005 - Data

Karel Räpp, Henry Enno Turu

28 11 2020

```
library(readr)
library(ggplot2)
library(knitr)
GDP_data <- read_csv("data/actualvsplot.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   Date = col_date(format = ""),
##   Actual = col_double(),
##   Extracted = col_double()
## )
```

1 Data

For the regression analysis, we (have) extract(ed) the following data:

- Quarterly 5-year Treasury Implied Volatility
- Dependent Variables:
    1. Quarterly GDP year-on-year growth rate
    2. Monthly industrial production
    3. Monthly non-farm payroll
    4. Monthly consumption
- Control Variables
    5. US treasury interest rates to construct term spreads
    6. credit spreads
    7. stock returns
    8. stock market implied volatility
    9. Housing Starts

As mentioned in the review of literature, YIV is constructed using Black model & deriving implied volatility through option prices, time-to-maturity, etc. The data regarding options

on treasury bond futures, however, is collected by CME and available only through accessing their database.

Due to the data price being behind the paywall, we had to resort to other measures to access the data. Firstly, we have contacted prof. Cremers, Gandhi & Fleckenstein, who are ready to share their data with us if the CME group gives their consent. Nevertheless, we cannot rely solely on this possibility, and thus found an alternative.

As researchers typically do not typically post underlying data with their research, various plot digitizers have seen an exponential increase in use. Drevon et al. (2017) researched intercoder reliability, during which over 3500 data points were extracted with WebPlotDigitizer from 36 different graphs. Nevertheless, they controlled the validity of the results and concluded that there was a near perfect correlation (r=0.989 with p-value <0.01) between extracted and actual data. Nevertheless, the limitations mentioned highlight coders previous experience with plot-digitizing tools.

Furthermore, Burda et al. (2017) also highlight that systematic reviewers often tend to have data constraints which is why plot digitizers are of a great help. They estimated data using WebPlotDigitizer and conclude that the extraction done by different coders was consistent; nevertheless, in the case of continuous data (compared to event data), the distribution varied more. Whatsoever, the intreclass coefficient for both types of plots was over 95%.

We also used the WebPlotDigitzer in our research and as validity test extracted GDP from the same graph as YIV time series & plotted it with actuals - see the graph below.

The data regarding GDP year-on-year growth rate was extracted on a quarterly basis from Archival FRED (2020). Furthermore, the vintage 2019-12-20 was extracted as it coincided the best with the data used in the research paper by Cremers et al. (2017). Other dependent variables (industrial production, non-farm payroll and consumption) & housing data was extracted from the same database but on a monthly basis. The data for risk free rates that will be used in constructing term spreads was extracted from the dataset of The Federal Reserve Board working paper. (Gurkaynak et al., 2006)

Archival FRED. (2020). *Dataset. (Vintage 2020-20-12).* https://alfred.stlouisfed.org/

Burda, B. U., O'Connor, E. A., Webber, E. M., Redmond, N., & Perdue, L. A. (2017). Estimating data from figures with a Web-based program: Considerations for a systematic review. *Research Synthesis Methods*, *8*(3), 258–262. https://doi.org/10.1002/jrsm.1232

Cremers, M., Fleckenstein, M., & Gandhi, P. (2017). Treasury Yield Implied Volatility and Real Activity. *Journal of Financial Economics (JFE), Forthcoming.* https://doi.org/10.2139/ssrn.3006473

Drevon, D., Fursa, S. R., & Malcolm, A. L. (2017). Intercoder Reliability and Validity of WebPlotDigitizer in Extracting Graphed Data. *Behavior Modification*, *41*(2), 323–339. https://doi.org/10.1177/0145445516673998

Gurkaynak, R. S., Sack, B. P., & Wright, J. H. (2006). *The U.S. Treasury Yield Curve: 1961 to the Present.* https://doi.org/10.2139/ssrn.920183

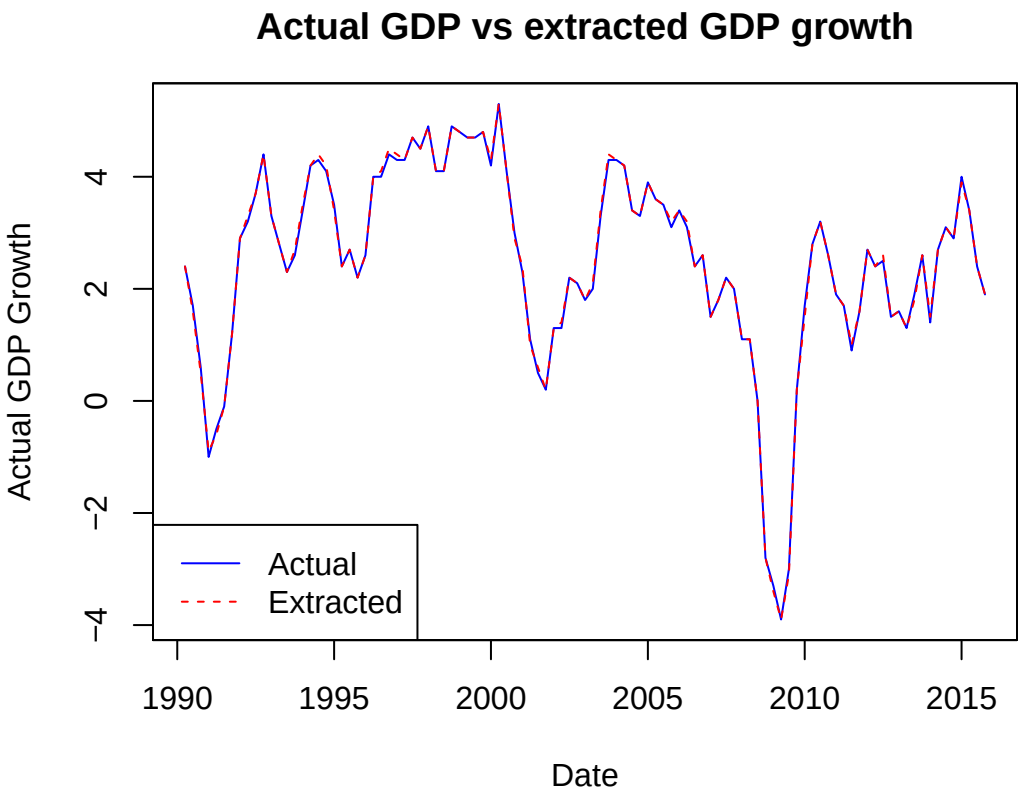

Figure 1.1: Actual GDP growth vs extracted GDP growth in % using WebPlotDigitizer