Assignment 2

Assignment 2

Task A. Plotting monthly averages of temparature, volume, price and wind production.

```r
# Create a data frame of monthly averages of the variables volume, price,
# temperature and wind_production in long format.
df_avg_long <-
  df %>%

    # Select the specified variables
    select(date, volume, temperature, price, wind_production) %>%

    # Calculate monthly averages for each variable
    mutate(year_month = floor_date(date, "month")) %>%
    group_by(year_month) %>%
    summarise(
      avg_vol = mean(volume, na.rm = TRUE),
      avg_temp = mean(temperature, na.rm = TRUE),
      avg_price = mean(price, na.rm = TRUE),
      avg_wind = mean(wind_production, na.rm = TRUE)
    ) %>%

    # Scale the values of volume and wind to make the plot readable
    mutate(
      avg_vol = avg_vol / 10^5,
      avg_wind = avg_wind / 10^3
    ) %>%

    # Pivot to long format
```

```
    pivot_longer(
      cols = -year_month,
      names_to = "variable",
      values_to = "scaled values"
    )

# Plot the specified variables
df_avg_long %>%
  ggplot(
    aes(
      x = year_month,
      y = `scaled values`,
      col = variable
    )
  ) +
  geom_line() +
  labs(
    x = "",
    y = ""
  ) +
  theme_classic()
```
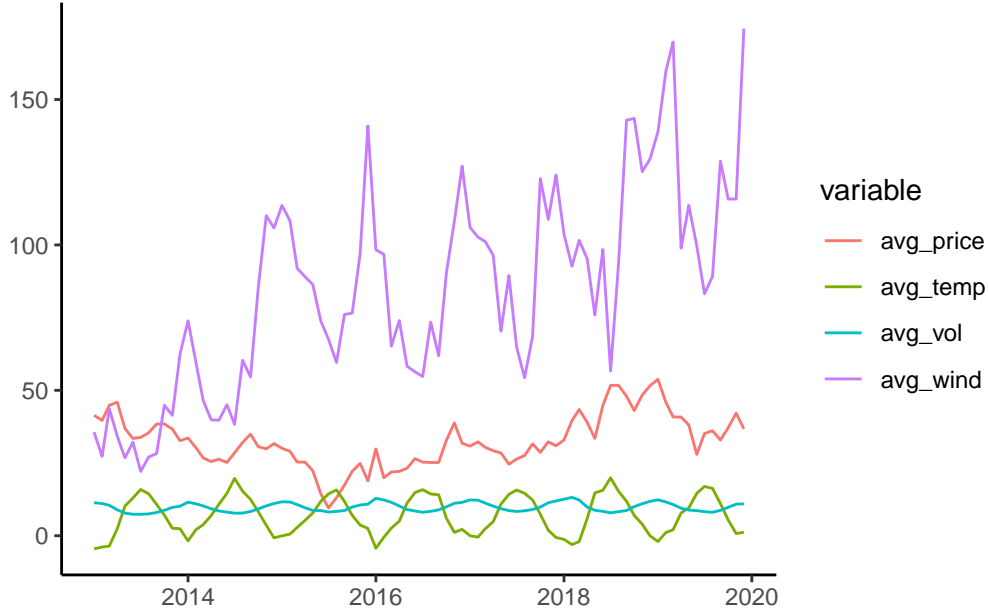
Figure 1: Monthly averages of temperature, volume, price and wind power production. The values have been scaled to fit into a single plot.

To discuss the relationships between temperature, volume, price and wind power production we chose The variables in the data set have vastly different scales. To surpass this issue when comparing the specified variables and in discussing their relationship, we chose to scale the values of the monthly averages of volume and wind power production by dividing their values of factors of 10^5 and 10^4 respectively.

We notice large seasonal variations in all variables. Especially in average temperatures and volume there is little variation between each year. The monthly average price had a dip in the summer of 2015, but increased steadily up until the Covid19-pandemic in the spring of 2019. The wind power production have strong seasonal variations, but we notice that the average wind production has increased over the period by a large margin.

When analyzing the relationship between the variables, we noticed what is likely a negative correlation between temperature and volume. This seems likely as the temperature decreases the demand for energy increases. Also we notice the negative correlation between wind production and temperature. This is likely due to the fact that the colder periods like winter and autumn is more windy. Lastly, we notice the positive correlation between volume and price. This is likely because in periods of high demand, the supply of energy is strained resulting in a higher price for electric energy.

There is likely more relationships between the variables that can be identified and discussed

using the plot we have provided. We chose to only comment on the most apparent of the relationships at this point.

Task B. Why will a OLS regression on quantity not provide an estimate of demand?

An OLS regression of quantity on price alone, even with controls, is not fitted to estimate demand due to many reasons:

1. Endogeneity exists as price and quantity in markets are simultaneously determined

2. Omitted variables like for example consumer income, advertising and seasonality. These can affect and lead to omitted variable bias.

3. Measurement errors in price and quantity can contribute to bias for the estimates for the parameters

4. OLS (or Gauss-Markow theorem) assumes a linear relationship. This may not hold if the demand curves are nonlinear.

5. Heteroskedasticity which is variability in the error term given any explanatory variable. In this case the variability in quantity varies across price levels which in return can affect standard erros.

6. The demand's dynamic behavior and seasonality may not be sufficiently captured. To estimate demand accurately, more advanced econometric methods are often required. These advanced methods usually contain instrumental variable regression and structural equation modelling. Controlling for relevant factors and improve our data quality can also make the estimation precision better.

Task C. Considering choice of instrument for price when estimating demand

When choosing a instrument for price when estimating demand, it's important to consider the validity of the instrument in terms of the requirements for a valid instrument. These requirements can be relevance, exogeneity, and exclusion restrictions:

1. **Why is temperature not a valid instrument for price when estimating demand?** The key requirements for exogeneity may not be satisfied if one were to use temperature as an instrument for price when estimating demand. (Exogeneity refers to the condition in which an independent variable is unrelated to the error term, indicating that it is not influenced by unobserved factors in the model.) Temperature is usually affected by factors that also may affect demand, like seasonality and consumer behavior. The instrument would not be valid due to endogeneity concerns if temperature is correlated with unobservable demand shocks.

2. **Why can magazine levels (or deviations) and wind power production potentially be good instruments for price when estimating demand?** Magazine levels and wind power production could potentially be sufficient instruments for price as they can fulfil the requirements for a valid instrument. They should be relevant in other terms they should be affecting the price. In addition, they should be exogenous, meaning that they should not be correlated with unobservable factors. But establishing their exogeneity requires thoughtful consideration and testing.

3. **Why is it necessary to control for seasonality (say, calendar month) and temperature?** It is necessary to control for seasonality and temperature to avoid omitted variable bias. (Omitted variable bias appear if a relevant variable that affects the dependent variable is excluded in a regression model. This leads to biased and estimates of the coefficients of the variables that are included.) Seasonality captures variation in the systematic demand throughout the year unrelated to price. Meanwhile temperature affects demand independently. When we control for these factors, we ensure precise demand estimates.

4. **How could controlling for weekday and year be useful?** Incorporating controls for weekday and year is beneficial to justify for demand patterns variation. Weekdays and weekends usually give different demand profiles given the consumers consumption. Meanwhile years would capture trends that are long-term and macroeconomic affects on demand. These controls strengthen the accuracy of demand estimates, while taking into account sources of variation unrelated to price.

D. Performing OLS regressions

```
# Create new variables in data frame containing the dummies for
# year, month and day in which the observation is made.
df <-
  df %>%
  mutate(
    year = as.factor(year(date)),
    month = as.factor(month(date)),
    weekday = as.factor(wday(date,
                    label = FALSE,
                    week_start = 1))
  )


# Define formulas to apply in the regressions
formulas <-
  list(
```

```r
    # Structural form: Demand
    volume ~
      price + temperature + as.factor(year) + as.factor(month) + as.factor(weekday),

    # Reduced form: Demand
    #volume ~
      #price

    # First stage: Demand

    # Structural form: Price
    price ~
      magazine_level + temperature + wind_production + year + month + weekday

    # Reduced form: Price
    #volume ~
      #price

    # First stage: Price

  )

# Perform regressions for all the specified formulas

# Empty reg list
reg_list <- list()

# Loop over all formulas
for (i in 1:length(formulas)) {

  # Perform regression with specified formula
  reg <- lm(formulas[[i]], df)

  # Save the regression
  reg_list[[i]] <- reg
}

# Create regression tables
stargazer(reg_list,
          type = "text",
```

```
        dep.var.labels = c("volume"),

        # Omit all the dummies to make regression table more readable
        keep = c("price", "temperature", "wind_production", "magazine_level", "Constant"
```

```
=============================================================================
                                 Dependent variable:
                    -----------------------------------------------------
                                      volume
                            (1)                           (2)
-----------------------------------------------------------------------------
price                   1,173.488***
                         (146.007)


magazine_level                                       -0.380***
                                                      (0.022)


temperature            -12,341.090***                -0.321***
                         (291.605)                    (0.036)


wind_production                                      -0.00003***
                                                      (0.00000)


Constant               1,064,506.000***              63.158***
                        (7,300.236)                   (1.287)


-----------------------------------------------------------------------------
Observations               2,556                       2,556
R2                         0.936                       0.717
Adjusted R2                0.935                       0.714
Residual Std. Error  42,387.110 (df = 2530)      5.285 (df = 2529)
F Statistic        1,468.102*** (df = 25; 2530) 246.224*** (df = 26; 2529)
=============================================================================
Note:                                       *p<0.1; **p<0.05; ***p<0.01
```

```
  # First stage
  first_stage <- lm(
    price ~ magazine_level + temperature + wind_production + year + month + weekday,
    data = df
  )
```

```r
# Reduced form
reduced_form <- lm(
  volume ~ magazine_level + temperature + wind_production + year + month + weekday,
  data = df
)

# IV/2SLS
iv_model <- ivreg(
  volume ~ price + temperature + year + month + weekday
  | magazine_level + temperature + wind_production + year + month + weekday,
  data = df
)

# OLS
ols_model <- lm(
  volume ~ price + temperature + year + month + weekday,
  data = df
)

# Regression table for first stage, reduced form and OLS
stargazer(
  title = "First stage, reduced form and OLS",
  first_stage,
  reduced_form,
  ols_model,
  type = "text",
  omit = c("year", "month", "weekday")
)
```

First stage, reduced form and OLS
===
                                         Dependent variable:
                   ----
                            price   volume
                             (1)   (2)   (3)
----
magazine_level   -0.380***   1,281.908***
                           (0.022)   (162.065)

price   1,173.488

temperature	-0.321***	-12,755.150***	-12,341.0
	(0.036)	(269.782)	(291.60
wind_production	-0.00003***	-0.356***	
	(0.00000)	(0.019)	
Constant	63.158***	1,062,206.000***	1,064,506.
	(1.287)	(9,653.483)	(7,300.2

Observations	2,556	2,556	2,556
R2	0.717	0.944	0.936
Adjusted R2	0.714	0.943	0.935
Residual Std. Error	5.285 (df = 2529)	39,647.870 (df = 2529)	42,387.110 (d
F Statistic	246.224*** (df = 26; 2529)	1,627.381*** (df = 26; 2529)	1,468.102*** (df

==

Note: *p<0.1; **p<0.0

```r
# Regression table for IV/2SLS
stargazer(
  title = "IV/2SLS",
  iv_model,
  type = "text",
  omit = c("year", "month", "weekday")
)
```

IV/2SLS
===

Dependent variable:
----
volume
----

price	1,831.056***
	(364.464)
temperature	-12,036.220***
	(331.135)

```
Constant                    1,037,639.000***
                            (15,478.790)


-----------------------------------------------
Observations                  2,556
R2                            0.935
Adjusted R2                   0.934
Residual Std. Error   42,556.680 (df = 2530)
===============================================
```

```r
# Test
stargazer(
  first_stage,
  reduced_form,
  iv_model,
  ols_model,
  type = "text",
  omit = c("year", "month", "weekday"),
  font.size = "footnotesize"
  #omit.table.layout = "sn"
)
```

		Dependent variable:	
	price		volume
	OLS	OLS	instrumental variable
	(1)	(2)	(3)
magazine_level	-0.380***	1,281.908***	
	(0.022)	(162.065)	
price			1,831.056**
			(364.464)
temperature	-0.321***	-12,755.150***	-12,036.220*
	(0.036)	(269.782)	(331.135)

wind_production	-0.00003***	-0.356***	
	(0.00000)	(0.019)	
Constant	63.158***	1,062,206.000***	1,037,639.000*
	(1.287)	(9,653.483)	(15,478.790)

```
--------------------------------------------------------------------------------
Observations              2,556                  2,556                  2,556
R2                        0.717                  0.944                  0.935
Adjusted R2               0.714                  0.943                  0.934
Residual Std. Error    5.285 (df = 2529)    39,647.870 (df = 2529)  42,556.680 (df =
F Statistic      246.224*** (df = 26; 2529) 1,627.381*** (df = 26; 2529)
================================================================================
```

Note: