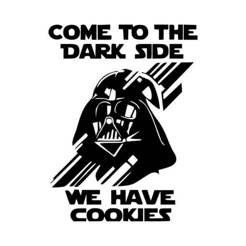
LaTeX for Economics and Business Administration

Thomas de Graaff

January 13, 2020



Introduction

Why this workshop?

- In the social sciences few attention to what tools to use (and why)
  - you just use what colleagues, friends or teachers used
  - huge fixed (and sometimes sunk) costs

Why this workshop?

- In the social sciences few attention to what tools to use (and why)
    - you just use what colleagues, friends or teachers used
    - huge fixed (and sometimes sunk) costs
- Increasing use of LaTeX
    - more user friendly (editors, online environments)
    - combination with markdown used on internet/blogs
    - tight connection with (statistical) software (`R`/`Python`/`Stata`)
    - combination with data science

What I want (and don't want) with this workshop

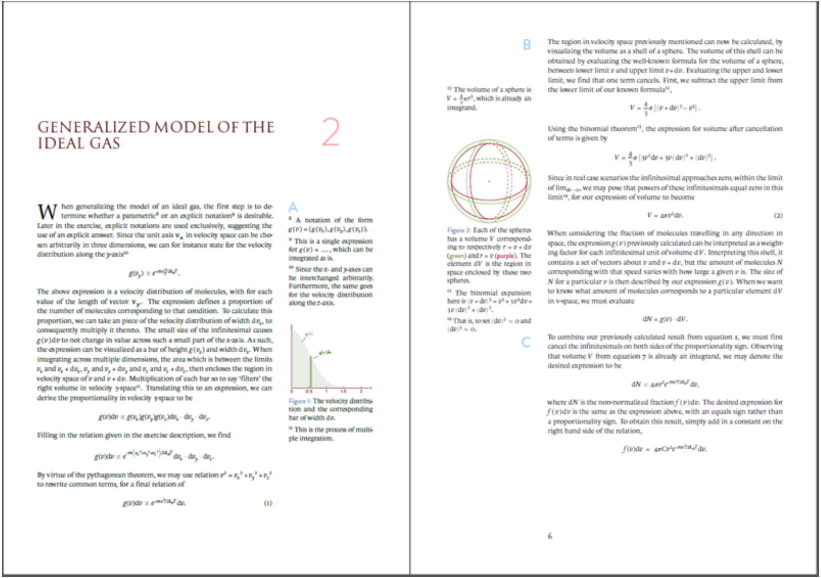
- Give a general introduction of why some tools work together
    - LaTeX
    - reference managers
    - (statistical) output
- Give an introduction to LaTeX
    - First vanilla basics (including references)
    - Next workshop: more advanced stuff
- What *I* do not want
    - Tell you what applications to use (you need to decide and make a well-informed decision)

Background

- TEX created by Donald Knuth (70's)
- LATEX is a set of macro's around TeX (1986)
- LATEX is a typesetting program, not a Word processor
    - So edit code that needs to be compiled
    - Editors
        - specific: TeXstudio, TeXshop, Rstudio
        - general: Sublime, Atom, Vim, Emacs

"Beware of bugs in the above code; I have only proved it correct, not tried it."

Showcase: Tufte lay-out style



GENERALIZED MODEL OF THE IDEAL GAS

2

A

When generalizing the model of an ideal gas, the first step is to determine whether a parametric[8] or an explicit notation[9] is desirable. Later in the exercise, explicit notations are used exclusively, suggesting the use of an explicit answer. Since the unit axis $\mathbf{v}_x$ in velocity space can be chosen arbitrarily in three dimensions, we can for instance state for the velocity distribution along the y-axis[10]

$$g(v_y) \propto e^{-mv_y^2/2k_BT}.$$

The above expression is a velocity distribution of molecules, with for each value of the length of vector $\mathbf{v}_y$. The expression defines a proportion of the number of molecules corresponding to that condition. To calculate this proportion, we can take an piece of the velocity distribution of width dv_x, to consequently multiply it thereto. The small size of the infinitesimal causes $g(v)dv$ to not change in value across such a small part of the x-axis. As such, the expression can be visualized as a bar of height $g(v_x)$ and width dv_x. When integrating across multiple dimensions, the area which is between the limits v_x and $v_x + dv_x$, v_y and $v_y + dv_y$ and v_z and $v_z + dv_z$, then encloses the region in velocity space of v and $v + dv$. Multiplication of each bar so to say 'filters' the right volume in velocity 3-space[11]. Translating this to an expression, we can derive the proportionality in velocity 3-space to be

$$g(v)dv \propto g(v_x)g(v_y)g(v_z)dv_x \cdot dv_y \cdot dv_z.$$

Filling in the relation given in the exercise description, we find

$$g(v)dv \propto e^{-m\left(v_x^2+v_y^2+v_z^2\right)/2k_BT}dv_x \cdot dv_y \cdot dv_z.$$

By virtue of the pythagorean theorem, we may use relation $v^2 = v_x^2 + v_y^2 + v_z^2$ to rewrite common terms, for a final relation of

$$g(v)dv \propto e^{-mv^2/2k_BT}dv. \qquad (1)$$

[8] A notation of the form $g(v) = (g(v_x), g(v_y), g(v_z))$.

[9] This is a single expression for $g(v) = \ldots$, which can be integrated as is.

[10] Since the x- and y-axes can be interchanged arbitrarily. Furthermore, the same goes for the velocity distribution along the z-axis.

Figure 1: The velocity distribution and the corresponding bar of width dv.

[11] This is the process of multiple integration.

B

The region in velocity space previously mentioned can now be calculated, by visualizing the volume as a shell of a sphere. The volume of this shell can be obtained by evaluating the well-known formula for the volume of a sphere, between lower limit v and upper limit $v+dv$. Evaluating the upper and lower limit, we find that one term cancels. First, we subtract the upper limit from the lower limit of our known formula[12],

$$V = \frac{4}{3}\pi\left[(v+dv)^3 - v^3\right].$$

Using the binomial theorem[13], the expression for volume after cancellation of terms is given by

$$V = \frac{4}{3}\pi\left[3v^2dv + 3v(dv)^2 + (dv)^3\right].$$

Since in real case scenarios the infinitesimal approaches zero, within the limit of $\lim_{dv\to 0}$, we may pose that powers of these infinitesimals equal zero in this limit[14], for our expression of volume to become

$$V = 4\pi v^2 dv. \qquad (2)$$

When considering the fraction of molecules travelling in any direction in space, the expression $g(v)$ previously calculated can be interpreted as a weighting factor for each infinitesimal unit of volume dV. Interpreting this shell, it contains a set of vectors about v and $v + dv$, but the amount of molecules N corresponding with that speed varies with how large a given v is. The size of N for a particular v is then described by our expression $g(v)$. When we want to know what amount of molecules corresponds to a particular element dV in v-space, we must evaluate

$$dN = g(v) \cdot dV.$$

C

To combine our previously calculated result from equation 1, we must first cancel the infinitesimals on both sides of the proportionality sign. Observing that volume V from equation 7 is already an integrand, we may denote the desired expression to be

$$dN \propto 4\pi v^2 e^{-mv^2/2k_BT}dv,$$

where dN is the non-normalized fraction $f(v)dv$. The desired expression for $f(v)dv$ is the same as the expression above, with an equals sign rather than a proportionality sign. To obtain this result, simply add in a constant on the right hand side of the relation,

$$f(v)dv = 4\pi C v^2 e^{-mv^2/2k_BT}dv.$$

[12] The volume of a sphere is $V = \frac{4}{3}\pi r^3$, which is already an integrand.

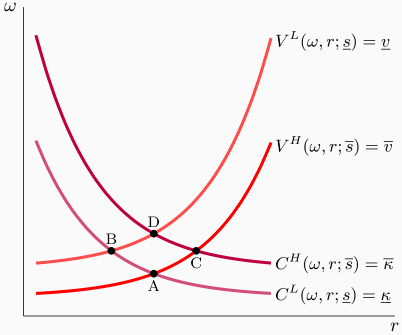
Figure 2: Each of the spheres has a volume V corresponding to respectively $r = v + dv$ (green) and $r = v$ (purple). The element dV is the region in space enclosed by these two spheres.

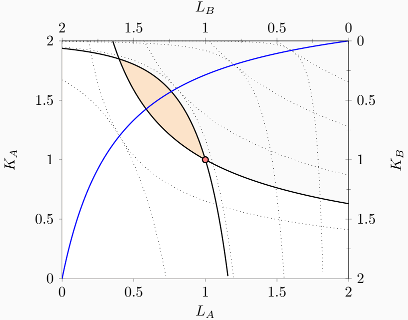
[13] The binomial expansion here is $(v+dv)^3 = v^3 + 3v^2dv + 3v(dv)^2 + (dv)^3$.

[14] That is, to set $(dv)^2 \approx 0$ and $(dv)^3 \approx 0$.

6

Showcase: tikz and PGFPlots





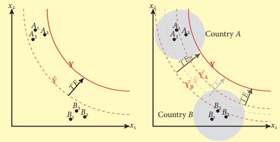
Showcase: posters

Stochastic frontier models with spatial dependence

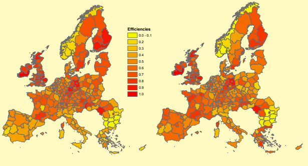
Thomas de Graaff tgraaff@feweb.vu.nl VU University Amsterdam & Netherlands Environmental Assessment Agency

The problem and research aim

Estimation of technical efficiencies may be biased in the presence of spatial dependence or unobserved spatial heterogeneity amongst regions. The aim is therefore to simultaneously model and consistently estimate a model that incorporates both technical inefficiencies and spatial dependence.



Technical inefficiencies in Europe's manufacturing: an application



Left: Standard technical efficiencies *Right:* with additional spatial lag

Stochastic production frontiers

Assume that regional production, y, can be modeled as:

$$y = f(\mathbf{X}; \beta^T) TE,$$

where $\mathbf{X}$ are regional production factors, β the parameters of the production function and TE is the regional specific technical efficiency. By assuming a Cobb-Douglas and that $TE = \exp(-u)$, we get:

$$\ln y = \ln(\mathbf{X})\beta - u + v,$$

Using skew-normal distributions

Let u and v be distributed as:

$$\begin{pmatrix} u \\ v \end{pmatrix} \sim N(0, \Omega^*), \qquad \Omega^* = \begin{pmatrix} 1 & -\delta^T \\ -\delta & \Omega \end{pmatrix}$$

We are interested in $\xi = \ln y - \ln \mathbf{X} = \Pr(v|u < 0)$ (via conditioning): leading to $\ln y \sim \mathrm{SN}(\ln(\mathbf{X})\beta, \Omega, \alpha)$; a multivariate skew-normal distribution with:

$$f(\ln y) = 2\phi(\ln y - \ln(\mathbf{X})\beta; \Omega)\Phi\left(\frac{\alpha}{\omega}(\ln y - \ln(\mathbf{X})\beta\right)$$

- ω is a scale parameter
- α is a measure of skewness
- $\alpha = \left(1 - \delta^T \Omega^{-1} \delta\right)^{-1/2} \Omega^{-1}\delta$

Introducing a spatial lag

Because multivariate skew-normal distributions are closed under affine transformations (similarly to normal distributions), we may write:

$$\mathbf{B} \ln(y) = \ln(\mathbf{X})\beta + \xi \qquad (1)$$

where $\mathbf{B} = (\mathbf{I} - \rho \mathbf{W})$ and with ξ again a multivariate skew normal distribution with $\Omega = \omega(\mathbf{B}'\mathbf{B})^{-1})$. This leads to the following loglikelihood:

$$-\frac{n}{2}\ln(\pi\omega^2) + \ln|\mathbf{B}| - \frac{e'e}{2\omega^2} + \sum \ln 2\Phi\left(\frac{\alpha}{\omega}e\right)$$

where e is the vector of residuals of model (1).

Finding technical inefficiencies

We need to find $TE = \exp(u)$ or $\mathbb{E}(u|\xi)$ given that $u < 0$. Because we can write as well $\xi = \delta|u| + \sqrt{(1-\delta^2)}v$ with $\delta < 0$ (via convolution), where $u \sim N(0,1)$ and $v \sim N(0, \Omega)$, the following general expression holds:

$$u|\xi \sim N^c\left(\left(\mathbf{D}'\Sigma^{-1}\mathbf{D} + \mathbf{I}\right)^{-1}\mathbf{D}'\Sigma^{-1}e, \left(\mathbf{D}'\Sigma^{-1}\mathbf{D} + \mathbf{I}\right)^{-1}\right)$$

where N^c indicates a normal distribution truncated at 0, $\mathbf{D}$ is a diagonal matrix with δ's on the diagonal and Σ equals $\sqrt{(\mathbf{I} - \mathbf{D}^2)}\Omega$. The expectation can now be readily derived.

Empirical specification

We estimate for the period 1991–2008 a neoclassical growth model of the manufacturing sector across 273 European NUTS-2 regions with the following specification:

$$\ln\frac{y(t)}{y(0)} = \beta_0 + \beta_1 \ln y(0) + \beta_2 \ln s + \beta_3 \ln(n + 0.05) + \xi$$

where s is the savings rate and $y(t)$ the GVA in manufacturing measured at time t, n the manufacturing working population growth rate, ξ is skew-normally distributed and the convergence rate across regions is calculated as: $\hat{\lambda} = -\ln(1 + \hat{\beta}_1)$.

Estimation results

Variable	Growth model	Frontier model	Spatial frontier
Constant	$0.30^{\dagger}$	0.68^{**}	0.07
$\ln(y_0)$	-0.50^{**}	-0.47^{**}	-0.41^{**}
$\ln s$	0.55^{**}	0.51^{**}	0.45^{**}
$\ln(n+g+\delta)$	-0.19^{**}	-0.18^{**}	-0.18^{**}
ω	0.30^{**}	0.44^{**}	0.36^{**}
α		-2.91^{**}	-2.57^{**}
ρ			0.91^{**}
Logl.	-0.21	0.21	0.36
$\hat{\lambda}$	0.053^{**}	0.049^{**}	0.041^{**}

Significance levels : † : 10% * : 5% ** : 1%

Conclusions

1. Spatial dependence and stochastic frontiers can be simulteneously and consistently estimated using multivariate skew-normal distribution functions
2. In the presence of spatial dependence, regional technical inefficiency differences can be significantly mitigated

Key references

[1] Abreu, M. *Spatial Determinants of Economic Growth and Technology Diffusion*. Thela Thesis Publishers, Amsterdam, 2005.

[2] Aigner, D. J., Lovell, C. A. K., and Schmidt, P. Formulation and Estimation of Stochastic Production Frontier Models. *Journal of Econometrics 6* (1977), 21–37.

[3] Azzalini, A., and Capitanio, A. Statistical Application of the Multivariate Skew-Normal distribution. *Journal of Royal Statistical Society 61* (1999), 579–602.

[4] Dominguez-Molina, J. A., Gonzalez-Farias, G., and Ramos-Quiroga, R. Skew-normality in stochastic frontier analysis. In *Skew-Elliptical distributions and their applications*, M. G. Genton, Ed. Chapman & Hall/CRC, 2004, ch. 13, pp. 223–242.

Disadvantages

- not WYSIWYG

Disadvantages

- not WYSIWYG
- you nead to learn (quite) some commands
  - Learning curve, but
  - hurray for cheat sheets and Google

Disadvantages

- not WYSIWYG
- you nead to learn (quite) some commands
  - Learning curve, but
  - hurray for cheat sheets and Google
- Difficult to cooperate with people from the other side

Disadvantages

- not WYSIWYG
- you nead to learn (quite) some commands
  - Learning curve, but
  - hurray for cheat sheets and Google
- Difficult to cooperate with people from the other side
- Basic LaTeX has difficulties with incorporating new fonts (Hoefler, minion pro)
  - XeTeX
  - For the purists: LaTeX does it right (LaTeX vs Word)

Disadvantages

- not WYSIWYG
- you nead to learn (quite) some commands
  - Learning curve, but
  - hurray for cheat sheets and Google
- Difficult to cooperate with people from the other side
- Basic LaTeX has difficulties with incorporating new fonts (Hoefler, minion pro)
  - XeTeX
  - For the purists: LaTeX does it right (LaTeX vs Word)
- Difficult to create unstructured and ugly documents

Advantages

- free (as in beer & in speach)

Advantages

- free (as in beer & in speach)
- WYSIWYM

Advantages

- free (as in beer & in speach)
- WYSIWYM
- consistent lay-out throughout the whole document (including tables, appendices, formulas, source code, etc)

Advantages

- free (as in beer & in speach)
- WYSIWYM
- consistent lay-out throughout the whole document (including tables, appendices, formulas, source code, etc)
- internal references are a breeze (citations, ToC, ToT . . .)

Advantages

- free (as in beer & in speach)
- WYSIWYM
- consistent lay-out throughout the whole document (including tables, appendices, formulas, source code, etc)
- internal references are a breeze (citations, ToC, ToT . . .)
- forced to structure documents

Advantages

- free (as in beer & in speach)
- WYSIWYM
- consistent lay-out throughout the whole document (including tables, appendices, formulas, source code, etc)
- internal references are a breeze (citations, ToC, ToT . . .)
- forced to structure documents
- macros around plain text, thus scriptable

Advantages

- free (as in beer & in speach)
- WYSIWYM
- consistent lay-out throughout the whole document (including tables, appendices, formulas, source code, etc)
- internal references are a breeze (citations, ToC, ToT ...)
- forced to structure documents
- macros around plain text, thus scriptable
- large community, thus a package for almost everything (books, articles, presentation, posters, exams, musicscores)

Advantages

- free (as in beer & in speach)
- WYSIWYM
- consistent lay-out throughout the whole document (including tables, appendices, formulas, source code, etc)
- internal references are a breeze (citations, ToC, ToT ...)
- forced to structure documents
- macros around plain text, thus scriptable
- large community, thus a package for almost everything (books, articles, presentation, posters, exams, musicscores)
- superior typography & output

Advantages

- free (as in beer & in speach)
- WYSIWYM
- consistent lay-out throughout the whole document (including tables, appendices, formulas, source code, etc)
- internal references are a breeze (citations, ToC, ToT ...)
- forced to structure documents
- macros around plain text, thus scriptable
- large community, thus a package for almost everything (books, articles, presentation, posters, exams, musicscores)
- superior typography & output
- many free LaTeX templates

LaTeX versus Markdown

- markdown:
    - lightweight markup language that can export to `.doc`, `.html`, and `.pdf`.
- much easier then LaTeX but less flexible
- used by writers/blogs even for complete websites
- interaction with LaTeX; if not only for formula's

Some thoughts on overleaf

pro

- great learning environment
- track changes
- rich text
- cooperation
- great documentation

Some thoughts on overleaf

pro

- great learning environment
- track changes
- rich text
- cooperation
- great documentation

con

- you need to be online
- proprietary software
- standalone
- backing-up

How does LaTeX work in practice?

- You edit a `.tex` file without thinking about how it looks
  - distraction free writing (yeah right)
- You then compile it
  - LaTeX is unforgiving: if there is an error, usually it does not compile
  - Typically, errors are missing brackets or parentheses.
- Typically, source `.tex` file is compiled into `.pdf` and many other (auxiliary) files

TeXstudio

TeXstudio: A quick tour

- Preferences
- Keyboard shortcuts
- LaTeX dropdown menu

Exercises

First: organize!

1. Create a specific workshop folder somewhere where you can find it.

2. Think about versioning system and a back-up system

3. E.g.: use dropbox and/or Time Machine

Exercise 1: Open from template and fill in!

```
\documentclass[]{article}
%opening
\title{}
\author{}

\begin{document}

\maketitle

\begin{abstract}

\end{abstract}

\section{}

\end{document}
```

OK; and now what?

1. Save your file in your folder (give it an appropriate name)
2. Press `F1` (or `F5`)
3. The editor now sends LaTeX the message that it should *compile* your file
4. LaTeX creates many new files

Exercise 2: Create a paper structure

```
\section{}
\subsection{}
\subsubsection{}

```

Note that the following are used for books

```
\part{}
\chapter{}
```

And for bigger projects:

```
\include{}
\input{}
```

Intermezzo: preamble

Part before `\begin document` is called preamble

```
\documentclass[]{article}

% This is where packages are loaded
% and specific commands are given that
% determine how the lay-out and desing
% of your document will look like
% including: references, tables,
% paragraphs, headers, etc.

\usepackage{graphicx}

\begin{document}
```

Intermezzo: white spaces and special characters

An empty line starts a new paragraph and consecutive white spaces are treated as one

```
One paragraph

Second     paragraph (just one white space)
```

The following characters are reserved # $ % ^ & _ { } ~ \ and should be used as follows

```
  \# \$ \% \^ \& \_ \{ \} \~ \\
```

So, with a backslash before except for the backslash (does this make sense?)

Exercise 3: Create a table of contents

More complex text structures are relatively easy, just insert (after `\begin document`)

```
\tableofcontents
\listoffigures
\listoftables
```

Lists

- Itemization

```
\begin{itemize}
        \item blue
        \item red
\end{itemize}
```

- Enumeration

```
\begin{enumerate}
        \item first item
        \item second item
\end{enumerate}
```

Exercise 4: Lists

Create the following mode choice list in your `.tex` document

1. Cycling
2. Walking
3. Driving
4. Public transport
    - Bus
    - Tram
    - Metro
    - Train

Further text control

- Bold

```
\textbf{bold}
```

- Emphasize

```
\textit{italics} or \emph{emphasized}
```

Formula's

Inline math `$ $`; displayed math `$$ $$`; for example:

```
$x^2$
$x_2$
$\sqrt{x}$
$$Y = K^\alpha L^{1-\alpha}$$
$$\sum_{i=1}^I$$
$$\frac{\partial x}{\partial y}$$
\begin{equation}
        E = mc^2
\end{equation}
```

Exercise 5: Create these formula's

1. Regression formula:

$$y_i = \alpha + \beta x_i + \epsilon_i$$

2. The mean

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

3. Optimal economic order quantity:

$$Q^* = \sqrt{\frac{2DK}{h}}$$

Figures

Figures/graphs and tables in a floating environment

```
\begin{figure}[h!]}
\center
        \includegraphics{ligatures_latex}
        \caption{A figures about ligatures}
        \label{fig:ligatures}
\end{figure}
```

Figures can be `.pdf`, `.jpg`, `.png` and a whole lot of other types (but not bitmaps!)

Tables

```
\begin{table}[t!]
        \caption{This is the caption}
        \begin{tabular}{|l|c|r|}
                \hline
                first & row & data \\
                second & row & data \\
                \hline
        \end{tabular}
        \label{tab:example}
\end{table}
```

Referencing

Internal references are a breeze

```
\label{}          % Label something
\ref{}         % Refer to that
\footnote{}         % Add footnote
\thanks{}          % For in title
```

Exercise 6: Create a table

Create the following table

Table 1: Average grades

First name	Surname	Grade
Sherlock	Holmes	7.9
John H.	Watson	8.1

And refer to it in text as such:

> *Table 1 gives the average grades for course solving crimes.*

BibTeX

Literature references (at the end)

```latex
\cite{} % cite something
% Now tell LaTeX where to find references
\bibliography{references.bib}
% and which citation style to use
\bibliographystyle{apalike}
```

Later, we dive into how to make this look good

Exercise 7: References

1. Search on Google Scholar for three references from Erik Verhoef and/or Wout Dullaert
2. Put those in a `.bib` file in the **same** directory as your `.tex` file
3. Refer to those in your `.tex` file
4. Create the reference list

Conclusion

Next workshop

- Use of packages
- Making things look better!
- Graphs
- Better tables with `Stata` and `R` output
- Slides