2,t-1} + u_{2t}
\end{aligned}
$$

The linear correlation between u_{1t} and u u_{2t} can be characterized by

$$
u_{1t} = \left(\frac{\sigma_{12}}{\sigma_{22}} \right) u_{2t} + \eta_{1t}.
$$

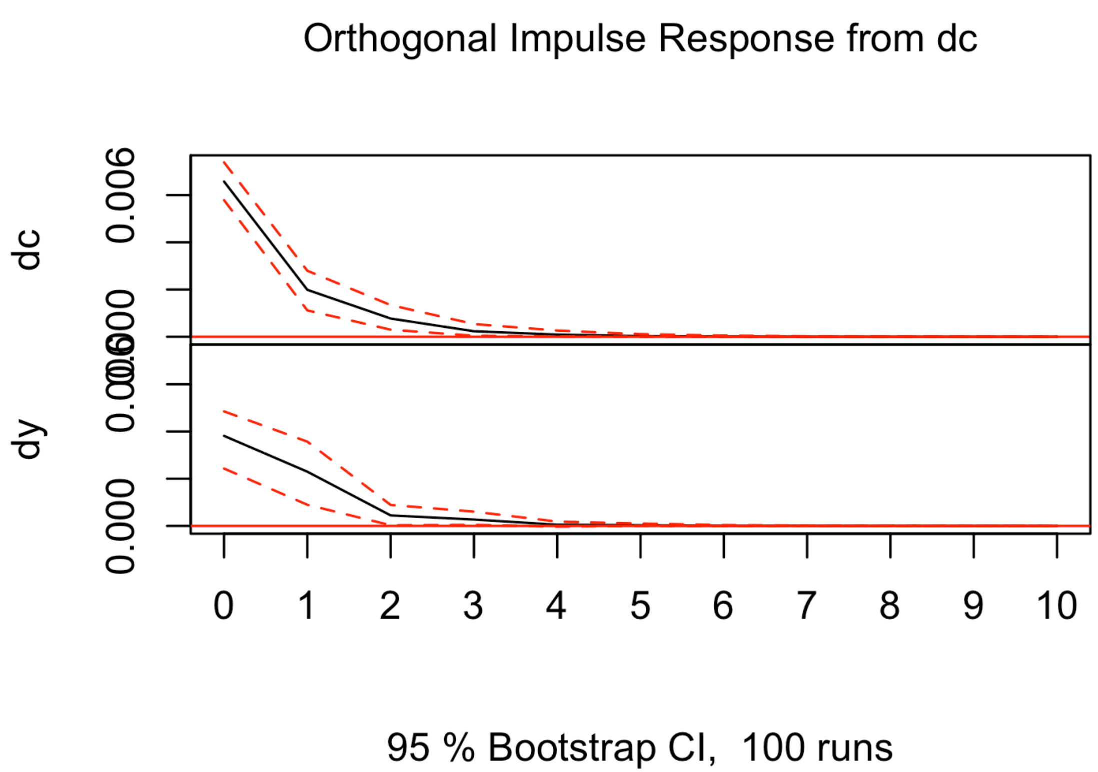
where η_{1t} has 0 correlation with u_{2t}.The structural VAR,

$$
\begin{aligned}
y_{1t} =& (\sigma_{12}/\sigma_{22})\, y_{2t} + \left(\phi_{11} - \frac{\sigma_{12}}{\sigma_{22}} \phi_{21} \right) y_{1,t-1} + \left(\phi_{12} - \frac{\sigma_{12}}{\sigma_{22}} \phi_{22} \right) y_{2,t-1} + \eta_{1t} \\
y_{2t} =& \phi_{21} y_{1,t-1} + \phi_{22} y_{2,t-1} + u_{2t}.
\end{aligned}
$$

The comtemporaneous effect of y_{2t} on y_{1t} does not imply causality (just a mechanical alternative representation), but it reflects the researcher's perspective.

A Shock u_{2t} (usually consider the magnitude of one standard deviation) affects both y_{1t} and

```r
impresp <- vars::irf(varfit, impulse = "dc")
plot(impresp)
```



Granger Causality

- In the VAR(p) system, does it make statistical significant difference if we completely shut down y_2 from the regression equation for y_1?

$$\mathbf{y}_{1t} = \mathbf{\Phi}_{11}\mathbf{y}_{1,t-1} + \mathbf{\Phi}_{12}\mathbf{y}_{2,t-1} + \mathbf{U}_{1t}$$
$$\mathbf{y}_{2t} = \mathbf{\Phi}_{21}\mathbf{y}_{1,t-1} + \mathbf{\Phi}_{22}\mathbf{y}_{2,t-1} + \mathbf{U}_{2t}$$

- The null hypothesis restricts $\mathbf{\Phi}_{12}$ as 0. Under the null, the Wald statistic asymptotically follows chi-sqaure distribution

- Caution: Granger causality is not "causal". Instead, it is merely a statistical predictive relationship

Granger Causality: Example

- Thurman W.N. & Fisher M.E. (1988), Chickens, Eggs, and Causality, or Which Came First?, *American Journal of Agricultural Economics*, 237-238
- 1930–1983 US chicken population and egg production

```r
data("ChickEgg", package="lmtest")
lmtest::grangertest(egg ~ chicken, order = 3, data = ChickEgg, test = "Chisq")
```

```
## Granger causality test
##
## Model 1: egg ~ Lags(egg, 1:3) + Lags(chicken, 1:3)
## Model 2: egg ~ Lags(egg, 1:3)
##   Res.Df Df  Chisq Pr(>Chisq)
## 1     44
## 2     47 -3 1.7748     0.6204
```

```r
lmtest::grangertest(chicken ~ egg, order = 3, data = ChickEgg, test = "Chisq")
```

```
## Granger causality test
##
## Model 1: chicken ~ Lags(chicken, 1:3) + Lags(egg, 1:3)
## Model 2: chicken ~ Lags(chicken, 1:3)
##   Res.Df Df  Chisq Pr(>Chisq)
```

Forecasting

Forecasting

- Forecasting target
    - Inflation
    - Financial markets
    - Cases of virus inflection
    - Airticket sales
    - Housing price
    - etc
- Types of forecasting
    - Ex ante forecasts: Use training data $\{y_1, \ldots, y_T\}$ to forecast (genuine) future values $y_{T+1}, y_{T+2}, \cdots$
    - Ex post forecast:
        - Use data $\{y_1, \ldots, y_{T-H}\}$ while hide $\{y_{T-H+1}, \ldots, y_T\}$
        - After obtaining forecasts $\{\hat{y}_{T-H+1}, \ldots, \hat{y}_T\}$, reveal $\{y_{T-H+1}, \ldots, y_T\}$ and

Forecasting

Clements and Hendry (1998) identify five sources of uncertainties for model-based forecasts:

- Mis-measurement of the data used for forecasting
- Misspecification of the model (or model uncertainty, including policy uncertainty)
- Future changes in the underlying structure of the economy
- The cumulation of future errors, or shocks, to the economy (or future uncertainty)
- Inaccuracies in the estimates of the parameters of a given model (or parameter uncertainty).

Forecasting with AR processes

- Consider an AR(1) model, $y_t = \phi_0 + \phi_1 y_{t-1} + u_t$, $u_t \sim WN(0, \sigma^2)$, $t = 1, 2, \cdots, T$.
- One-period-ahead: Extrapolate for one period to $T+1$, $y_{T+1} = \rho_0 + \rho_1 y_T + u_{T+1}$
- A natural forecast is $\hat{y}_{T+1} = \hat{\phi}_0 + \hat{\phi}_1 y_T$
- Multiple period ahead:

$$\hat{y}_{T+h} = \hat{\phi}_0 \left(\frac{1 - \hat{\phi}_1^h}{1 - \hat{\phi}_1} \right) + \hat{\phi}_1^h y_T$$

obtained by iteratively evaluate $\hat{y}_{T+h} = \hat{\phi}_0 + \hat{\phi}_1 \hat{y}_{T+h-1}$

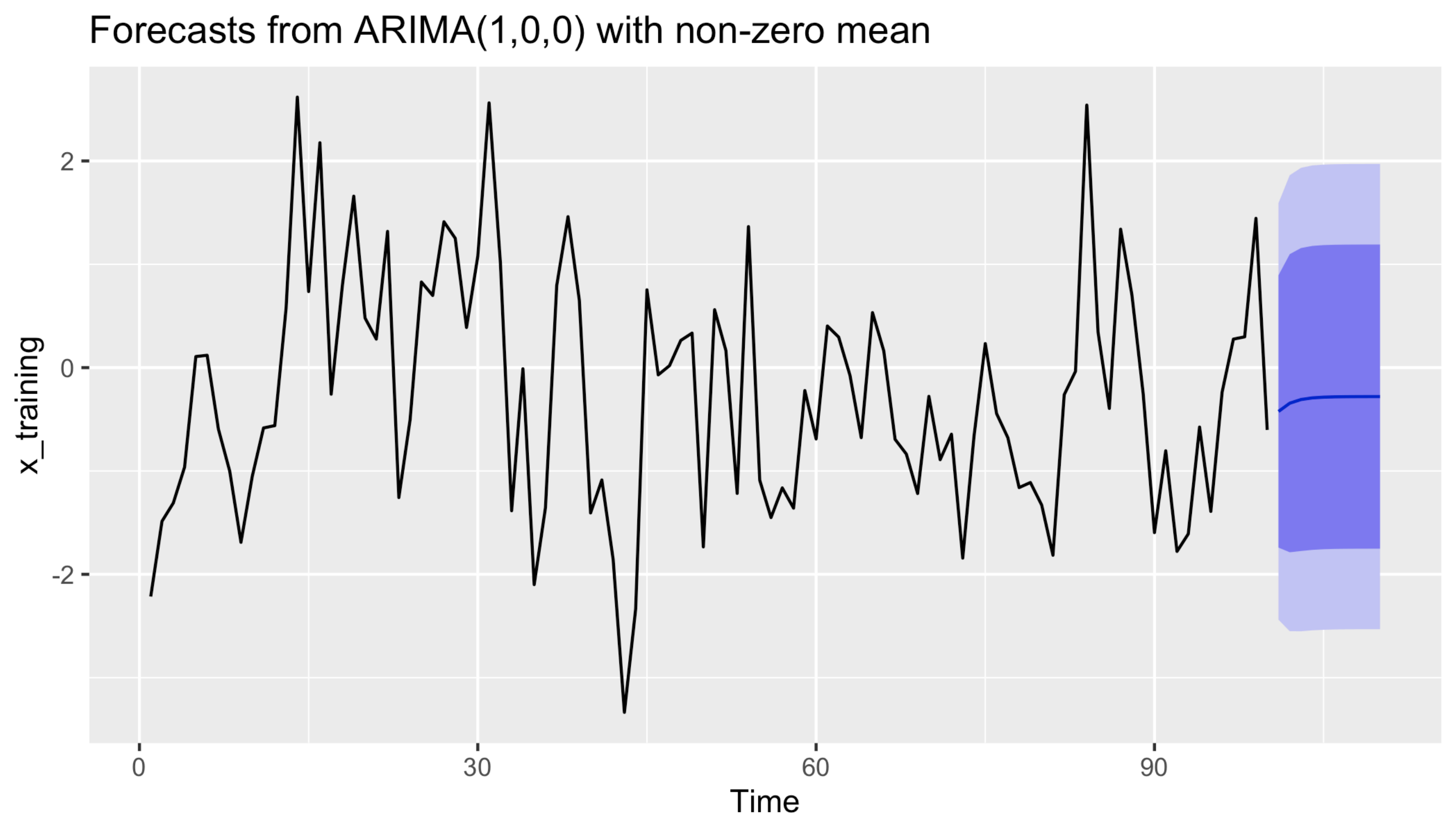
```r
library(forecast)
n = 100
H = 10
x ← arima.sim(model=list(ar = 0.5), n+H)
x_training ← x[1:n]
fit1 ← arima(x_training, order = c(1,0,0))
f1 ← forecast(fit1, h = H)
```

Forecasting with AR processes

```
##
## Forecast method: ARIMA(1,0,0) with non-zero mean
##
## Model Information:
##
## Call:
## arima(x = x_training, order = c(1, 0, 0))
##
## Coefficients:
##          ar1  intercept
##       0.4456    -0.2799
## s.e.  0.0903     0.1841
##
## sigma^2 estimated as 1.057:  log likelihood = -144.76,  aic = 295.52
##
## Error measures:
##                      ME     RMSE       MAE      MPE     MAPE      MASE
## Training set 0.01065144 1.027961 0.8110937 10.19605 188.8015 0.8402995
```

Forecasting with AR processes

```
autoplot(forecast(fit1))
```



Decision theory

- let $y^*_{t+1|t}$ be point forecast, and the error is $e_{t+1} = y_{t+1} - y^*_{t+1|t}$
- Need to choose a loss function $L\left(y_{t+1}, y^*_{t+1|t}\right)$
  - Commonly choice: squared loss function

$$L_q\left(y_{t+1}, y^*_{t+1|t}\right) = Ae^2_{t+1} = A\left(y_{t+1} - y^*_{t+1|t}\right)^2$$

- **Risk**: expected loss conditional on the information available at time t

$$E\left[L_q\left(y_{t+1}, y^*_{t+1|t}\right) \mid \Omega_t\right]$$

- **Optimal Forecast**:

$$\underset{y^*_{t+1|t}}{\operatorname{argmin}}\left\{E\left[L\left(y_{t+1}, y^*_{t+1|t}\right) \mid \Omega_t\right]\right\}$$

  - With squared loss, the optimal forecast is simply

$$y^*_{t+1|t} = E\left(y_{t+1} \mid \Omega_t\right),$$

Decision theory

- Other loss? An example can be an asymmetric loss function that is a simple version of the linear exponential (LINEX) function

$$L_a\left(y_{t+1}, y^*_{t+1|t}\right) = \frac{2\left[\exp(\alpha e_{t+1}) - \alpha e_{t+1} - 1\right]}{\alpha^2}.$$

  - The optimal forecast is

$$y^*_{t+1|t} = E\left(y_{t+1} \mid \Omega_t\right) + \frac{\alpha}{2}\mathrm{Var}(y_{t+1} \mid \Omega_t),$$

    when the conditional density is normal.

Metrics of forecast evaluation

- In the ex post forecast practice, we can use the following metrics to evaluate the performance
  - Mean absolute error (MAE)

$$\mathrm{MAE} = \frac{1}{\mathrm{H}} \sum_{\mathrm{h}=1}^{\mathrm{H}} |\mathrm{e}_{\mathrm{T}+\mathrm{h}}|$$

  - Mean sqaured error (MSE)

$$\mathrm{MSE} = \frac{1}{\mathrm{H}} \sum_{\mathrm{h}=1}^{\mathrm{H}} \mathrm{e}^2_{\mathrm{T}+\mathrm{h}}$$

  - etc

```
accuracy(f1)
```

```
##                      ME     RMSE       MAE      MPE     MAPE      MASE
## Training set 0.01065144 1.027961 0.8110937 10.19605 188.8015 0.8402995
```

Predictability Tests

To determine which model produces better forecasts, we may test the null hypothesis

$$H_0 : E\left[L\left(y_{t+h}, y^{*1}_{t+h|t}\right)\right] - E\left[L\left(y_{t+h}, y^{*2}_{t+h|t}\right)\right] = 0$$

against

$$H_1 : E\left[L\left(y_{t+h}, y^{*1}_{t+h|t}\right)\right] - E\left[L\left(y_{t+h}, y^{*2}_{t+h|t}\right)\right] \neq 0$$

Diebold and Mariano (1995) have proposed a test based on the loss-differential

$$d_t = L\left(y_{t+h}, y^{*1}_{t+h|t}\right) - L\left(y_{t+h}, y^{*2}_{t+h|t}\right).$$

The test statistic

$$DM = \frac{T^{1/2}\bar{d}}{(\widehat{\text{Var}}(\bar{d}\,))^{1/2}}$$

```
dm.test( e1 = e.ar12, e2 = e.ma12, alternative = "two.sided", h = 1, power = 1 )
```

Predictability Tests

- Giacomini and White (2006) (GW) have focused on a test for the null hypothesis of equal conditional predictive ability

$$H_0 : E\left[L\left(y_{t+h}, \hat{y}^{*1}_{t+h|t}\right) \mid \Omega_t\right] - E\left[L\left(y_{t+h}, \hat{y}^{*2}_{t+h|t}\right) \mid \Omega_t\right] = 0.$$

- More recent development: Li, J., Liao, Z., & Quaedvlieg, R. (2022). Conditional superior predictive ability. *The Review of Economic Studies*, 89(2), 843-875.

Forecast Combination

- Combine forecasts from models
- Combine opinions of individuals
- Examples in the USA
  - Surveys of Consumers (University of Michigan)
  - Livingston Survey (FRED of Philadelphia)
- Example in Europe
  - European Central Bank's surveys of professional forecasters
  - CPI, 1-year-ahead or 2-year-ahead
  - Data: 1999Q1–2018Q4 (20 years), about 120 forecasters
  - Unbalanced, 30 forecasters of complete record

Optimal forecast combination

- Bates and Granger (1969)
- Forecast error $\mathbf{e}_\mathrm{t} = (\mathrm{e}_{1\mathrm{t}}, \ldots, \mathrm{e}_{\mathrm{Nt}})'$ with $e_{it} = y_{t+1} - f_{it}$
- Sample variance-covariance $\widehat{\Sigma} := T^{-1} \sum_{\mathrm{t}=1}^{\mathrm{T}} \mathbf{e}_\mathrm{t} \mathbf{e}_\mathrm{t}'$
- The weights:

$$\min_{\mathbf{w} \in \mathbb{R}^N} \frac{1}{2} \mathbf{w}' \widehat{\boldsymbol{\Sigma}} \mathbf{w} \text{ subject to } \mathbf{w}' \mathbf{1}_\mathrm{N} = 1.$$

- When $\hat{\Sigma}$ is invertible,

$$\widehat{\boldsymbol{w}} = \frac{\widehat{\boldsymbol{\Sigma}}^{-1} \mathbf{1}_\mathrm{N}}{\mathbf{1}_\mathrm{N}' \widehat{\boldsymbol{\Sigma}}^{-1} \mathbf{1}_\mathrm{N}}$$

- R package `ForecastComb::comb_BG`
- High-dimensional case: Zhentao Shi, Liangjun Su and Tian Xie (2022): “L2-Relaxation: With Applications to Forecast Combination and Portfolio Analysis,” *Review of Economics and Statistics*

Optimal forecast combination

- Regression approach (Granger and Ramanathan, 1984)
- Run OLS regression

$$y_t = \sum_{i=1}^{N} w_i f_{it} + v_t$$

- `ForecastComb::comb_OLS`
- If the restriction $\sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{W}_{\mathrm{i}} = 1$ is imposes, the regression approach is equivalent to the restricted optimization approach.

Simple Average

- It is a myth that the simplest weight boasts robust performance in empirical examples and simulation exercises
- DeMiguel, V., L. Garlappi, and R. Uppal (2007). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? The Rview of Financial Studies 22(5), 1915–1953.
- Reasons:
  - Variance-bias trade-off
  - Parameter instability
  - Similar variances
- `ForecastComb::comb_SA`