workshop.nectric.com.au/tidyfc2024

Time Series Analysis & Forecasting Using R

9. Dynamic regression

Outline

1 Regression with ARIMA errors
2 Some useful predictors
3 Dynamic harmonic regression
4 Lab Session 16
5 Forecasting with regressors
6 Lab Session 17

Outline

1 Regression with ARIMA errors

2 Some useful predictors

3 Dynamic harmonic regression

4 Lab Session 16

5 Forecasting with regressors

6 Lab Session 17

Time series regression

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
- In regression, we assume that ε_t is white noise.

Time series regression

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
- In regression, we assume that ε_t is white noise.

Specify this model with `TSLM()`.

Much like `lm()`, regressors are specified on the formula's right.

Regression with ARIMA errors

RegARIMA models

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

- Residuals are from ARIMA model.
- Estimate model in one step using MLE
- Select model with lowest AICc value.

Simply add regression terms to the `ARIMA()` formula's right.

Outline

1 Regression with ARIMA errors
2 Some useful predictors
3 Dynamic harmonic regression
4 Lab Session 16
5 Forecasting with regressors
6 Lab Session 17

Some useful predictors

Linear trend

The time index is an effective predictor for trends.

It can be added to models as a regressor with `trend()`.

Some useful predictors

Linear trend

The time index is an effective predictor for trends.
It can be added to models as a regressor with `trend()`.

Piecewise linear trend

Trends often change over time.
We can add changepoints in the trend with
`trend(knots = <times>)`.

Some useful predictors

Dummy variables

Identify categories of observations with $\{0, 1\}$ indicators.
Useful for public holidays, special events & policy changes.

Some useful predictors

Dummy variables

Identify categories of observations with $\{0, 1\}$ indicators.
Useful for public holidays, special events & policy changes.

Dummy seasonality

Use dummy variables for each season.
It can be added to models as a regressor with `season()`.

Some useful predictors

Fourier seasonality

Fourier terms use sine and cosine harmonics to model seasonality.

It offers key advantages over dummy seasonality:

- Reduce model complexity
- Non-integer seasonality

Use `fourier(K = ???)` to add fourier seasonality with `K` harmonics to the model.

Outline

1 Regression with ARIMA errors

2 Some useful predictors

3 Dynamic harmonic regression

4 Lab Session 16

5 Forecasting with regressors

6 Lab Session 17

Dynamic harmonic regression

Capture seasonality with fourier terms instead of ARIMA PDQ().

Advantages

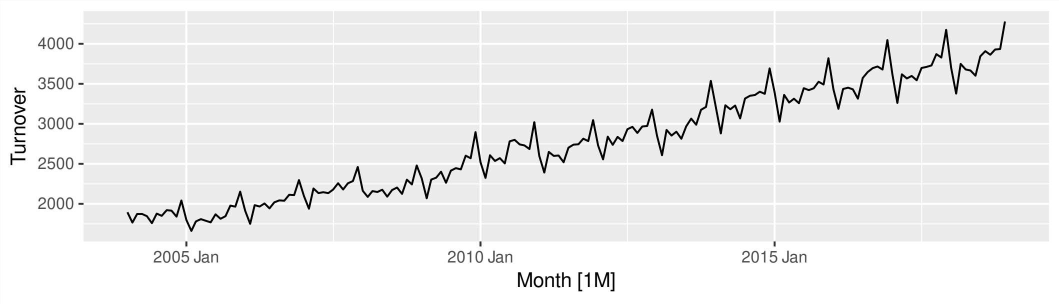
- all the benefits of fourier terms;
- supports multiple seasonality via multiple fourier terms;
- capture remaining dynamics with a simple ARMA model.

Disadvantages

- seasonality cannot change over time.

Eating-out expenditure

```r
aus_cafe <- aus_retail |>
  filter(
    Industry == "Cafes, restaurants and takeaway food services",
    year(Month) %in% 2004:2018
  ) |>
  summarise(Turnover = sum(Turnover))
aus_cafe |> autoplot(Turnover)
```

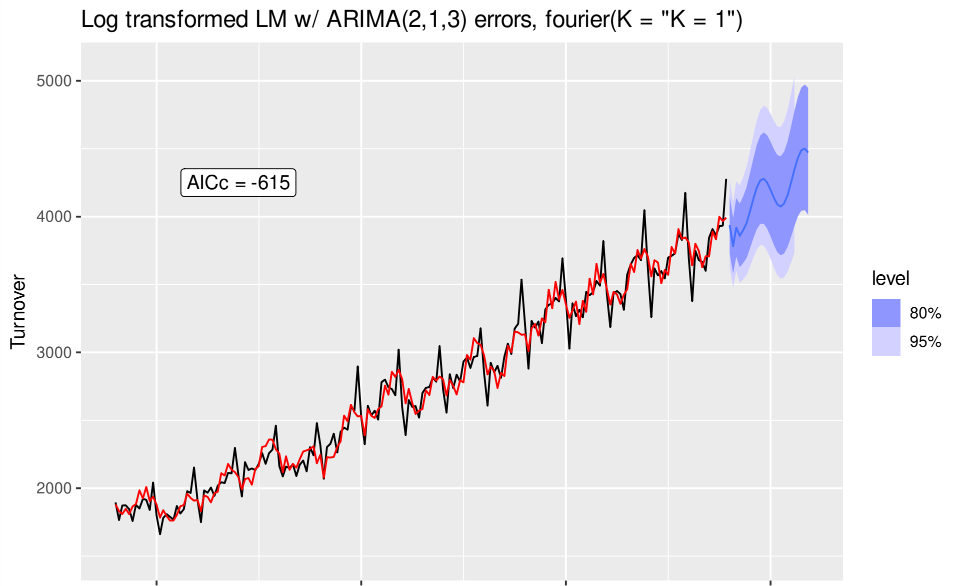


Eating-out expenditure

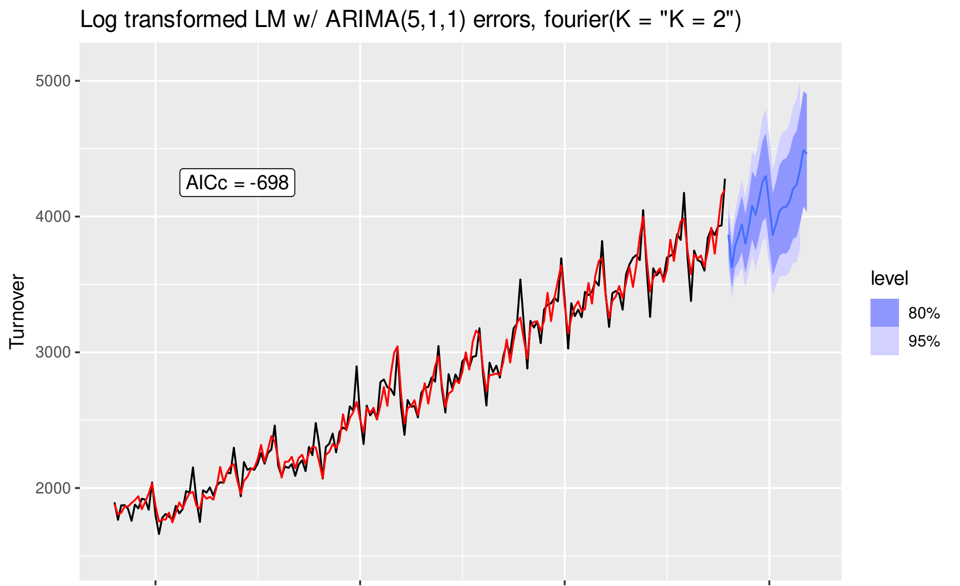
```
fit <- aus_cafe |> model(
  `K = 1` = ARIMA(log(Turnover) ~ fourier(K = 1) + PDQ(0, 0, 0)),
  `K = 2` = ARIMA(log(Turnover) ~ fourier(K = 2) + PDQ(0, 0, 0)),
  `K = 3` = ARIMA(log(Turnover) ~ fourier(K = 3) + PDQ(0, 0, 0)),
  `K = 4` = ARIMA(log(Turnover) ~ fourier(K = 4) + PDQ(0, 0, 0)),
  `K = 5` = ARIMA(log(Turnover) ~ fourier(K = 5) + PDQ(0, 0, 0)),
  `K = 6` = ARIMA(log(Turnover) ~ fourier(K = 6) + PDQ(0, 0, 0))
)
glance(fit)
```

.model	sigma2	log_lik	AIC	AICc	BIC
K = 1	0.002	317	-616	-615	-588
K = 2	0.001	362	-700	-698	-661
K = 3	0.001	394	-763	-761	-725
K = 4	0.001	427	-822	-818	-771
K = 5	0.000	474	-919	-917	-875
K = 6	0.000	474	-920	-918	-875

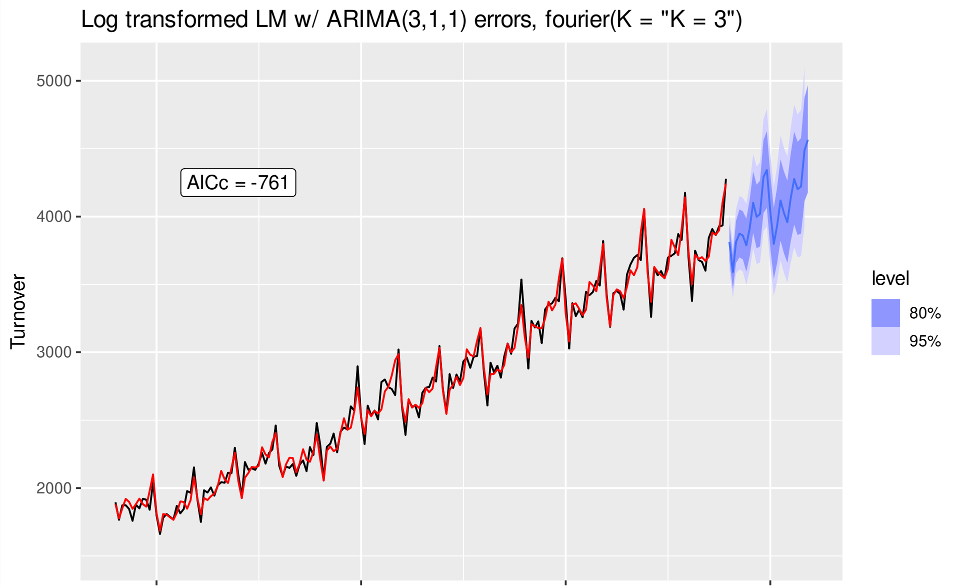
Eating-out expenditure



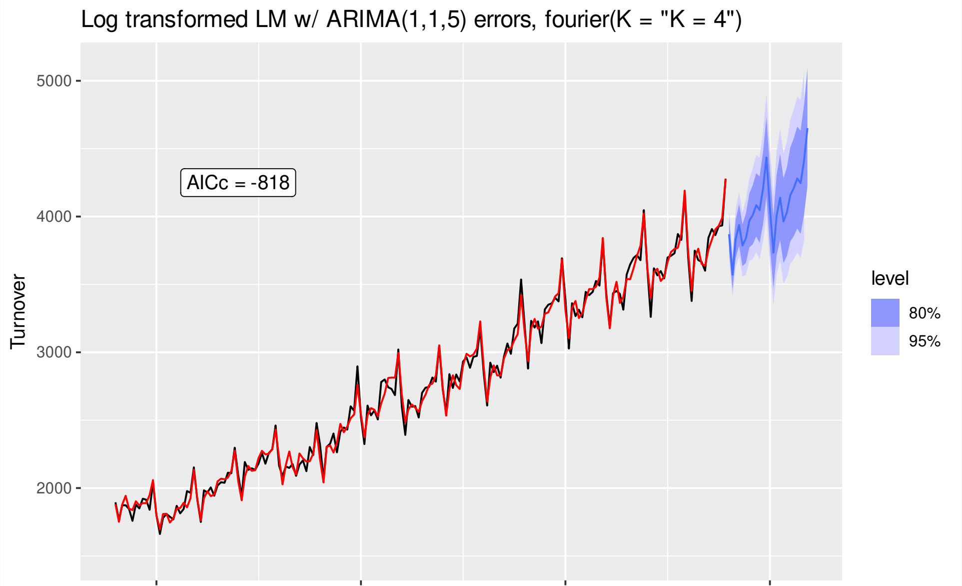
Eating-out expenditure



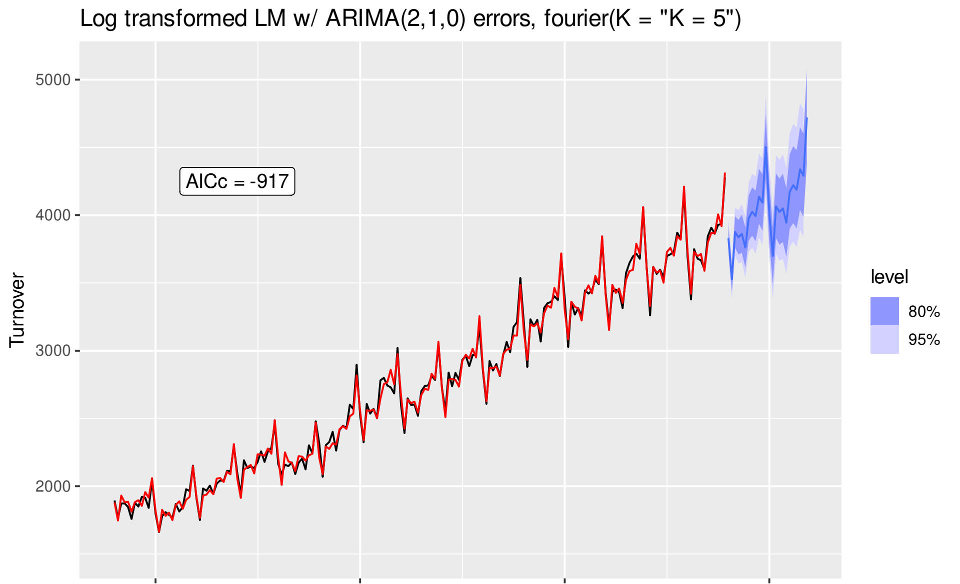
Eating-out expenditure



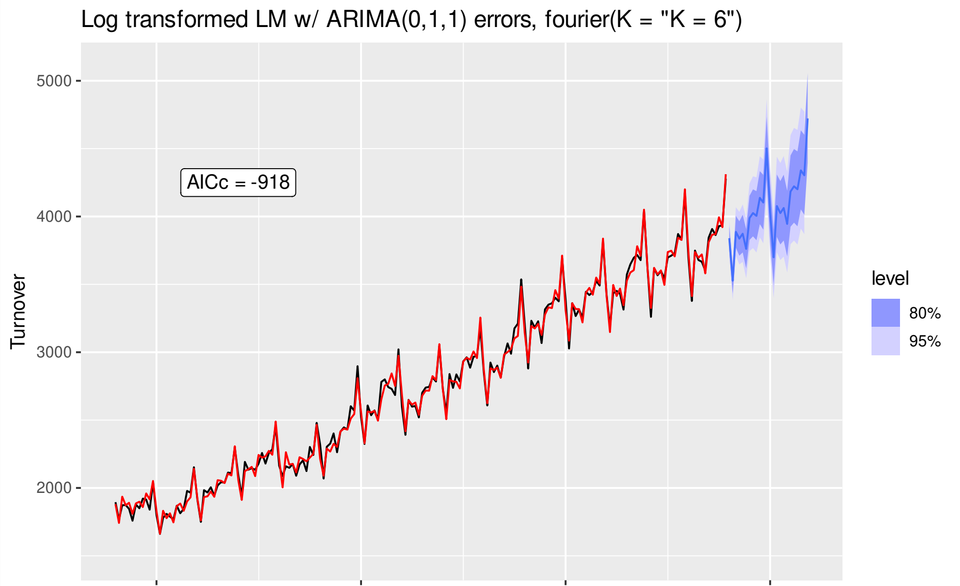
Eating-out expenditure



Eating-out expenditure



Eating-out expenditure



Outline

1 Regression with ARIMA errors

2 Some useful predictors

3 Dynamic harmonic regression

4 Lab Session 16

5 Forecasting with regressors

6 Lab Session 17

Lab Session 16

Produce forecasts of preschool and school education jobs from the ABS payroll data (`payroll_education`).

```
payroll_education |>
  filter(Industry == "Preschool and School Education")
```

1. Estimate a TSLM model with appropriate regression terms (trend, fourier harmonics, ...)
2. Produce and visualise forecasts from this model
3. Perform residual diagnostic checks on the model
4. Instead use dynamic harmonic regression
5. Do the residuals and forecasts look better?

Outline

1 Regression with ARIMA errors

2 Some useful predictors

3 Dynamic harmonic regression

4 Lab Session 16

5 Forecasting with regressors

6 Lab Session 17

Forecasting with regressors

> **Additional regressors**
>
> Using additional information from other variables is a great way to enhance your time series model.
> Add them the the formula just like `lm()`.

Additional regressors in forecasting models can make it harder to produce forecasts. Why?

Forecasting with regressors

! Future values

Future regressor values need to be given for forecasting.

Forecasting with regressors

> **!** Future values
>
> Future regressor values need to be given for forecasting.

Advantages

- Future values could be known in advance.
- Forecasts under different scenarios can be compared.

Disadvantages

- Unknown future values also need forecasting.
- Forecasts ignore the uncertainty in predictors.

Lagged predictors

Sometimes a change in x_t does not affect y_t instantaneously

- y_t = sales, x_t = advertising.
- y_t = stream flow, x_t = rainfall.
- y_t = size of herd, x_t = breeding stock.

Lagged predictors

Sometimes a change in x_t does not affect y_t instantaneously

- y_t = sales, x_t = advertising.
- y_t = stream flow, x_t = rainfall.
- y_t = size of herd, x_t = breeding stock.

- These are dynamic systems with input (x_t) and output (y_t).
- x_t is often a leading indicator.
- There can be multiple predictors.

Lagged predictors

The model include present and past values of predictor: $x_t, x_{t-1}, x_{t-2}, \ldots.$

$$y_t = a + \nu_0 x_t + \nu_1 x_{t-1} + \cdots + \nu_k x_{t-k} + \eta_t$$

where η_t is an ARIMA process.

Lagged predictors

The model include present and past values of predictor: $x_t, x_{t-1}, x_{t-2}, \dots .$

$$y_t = a + \nu_0 x_t + \nu_1 x_{t-1} + \dots + \nu_k x_{t-k} + \eta_t$$

where η_t is an ARIMA process.

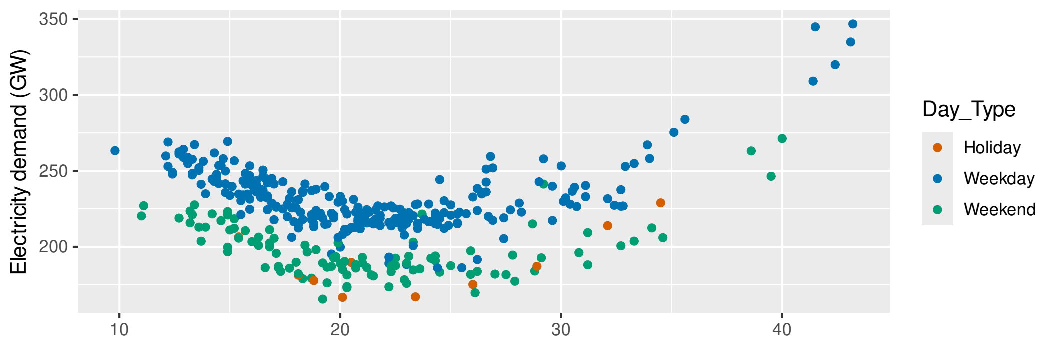
- x can influence y, but y is not allowed to influence x.

Use `lag()` on model regressors to lag them.

Daily electricity demand

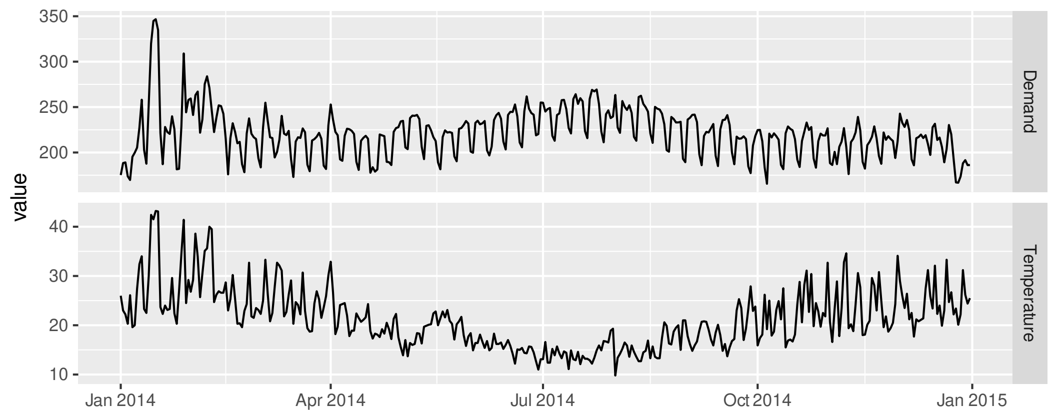
Model daily electricity demand as a function of temperature using quadratic regression with ARMA errors.

```
vic_elec_daily |>
  ggplot(aes(x = Temperature, y = Demand, colour = Day_Type)) +
  geom_point() +
  labs(x = "Maximum temperature", y = "Electricity demand (GW)")
```



Daily electricity demand

```
vic_elec_daily |>
  pivot_longer(c(Demand, Temperature)) |>
  ggplot(aes(x = Date, y = value)) +
  geom_line() +
  facet_grid(vars(name), scales = "free_y")
```



Daily electricity demand

```r
fit <- vic_elec_daily |>
  model(fit = ARIMA(Demand ~ Temperature + I(Temperature^2) +
    (Day_Type == "Weekday")))
report(fit)
```

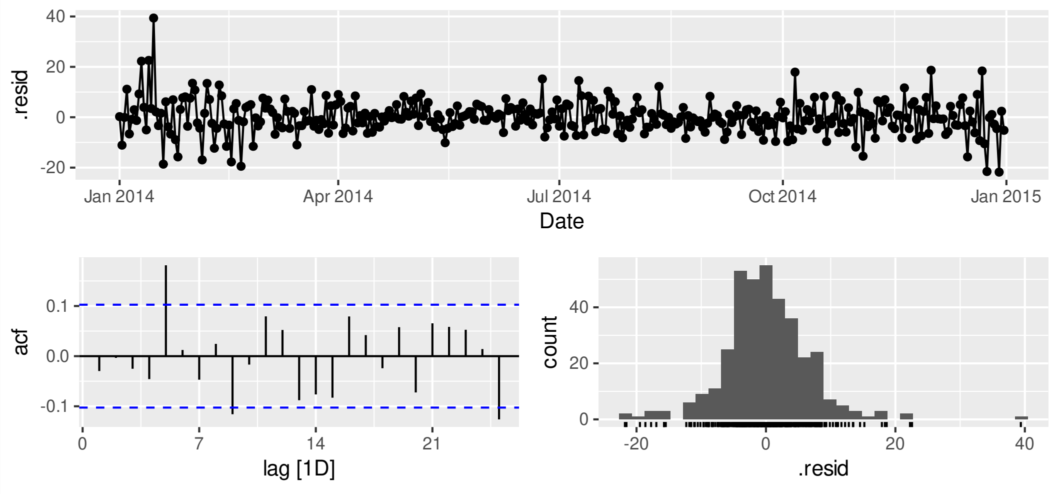
```
Series: Demand
Model: LM w/ ARIMA(2,1,2)(2,0,0)[7] errors

Coefficients:
          ar1     ar2      ma1      ma2    sar1   sar2  Temperature
      -0.1093  0.7226  -0.0182  -0.9381  0.1958  0.417       -7.614
s.e.   0.0779  0.0739   0.0494   0.0493  0.0525  0.057        0.448
      I(Temperature^2)  Day_Type == "Weekday"TRUE
                0.1810                      30.40
s.e.            0.0085                       1.33

sigma^2 estimated as 44.91:  log likelihood=-1206
AIC=2432   AICc=2433   BIC=2471
```

Daily electricity demand

```
augment(fit) |>
  gg_tsdisplay(.resid, plot_type = "histogram")
```



Daily electricity demand

```r
fit |>
  forecast(h = 14)
```

```
Error in `mutate()`:
i In argument: `fit = (function (object, ...) ...`.
Caused by error in `value[[3L]]()`:
! object 'Temperature' not found
  Unable to compute required variables from provided `new_data`.
  Does your model require extra variables to produce forecasts?
```

! More information needed

Our model depends on `Temperature` and `Day_Type`.
To produce forecasts, we need to also provide future values for these variables.

Daily electricity demand

```r
# Forecast one day ahead.
vic_next_day <- new_data(vic_elec_daily, 1) |>
  mutate(Temperature = 26, Day_Type = "Holiday")
forecast(fit, vic_next_day)
```

```
# A fable: 1 x 6 [1D]
# Key:     .model [1]
  .model Date            Demand .mean
  <chr>  <date>          <dist> <dbl>
1 fit    2015-01-01 N(161, 45)  161.
# i 2 more variables: Temperature <dbl>, Day_Type <chr>
```

Daily electricity demand

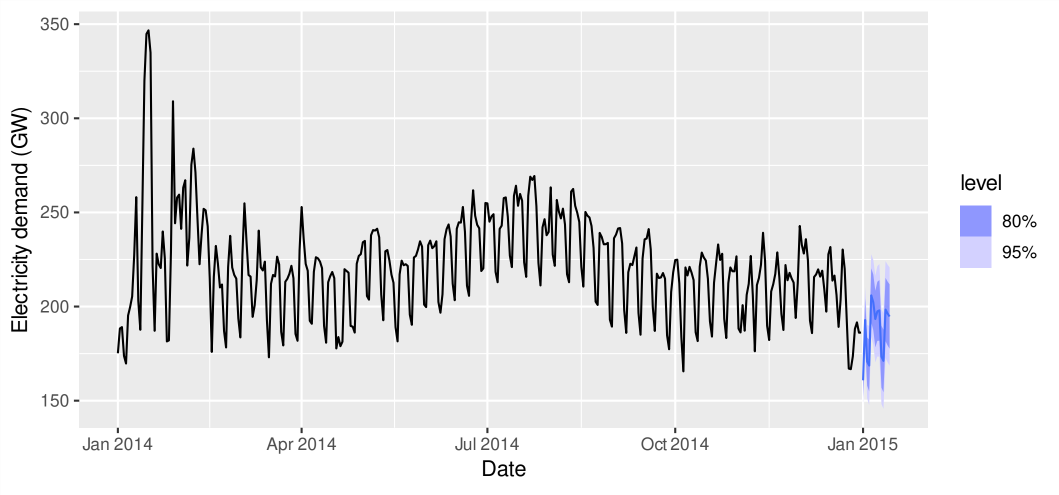
```r
# Forecast two weeks ahead.
vic_elec_future <- new_data(vic_elec_daily, 14) |>
  mutate(
    Temperature = 26,
    Holiday = c(TRUE, rep(FALSE, 13)),
    Day_Type = case_when(
      Holiday ~ "Holiday",
      wday(Date) %in% 2:6 ~ "Weekday",
      TRUE ~ "Weekend"
    )
  )
```

Scenario forecasting

Instead of forecasting most-likely values for regressors, it can be worth-while forecasting worst-case scenarios to adequately prepare.

Daily electricity demand

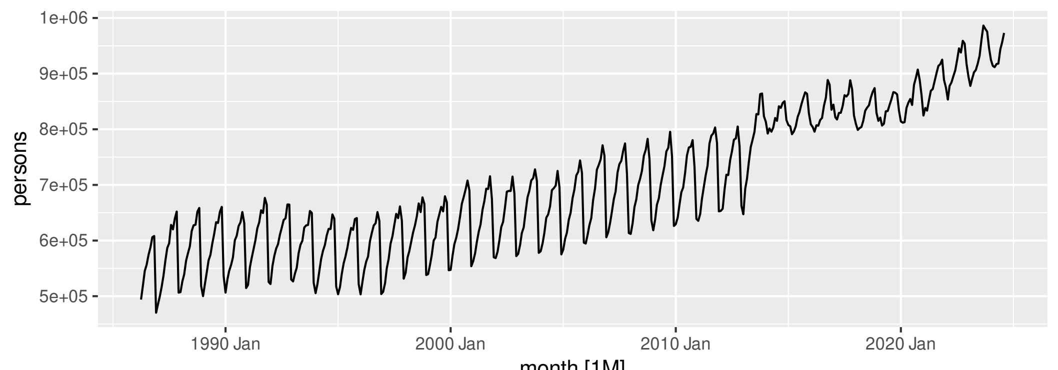
```r
forecast(fit, vic_elec_future) |>
  autoplot(vic_elec_daily) + labs(y = "Electricity demand (GW)")
```



Scenarios with policy decisions

Consider the total school students aged 15-19 in Australia.

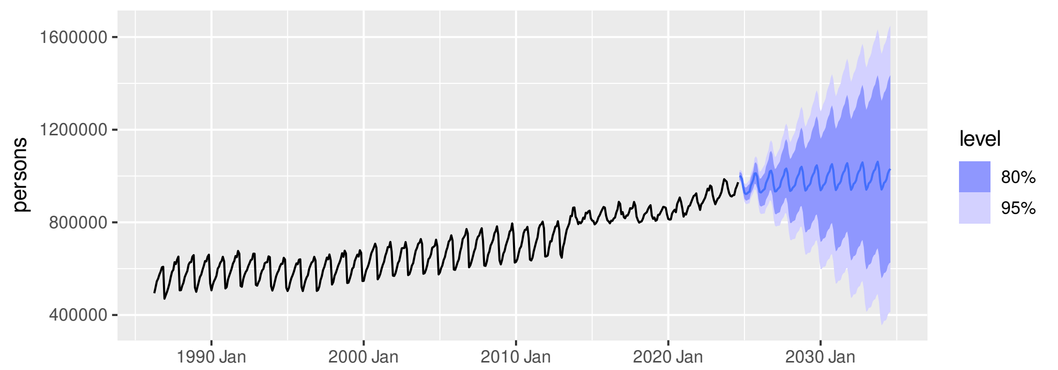
```
working_age_school_students <- student_labour |>
  filter(attendance == "Attending school (aged 15-19 years)") |>
  summarise(persons = sum(persons))
working_age_school_students |> autoplot(persons)
```



Scenarios with policy decisions

Without capturing the policy change, the forecasts are biased.

```
working_age_school_students |>
  model(ARIMA(persons ~ fourier(K = 3))) |>
  forecast(h = "10 years") |>
  autoplot(working_age_school_students)
```



Scenarios with policy decisions

Add dummy variable for the change in 2013 that interacts with the seasonality.

```
fit_policy <- working_age_school_students |>
  mutate(new_policy = month >= yearmonth("2013 Jan")) |>
  model(ARIMA(persons ~ new_policy*fourier(K = 3)))
report(fit_policy)
```

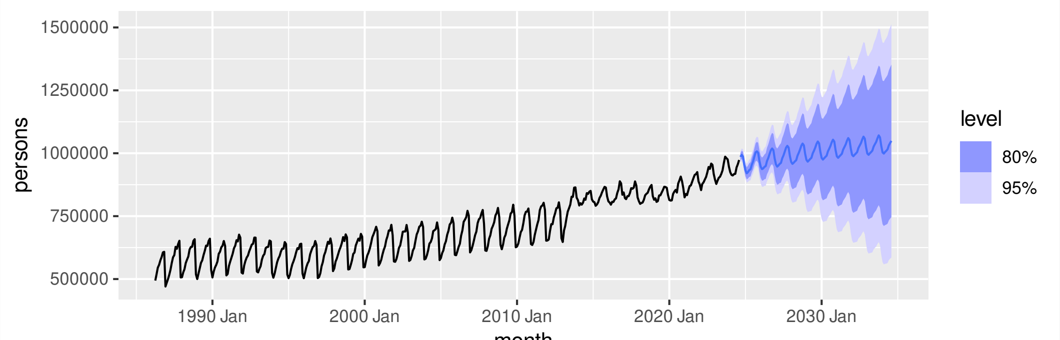
```
Series: persons
Model: LM w/ ARIMA(0,1,1)(2,0,0)[12] errors

Coefficients:
          ma1    sar1    sar2  new_policyTRUE  fourier(K = 3)C1_12
      -0.3537  0.4544  0.3826          -38178               -47610
s.e.   0.0463  0.0436  0.0445           10015                 5302
      fourier(K = 3)S1_12  fourier(K = 3)C2_12  fourier(K = 3)S2_12
                   -41123               -24555                 2573
s.e.                 5346                 3368                 3371
```

Scenarios with policy decisions

The policy is expected to continue into the future.

```
future_policy <- new_data(working_age_school_students, 120) |>
  mutate(new_policy = TRUE)
fit_policy |>
  forecast(new_data(working_age_school_students, 120) |> mutate(new_policy = TRUE))
  autoplot(working_age_school_students)
```



Outline

1 Regression with ARIMA errors
2 Some useful predictors
3 Dynamic harmonic regression
4 Lab Session 16
5 Forecasting with regressors
6 Lab Session 17

Lab Session 17

What if the 2013 policy was later reverted, what would you expect the forecasts to be?

1. Produce forecasts from a scenario in which this policy was reverted in 2030.

   Hint: Use `new_data()` to create the future time points, and `mutate()` a date comparison to create the future dummy variable values.

2. Visualise the forecasts, are they realistic?

Forecasting unkown scenarios

ABC NEWS

Search...

Log In

Just In | Watch Live | Politics | World | Business | Analysis | Sport | Science | More

International student commencements to be capped at 270,000 next year

By political reporter Maani Truu

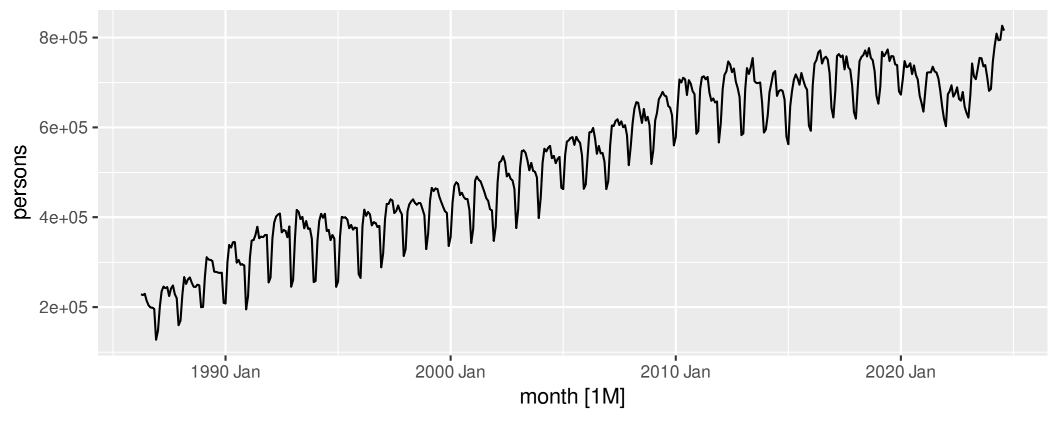
Posted Tue 27 Aug 2024 at 12:05pm, updated Tue 27 Aug 2024 at 4:18pm

abc.net.au/news/international-student-ca... Copy link Share article

- **In short:** New international student commencements will be capped at 270,000 across Australian higher education and vocational providers for the 2025 calendar year.
- Education Minister Jason Clare said the changes — if passed by parliament — would make the international student sector "better and fairer".
- **What's next?** The federal government will consult with universities about individual caps.

Forecasting unkown scenarios

```
student_labour |>
  filter(attendance == "Attending tertiary educational institution full-time") |>
  summarise(persons = sum(persons)) |>
  autoplot(persons)
```



Forecasting unkown scenarios

How would we forecast something without history?

- Judgemental forecasting with expert opinions
- Incorporate limits into the model
  (will these limits be met by demand?)
- Forecast more disaggregated data separately