

# ETC3550/ETC5550 Applied forecasting

**Week 11: Dynamic regression models**

**[af.numbat.space](http://af.numbat.space)**



# Regression with ARIMA errors

## Regression models

$$y_t = \beta_0 + \beta_1 X_{1,t} + \dots + \beta_k X_{k,t} + \varepsilon_t,$$

- $y_t$  modeled as function of  $k$  explanatory variables  $X_{1,t}, \dots, X_{k,t}$ .
- In regression, we assume that  $\varepsilon_t$  is WN.
- Now we want to allow  $\varepsilon_t$  to be autocorrelated.

# Regression with ARIMA errors

## Regression models

$$y_t = \beta_0 + \beta_1 X_{1,t} + \cdots + \beta_k X_{k,t} + \varepsilon_t,$$

- $y_t$  modeled as function of  $k$  explanatory variables  $X_{1,t}, \dots, X_{k,t}$ .
- In regression, we assume that  $\varepsilon_t$  is WN.
- Now we want to allow  $\varepsilon_t$  to be autocorrelated.

## Example: ARIMA(1,1,1) errors

$$y_t = \beta_0 + \beta_1 X_{1,t} + \cdots + \beta_k X_{k,t} + \eta_t,$$
$$(1 - \phi_1 B)(1 - B)\eta_t = (1 + \theta_1 B)\varepsilon_t,$$

# Estimation

If we minimize  $\sum \eta_t^2$  (by using ordinary regression):

- 1 Estimated coefficients  $\hat{\beta}_0, \dots, \hat{\beta}_k$  are no longer optimal as some information ignored;
- 2 Statistical tests associated with the model (e.g., t-tests on the coefficients) are incorrect.
- 3 AIC of fitted models misleading.

# Estimation

If we minimize  $\sum \eta_t^2$  (by using ordinary regression):

- 1 Estimated coefficients  $\hat{\beta}_0, \dots, \hat{\beta}_k$  are no longer optimal as some information ignored;
- 2 Statistical tests associated with the model (e.g., t-tests on the coefficients) are incorrect.
- 3 AIC of fitted models misleading.
  - Minimizing  $\sum \varepsilon_t^2$  avoids these problems.
  - Maximizing likelihood similar to minimizing  $\sum \varepsilon_t^2$ .

# Regression with ARIMA errors

Any regression with an ARIMA error can be rewritten as a regression with an ARMA error by differencing all variables.

# Regression with ARIMA errors

Any regression with an ARIMA error can be rewritten as a regression with an ARMA error by differencing all variables.

## Original data

$$y_t = \beta_0 + \beta_1 X_{1,t} + \cdots + \beta_k X_{k,t} + \eta_t$$

where  $\phi(B)(1 - B)^d \eta_t = \theta(B)\varepsilon_t$

# Regression with ARIMA errors

Any regression with an ARIMA error can be rewritten as a regression with an ARMA error by differencing all variables.

## Original data

$$y_t = \beta_0 + \beta_1 x_{1,t} + \cdots + \beta_k x_{k,t} + \eta_t$$

$$\text{where } \phi(B)(1-B)^d \eta_t = \theta(B)\varepsilon_t$$

## After differencing all variables

$$y'_t = \beta_1 x'_{1,t} + \cdots + \beta_k x'_{k,t} + \eta'_t$$

$$\text{where } \phi(B)\eta'_t = \theta(B)\varepsilon_t,$$

$$y'_t = (1-B)^d y_t, \quad x'_{i,t} = (1-B)^d x_{i,t}, \quad \text{and } \eta'_t = (1-B)^d \eta_t$$



## Regression with ARIMA errors

- In R, we can specify an  $ARIMA(p, d, q)$  for the errors, then  $d$  levels of differencing will be applied to all variables  $(y, x_{1,t}, \dots, x_{k,t})$  during estimation.
- Check that  $\varepsilon_t$  series looks like white noise.
- AICc can be calculated for final model.
- Repeat procedure for all subsets of predictors to be considered, and select model with lowest AICc value.

# Forecasting

- To forecast a regression model with ARIMA errors, we need to forecast the regression part of the model and the ARIMA part of the model and combine the results.
- Some predictors are known into the future (e.g., time, dummies).
- Separate forecasting models may be needed for other predictors.
- Forecast intervals ignore the uncertainty in forecasting the predictors.

# Your turn

- 1 Fit a regression model with a piecewise linear trend and Fourier terms for the US leisure employment data.

```
leisure <- us_employment |>
  filter(Title == "Leisure and Hospitality", year(Month) > 2001) |>
  mutate(Employed = Employed / 1000) |>
  select(Month, Employed)
```

- 2 Add a dynamic regression model with the same predictors.
- 3 How do the models compare on AICc?
- 4 Does the additional ARIMA component fix the residual autocorrelation problem in the regression model?
- 5 How different are the forecasts from each model?