Charles University in Prague
Faculty of Social Sciences
Institute of Economic Studies

Home assignment 1

Marek Chadim, Jakub Strašlipka

Econometrics II – JEB110
Winter semester 2022/2023

Problem 1.1

First, let us note that this problem is a special case of the Frisch-Waugh-Lovell theorem. Our proof losely follows paper of Lovell (2008).

Proof. To simplify the notation as well as the computation, let us express the intercept term as a variable $\mathbf{X_1} = (x_{11}, x_{21}, \dots, x_{k1})^\top = (1, 1, \dots, 1)^\top$ and consider the following regression containing k explanatory variables plus the trend variable

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + \delta t + u_t. \tag{1}$$

Now, define the detrended regression as

$$\ddot{y}_t = \ddot{\beta}_1 \ddot{x}_{t1} + \ddot{\beta}_2 \ddot{x}_{t2} + \dots + \ddot{\beta}_k \ddot{x}_{tk} + e_t, \tag{2}$$

where $\ddot{y}_t$ is a residual from regression

$$y_t = \alpha_1 t + \ddot{y}_t, \tag{3}$$

and $\ddot{x}_{ti}$ for $i = 1, \dots, k$ are residuals from regressions

$$x_{ti} = \alpha_{1i} t + \ddot{x}_{ti}. \tag{4}$$

By plugging expressions (3) and (4) into (1), we obtain

$$\begin{aligned}
\alpha_1 t + \ddot{y}_t &= \beta_1(\alpha_{11} t + \ddot{x}_{t1}) + \beta_2(\alpha_{12} t + \ddot{x}_{t2}) + \dots + \beta_k(\alpha_{1k} t + \ddot{x}_{tk}) + \delta t + u_t, \\
\ddot{y}_t &= \left(\sum_{i=1}^{k} \beta_i \ddot{x}_{ti}\right) + \left(\delta - \alpha_1 t + \sum_{i=1}^{k} \beta_i \alpha_{ti}\right) t + u_t.
\end{aligned}$$

Now, we realize that the residuals $\ddot{y}_t, \ddot{x}_{t1}, \dots, \ddot{x}_{tk}$ are uncorrelated with the explanatory trend variable t (as they are from regressions where t was an explanatory variable). Using the previous and the fact that explanatory variables that are uncorrelated with both the dependent and other explanatory variables have zero coefficients, we set $\left(\delta - \alpha_1 t + \sum_{i=1}^{k} \beta_i \alpha_{ti}\right) = 0$ and therefore get

$$\ddot{y}_t = \left(\sum_{i=1}^{k} \beta_i \ddot{x}_{ti}\right) + u_t.$$

This, when compared with (2), instantly implies that $\beta_i = \ddot{\beta}_i$ for $i = 1, \dots, k$, in plain words that the coefficients from the detrended regression (2) are identical to the coefficients of the regression (1) containing the trend variable.

□

Problem 1.2

(a) In order to be covariance (weakly) stationary, it must hold for the stochastic process $\{y_t; t \in \mathbb{N}\}$ that $\forall t \in \mathbb{N} : \mathbb{E}(y_t) = \mu$, where μ is a constant. Thus $\mathbb{E}(y_t) = \mathbb{E}(y_{t-1}) = \mathbb{E}(y_{t-2})$.

Hence,

$$\begin{aligned}
\mathbb{E}(y_t) &= \mathbb{E}\left(\alpha + \gamma_1 y_{t-1} + \gamma_2 y_{t-2} + \theta_1 \epsilon_{t-1} + \epsilon_t\right) \\
&= \alpha + \gamma_1 \mathbb{E}(y_{t-1}) + \gamma_2 \mathbb{E}(y_{t-2}) + \theta_1 \mathbb{E}(\epsilon_{t-1}) + \mathbb{E}(\epsilon_t) \\
&= \alpha + \gamma_1 \mathbb{E}(y_t) + \gamma_2 \mathbb{E}(y_t),
\end{aligned}$$

where we used expectation calculus and the i.i.d. distribution of $\{\epsilon_t\}$ with zero mean. After simple algebraic manipulation, we get

$$\mathbb{E}(y_t) = \frac{\alpha}{1 - \gamma_1 - \gamma_2}.$$

The unbiasedness assumptions are, in general, the assumption that the stochastic process actually follows the linear model as written in the task, the strict exogeneity assumption, and the no perfect colinearity between explanatory variables assumption. The ARMA model, however, fails the strict exogeneity assumption as any model containing a lagged dependent variable does not satisfy the strict exogeneity (see Woolridge (2015, p. 387)). Thus, we cannot guarantee unbiasedness of the gamma coefficients under our developed theory.

(b.i) These estimates are for the AR(2) part of the ARMA(2,1) model. We can expect the true estimates of the model to be somewhat similar to these, but the epsilon error terms might affect them marginally.

(b.ii) We need to check whether $\{y_t\}$ is a stationary process. There should be no problem with the error terms as they follow a MA(1) process, which is weakly dependent and stationary. Furthermore, we should determine, whether the errors and the explanatory variables are contemporaneously exogenous. If so, then our estimates will be consistent, if they are neither contemporaneously nor strictly exogenous, then our estimates will not be consistent.

(b.iii) Inclusion of the second lag term may help in case when y_t is more "persistent" (yet still weakly dependent in the long run) and its values from more than one period ago affect current values. Another instance when adding the second lag term into the regression might be beneficial is when the variable is recorded frequently, thus we have plenty observations, and again, we suspect that the variable is somewhat persistent beyond one time period. Generally speaking, if one has a large sample of observations, adding a second lag term should cause little harm to the model (as long as one has plenty degrees of freedom); in the worst case, it will turn out to be insignificant.

References

Lovell, M. C. (2008). A Simple Proof of the FWL Theorem. The Journal of Economic Education, 39(1), 88–91. http://www.jstor.org/stable/41426805

Wooldridge, J. M. (2015). Introductory econometrics: A modern approach. Cengage learning.

Problem 2 - Solution

Methodology and data

In this report, we study the impact of fluctuations in global oil prices on inflation. For the purpose of finding out the effect of oil price on inflation rate, we have compared the data for the US, controlling for exchange, interest and growth rates, imports and exports, and money supply. (source) We will estimate a single equation model by OLS method.

$$y = f(x_1, \ldots, x_8)$$

Dependent Variable Y - Growth rate of consumer price index (CPI)

Explanatory Variables X1 - Crude oil price in USD per barrel (WTI)

X2 - Unemployment rate,

X3 - GDP growth rate (% per month)

X4 – Exports (in US dollars)

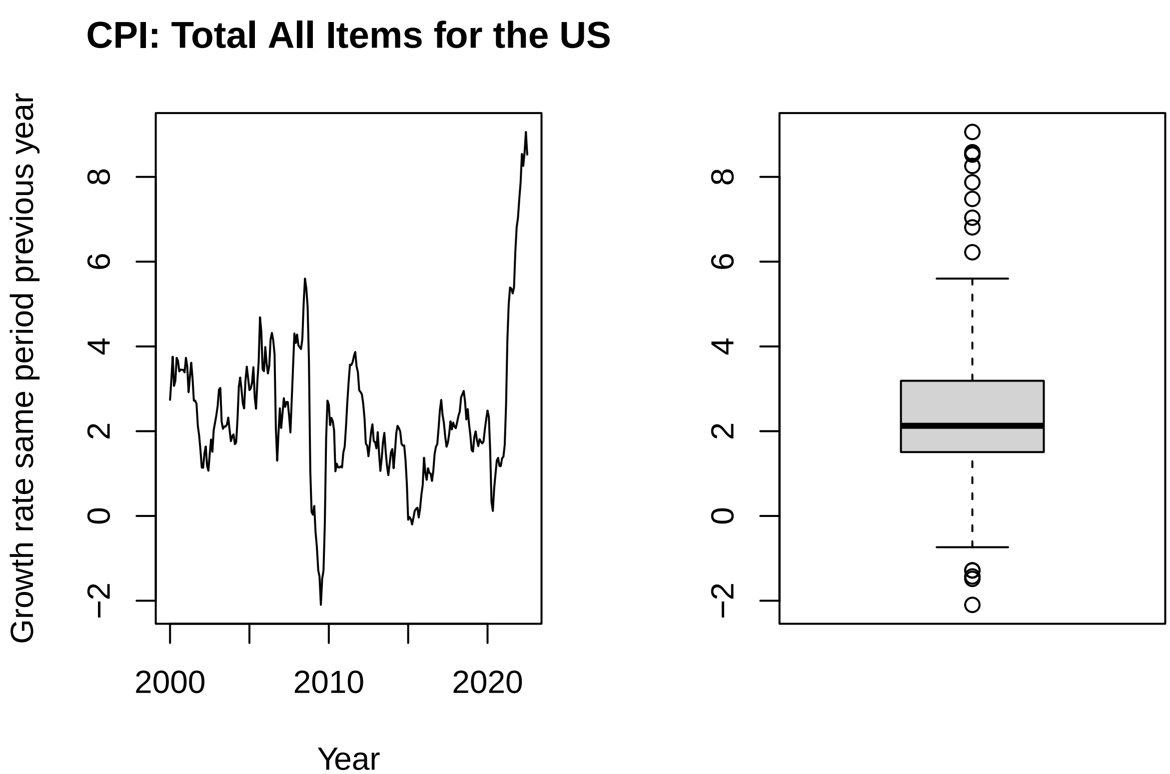
X5 - Imports (in US dollars)

X6 – Money supply (M3) (in national currency)

X7 - Nominal exchange rate

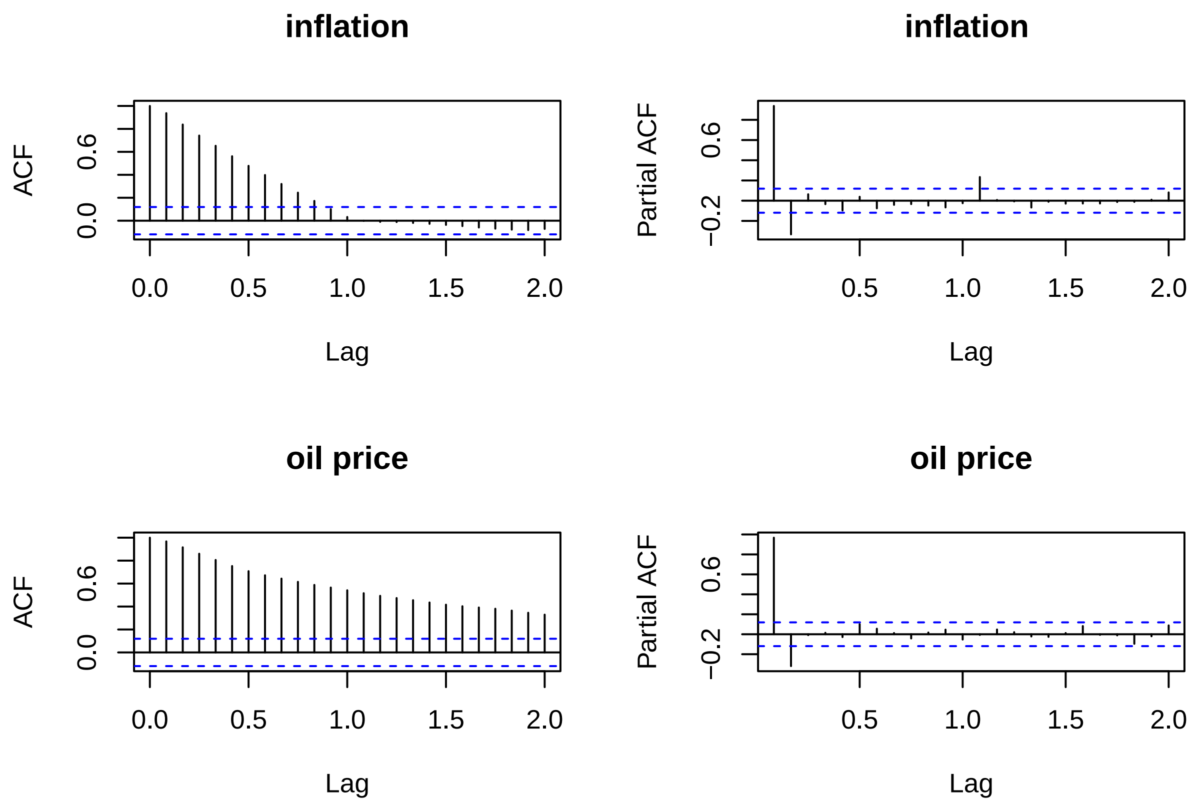
X8 - Short term Interestrate

The analysis covers monthly data for period between 2000 Q1 and 2022 Q1. Outliers were found in inflation data, however, we have not removed them. Explanatory variables selection was based on their potential to be significant determinants for the US inflation rate. Positive effect of oil price on inflation rate is expected.

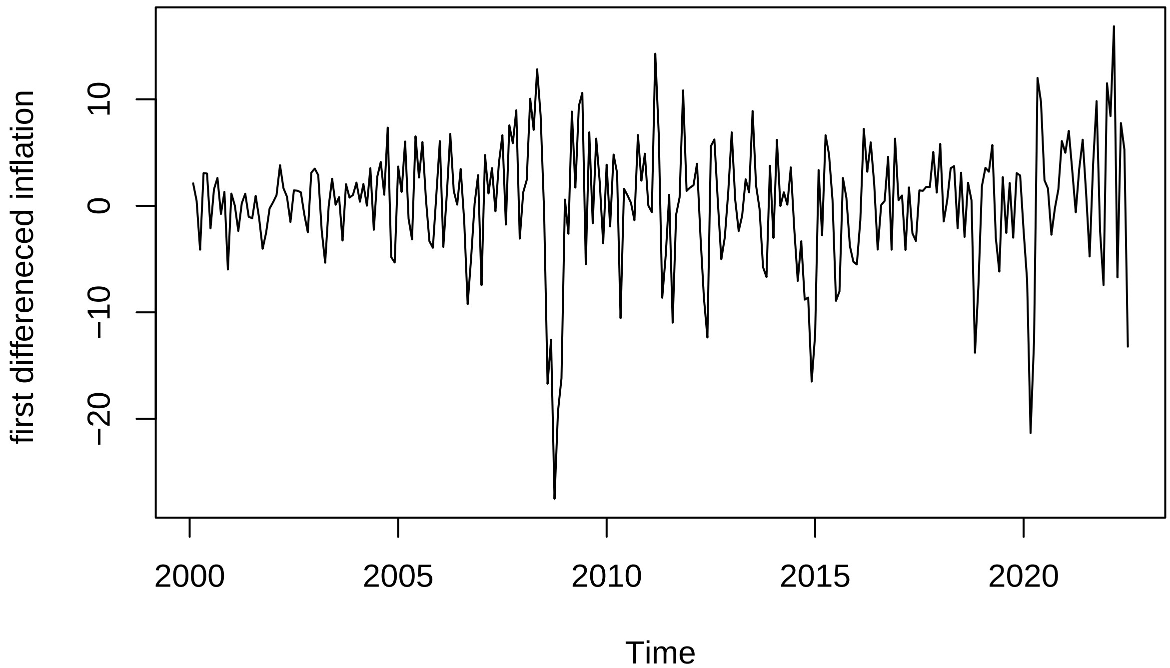


Looking at each of the time series, they all appear non-stationary except gdp growth. Unlike cross-sectional data, the data involving time series is increasingly susceptible to problems arising from the persistent nature of some series over a long period. To avoid this spurious regression problem, any highly persistent process must be identified and transformed into a weakly dependent time series, as conventional OLS inference procedures are valid and justified only under the assumption of weakly dependence. Correlograms appear to show evidence of AR(1) serial correlation.

Conducting a test for unit roots for each of the series we find that in all cases (including a time trend for those series that are clearly trending), that the p value against the appropriate Dickey-Fuller distribution is above the 5% critical value. Hence we cannot reject the null hypothesis of each series having a unit root. Hence the series themselves are highly persistent, and non-stationary.



We find that for all first differenced series except we reject the null of the series having a unit root. Hence all series rate appear to be I(1), since although the levels are non-stationary, the first differences are stationary.



We started with the following model to test the relationship between inflation and oil price, bearing in mind that testing for a relationship in differences between variables is not equivalent to testing for a long run relationship in levels.

$$\Delta inf = \beta_0 + \beta_1 \Delta oilprice + \beta_2 \Delta unem + \beta_3 gdpgrowth + \beta_4 \Delta exp + \beta_5 \Delta imp + \beta_6 \Delta m3 + \beta_7 \Delta fxrate + \beta_8 \Delta intrate$$

Further refinements were made after considering the possibility of a finite distributed lag model as it may be the case that some of the included variables have a lagged effect on inflation. Moreover, we have included the economic idea of inflation's self fullfiling expectations. As we examined serial correlations in the errors, we applied robust standard errors in our model, which also make it resistant to heteroscedasticity.

Ordinary least squares are used by estimating them with robust standard errors. The expected statistically significant positive effect of oil price on inflation rate was observed.For the US, we found that an immediate effect of a 10% increase in the difference of global oil price increases the difference in domestic inflation by about 3 percentage points. Interestingly, considering the total effect, the coefficient of the one year lag was estimated to be negative, while 18 month lagged effect is positive again. The self-fullfiling hypothesis seems to play its role as the lagged differenced inflation's estimates are all significant. The significant explanatory differenced variables are, in descending order by significance, monetary supply, oil price, imports, lagged oil price and gdp growth. The overall R squared is quite high, which may be partly due to the explanatory power of the model, but we have to be cautious as a lot of the variability is likely to be explained by the lagged dependent variable.

Empirical results

```
##
## Call:
## lmrob(formula = d_inf ~ d_op + lagpad(d_op, 6) + lagpad(d_op, 12) + lagpad(d_op,
##     18) + gdpg[-1] + diff(ex, 1) + diff(im, 1) + diff(fxr, 1) + diff(intr,
##     1) + lagpad(gdpg[-1], 12) + lagpad(diff(m3, 1), 12) + lagpad(diff(unr,
##     1), 12) + lagpad(d_inf, 12) + lagpad(d_inf, 24) + lagpad(d_inf, 36),
##     cov = ".vcov.w")
##  \--> method = "MM"
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -0.87708 -0.14486 -0.00338  0.15444  0.94070
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -6.069e-02  2.394e-02  -2.535  0.01194 *
## d_op                     2.747e-02  3.395e-03   8.090 4.15e-14 ***
## lagpad(d_op, 6)          2.009e-03  3.053e-03   0.658  0.51119
## lagpad(d_op, 12)        -6.764e-03  3.798e-03  -1.781  0.07635 .
## lagpad(d_op, 18)         7.221e-03  3.082e-03   2.343  0.02002 *
## gdpg[-1]                -2.638e-03  3.246e-03  -0.813  0.41728
## diff(ex, 1)              1.035e-05  7.050e-06   1.468  0.14343
## diff(im, 1)              1.143e-05  4.334e-06   2.637  0.00897 **
## diff(fxr, 1)             9.688e-01  6.818e-01   1.421  0.15677
## diff(intr, 1)            1.213e-01  1.092e-01   1.111  0.26770
## lagpad(gdpg[-1], 12)     6.506e-03  3.492e-03   1.863  0.06380 .
## lagpad(diff(m3, 1), 12)  9.454e-13  2.009e-13   4.707 4.47e-06 ***
## lagpad(diff(unr, 1), 12) 3.671e-02  3.453e-02   1.063  0.28891
## lagpad(d_inf, 12)       -6.324e-01  6.234e-02 -10.144  < 2e-16 ***
## lagpad(d_inf, 24)       -4.576e-01  6.388e-02  -7.163 1.20e-11 ***
## lagpad(d_inf, 36)       -1.480e-01  5.247e-02  -2.820  0.00525 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.2351
##   (36 observations deleted due to missingness)
## Multiple R-squared:  0.6843, Adjusted R-squared:  0.6625
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
##  26 weights are ~= 1. The remaining 208 ones are summarized as
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.07336 0.86060 0.95210 0.88910 0.98470 0.99890
## Algorithmic parameters:
##        tuning.chi                bb        tuning.psi        refine.tol
##         1.548e+00         5.000e-01         4.685e+00         1.000e-07
##           rel.tol         scale.tol         solve.tol       eps.outlier
##         1.000e-07         1.000e-10         1.000e-07         4.274e-04
##             eps.x warn.limit.reject warn.limit.meanrw
##         1.844e+00         5.000e-01         5.000e-01
##      nResample         max.it       best.r.s       k.fast.s          k.max
##            500             50              2              1            200
##    maxit.scale      trace.lev            mts     compute.rd fast.s.large.n
##            200              0           1000              0           2000
##                   psi           subsampling                   cov
##            "bisquare"         "nonsingular"          ".vcov.w"
## compute.outlier.stats
##                  "SM"
## seed : int(0)
```

Based on the shape of the series, we chose the years indicating The Great Financial Crisis during 2007 and 2008 as it had a visible impact on the shape of the graphs. Moreover, the invasion of Ukraine was considered.

```
##
## Call:
## lmrob(formula = d_inf ~ d_op + lagpad(d_op, 6) + lagpad(d_op, 12) + lagpad(d_op,
##     18) + gdpg[-1] + diff(ex, 1) + diff(im, 1) + diff(fxr, 1) + diff(intr,
##     1) + lagpad(gdpg[-1], 12) + lagpad(diff(m3, 1), 12) + lagpad(diff(unr,
##     1), 12) + lagpad(d_inf, 12) + lagpad(d_inf, 24) + lagpad(d_inf, 36) +
```

```
##     dummy_7[-1] + dummy_22[-1], cov = ".vcov.w")
##  \--> method = "MM"
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.8750202 -0.1456868  0.0004981  0.1544403  0.9407805
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -5.947e-02  2.482e-02  -2.396  0.01743 *
## d_op                     2.730e-02  3.447e-03   7.919 1.24e-13 ***
## lagpad(d_op, 6)          1.827e-03  3.085e-03   0.592  0.55439
## lagpad(d_op, 12)        -6.660e-03  3.833e-03  -1.737  0.08374 .
## lagpad(d_op, 18)         7.166e-03  3.129e-03   2.290  0.02297 *
## gdpg[-1]                -2.236e-03  3.369e-03  -0.664  0.50768
## diff(ex, 1)              9.642e-06  7.264e-06   1.327  0.18578
## diff(im, 1)              1.161e-05  4.381e-06   2.651  0.00862 **
## diff(fxr, 1)             1.007e+00  6.897e-01   1.459  0.14590
## diff(intr, 1)            1.077e-01  1.180e-01   0.913  0.36245
## lagpad(gdpg[-1], 12)     6.179e-03  3.561e-03   1.735  0.08412 .
## lagpad(diff(m3, 1), 12)  8.971e-13  2.162e-13   4.150 4.79e-05 ***
## lagpad(diff(unr, 1), 12) 3.726e-02  3.488e-02   1.068  0.28658
## lagpad(d_inf, 12)       -6.387e-01  6.344e-02 -10.068  < 2e-16 ***
## lagpad(d_inf, 24)       -4.590e-01  6.466e-02  -7.099 1.79e-11 ***
## lagpad(d_inf, 36)       -1.485e-01  5.309e-02  -2.797  0.00562 **
## dummy_7[-1]              6.683e-03  7.952e-02   0.084  0.93310
## dummy_22[-1]             7.803e-02  1.213e-01   0.643  0.52062
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.2377
##   (36 observations deleted due to missingness)
## Multiple R-squared:  0.6833, Adjusted R-squared:  0.6584
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
##  28 weights are ~= 1. The remaining 206 ones are summarized as
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.08203 0.86210 0.95450 0.89050 0.98410 0.99890
## Algorithmic parameters:
##        tuning.chi                bb        tuning.psi        refine.tol
##         1.548e+00         5.000e-01         4.685e+00         1.000e-07
##           rel.tol         scale.tol         solve.tol       eps.outlier
##         1.000e-07         1.000e-10         1.000e-07         4.274e-04
##             eps.x warn.limit.reject warn.limit.meanrw
##         1.844e+00         5.000e-01         5.000e-01
##      nResample         max.it       best.r.s       k.fast.s          k.max
##            500             50              2              1            200
##    maxit.scale      trace.lev            mts     compute.rd fast.s.large.n
##            200              0           1000              0           2000
##                   psi           subsampling                   cov
##            "bisquare"         "nonsingular"             ".vcov.w"
## compute.outlier.stats
##                  "SM"
## seed : int(0)
```

The addition of significant events did not lead to an improvement in our model.

CLM assumptions

During our analysis, we applied both formal and informal methods of testing for possible violations of the asymptotic Gauss-Markov assumptions as our 271 observations can be considered sufficient to approximate infinite sample size.(By adding the lagged dependent variable, we violated the strict exogeneity and we did not check normality of the disturbances).

To elaborate, the first assumption of linearity in parameters, stationary and weak dependence should be satisfied by difference first order integrated processes. We expect no perfect collinearity between our explanatory variables. but We tried to make achieve contemporaneous exogeneity and corresponding consistency of our estimators by having a rich model and preventing missing variable bias. However, it is possible that measurment error or functional form misspecification is present, leading to bias. We rejected the null hypothesis of no autocorrelation for higher order residuals by F-test. This implies that the no autocorrelation of disturbances hypothesis is not satisfied and our estimates are not efficient. In the case of serial correlation, OLS is not BLUE; standard errors and test statistic are invalid, and inference and hypothesis testing are impossible. The fifth assumption of homoscedasticity is likely to be violated too, as indicated by the White test.