COVid-19 Empirical Research Webinar

Notes from the presentation “Real time forecasting of Covid-19 intensive care units demand” and testing of the model proposed in the Danish context

Gregorio Luigi Saporito

Introduction

This report includes:

- Notes from the presentation "Real time forecasting of Covid-19 intensive care units demand" held the 30th of October 2020 as part of the "COVid-19 Empirical Research Webinar" by the Centre of Excellence in Economics and Data Science (CEEDS). For more details on the event and the program see https://ceeds.unimi.it/webinar-cover/.
- The implementation of the model suggested by Berta P., Lovaglio P. G., Paruolo P., Verzillo S. to forecast Covid-19 intensive care units demand in the Danish context.

Vector Error correction model (VECM)

The model proposed by the authors exploits the cointegrating relationship (in the sense of Engle R. F. and Granger C. W. J. (1987)) between the time series of hospitalised patients and intensive care unit (IC) occupancy. While this approach is quite popular in the Time Series Econometrics Literature, its application in the context of forecasting IC unit demand is less established.

The application of the VEC model in this context has some great potential because modelling time series in differences allows to capture just the short-term relationship between variables. The VEC model instead allows the possibility to include some aspects of the lon-run relationship if a cointegrating relationship exists. Essentially the presence of a cointegrating relationships indicates the presence of a long-run equilibrium relationship between two variables.

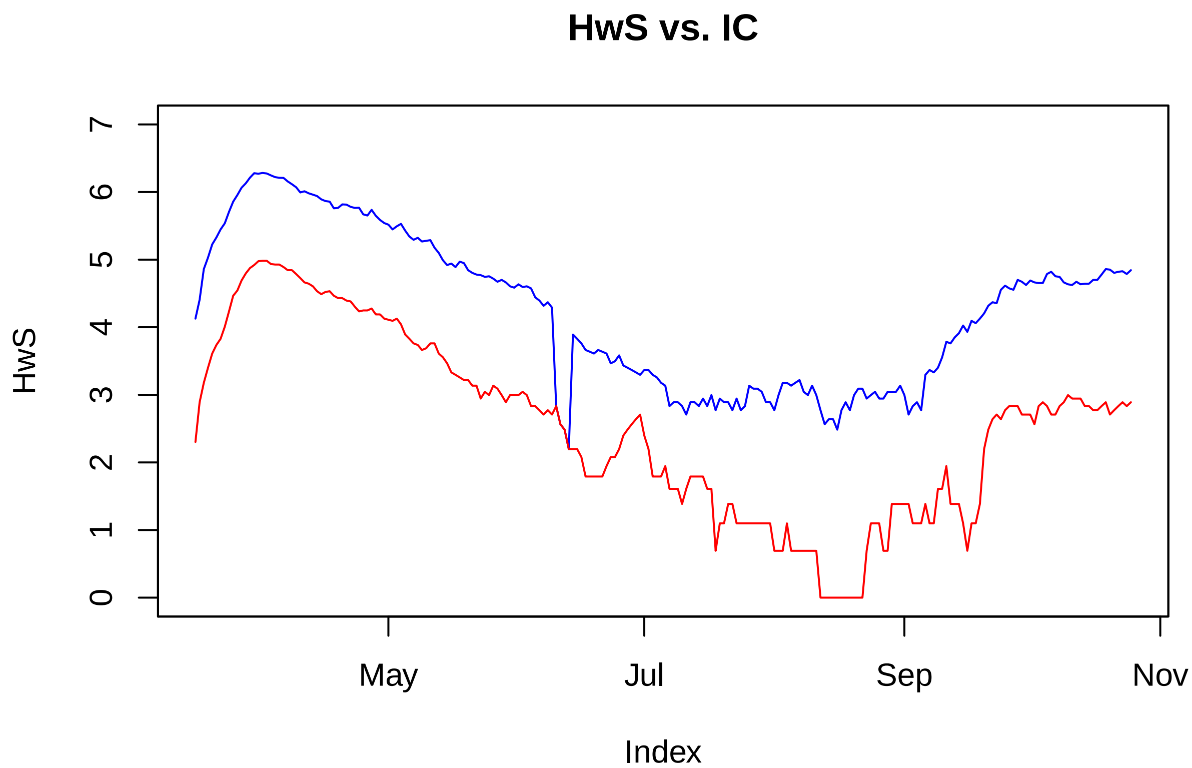
As Berta P., Lovaglio P. G., Paruolo P., Verzillo S. suggest, the presence of such cointegrating relationship oftentimes exists between the number of hospitalised patients and the demand for intensive care beds and it can be exploited to obtain better forecasts. In other words, leveraging this relationship can turn out to be more powerful than just modelling the time series in differences.

This methodology was successfull validated in regions of Italy, Spain, and Switzerland (Berta et al. (2020)). Based on this evidence, I decided to investigate if this approach can work in other contexts too. In particular, I chose to test it on Denmark.

Application in the Danish context

The dataset was retrieved from the European Centre for Disease Prevention and Control (ECDC) (see https://www.ecdc.europa.eu/en/publications-data/download-data-hospital-and-icu-admission-rates-and-current-occupancy-covid-19) and it contains the daily time series of hospitalisation and Intensive Care Unit occupancy.

A total of 9 missing data points were interpolated using `auto.arima`. This is how the plot in logs of hospitalised with Covid and in intensive care looks like.



As can be seen from the plot above, there appears to be a systemic long-run relationship between the two time series and it is worth investigating if a cointegrating relationship exists.

Unit Root Test

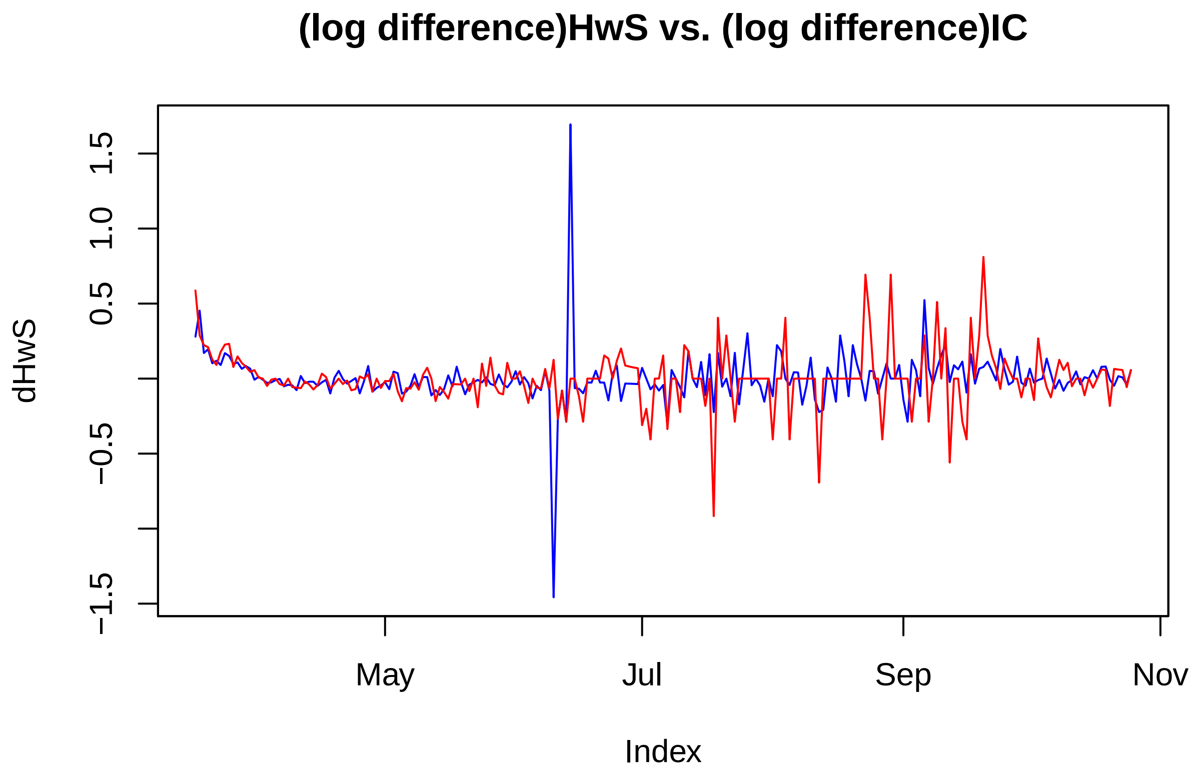
The results below show that we can reject the null of no unit root (the test statistic has to be larger than the critical value).

test	variable	test statistic	1pct	5pct	10pct
Elliot, Rothenberg and Stock P-test	HwS	12.29425	1.99	3.26	4.48
Elliot, Rothenberg and Stock P-test	IC	10.16641	1.99	3.26	4.48

Having evidence of a unit root in levels, we now test for a unit root in differences.

test	variable	test statistic	1pct	5pct	10pct
Elliot, Rothenberg and Stock P-test	diff(HwS)	0.1004807	1.99	3.26	4.48
Elliot, Rothenberg and Stock P-test	diff(IC)	2.6291726	1.99	3.26	4.48

Based on this result, we can say that both time series are I(1).



Cointegration analysis

According to the Akaike information criterion the number of lags to include in the cointegration analysis is 5. We then proceed with the cointegration test. r=0 tests for the presence of cointegration. The test statistic exceeds the significance level and we can reject the null hypothesis of no cointegration.

```
##
## ######################
## # Johansen-Procedure #
## ######################
##
## Test type: maximal eigenvalue statistic (lambda max) , without linear trend and constant in cointegra
##
## Eigenvalues (lambda):
## [1] 7.104149e-02 7.005939e-03 1.466825e-18
##
## Values of teststatistic and critical values of test:
##
##           test 10pct  5pct  1pct
## r <= 1 |  1.49  7.52  9.24 12.97
## r = 0  | 15.62 13.75 15.67 20.20
##
## Eigenvectors, normalised to first column:
## (These are the cointegration relations)
##
```

```
##              HwS.l1      IC.l1  constant
## HwS.l1    1.0000000  1.0000000  1.000000
## IC.l1    -0.8523499  0.2618179 -1.990459
## constant -2.0945173 -4.6743596 11.839450
##
## Weights W:
## (This is the loading matrix)
##
##             HwS.l1        IC.l1      constant
## HwS.d -0.005023771 -0.009697439 -1.156890e-18
## IC.d   0.133315174 -0.002200217  1.101521e-17
```

This is essentially the same as breaking down the test into two parts

- The two series are integrated of order 1 (I(1)) (as proven above)
- Their residuals from OLS are I(0) (the test statistic 0.8359569 is lower than the 5% significance level 3.26)

These conditions are satisfied, the two series are hence cointegrated and we can exploit this long-run equilibrium by including it in our model.

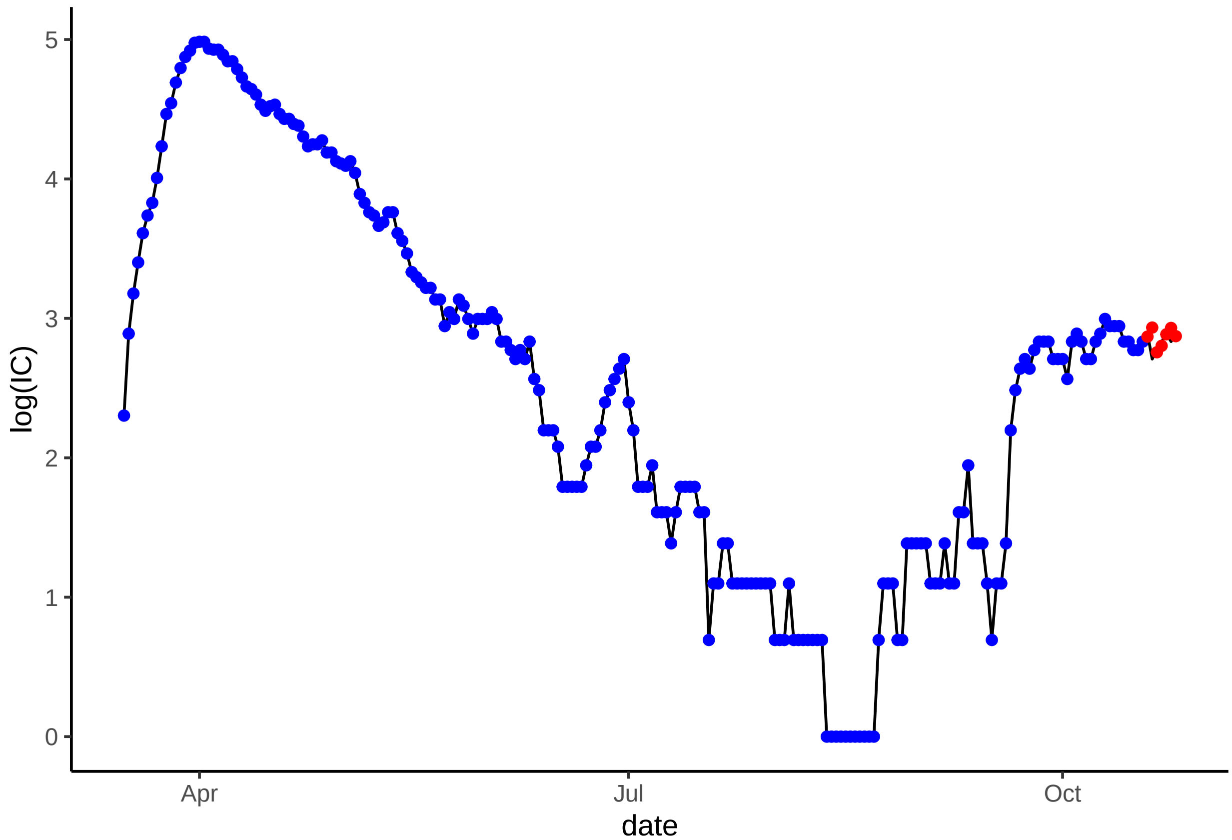
We then estimate the unrestriced VEC model and use the restricted sample to estimate a bivariate VEC model. These adjustment parameters α_1 and α_2 for HwS and IC are -0.0050238 and 0.1333152 respectively. While α_2 is significan α_1 is not. We therefore proceed with the restricted adjustment of the parameter α setting α_1 to zero. This is essentially like saying that HwS is a pure random walk and all the adjustment occurs in IC.

```
##
## ######################
## # Johansen-Procedure #
## ######################
##
## Estimation and testing under linear restrictions on beta
##
## The VECM has been estimated subject to:
## beta=H*phi and/or alpha=A*psi
##
##      [,1]
## [1,]    0
## [2,]    1
##
## Eigenvalues of restricted VAR (lambda):
## [1] 0.0709 0.0000 0.0000
##
## The value of the likelihood ratio test statistic:
## 0.02 distributed as chi square with 1 df.
## The p-value of the test statistic is: 0.88
##
## Eigenvectors, normalised to first column
## of the restricted VAR:
##
##                [,1]
## RK.HwS.l1    1.0000
## RK.IC.l1    -0.8554
```

```
## RK.constant -2.0873
##
## Weights W of the restricted VAR:
##
##        [,1]
## [1,] 0.0000
## [2,] 0.1339
```

One week ahead forecast

In red the out of sample estimates of IC occupancy.



Comparison with the random walk and performance

The VEC model does outperform the random walk according to the MAE (mean absolute error) and MAPE (mean absolute percentage error) indices.

	MAE	MAPE
VEC	0.0591286	2.032092
Random_Walk	0.1036886	138.839070

$$MAE = (\frac{1}{n})\sum_{i=1}^{n}|F_i - A_i|\, ;\; MAPE = (\frac{1}{n})\sum_{i=1}^{n}\left|\frac{A_i - F_i}{A_i}\right|$$

Where A indicates the actual and F the forecast

Conclusion

The VEC model did outperform the random walk, which validates the approach suggested by Berta P., Lovaglio P. G., Paruolo P., Verzillo S. in their paper. The VEC has the advatage of capturing the long-run relationship between the variables allowing for temporary deviations from the equilibrium. Furthermore, its parameters have a clear interpretation (i.e. no black box). However, the main disadvantage is its flexibility: a cointegrating relationship might not always exists and, in such circumstances, other forecasting techniques have to be used. Nonetheless, when then data generating process gives evidence of a cointegrating relationship the VEC model can be a powerful candidate to obtain robust forecasts.

References

Berta, P., Lovaglio, P. G., Paruolo, P., & Verzillo, S. (2020). Real time forecasting of Covid-19 intensive care units demand (No. 2020-08).

Berta, P., Paruolo, P., Verzillo, S., & Lovaglio, P. G. (2020). A bivariate prediction approach for adapting the health care system response to the spread of COVID-19. Plos one, 15(10), e0240150.

Engle, R. F., & Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing. Econometrica: journal of the Econometric Society, 251-276.