Introduction to LaTeX

Carl-Emil Pless

8th of October 2021

Contents

1 A thrilling introduction

For very good reasons, we want to look at what determines the hourly wage in the U.S. In order to do so, we want to use `Overleaf`. This document is meant as a reference that you can use when working on your own projects that provides a brief introduction (through examples) to LaTeX. In order to compile your document, you want to press `CTRL (CMD) + ENTER`, `CTRL (CMD) + S`, or the big green button that says *recompile*.

2 Writing a bit of rather arbitrary math

In order to estimate a linear model of hourly wage as a function of various covariates, a range of assumptions need to be in place. We assume that the data is randomly sampled and follows the model for the population:

$$y_i = x_i^{'}\beta + u_i,$$

for observations $i = 1, 2, \ldots, N$. Let β be a $k \times 1$ vector of parameters and u_i is an asymptotically normal error term, such that $u_i \sim i.i.d.N(0, \sigma^2)$. Let $x_{0i} = 1$, such that β_0 corresponds to the population intercept and the model can be described as:

$$y_i = x_i^{'}\beta + u_i = x_{0i}\beta_0 + x_{1i}\beta_1 + \ldots + x_{ik}\beta_k + u_i.$$

Using the Law of Iterated Expectations we assume that x_i and u_i are uncorrelated such that $E[x_i u_i] = E[E[x_i u_i \mid x_I]] = E[x_i E[u_i \mid x_i]] = 0$. This implies that if the $k \times k$ matrix $E[x_i x_i^{'}]$ is non-singular (and hence invertible) we can derive the OLS estimator as:

$$\beta = E[x_i x_i^{'}]^{-1} E[x_i y_i], \tag{1}$$

which is the population parameter.

2.1 The OLS estimator

From a finite sample of y_i and x_i, we cannot compute the expectations and (1) is infeasible. Therefore, we replace expectations with sample averages in (1), such that:

$$\hat{\beta} = \left(N^{-1} \sum_{i=1}^{N} x_i x_i^{'} \right)^{-1} \left(N^{-1} \sum_{i=1}^{N} x_i y_i \right). \tag{2}$$

In order for the steps from (1) to (2) to work, we need a law of large numbers (LLN) to apply, such that the sample averages converge to the expectations for increasing sample sizes. In other words:

$$N^{-1} \sum_{i=1}^{N} x_i y_i \rightarrow E[x_i y_i] \quad \text{and} \quad N^{-1} \sum_{i=1}^{N} x_i x_i^{'} \rightarrow E[x_i x_i^{'}].$$

Under these assumptions it holds that $\hat{\beta}$ is defined and (2) converges to the true value in (1). In addition, we note that this estimator is consistent and unbiased when $u_i \sim i.i.d.N(0, \sigma^2)$. In fact, our estimator is the minimum variance unbiased (even nonlinear) estimator, under these conditions.

2.1.1 Making a subsubsubsection

You make subsections by adding the proper amount of *sub* prefixes to your section. If you include a star (`subsubsection*`), then it will be unnumbered. These are automatically included in the table of contents.

2.1.2 Making a list

Making a list is quite simple. If you want it to be numerated (or alphabetical) you want to be using the `enumerate` or `enumerate[label=\alph*]` environment like so:

1. My first item
2. My second item
3. .
4. .
5. .
6. My last item

My alphabetized version:

a My first (alphabetized) item

b My second (alphabetized) item

c .

d .

e .

f My last (alphabetized) item

3 Functionality with *R*

In order to compile *R* code directly in `Overleaf`, all you need to do is rename your file to have a `.Rtex` ending. If you have not changed the file name already, then this would be from `main.tex` to `main.Rtex`. Then you can freely write *R* code directly in `Overleaf` by wrapping your code in `<<>>` and ending in `@`.

You should note that it is not possible to use external packages, as you do not have an actual working library. A possible workaround is to simply specify the functions you need manually by finding the relevant source code. However, it is rare that you actually need to compile more advanced `R`-code directly, rather than just include your results.

3.1 A working example

In the introduction, we thoroughly discussed why looking at wage determinants was interesting. Therefore, we decide to estimate the following:

$$\begin{aligned}\log(educ) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure \\ + \beta_4 nonwhite + \beta_5 female + \beta_6 married + u,\end{aligned}$$

where $u \sim i.i.d.N(0, \sigma^2)$. Before we can do this, we have to upload the relevant data to `Overleaf`. See the R script I provided in the appendix on how to save the data set `wage1` from the `wooldridge` package. Uploading it to `Overleaf` should be relatively intuitive. Once you have done this, we need to load the data set:

```
load("data.RData")
```

Now that we have loaded the data set, we may be interested in looking at some generic information on parts of the data set. For instance, what variables are included:

```
names(data)[1:6]

## [1] "wage"     "educ"     "exper"    "tenure"   "nonwhite" "female"
```

Also, it may be useful to know what types of random variables they are:

```
str(data[1:7])

## 'data.frame': 526 obs. of  7 variables:
##  $ wage    : num  3.1 3.24 3 6 5.3 ...
##  $ educ    : int  11 12 11 8 12 16 18 12 12 17 ...
##  $ exper   : int  2 22 2 44 7 9 15 5 26 22 ...
##  $ tenure  : int  0 2 0 28 2 8 7 3 4 21 ...
##  $ nonwhite: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ female  : int  1 1 0 0 0 0 0 1 1 0 ...
##  $ married : int  0 1 0 1 1 1 0 0 0 1 ...
```

4 Making tables and inserting figures

4.1 An important table

Now I want to make a table about the information I just gathered. Why? Well sometimes you just wanna make a table. The basic principle behind tables in LaTeXis that you should get an external program to make them. Mostly because this is an *extremely* tedious process. But for the purpose of this introduction, we will do it anyway. Personally, I use `stargazer` or something like tablesgenerator.com.

Table 1: My very first LaTeX table

Variable	Description	Type
`wage`	Wage in USD per hour	Numeric
`educ`	Years of education	Integer
⋮	⋮	⋮
⋮	⋮	⋮
`married`	Marital status	Binary

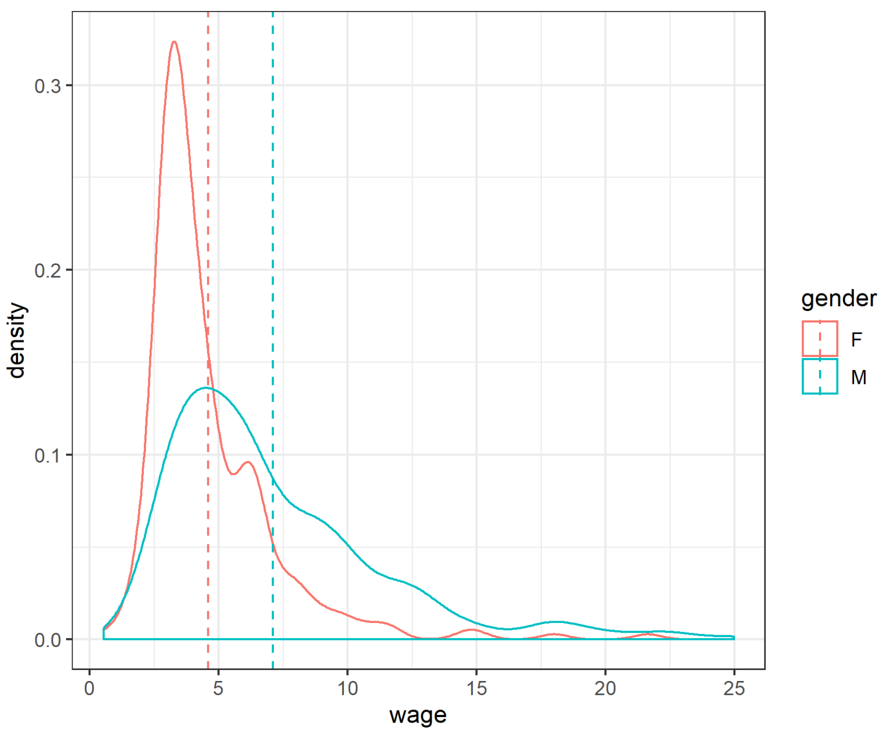
Once you get the hang of making these tables, it is pretty straight forward, but mostly annoying. The possibilities are almost endless for making very nice tables.

***Note**: There is a range of different options for the knitr package. A good place to start is the introduction Overleaf have made: www.overleaf.com/learn/latex/Knitr*

4.2 A brilliant plot

Lets suppose I made an awesome plot of wage and gender in `ggplot2` that I want to include. First, I need to upload the plot to `Overleaf`. And then I can just include it:

Figure 1: My nice figure



5 stargazer

Now I want to estimate my the model I specified:

```
wage_model <- lm(log(wage) ~ educ + exper + tenure +
                nonwhite + female + married,
                data=data)
```

Looking at the summary, I note that I should drop *exper* and *nonwhite*:

```
wage_model2 <- lm(log(wage) ~ educ + tenure +
                female + married,
                data=data)
summary(wage_model2)
```

```
##
## Call:
## lm(formula = log(wage) ~ educ + tenure + female + married, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.89826 -0.27087 -0.03548  0.24512  1.24687
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.570247   0.091707   6.218 1.03e-09 ***
## educ          0.079504   0.006576  12.090  < 2e-16 ***
## tenure        0.019441   0.002619   7.423 4.68e-13 ***
## female       -0.280364   0.037249  -7.527 2.30e-13 ***
## married       0.146801   0.038445   3.818  0.00015 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4134 on 521 degrees of freedom
## Multiple R-squared:  0.3996,Adjusted R-squared:  0.395
## F-statistic:  86.7 on 4 and 521 DF,  p-value: < 2.2e-16
```

Having estimated two models, I would now like to report my findings. The package `stargazer` (that most of you are probably familiar with) creates tables of estimation output for LaTeX natively:

```
stargazer(wage_model, wage_model2,
          covariate.labels= c("Intercept","Education", "Experience",
```

```
                              "Tenure", "Non-white",
                              "Female", "Married"),
          dep.var.labels = "log(wage) per hour",
          star.cutoffs = c(0.05, 0.01, 0.001),
          intercept.bottom=FALSE,
          single.row = TRUE,
          title="Estimation results")
```

If you copy the results in your *R*-console (when running the above code) into `Overleaf`, the following table should appear: Which was multitudes more efficient that creating it by

Table 2: Estimation results

	Dependent variable:	
	log(wage) per hour	
	(1)	(2)
Intercept	0.491*** (0.102)	0.570*** (0.092)
Education	0.084*** (0.007)	0.080*** (0.007)
Experience	0.003 (0.002)	
Tenure	0.017*** (0.003)	0.019*** (0.003)
Non-white	−0.003 (0.060)	
Female	−0.286*** (0.037)	−0.280*** (0.037)
Married	0.126** (0.040)	0.147*** (0.038)
Observations	526	526
R^2	0.404	0.400
Adjusted R^2	0.397	0.395
Residual Std. Error	0.413 (df = 519)	0.413 (df = 521)
F Statistic	58.540*** (df = 6; 519)	86.699*** (df = 4; 521)
Note:		*p<0.05; **p<0.01; ***p<0.001

hand. In general, there are many tools that allow for other programs to export tables into LaTeX. This is also possible with other statistics programs such as Stata, SAS, OxMetrics, and Python.

6 BibTeX

BibTeX is a reference handling tool built specifically for use with LaTeX. It looks and seems a bit intimidating at first, but as with most coding, it is rather simple once you get the hang of it. Almost all publishers provide a pre-made BibTeX reference that you can just steal. If not, you have to make one yourself. Say I want to reference Verbeeks's book on econometrics. This requires, that I create a new file in my `Overleaf` project with the extension name `.bib`. This tells `Overleaf` that it is working with a BibTeX file. In that file, you want to define your reference as:

```
@Book{Verbeek2017,
  title     = {A Guide to Modern Econometrics},
  publisher = {John Wiley \& Sons},
  year      = {2017},
  author    = {Verbeek, M.},
  address   = {Rotterdam School of Management, Erasmus University, Rotterdam},
  edition   = {Fifth},
}
```

Note that you can specify a range of different items using the @ command. For instance, `@Article` if it is a journal article you are referencing. It is important to remember the ”tag” you give each entry on your references. This is the `Verbeek2017` at the start of entry, which can be whatever you want. This is how you access your references when writing. So if I want to reference the book passively I can simply use the `\citep{Verbeek2017}` which will look like this: (Verbeek, 2017). The active equivalent is simply `\cite{Verbeek2017}`, which provides the following text output Verbeek (2017). Making a list of your references is then simply a matter of writing `\bibliography{references}`, where `references` should be the name of your `.bib` file. This will automatically update as you include more or less references. The preamble I provided uses standard author-year references, but you can specify this to whichever citation style you need. All in all, writing:

```
\bibliographystyle{apalike}
\bibliography{references}
```

will produce the desired output:

References

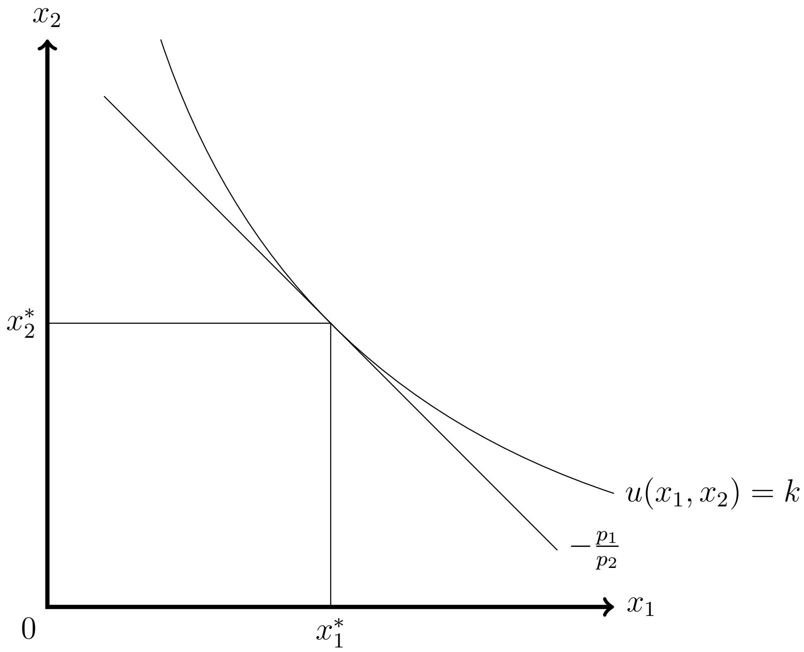
Verbeek, M. (2017). *A Guide to Modern Econometrics*. John Wiley & Sons, Rotterdam School of Management, Erasmus University, Rotterdam, fifth edition.

6.1 Important note on handling references

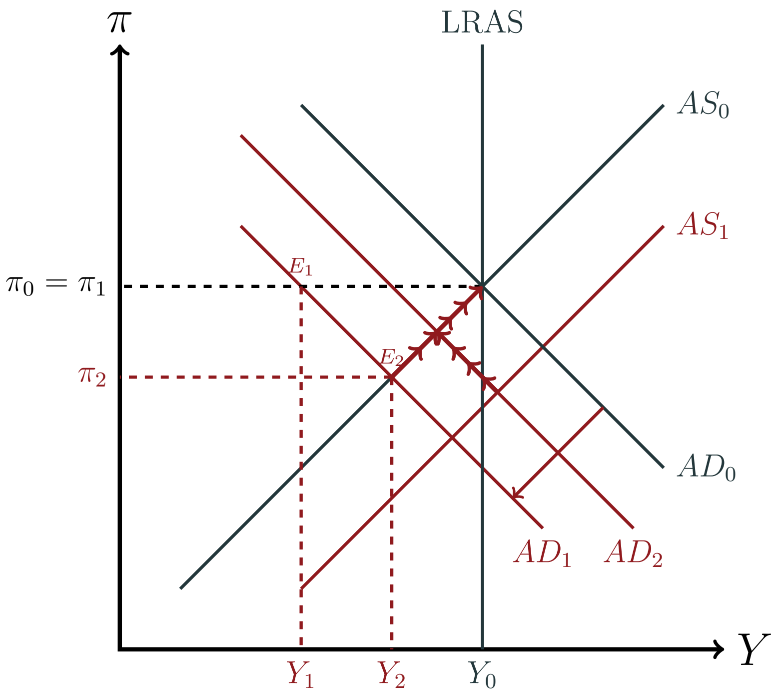
In practice, the by far easiest solution is using a reference management program, where both Zotero and Mendeley can produce `.bib` outputs. After producing your desired `.bib` file, all you have to do is upload it to `Overleaf`. JabRef is an example of a program dedicated exclusively to making BibTeX more applicable. Personally, I use Zotero.

7 TikZ (making your own figures)

Another ingenious part of LaTeXis the `TikZ` package that allows you to create your own completely customizable plots. This package is in no way whatsoever intuitive, but makes some extremely neat plots. And again, as with all coding, some nice person on the internet already made it for you. In the case of `TikZ` the best guide I have come across is the following: TikZ Cookbok. This allows you to make figures such as the textbook optimal choice:



Or slightly more advanced (read: confusing) figures:



The possibilities with `TikZ` are almost endless, as you define and control everything. In addition, it allows you to reproduce your results endlessly, which is a surprisingly large advantage - just like with *R*.

8 *R*-code

```
# New script
rm(list=ls())
# Working directory
setwd("")

# Installing relevant packages
install.packages("wooldridge") # Don't run this if already installed
install.packages("ggplot2")    # Don't run this if already installed
install.packages("plyr")       # Don't run this if already installed
install.packages("stargazer")  # Don't run this if already installed

#Retrieving them
library(ggplot2)
library(wooldridge)
library(plyr)
library(stargazer)

# We wanna use tha dataset "wage1" in wooldridge
data<-wage1

# Save it a place you can find it, as a .RData extension
save(data,file="data.RData")

# Some generic information about the data set
names(data)[1:7]
str(data[1:7])

# A nice plot about the data
data$gender <- "M"
data$gender[data$female==1] <- "F"
mu <- ddply(data, "gender", summarise, grp.mean=mean(wage))

wageplot <- ggplot(data, aes(x=wage, color=gender)) +
        geom_density() +
```

```r
        geom_vline(data=mu, aes(xintercept=grp.mean, color=gender),
                   linetype="dashed") +
        theme_bw()
wageplot

# Save the plot somewhere you can find it
ggsave("plot.png", plot=wageplot)

# Making a linear model
wage_model <- lm(log(wage) ~ educ + exper + tenure
                 + nonwhite + female + married,
                 data=data)

summary(wage_model)

# Making another linear model
wage_model2 <- lm(log(wage) ~ educ + tenure
                 + female + married,
                 data=data)

summary(wage_model2)

# Making a hot table for Overleaf
stargazer(wage_model, wage_model2,
          covariate.labels= c("Intercept","Education", "Experience",
                              "Tenure", "Non-white",
                              "Female", "Married"),
          dep.var.labels = "log(wage) per hour",
          star.cutoffs = c(0.05, 0.01, 0.001),
          intercept.bottom=FALSE,
          single.row = TRUE,
          title="Estimation results")
```