

Python Generate Latex Regression Tables from CSV/TXT

Fan Wang

2020-12-01

Contents

1 CVS to Latex	1
1.1 Store a Specially Formatted Text File	1
1.2 Latex Table Template	2

1 CVS to Latex

Go to the [RMD](#), [PDF](#), or [HTML](#) version of this file. Go back to [Python Code Examples](#) Repository ([bookdown site](#)) or the [pyfan](#) Package ([API](#)).

```
import numpy as np
```

1.1 Store a Specially Formatted Text File

Below we have the text file, for processing by the XXX function in pyfan to convert a specially formatted text file to a latex table. The text file is saved to the *_file* folder

The text format has these rules:

1. Table heading starts with single pound
2. Each grouping-heading starts with double pound
 - if this is directly followed by no pound results, this is a variable by itself.
3. Each sub-heading with triple bpound corresponds to a particular variable
4. Right arrow indicates text to be copy-pasted directly as latex
5. Assume that point estimates have possibly stars, are in the form of XXX.XX***
6. Assume that standard errors have brackets: (YY.YY)

```
stf_txt_contents_regq = """
# Quantile Regression Estimates

> \\toprule

> & \\multicolumn{5}{l}{Estimates at conditional quantiles}\\\\\\
> \\cmidrule(l{5pt}r{5pt}){2-6}
> Variable & P10 & P25 & P50 & P75 & P90\\\\\\

> \\midrule
## Environmental Exposure Variables
# PM10
-33.633***, -21.757***, -17.121***, -12.970***, -15.747***
(3.305), (2.396), (1.862), (2.223), (2.663)
# Extreme hot
```

```

-50.202**, -32.292**, -30.162***, -28.485**, -30.271*
(21.322), (13.677), (10.892), (12.193), (16.258)

## Education and Environmental Exposure Interactions
# College completion
-252.719***, -133.758**, -26.323, -70.733, -39.212
(80.041), (54.355), (47.235), (58.473), (80.659)
# College x PM10
3.858***, 2.153***, 0.705, 1.072, 0.684
(1.065), (0.704), (0.592), (0.758), (1.049)
# College x extreme hot
9.134, 2.405, -10.92, -8.465, -7.021
(14.246), (8.026), (8.823), (8.392), (11.379)

## Control Variables
# Male
84.506***, 91.758***, 107.950***, 115.879***, 133.560***
(7.796), (4.848), (4.111), (5.044), (6.961)
# Mother's age
68.180***, 47.921***, 47.127***, 51.366***, 41.764***
(8.986), (5.368), (4.995), (5.099), (8.637)

##> Intercept
1180.908***, 1434.431***, 1891.264***, 1923.296***, 2460.865***
(327.266), (253.204), (239.577), (244.979), (333.483)

> \midrule

> $R(p)$ & MISSING & MISSING & MISSING & MISSING & MISSING\\

> \\bottomrule
"""
# Relative file name
srt_regq_txt = "_file/"
sna_regq_txt = "test_regquant_fan"
srn_regq_txt = srt_regq_txt + sna_regq_txt + ".txt"
srn_regq_tex = srt_regq_txt + sna_regq_txt + ".tex"
# Open new file
fl_txt_contents_regq = open(srn_regq_txt, 'w')
# Write to File
fl_txt_contents_regq.write(stf_txt_contents_regq)

## 1305

fl_txt_contents_regq.close()

print(stf_txt_contents_regq.split("\n"))

```

```
## [' ', '# Quantile Regression Estimates', ' ', '> \\toprule', ' ', '> & \\multicolumn{5}{l}{Estimates at
```

1.2 Latex Table Template

Create a latex file from the text file we just created, with minimal formatting.

First, need to specify several additional parameters, including the number of columns, and column widths for

the label column and coefficient columns.

```
# Define columns and width
it_col_count = 6
fl_adj_box_maxwidth = 1
# Define format decimal
it_round_decimal = 2
# formatting width
fl_col_label_width_cm = 5
fl_col_coef_width_cm = 2
# Indentation
fl_indent_pound1_mm = 0
fl_indent_pound2_mm = 0
fl_indent_pound3_mm = 6
```

Second, load in the text file and read it out line by line.

```
fl_txt_regq_contents = open(srn_regq_txt)
ls_st_txt_regq_contents = fl_txt_regq_contents.readlines()
fl_txt_regq_contents.close()
print(ls_st_txt_regq_contents)
```

```
## ['\n', '# Quantile Regression Estimates\n', '\n', '> \\toprule\n', '\n', '> & \\multicolumn{5}{l}{Es
```

Third, loop over the lines in the list, and interpret each line differently based on what each line starts with:

```
# Return list
ls_st_tex_returns = []
# Parse lines one by one
for st_txt_regq_line in ls_st_txt_regq_contents:

    # delete linebreak at end of line
    st_txt_regq_line = st_txt_regq_line.replace("\n", "")

    # Check start of line for special characters
    st_N2_char = st_txt_regq_line[0:2]
    st_N3_char = st_txt_regq_line[0:3]
    st_N4_char = st_txt_regq_line[0:4]

    # Conditional processing
    st_text_out = ''
    if st_N4_char == '# ':
        # Variable heading, with 3rd level indent, to be connected with coefficients etc
        st_txt_regq_line = st_txt_regq_line.replace("# ", "")
        st_text_out = "\\hspace*{" + str(fl_indent_pound3_mm) + "mm}" + st_txt_regq_line
    elif st_N4_char == '##> ':
        # Variable heading, with 2nd level indent, to be connected with coefficients etc
        st_txt_regq_line = st_txt_regq_line.replace("##> ", "")
        st_text_out = "\\hspace*{" + str(fl_indent_pound2_mm) + "mm}" + st_txt_regq_line
    elif st_N3_char == '## ':
        # Group of Coefficients
        st_txt_regq_line = st_txt_regq_line.replace("## ", "")
        st_text_out = "\\addlinespace\n"
        st_text_out = st_text_out + "\\multicolumn{" + str(it_col_count) + "}{l}{\\hspace*{" + str(fl_ind
        st_text_out = st_text_out + '\\addlinespace\n'
    elif st_N2_char == '# ':

```

```

# title line and initialize file lines
st_txt_regq_line = st_txt_regq_line.replace("# ", "")
st_text_out = st_text_out + "\\begin{table}[htbp]\n"
st_text_out = st_text_out + "\centering\n"
st_text_out = st_text_out + "\caption{\hspace*{" + str(fl_indent_pound1_mm) + "mm}" + st_txt_regq_line + "\n"
st_text_out = st_text_out + "\\begin{adjustbox}{max width=" + str(fl_adj_box_maxwidth) + "\textheight=" + str(fl_adj_box_maxheight) + "cm}\n"
st_text_out = st_text_out + \
    "\\begin{tabular}{m{" + str(fl_col_label_width_cm) + "cm}" \
    "*{" + str(it_col_count-1) + "}>{\centering\\arraybackslash}" \
    "m{" + str(fl_col_coef_width_cm) + "cm}}}\n"

elif st_N2_char == '> ':
    # latex code line, include as it appears
    st_txt_regq_line = st_txt_regq_line.replace("> ", "")
    st_text_out = st_txt_regq_line + "\n"
else:
    # Assume no headline/code row is comma separated
    ls_st_estimates = st_txt_regq_line.split(",")

    # Loop over each value separated by commas
    for it_esti_ctr, st_esti in enumerate(ls_st_estimates):

        # Default update is to keep current
        st_esti_update = st_esti

        # If estimates, might have stars, first check star count
        # delete stars, for numeric conversion and rounding
        it_star_count = np.nan
        if "***" in st_esti:
            it_star_count = 3
            st_esti = st_esti.replace("***", "")
        elif "**" in st_esti:
            it_star_count = 2
            st_esti = st_esti.replace("**", "")
        elif "*" in st_esti:
            it_star_count = 1
            st_esti = st_esti.replace("*", "")
        else:
            it_star_count = 0

        # Check if has brackets
        it_bracket_count = np.nan
        if "(" in st_esti:
            it_bracket_count = 1
            st_esti = st_esti.replace("(", "")
            if ")" in st_esti:
                it_bracket_count = 2
                st_esti = st_esti.replace(")", "")

        # Decimal Rounding
        try:
            # numerical
            fl_esti_rounded = round(float(st_esti), it_round_decimal)

```

```

        fl_esti_rounded = f'{fl_esti_rounded:.{it_round_decimal}f}'
        st_esti_rounded = str(fl_esti_rounded)
    except Exception:
        # Might be non-numeric
        st_esti_rounded = st_esti

    # Conditional Processing for Point Estimates and SE
    if it_bracket_count is np.nan:

        # A. No brackets, these are point estimates
        # Convert Estimate
        st_esti_starred = st_esti_rounded
        if it_star_count == 3:
            st_esti_starred = st_esti_rounded + "\sym{***}"
        elif it_star_count == 2:
            st_esti_starred = st_esti_rounded + "\sym{**}"
        elif it_star_count == 1:
            st_esti_starred = st_esti_rounded + "\sym{*}"

        st_esti_update = st_esti_starred

    else:
        # B. brackets, these are standard errors

        if it_bracket_count == 2:
            st_esti_update = "(" + st_esti_rounded + ")"
        else:
            raise TypeError(f'{ls_st_estimates=} and {st_esti=}, missing bracket')

    # Update List
    ls_st_estimates[it_esti_ctr] = st_esti_update

    # Flatten comman
    st_text_out = ' & '.join(ls_st_estimates)
    if len(st_text_out) > 0:
        # add ampersand front
        st_text_out = ' & ' + st_text_out + '\\\\n\\addlinespace\n'

    # st_text_out = st_txt_regq_line
    # st_text_out = st_text_out.replace("***", "\sym{+++}")
    # st_text_out = st_text_out.replace("**", "\sym{==}")
    # st_text_out = st_text_out.replace("*", "\sym{*}")
    # st_text_out = st_text_out.replace("\sym{==}", "\sym{**}")
    # st_text_out = st_text_out.replace("\sym{+++}", "\sym{***}")

    if len(st_text_out) > 0:
        ls_st_tex_returns.append(st_text_out)

    # close connection
    ls_st_tex_returns.append("\end{tabular}\n")
    ls_st_tex_returns.append("\end{adjustbox}\n")
    ls_st_tex_returns.append("\end{table}\n")

```

Fourth, write list of strings to latex file:

```

# Open tex file
fl_tex = open(srn_regq_tex, "w")
# Read list of strings line by line and write to tex
for st_tex_return in ls_st_tex_returns:
    fl_tex.write(st_tex_return)
# close tex file

```

```

## 189
## 9
## 59
## 30
## 41
## 9
## 96
## 164
## 107
## 62
## 171
## 102
## 67
## 113
## 178
## 81
## 67
## 174
## 70
## 62
## 181
## 56
## 64
## 81
## 164
## 105
## 62
## 172
## 102
## 62
## 22
## 112
## 72
## 9
## 59
## 12
## 14
## 16
## 12

```

```

fl_tex.close()

```

Fifth, read from latex file:

```

# load latex file and print
fl_tex_regq_contents = open(srn_regq_tex)
st_txt_regq_contents = fl_tex_regq_contents.read()
fl_txt_regq_contents.close()

```

```
print(st_txt_regq_contents)
```

```
## \begin{table}[htbp]
## \centering
## \caption{\hspace*{0mm}Quantile Regression Estimates}
## \begin{adjustbox}{max width=1\textwidth}
## \begin{tabular}{m{5cm}*{5}{>{\centering\arraybackslash}m{2cm}}}}
## \toprule
## & \multicolumn{5}{l}{Estimates at conditional quantiles}\\
## \cmidrule{1{5pt}r{5pt}}{2-6}
## Variable & P10 & P25 & P50 & P75 & P90\\
## \midrule
## \addlinespace
## \multicolumn{6}{l}{\hspace*{0mm}Environmental Exposure Variables}\\
## \addlinespace
## \begin{table}[htbp]
## \centering
## \caption{\hspace*{0mm}PM10}
## \begin{adjustbox}{max width=1\textwidth}
## \begin{tabular}{m{5cm}*{5}{>{\centering\arraybackslash}m{2cm}}}}
## & -33.63\sym{***} & -21.76\sym{***} & -17.12\sym{***} & -12.97\sym{***} & -15.75\sym{***}\\
## \addlinespace
## & (3.31) & (2.40) & (1.86) & (2.22) & (2.66)\\
## \addlinespace
## \begin{table}[htbp]
## \centering
## \caption{\hspace*{0mm}Extreme hot}
## \begin{adjustbox}{max width=1\textwidth}
## \begin{tabular}{m{5cm}*{5}{>{\centering\arraybackslash}m{2cm}}}}
## & -50.20\sym{**} & -32.29\sym{**} & -30.16\sym{***} & -28.48\sym{**} & -30.27\sym{*}\\
## \addlinespace
## & (21.32) & (13.68) & (10.89) & (12.19) & (16.26)\\
## \addlinespace
## \addlinespace
## \multicolumn{6}{l}{\hspace*{0mm}Education and Environmental Exposure Interactions}\\
## \addlinespace
## \begin{table}[htbp]
## \centering
## \caption{\hspace*{0mm}College completion}
## \begin{adjustbox}{max width=1\textwidth}
## \begin{tabular}{m{5cm}*{5}{>{\centering\arraybackslash}m{2cm}}}}
## & -252.72\sym{***} & -133.76\sym{**} & -26.32 & -70.73 & -39.21\\
## \addlinespace
## & (80.04) & (54.35) & (47.23) & (58.47) & (80.66)\\
## \addlinespace
## \begin{table}[htbp]
## \centering
## \caption{\hspace*{0mm}College x PM10}
## \begin{adjustbox}{max width=1\textwidth}
## \begin{tabular}{m{5cm}*{5}{>{\centering\arraybackslash}m{2cm}}}}
## & 3.86\sym{***} & 2.15\sym{***} & 0.70 & 1.07 & 0.68\\
## \addlinespace
## & (1.06) & (0.70) & (0.59) & (0.76) & (1.05)\\
## \addlinespace
```

```

## \begin{table}[htbp]
## \centering
## \caption{\hspace*{0mm}College x extreme hot}
## \begin{adjustbox}{max width=1\textwidth}
## \begin{tabular}{m{5cm}*{5}{>{\centering\arraybackslash}m{2cm}}}}
## & 9.13 & 2.40 & -10.92 & -8.46 & -7.02\\
## \addlinespace
## & (14.25) & (8.03) & (8.82) & (8.39) & (11.38)\\
## \addlinespace
## \addlinespace
## \multicolumn{6}{l}{\hspace*{0mm}Control Variables}\\
## \addlinespace
## \begin{table}[htbp]
## \centering
## \caption{\hspace*{0mm}Male}
## \begin{adjustbox}{max width=1\textwidth}
## \begin{tabular}{m{5cm}*{5}{>{\centering\arraybackslash}m{2cm}}}}
## & 84.51\sym{***} & 91.76\sym{***} & 107.95\sym{***} & 115.88\sym{***} & 133.56\sym{***}\\
## \addlinespace
## & (7.80) & (4.85) & (4.11) & (5.04) & (6.96)\\
## \addlinespace
## \begin{table}[htbp]
## \centering
## \caption{\hspace*{0mm}Mother's age}
## \begin{adjustbox}{max width=1\textwidth}
## \begin{tabular}{m{5cm}*{5}{>{\centering\arraybackslash}m{2cm}}}}
## & 68.18\sym{***} & 47.92\sym{***} & 47.13\sym{***} & 51.37\sym{***} & 41.76\sym{***}\\
## \addlinespace
## & (8.99) & (5.37) & (5.00) & (5.10) & (8.64)\\
## \addlinespace
## \hspace*{0mm}Intercept & 1180.91\sym{***} & 1434.43\sym{***} & 1891.26\sym{***} & 1923.30\sym{***} & \\
## \addlinespace
## & (327.27) & (253.20) & (239.58) & (244.98) & (333.48)\\
## \addlinespace
## \midrule
## $R(p)$ & MISSING & MISSING & MISSING & MISSING & MISSING\\
## \bottomrule
## \end{tabular}
## \end{adjustbox}
## \end{table}

```