

CHRISTIAN MICHELS  
NIELS BOHR INSTITUTE  
UNIVERSITY OF COPENHAGEN

A PHYSICIST'S  
APPROACH TO  
MACHINE LEARNING  
—  
UNDERSTANDING  
THE BASIC BRICKS

SUPERVISOR:  
TROELS PETERSEN  
NIELS BOHR INSTITUTE  
UNIVERSITY OF COPENHAGEN

Copyright © 2019

Christian Michelsen

[HTTPS:/ / GITHUB.COM / CHRISTIANMICHELSEN](https://github.com/CHRISTIANMICHELSEN)

Licensed under the Apache License, Version 2.0 (the “License”); you may not use this file except in compliance with the License. You may obtain a copy of the License at <http://www.apache.org/licenses/LICENSE-2.0>. Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS” BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

*First printing, November 2019*

# *Contents*

1	<i>Abstract</i>	1
2	<i>Introduction</i>	3
3	<i>Housing Prices Analysis</i>	7
4	<i>Particle Physics and LEP</i>	17
5	<i>Quark Gluon Analysis</i>	19
6	<i>Discussion and Outlook</i>	33
7	<i>Conclusion</i>	35
	<i>Bibliography</i>	39
	<i>Index</i>	41



# *List of Figures*

2.1	Illustration of the empirical risk as a function of model complexity.	4
2.2	Comparison of different objective functions.	4
2.3	Comparison of different objective functions zoom in.	4
2.4	My TikZ picture 1	5
2.5	My TikZ picture 2	5
2.6	Feynman diagram for the jet production at LEP	5
3.1	Geographic overview of square meter price in Denmark	7
3.2	Histogram of prices of houses and apartments sold in Denmark	7
3.3	Input paramater distributions for the housing prices dataset	8
3.4	Percentage of valid counts for each variable	8
3.5	Linear correlation between variables and price	8
3.6	Non-linear correlation between variables and price	9
3.7	Registration of property	9
3.8	Prophet Forecast for apartments	9
3.9	Prophet Trends	9
3.10	Overview of initial hyperparamater optimization of the housing model for apartments	10
3.11	Performance of LGB-model on apartment prices	10
3.12	Feature importance of apartments prices using LGB	11
3.13	Total feature importance of apartment prices using LGB	12
3.14	SHAP Prediction Explanation for apartment	12
3.15	Hyperparameter optimization: random search results	12
3.16	Hyperparameter optimization: Bayesian optimization results	12
3.17	Early Stopping results	13
3.18	Comparison of different objective functions	13
3.19	Comparison of different half times weights	13
3.20	2018 LGB Forecast	14
5.1	Histograms of the vertex variables	19
5.2	UMAP vizualisation of vertex variables	20
5.3	b-tag scores in 3-jet events	20
5.4	ROC curve for b-tag in 4-jet events	20
5.5	g-tag scores in 4-jet events	21
5.6	g-tag scores in 4-jet events for signal and background	21
5.7	ROC curve for g-tag in 4-jet events	21
5.8	1D Sum Model Cuts for 4-jets	22
5.9	1D Sum Models Predictions and Signal Fraction for 4-jets	22

5.10 Hyperparameter Optimization of b- and g-tagging	22
5.11 Overview of Hyperparamaters of g-tagging for 3-jet shuffled events	22
5.12 SHAP Prediction Explanation for b-like jet	23
5.13 Monte Carlo – Data bias for b-tags and jet energy	23
5.14 b-Tagging Efficiency $\varepsilon_b^{b-\text{sig}}$ as a function of jet energy	23
5.15 b-Tagging Efficiency $\varepsilon_b^{g-\text{sig}}$ as a function of jet energy	23
5.16 b-Tagging Efficiency $\varepsilon_g^{g-\text{sig}}$ as a function of jet energy	24
5.17 b-Tagging Efficiency $\varepsilon_g^{b-\text{sig}}$ as a function of jet energy	24
5.18 g-Tagging proxy efficiency for $b\bar{b}g$ -events as function of the mean invariant mass	24
5.19 g-Tagging proxy efficiency for $b\bar{b}g$ -events as function of g-tag	24
5.20 g-Tagging efficiency for 4-jet events in MC as a function of normalized gluon gluon jet energy difference	25
5.21 Closure plot between MC Truth and the corrected g-tagging model in 4-jet events for the normalized gluon gluon jet energy difference	25
5.22 R kt CA overview XXX <b>TODO!</b>	25
5.23 R kt CA cut region A XXX <b>TODO!</b>	25
5.24 This is a margin figure. The helix is defined by $x = \cos(2\pi z)$ , $y = \sin(2\pi z)$ , and $z = [0, 2.7]$ . The figure was drawn using Asymptote ( <a href="http://asymptote.sourceforge.net/">http://asymptote.sourceforge.net/</a> ).	26
5.25 This graph shows $y = \sin x$ from about $x = [-10, 10]$ . Notice that this figure takes up the full page width.	27
5.26 Hilbert curves of various degrees $n$ .	27

## *List of Tables*

3.1	RMSE.	10
3.2	LogCosh.	10
3.3	Heading styles used in <i>Beautiful Evidence</i> .	15
5.1	Here are the dimensions of the various margins used in the Tufte-handout class.	27



## *1. Abstract*

This sample book discusses the design of Edward Tufte's books and the use of the `tufte-book` and `tufte-handout` document classes.



## 2. Introduction

*“Begin at the beginning,” the King said, gravely, “and go on till you come to an end; then stop.”*

---

— Lewis Carroll, *Alice in Wonderland*

NOT ONLY is the title of this project fairly broad, so are the subjects covered in this thesis. The overall goal of this project is to apply machine learning to different datasets and see how well these comparatively new tools might improve classical statistical methods. The project have dealt with two (seemingly) very different datasets: Danish housing prices and Quark-Gluon discrimination in particle physics, and the aim of this section is to provide an initial overview of the scope and relationship of the two sub-projects; two sub-projects which are covered in each part of this book.

The first part of the thesis deals with the problem of estimating housing prices as precisely and accurately as possible. This was the sub-project that was worked on in the beginning of the overall project and worked as an initial introduction to the application of machine learning to real-life datasets. The housing prices dataset thus became the playground in which the subtleties of these new modern tools were examined, where the difference between real life datasets with all its quirks, outliers and bad formatting, and curated toy datasets that works out of the box (such as the famous Iris dataset<sup>1</sup>) were experienced first hand. Since the project started the dataset changed due to a new collaboration with the Danish housing agency [Boligsiden](#) where the agreement was, stated shortly, that we would get their data and they would get our results. Boligsiden is a natural collaborator since they are the biggest on the market<sup>2</sup> and have been very helpful the continuos process of providing data but it also should also be noted that they have had no say on the results presented in this thesis.

Notation:

$\mathcal{X}$

$\mathcal{Y}$

$X \in \mathcal{X}$

$(X, Y) \in (\mathcal{X} \times \mathcal{Y})$

$S = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$

$h : \mathcal{X} \rightarrow \mathcal{Y}$

$\mathcal{H} = \{h_1, h_2, \dots, h_M\}$

<sup>1</sup> Edgar Anderson. The Species Problem in Iris. *23(3):457–509*. ISSN 00266493. DOI: [10.2307/2394164](https://doi.org/10.2307/2394164). URL [www.jstor.org/stable/2394164](http://www.jstor.org/stable/2394164); and R. A. Fisher. The Use of Multiple Measurements in Taxonomic Problems. *7(2):179–188*. ISSN 2050-1439. DOI: [10.1111/j.1469-1809.1936.tb02137.x](https://doi.org/10.1111/j.1469-1809.1936.tb02137.x). URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-1809.1936.tb02137.x>

<sup>2</sup> Due to being owned by the “Dansk Ejendomsmæglerforening”, The Danish Association of Chartered Estate Agents

$$\ell(Y', Y)$$

$$\hat{L}(h, S) = \frac{1}{2} \sum_{i=1}^n \ell(h(X_i, Y_i))$$

$$L(h) = \mathbb{E} [\ell(h(X), Y)]$$

$$\mathbb{P} [|\nu - \mu| > \epsilon] \leq 2e^{-2\epsilon^2 N} \text{ for any } \epsilon > 0$$

$$\mathbb{P} [|E_{in}(g) - E_{out}(g)| > \epsilon] \leq 2Me^{-2\epsilon^2 N} \text{ for any } \epsilon > 0$$

$$E_{out}(g) \leq E_{in}(g) + \sqrt{\frac{1}{2N} \ln \frac{2M}{\delta}}$$

$$E_{out}(g) \leq E_{in}(g) + \sqrt{\frac{8}{2N} \ln \frac{4m_H(2N)}{\delta}}$$

$$E_{out}(g^{(\mathcal{D})}) = \mathbb{E}_{\mathbf{x}} \left[ \left( g^{(\mathcal{D})}(\mathbf{x}) - f(\mathbf{x}) \right)^2 \right]$$

$$\mathbb{E}_{\mathcal{D}} \left[ E_{out} \left( g^{\mathcal{D}} \right) \right] = \mathbb{E}_{\mathbf{x}} \left[ \underbrace{\mathbb{E}_{\mathcal{D}} \left[ g^{\mathcal{D}}(\mathbf{x}) \right] - \bar{g}(\mathbf{x})^2}_{\mathbb{E}_{\mathcal{D}}[g^{\mathcal{D}}(\mathbf{x})] - \bar{g}(\mathbf{x})^2} + \underbrace{\bar{g}(\mathbf{x})^2 - 2\bar{g}(\mathbf{x})f(\mathbf{x}) + f(\mathbf{x})^2}_{(\bar{g}(\mathbf{x}) - f(\mathbf{x}))^2} \right]$$

$$\prod_{n=1}^N P(y_n|x_n)$$

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(M-|S|-1)!}{M!} [f_x(S \cup \{i\}) - f_x(S)]$$

$$g(z') = \phi_0 + \sum_{i=1}^M \phi_i z'_i$$

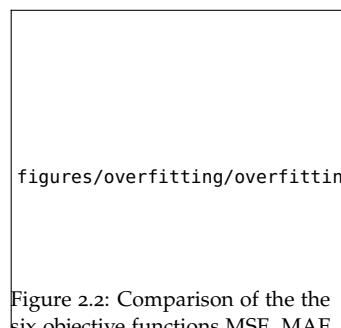
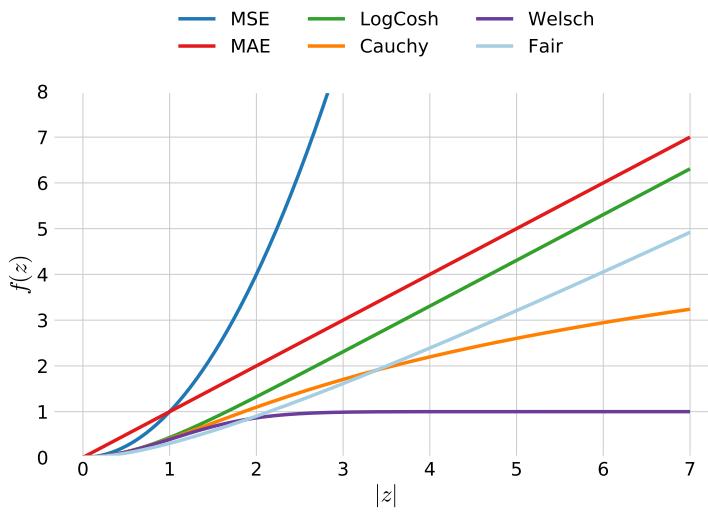


Figure 2.2: Comparison of the the six objective functions MSE, MAE,

LogCosh, Cauchy, Welsch, and Fair. Figure 2.4: Illustration of the empirical risk as a function of  $|z|$ . In the plot **MSE** is shown in blue, **MAE** in red, **LogCosh** in green, **Cauchy** in orange, **Welsch** in purple, and **Fair** in light blue. For the Cauchy and Welsch functions  $C$  is set to 1 and for the Fair function  $c$  is set to 1. For a zoom in of the inner region where  $|z| < 2$  see Figure 2.3.

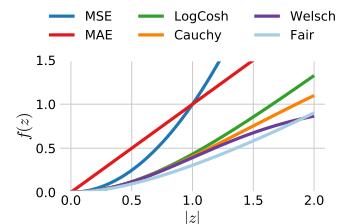


Figure 2.3: Zoom in of Figure 2.2.

Figure 2.4: My TikZ picture 1

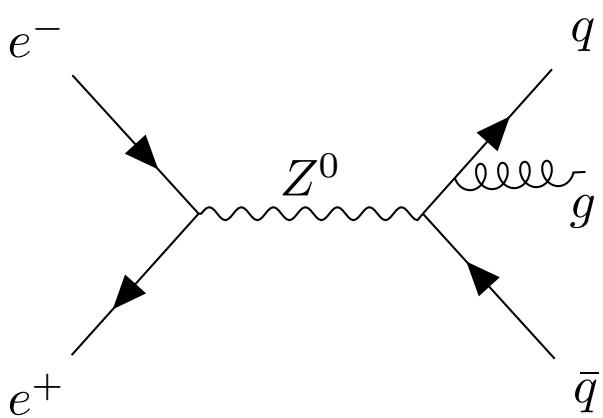


Figure 2.5: My TikZ picture 2  
Figure 2.6: Feynman diagram showing the  $e^+e^- \rightarrow Z^0$  production at LEP. The  $Z^0$  has several decay modes where the  $Z \rightarrow q\bar{q}g$  is shown here.



### 3. Housing Prices Analysis

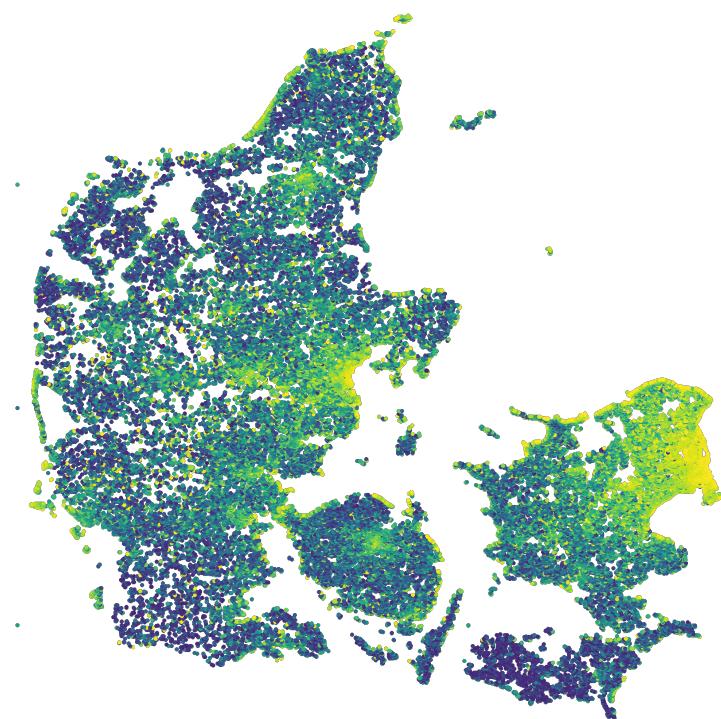


Figure 3.1: Geographic overview of square meter prices for houses and apartments in Denmark (excluding Bornholm for visual purposes). Notice the strong correlation with the major cities and the shore line. Also notice the three outliers west of Jutland.

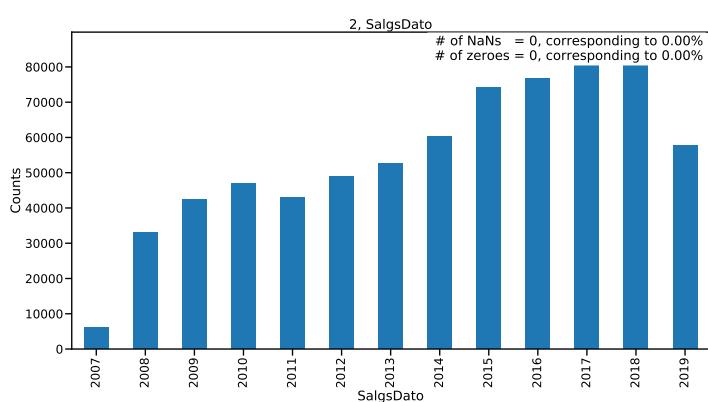


Figure 3.2: Histogram of prices of houses and apartments sold in Denmark.

Outside referencing isn't different: In fig. 3.3: Two images — fig. (a) shows an "A", fig. (b) shows a "B".

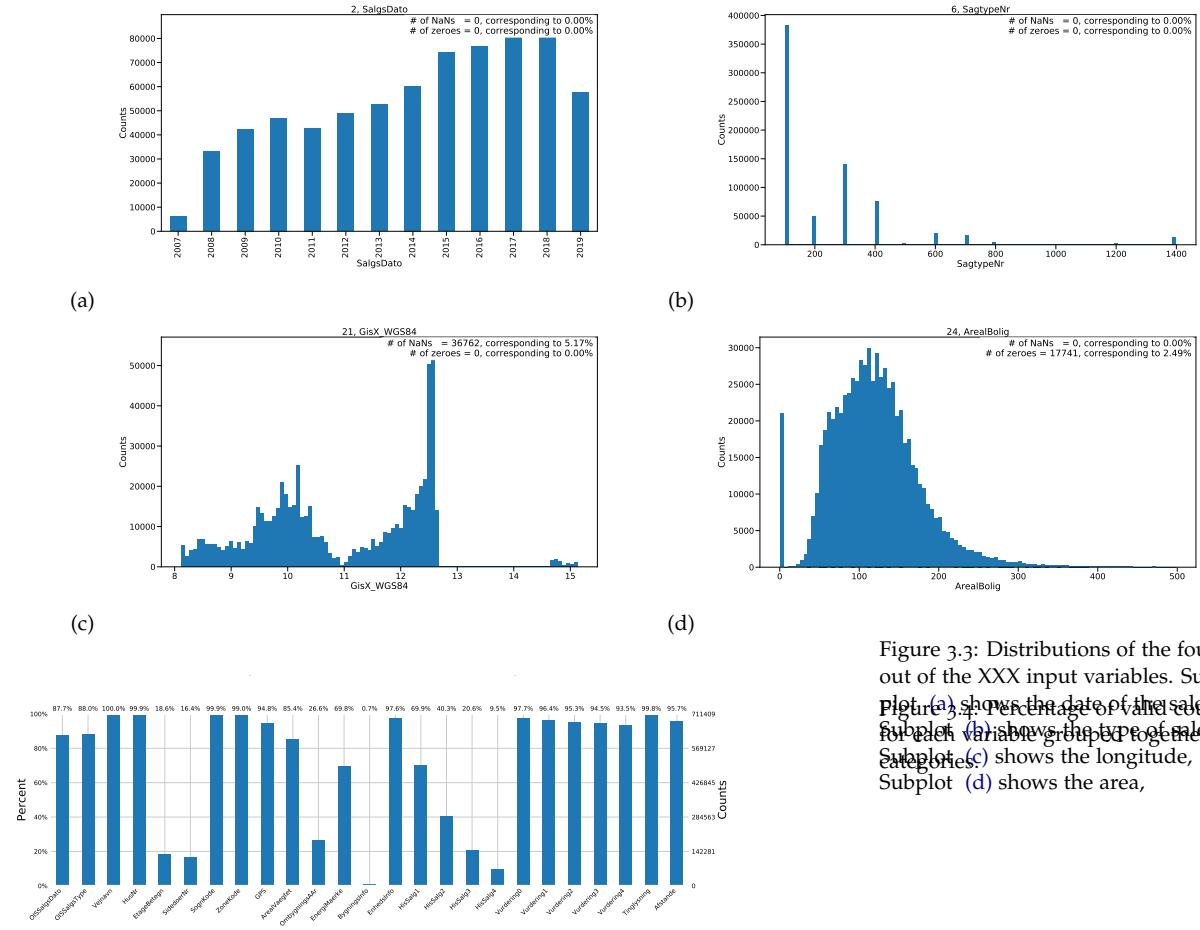


Figure 3.3: Distributions of the four out of the XXX input variables. Subplot (a) shows the date of the students' birth, subplot (b) shows the type of male categories, (c) shows the longitude, Subplot (d) shows the area,

### 3.1 Headings

Tufte's books include the following heading levels: parts, chapters,<sup>1</sup> sections, subsections, and paragraphs. Not defined by default are: sub-subsections and subparagraphs.

<sup>1</sup> Parts and chapters are defined for the `tufte-book` class only.

*Paragraph* Paragraph headings (as shown here) are introduced by italicized text and separated from the main paragraph by a bit of space.

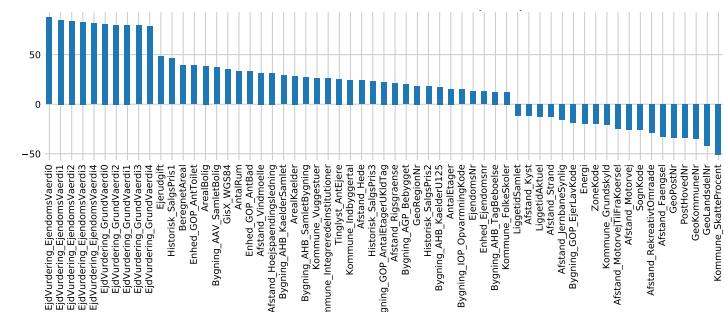


Figure 3.5: Linear correlation between variables and price for variables where the correlation coefficient  $\rho$  is  $|\rho| > 10\%$ .

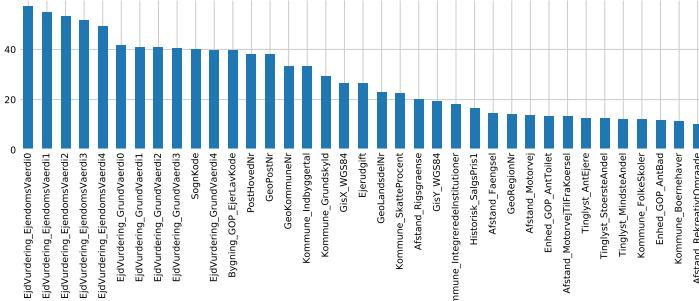


Figure 3.6: Non-linear correlation between variables and price using Maximal Information Coefficient (MIC) for variables where MIC > 10%.

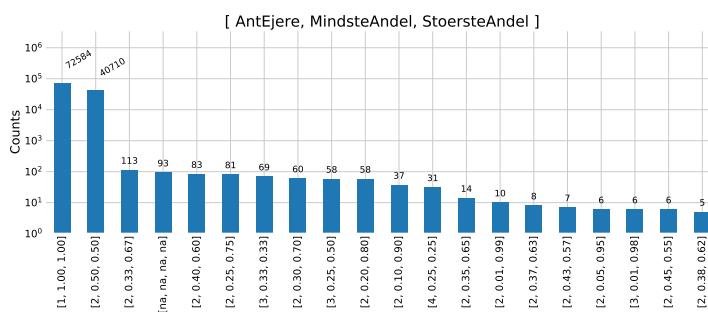


Figure 3.7: Overview of registration of property as a function of amount of owners (`AntEjere`), lowest share (`MindsteAndel`) and biggest share (`StoersteAndel`) written as `[AntEjere, MindsteAndel, StoersteAndel]`.

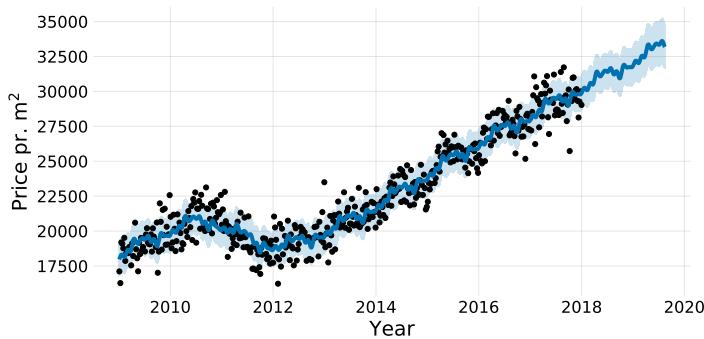


Figure 3.8: The predictions of the Facebook Prophet model trained on square meter prices for apartments sold before January 1st, 2018. The data is down-sampled to weekly bins where the median of each week is used as input to the Prophet model. This can be seen as black dots in the figure. The model's forecasts for 2018 and 2019 are shown in blue with a light blue error band showing the 1-sigma confidence interval.

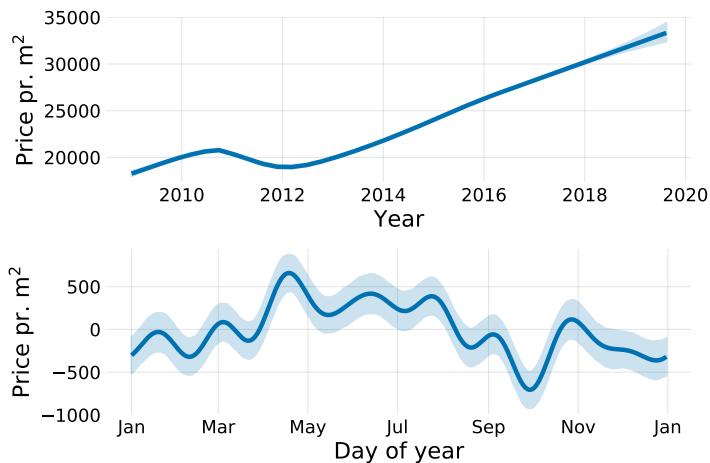


Figure 3.9: The trends of the Facebook Prophet model trained on square meter prices for apartments sold before January 1st, 2018. In the top plot is the overall trend as a function of year and in the bottom plot is the yearly variation as a function of day of year. It can be seen that the square meter price is higher during the Summer months compared to the Winter months, however, compared to the overall trend this effect is minor (< 10%).

Halflife	$\log_{10}$	$N_{\text{trees}}$	Time [s]	MAD
2.5	True	210	81	0.1651
2.5	False	217	64	0.1791
5	True	209	78	0.1647
5	False	100	41	0.1950
<b>10</b>	<b>True</b>	<b>375</b>	<b>118</b>	<b>0.1563</b>
10	False	411	97	0.1739
20	True	333	106	0.1639
20	False	125	44	0.1925
$\infty$	True	371	112	0.1596
$\infty$	False	79	30	0.2062

Table 3.1: RMSE.

Halflife	$\log_{10}$	$N_{\text{trees}}$	Time [s]	MAD
2.5	True	180	57	0.1650
<b>2.5</b>	<b>False</b>	<b>682</b>	<b>111</b>	<b>0.1500</b>
5	True	198	62	0.1647
5	False	892	139	0.1514
10	True	244	68	0.1627
10	False	381	65	0.1608
20	True	396	95	0.1612
20	False	302	53	0.1626
$\infty$	True	239	60	0.1639
$\infty$	False	303	50	0.1620

Table 3.2: LogCosh.

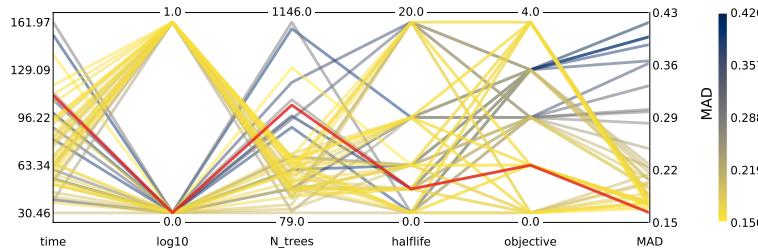


Figure 3.10: Hyperparameter optimization results of the housing model for apartments. The results are shown as parallel coordinates with each hyperparameter along the x-axis and the value of that parameter on the y-axis. Each line is an event in the 4-dimensional space colored according to the performance of that hyperparameter as measured by MAD from highest MAD in dark blue to lowest AUC in yellow. The **single best hyperparameter** is shown in red.

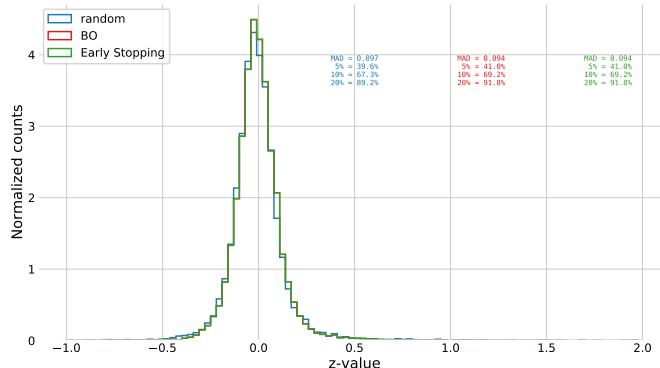


Figure 3.11: Histogram of z-values of the BO algorithm on apartments. The performance function RMSE (0), optimization (1), Cauchy (2), Welch (3), and Fail (4) are mapped to the integers 0, 1, 2, 3, and 4 respectively. The performance of the function after hyperparameter optimization (HPO) using Random Search (RS) is shown in blue for random (0), BO in red for Bayesian Optimization (BO) in red. After finding the best model, the model is retrained using early stopping, the performance of which is shown in green.

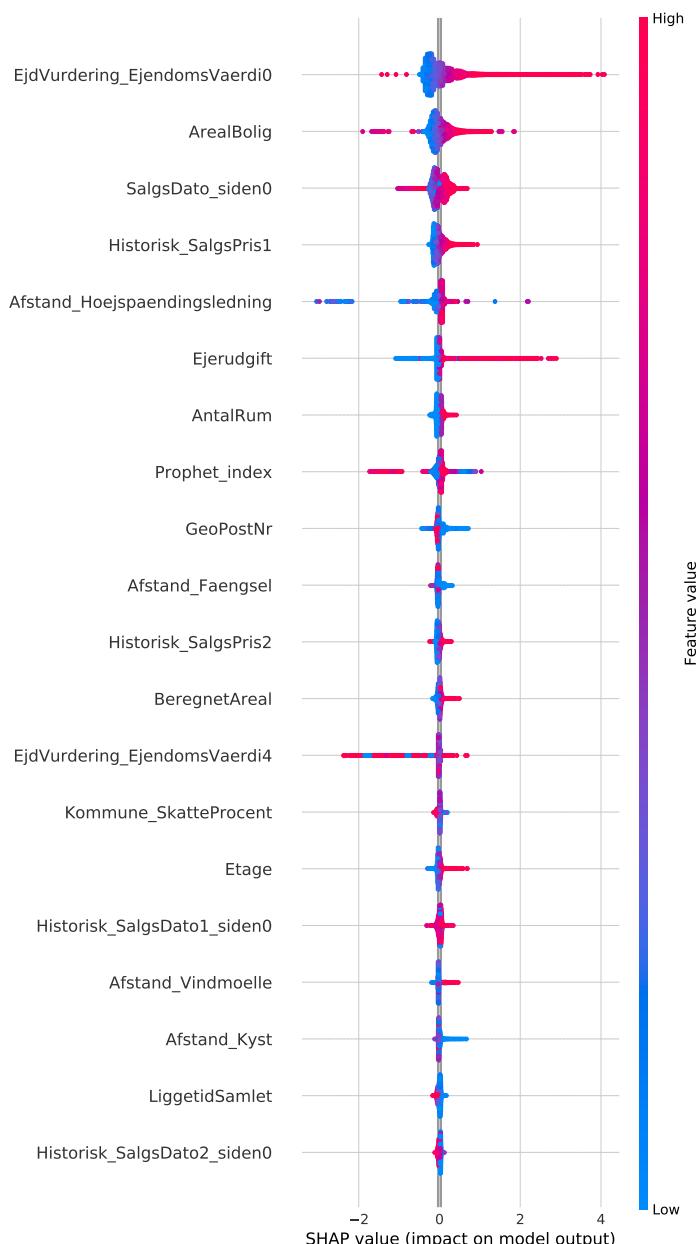


Figure 3.12: Feature importance of apartment prices using the LGB-model. The feature importance is measured using SHAP-values. The variables are sorted top to bottom according to their overall feature importance, i.e. the previous public property valuation `EjdVurdering_EjendomsVaerdi0` is the most important single feature. Along the x-axis is the impact on model output, in this example the price in XXX. This axes is colored by the value of the feature, from **low** (blue) to **high** (red). In this particular example we see that high values of the previous public property valuation has high, positive impact on the model prediction – exactly as expected. This is exactly opposite the total days on market (DOM) described by the variable `LiggetidSamlet` where a high value has a negative impact.

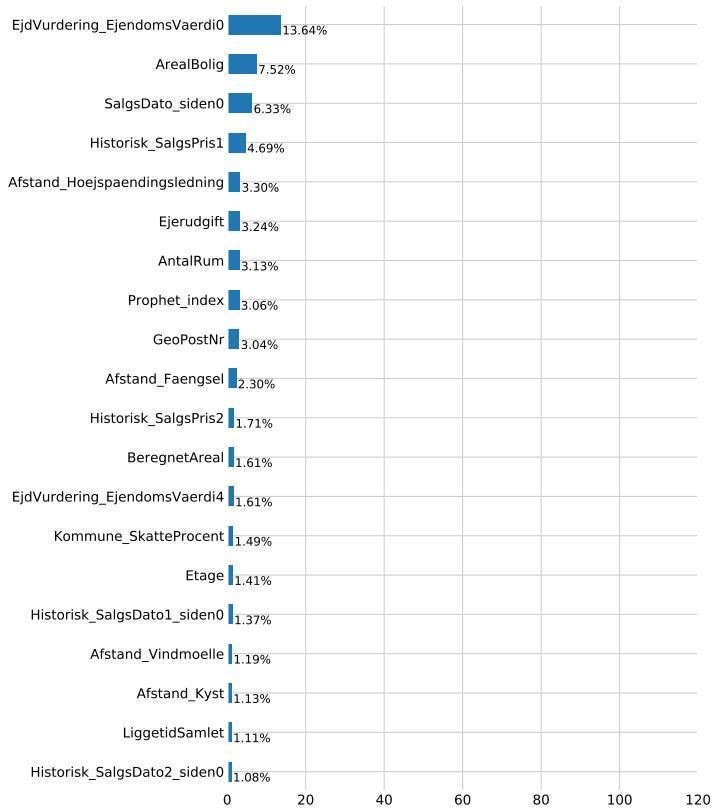


Figure 3.13: Total feature importance of apartment prices using the LGB-model. The total feature importance is measured using SHAP-values, more specifically the mean of the absolute SHAP-values for each variable. This allows a ranking of all of the features ordered after the total importance. Here the total feature importance is scaled such that they sum to 1 when summing over all variables. The absolute total feature importance is in itself not important, but the relative values indicate the difference in feature importance.

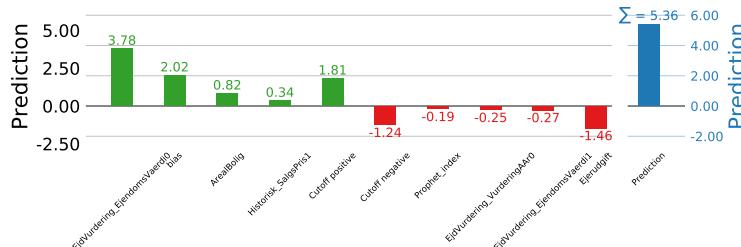


Figure 3.14: Model explanation for LGB model for a specific apartment. The bars are the variables in the dataset that the model found most important sorted after their importance for this particular apartment. The bias bar refers to the expected value of the model, which is simply the mean of the training set which acts as the naive prediction baseline. The “cutoff positive (negative)” bars are the sum of the remaining positive (negative) Figure 3.15: The results of running values that are not shown. On the random search (RS) as hyperparameter random search (HPO) on apartments model prediction shown the model using the LGB model. The **minimum prediction** is the sum of all of the bars (mean) loss along with its uncertainty in the left part (5 to 10 in this example) is shown in red, the **means** for the calculated in 100 DKK. The **negative different iterations of RS** in blue, and values are shown in red, **positive** ones in light blue bands are the **one (and two) standard deviation(s)** of the **means (SDOM)**, all as a function of iteration number.

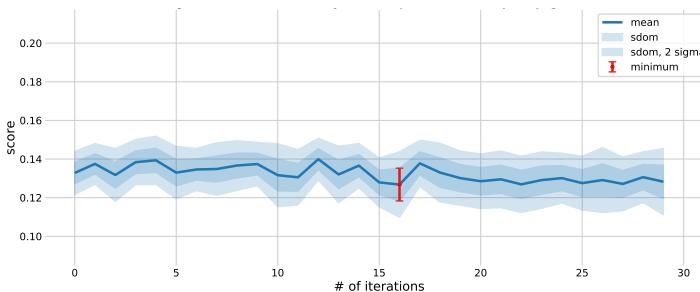
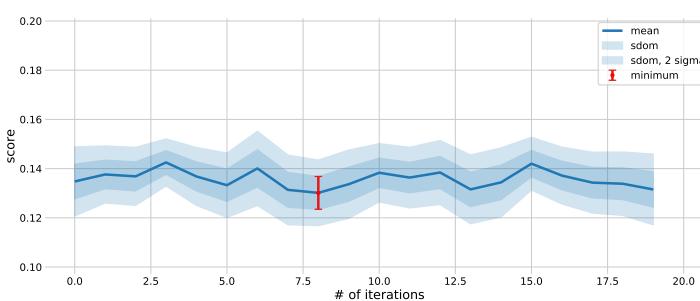


Figure 3.16: The results of running Bayesian optimization (BO) as hyperparameter optimization (HPO) on apartments using the LGB-model. The **minimum (mean) loss** along with its uncertainty is shown in red, the **means** for the different iterations of RS in blue, and as light blue bands are the **one (and two) standard deviation(s) of the means (SDOM)**, all as a function of iteration number. The first XXX iterations are run as RS for use as input to the internal optimization algorithm, the last XX iterations are thus the “smart” BO guesses.

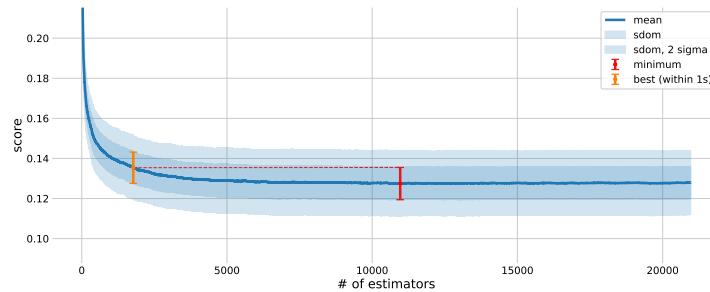


Figure 3.17: The results of early stopping on apartments using the LGB-model. The **minimum (mean)** loss along with its uncertainty is shown in red, the **means** for the different iterations of RS in blue, and as light blue bands are the **one (and two) standard deviation(s) of the means (SDOM)**, all as a function of number of estimators (trees). In orange the **“best” number of estimators** is shown, defined as the lowest number of estimators which are still within 1 SDOM of the minimum value. This leads to a model that has a performance that is within 1 SDOM of the best model, but a lot simpler and faster.

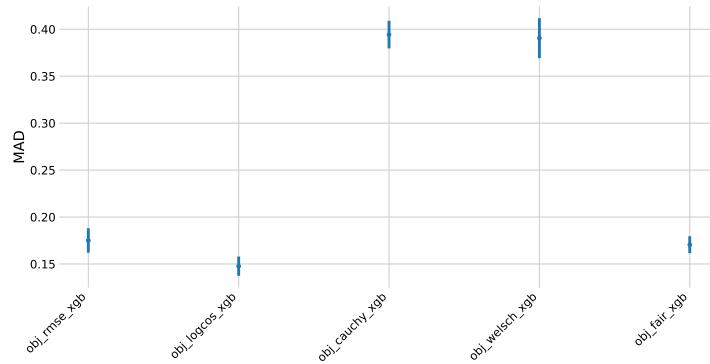


Figure 3.18: Comparison of performance (measured using the median of the absolute deviation, MAD, for apartments) of the five different objective functions: RMSE, LogCosh, Cauchy, Welsch, and Fair. We see that the default objective function, RMSE, does a reasonable job, however, the LogCosh function is a clear improvement.

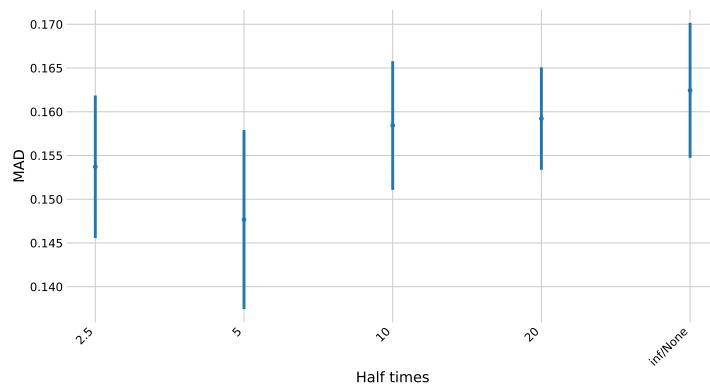


Figure 3.19: Comparison of performance (measured using the median of the absolute deviation, MAD, on apartments) of the five different half time weights: 2.5, 5, 10, 20, inf/None. We see that using no half time weights does has the worst performance, whereas a 5-year half time weight has the best performance (although with a very high uncertainty).

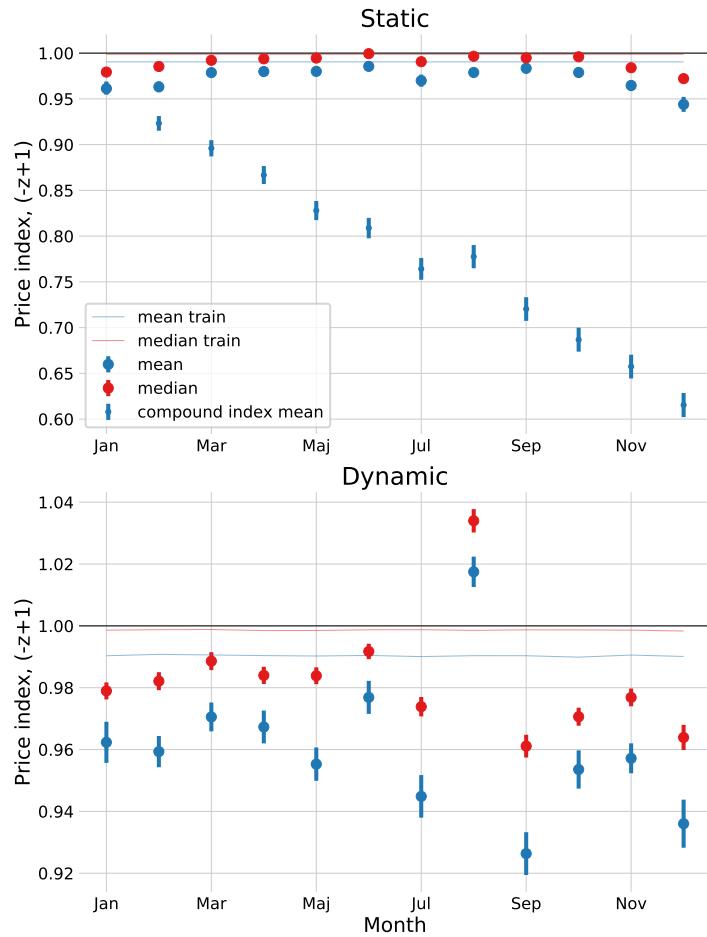


Figure 3.20: Performance of 1-month forecasts for 2018 for apartments. For both plots the LGM model is trained on data up to (but excluding) 2018. Top) The performance of the static model's prediction on sales in the individual months of 2018 is shown for both the [mean](#) and [median](#) of the  $z$ -scores. Bottom) Same as above, however this time based a dynamic model, i.e. a model which is retrained after every month to include the previous month's sales.

Heading	Style	Size
Part	roman	24/36×40 pc
Chapter	italic	20/30×40 pc
Section	italic	12/16×26 pc
Subsection	italic	11/15×26 pc
Paragraph	italic	10/14

Table 3.3: Heading styles used in *Beautiful Evidence*.



## 4. Particle Physics and LEP

The Tufte-L<sup>A</sup>T<sub>E</sub>X document classes define a style similar to the style Edward Tufte uses in his books and handouts. Tufte's style is known for its extensive use of sidenotes, tight integration of graphics with text, and well-set typography. This document aims to be at once a demonstration of the features of the Tufte-L<sup>A</sup>T<sub>E</sub>X document classes and a style guide to their use.

### 4.1 Page Layout

#### 4.1.1 Headings

This style provides **a**- and **b**-heads (that is, `\section` and `\subsection`), demonstrated above.

If you need more than two levels of section headings, you'll have to define them yourself at the moment; there are no pre-defined styles for anything below a `\subsection`. As Bringhurst points out in *The Elements of Typographic Style*, you should "use as many levels of headings as you need: no more, and no fewer."

The Tufte-L<sup>A</sup>T<sub>E</sub>X classes will emit an error if you try to use `\subsubsection` and smaller headings.

IN HIS LATER BOOKS, Tufte starts each section with a bit of vertical space, a non-indented paragraph, and sets the first few words of the sentence in **SMALL CAPS**. To accomplish this using this style, use the `\newthought` command:

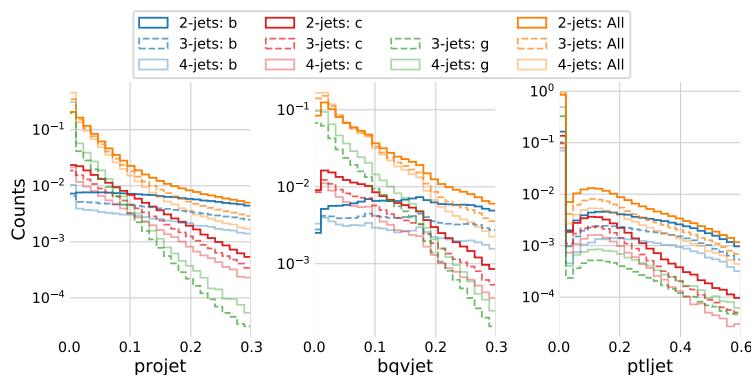
```
\newthought{In his later books}, Tufte starts...
```



# 5. Quark Gluon Analysis

## 5.1 Sidenotes

One of the most prominent and distinctive features of this style is the extensive use of sidenotes. There is a wide margin to provide ample room for sidenotes and small figures. Any `\footnotes` will automatically be converted to sidenotes.<sup>1</sup> If you'd like to place ancillary information in the margin without the sidenote mark (the superscript number), you can use the `\marginnote` command.



The specification of the `\sidenote` command is:

```
\sidenote[⟨number⟩][⟨offset⟩]{Sidenote text.}
```

Both the `⟨number⟩` and `⟨offset⟩` arguments are optional. If you provide a `⟨number⟩` argument, then that number will be used as the sidenote number. It will change of the number of the current sidenote only and will not affect the numbering sequence of subsequent sidenotes.

Sometimes a sidenote may run over the top of other text or graphics in the margin space. If this happens, you can adjust the vertical position of the sidenote by providing a dimension in the `⟨offset⟩` argument. Some examples of valid dimensions are:

```
1.0in    2.54cm    254mm    6\baselineskip
```

If the dimension is positive it will push the sidenote down the page; if the dimension is negative, it will move the sidenote up the page.

While both the `⟨number⟩` and `⟨offset⟩` arguments are optional, they must be provided in order. To adjust the vertical position of the sidenote while leaving the sidenote number alone, use the following syntax:

<sup>1</sup> This is a sidenote that was entered using the `\footnote` command.

This is a margin note. Notice that there isn't a number preceding the note, and there is no number in the main text. Where this note was written three vertex variables, `project`, `bqvjet`, and `ptljet`, used as input variables in the b-tagging models. In blue colors the variables are shown for **true b-jets**, in red for **true c-jets**, in green for **true g-jets**, and in orange for **all of the jets** (including non q-matched). In fully opaque color are shown the distributions for 2-jet events, in dashed (and lighter color) 3-jet events, and in semi-transparent 4-jet events. Notice the logarithmic y-axis, that there are no g-jets for 2-jet events (as expected), and that all of the distributions are very similar not matter how many jets.

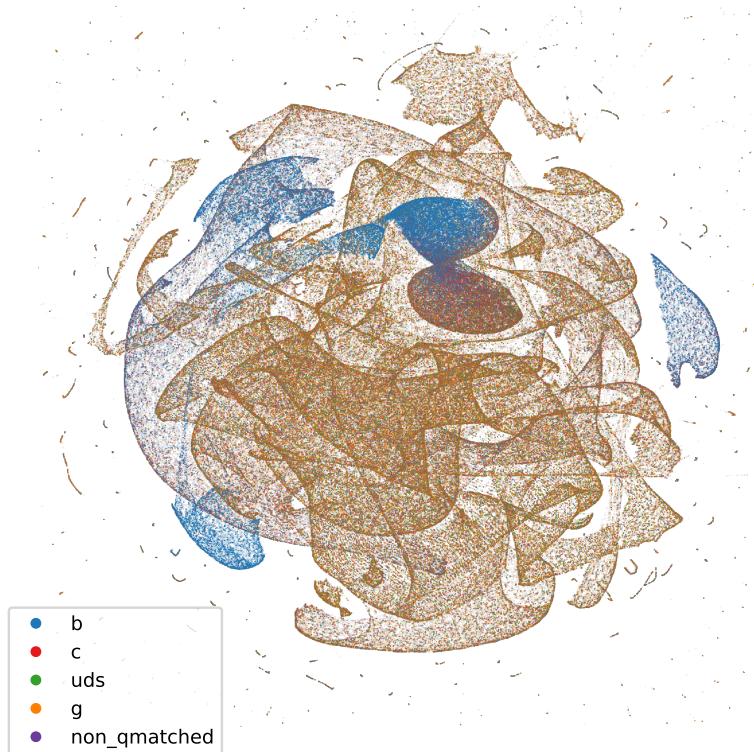


Figure 5.2: Vizualisation of the vertex variables for the different categories: **true b-jets** in blue, **true c-jets** in red, **true uds-jets** in green, **true g-jets** in orange, and **non q-matched**. The clustering is performed with the UMAP algorithm which outputs a 2D-projection. This projection is then visualized using the Datashader which takes takes care of point size, avoids over- and under-plotting, and color intensity.

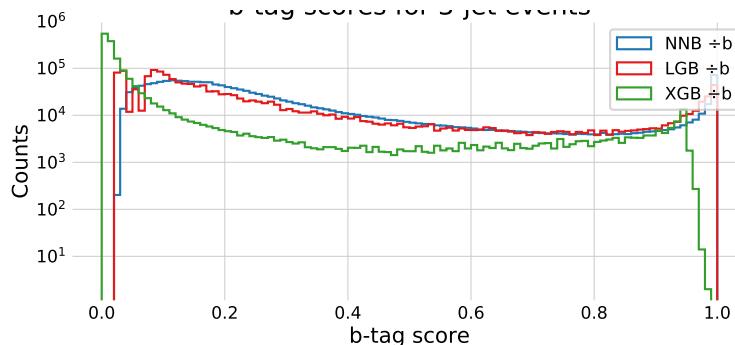


Figure 5.3: Histogram of b-tag scores (model prediction) in 3-jet events for **NNB** (the neural network trained by ATLAS, also called `nnbjet`) in blue, **LGB** in red, and **XGB** in green. We see that the LGB predictions closely match those of NNB which is a good confirmation of a successful fit.

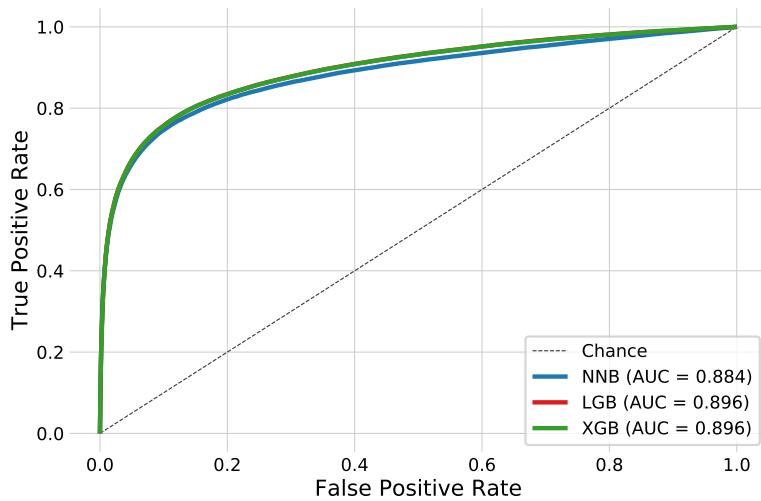


Figure 5.4: ROC curve of the three b-tag models in 3-jet events for **NNB** (the neural network trained by ATLAS, also called `nnbjet`) in blue, **LGB** in red, and **XGB** in green. In the legend the Area Under Curve (AUC) is also shown. Notice that the LGB and XGB models share performance and it is thus due to overplotting that only the green line for XGB can be seen. In the particle physics community False Positive Rate (FPR) is sometimes better known as background efficiency and True Positive Rate (TPR) as signal efficiency.

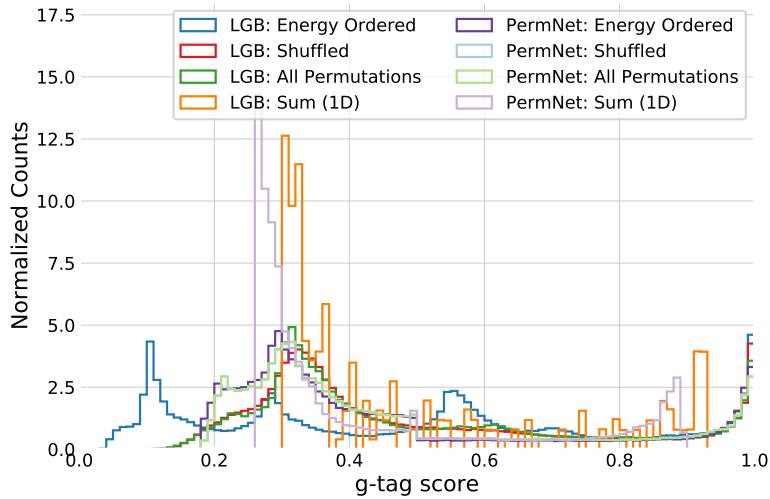


Figure 5.5: Histogram of g-tag scores (model prediction) in 4-jet events for LGB: Energy Ordered in blue, LGB: Shuffled in red, LGB: All Permutations in green, LGB: Sum 1D in orange, PermNet: Energy Ordered in purple, PermNet: Shuffled in light-blue, PermNet: All Permutations in light-green, PermNet: Sum 1D in light-purple. Here LGB and PermNet are the two different type of models and “Energy Ordered”, “Shuffled”, “All Permutations”, and “Sum 1D” are the different methods used for making the input data permutation invariant.

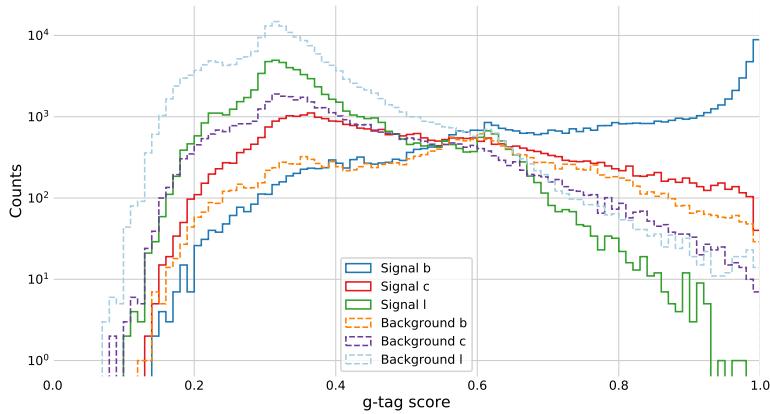


Figure 5.6: Histogram of g-tag scores (model prediction) from the LGB-model in 4-jet events for b signal in blue, c signal in red, l signal in green, b background in orange, c background in purple, l background in light-blue.

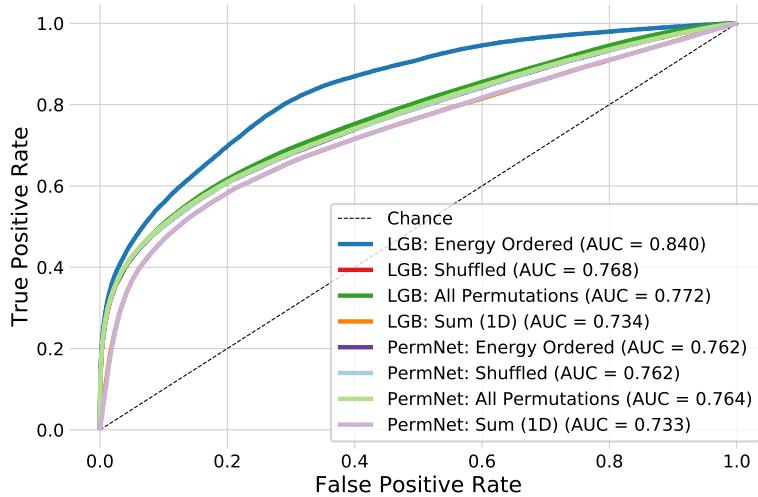


Figure 5.7: ROC curve of the eight g-tag models in 4-jet events. First one in dashed black is the ROC curve that you get by random chance. The colors are the same as in Figure 5.5 and in the legend also the Area Under the ROC curve (AUC) is shown. Notice that the LGB model which uses the energy ordered data produced the best model, however, this model is not permutation invariant. Of the permutation invariant models (the rest), the LGB model trained on all permutations of the b-tags performs highest. The lowest performing models are the two models trained only on the 1-dimensional sum of b-tags, as expected, however, still with a better performance than expected by the author.

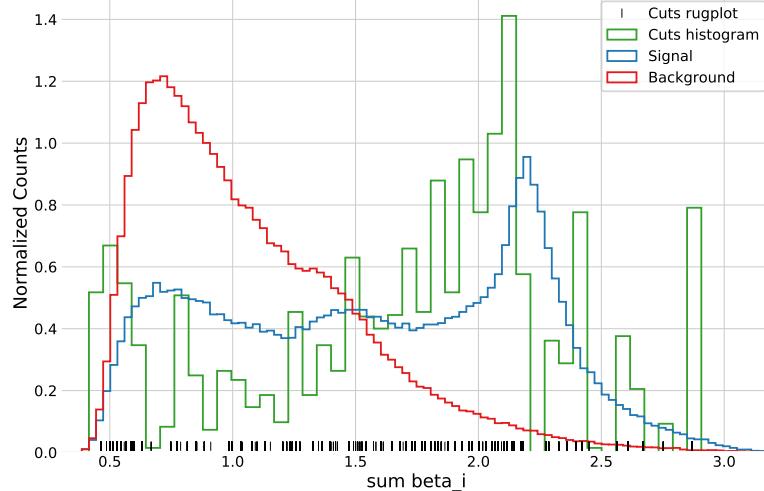


Figure 5.8: Histogram of the distribution of **signal** in blue and **background** in red for 1-dimensional sum of b-tags training data. A histogram of the **cut values** from the LGB model trained on this data is shown in green together with a rug plot of the cut values in black. Notice how most of the cuts match up with the signal peak at around a  $\sum \beta_i \sim 2.1$ , however, there are also quite a lot of cuts around  $\sum \beta_i \sim 0.5$ .

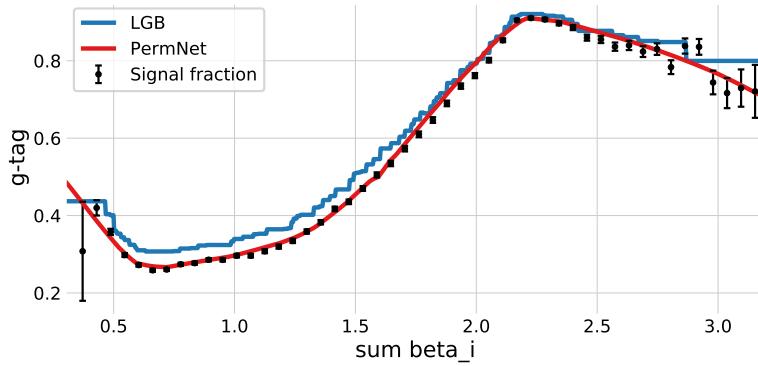


Figure 5.9: Plot of the (1D) g-tag scores as a function of  $\sum \beta_i$  for the **LGB** model in blue and the **PermNet** model in red. Here the g-tag scores are just the models' output values when input a uniformly spaced grid of  $\sum \beta_i$  values between 0 and 4. The signal fraction (based on the signal and background histograms in Figure 5.8) is plotted as black error bars where the size of the error bars is based on the propagated uncertainties of the signal and background histogram assuming Poissonian statistics. Notice how both models capture the overall trend of the signal fraction with the PermNet being **slightly** Hyperparameter Optimization (HPO) results after running 100 iterations of Random Search (only 10 for XGB). In the top row are the results of the 3-jet models and in the bottom row the results of the 4-jet models. From left to right, we have first) the b-tagging results of LGB, second) the b-tagging results of XGB using only 10 iterations of RS, third) the g-tagging results of LGB fit on the Energy Ordered b-tags, and forth) the g-tagging results of LGB fit on the shuffled b-tags. Notice the different ranges on the y-axes.

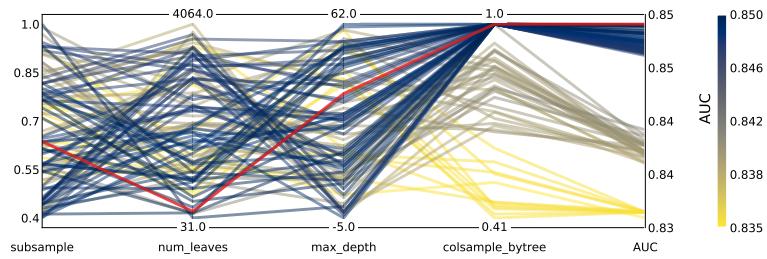
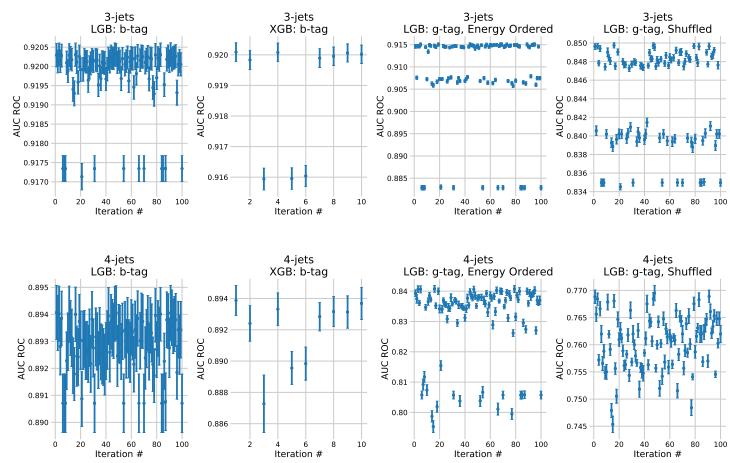


Figure 5.11: Hyperparameter optimization results of g-tagging for 3-jet shuffled events. The results are shown as parallel coordinates with each hyperparameter along the x-axis and the value of that parameter on the y-axis. Each line is an event in the 4-dimensional space colored according to the performance of that hyperparameter as measured by AUC from highest AUC in dark blue to lowest AUC in yellow. The **single best hyperparameter** is shown in red.

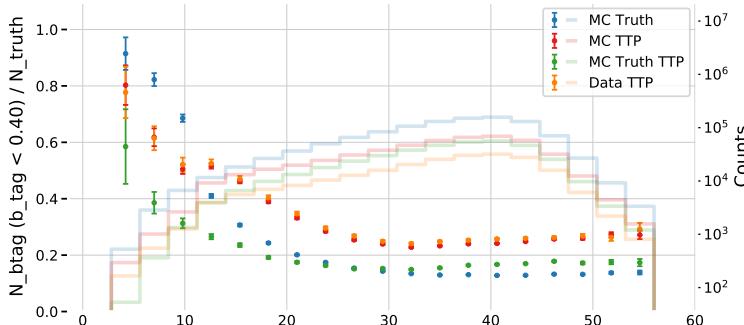
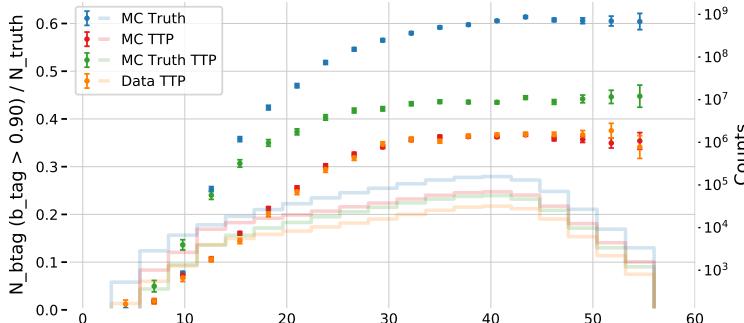
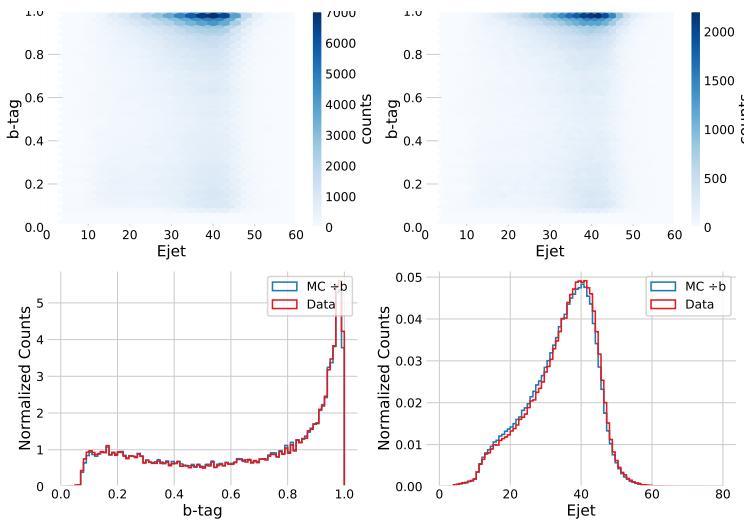
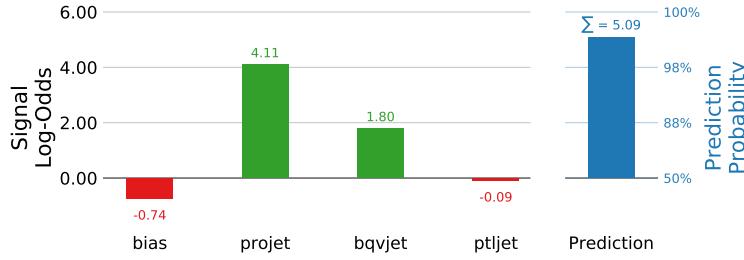


Figure 5.12: Model explanation for the 3-jet b-tagging model for a b-like jet. The first column is the bias of the training set which acts as the naive prediction baseline, the rest are the input data variables. On the right hand side of the plot is the model prediction shown. The left part of the plot is shown in log-odds space, the right part in probability space. The model prediction is the sum of the log-odds (5.09 in this example) transformed into probability space. The negative log-odd values are shown in red.

Figure 5.13: Comparison of the b-tag ones, positive prediction value and jet energy ( $E_{jet}$ ) distributions for Monte Carlo (MC) versus data. In the top row the 2D-distributions are shown for MC on the left (without the extra MC<sub>b</sub> samples) and data on the right. In the bottom row the 1D marginal distributions are shown for the b-tag and the jet energy with data in red and Monte Carlo ones in blue. Notice the almost identical distributions in b-tag.

Figure 5.14: Efficiency of the b-tags for b-jets in the b-signal region for 3-jet events,  $\varepsilon_b^{b-\text{sig}}$ , as a function of jet energy  $E_{jet}$ . The b-signal region is defined as  $\beta > 0.9$ . In the plot the efficiencies are shown for MC Truth in blue, MC TTP in red, MC Truth TTP in green, and Data TTP in orange. The efficiencies (the errorbars) can be read off on the left y-axis and the counts (histograms) on the right y-axis. The abbreviation TTP is short for "Tag, Tag, Probe" where two jets in a event are used as tags and the probe is then used for further analysis. Notice how both MC TTP and Data TTP follow each other closely.

Figure 5.14: Efficiency of the b-tags for b-jets in the g-signal region for 3-jet events,  $\varepsilon_b^{g-\text{sig}}$ , as a function of jet energy  $E_{jet}$ . The g-signal region is defined as  $\beta < 0.4$ . In the plot the efficiencies are shown for MC Truth in blue, MC TTP in red, MC Truth TTP in green, and Data TTP in orange. The efficiencies (the errorbars) can be read off on the left y-axis and the counts (histograms) on the right y-axis. The abbreviation TTP is short for "Tag, Tag, Probe" where two jets in a event are used as tags and the probe is then used for further analysis. Notice how both MC TTP and Data TTP follow each other closely.

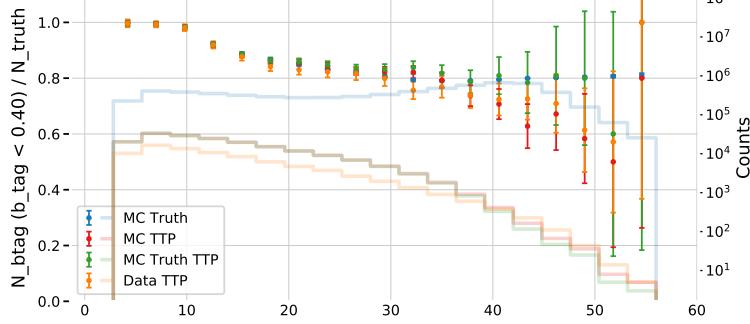


Figure 5.16: Efficiency of the b-tags for g-jets in the g-signal region for 3-jet events,  $\varepsilon_g^{g-\text{sig}}$ , as a function of jet energy  $E_{\text{jet}}$ . The g-signal region is defined as  $\beta < 0.4$ . In the plot the efficiencies are shown for MC Truth in blue, MC TTP in red, MC Truth TTP in green, and Data TTP in orange. The efficiencies (the errorbars) can be read off on the left y-axis and the counts (histograms) on the right y-axis. The abbreviation TTP is short for “Tag, Tag, Probe” where two jets in a event are used as tags and the probe is then used for further analysis. Notice how both MC TTP and Data TTP follow each other closely.

Figure 5.17: Efficiency of the b-tags for g-jets in the b-signal region for 3-jet events,  $\varepsilon_g^{b-\text{sig}}$ , as a function of jet energy  $E_{\text{jet}}$ . The b-signal region is defined as  $\beta > 0.9$ . In the plot the efficiencies are shown for MC Truth in blue, MC TTP in red, MC Truth TTP in green, and Data TTP in orange. The efficiencies (the errorbars) can be read off on the left y-axis and the counts (histograms) on the right y-axis. The abbreviation TTP is short for “Tag, Tag, Probe” where two jets in a event are used as tags and the probe is then used for further analysis.

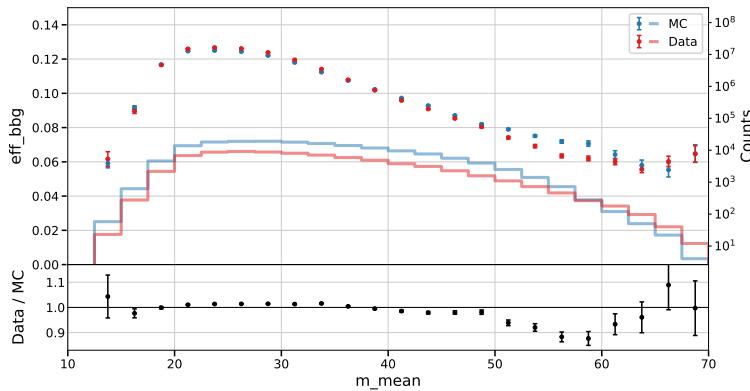
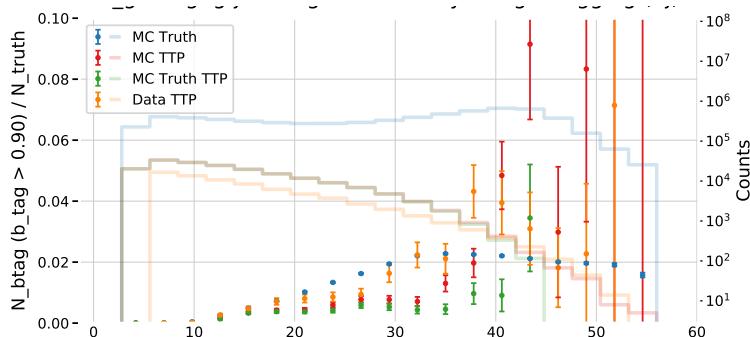


Figure 5.18: Proxy efficiency of the g-tags for  $b\bar{b}g$  3-jet events as a function of the mean of the two invariant masses  $m_{bb}$  and  $m_{b\bar{g}}$ . The proxy efficiency  $\varepsilon_{bb\bar{g}}$  is measured by finding  $b\bar{b}g$ -events where  $\beta_b > 0.9$ ,  $\beta_{\bar{b}} > 0.9$ , and  $\beta_g < 0.4$ . and then calculating  $\varepsilon_{bb\bar{g}} = \varepsilon_b^{b-\text{sig}} \cdot \varepsilon_{\bar{b}}^{b-\text{sig}} \cdot \varepsilon_g^{g-\text{sig}}$ . In the top plot  $\varepsilon_{bb\bar{g}}$  is shown for MC in blue and Data in red where the counts in each bin can be read on right y-axis. In the bottom plot the ratio between Data and MC is shown.

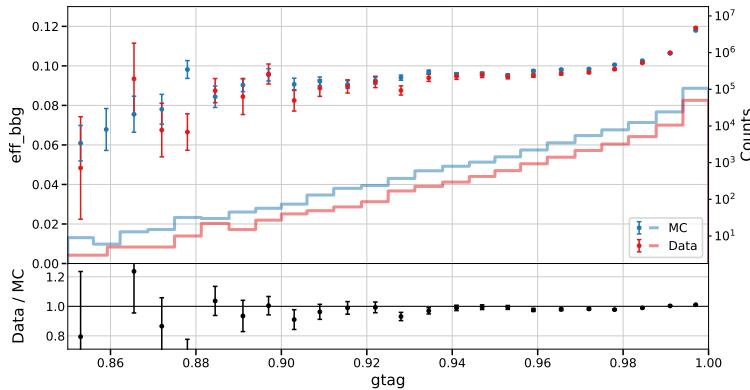


Figure 5.19: Proxy efficiency of the g-tags for  $b\bar{b}g$  3-jet events as a function of the event's g-tag. The proxy efficiency  $\varepsilon_{bb\bar{g}}$  is measured by finding  $b\bar{b}g$ -events where  $\beta_b > 0.9$ ,  $\beta_{\bar{b}} > 0.9$ , and  $\beta_g < 0.4$ . and then calculating  $\varepsilon_{bb\bar{g}} = \varepsilon_b^{b-\text{sig}} \cdot \varepsilon_{\bar{b}}^{b-\text{sig}} \cdot \varepsilon_g^{g-\text{sig}}$ . In the top plot  $\varepsilon_{bb\bar{g}}$  is shown for MC in blue and Data in red where the counts in each bin can be read on right y-axis. In the bottom plot the ratio between Data and MC is shown.

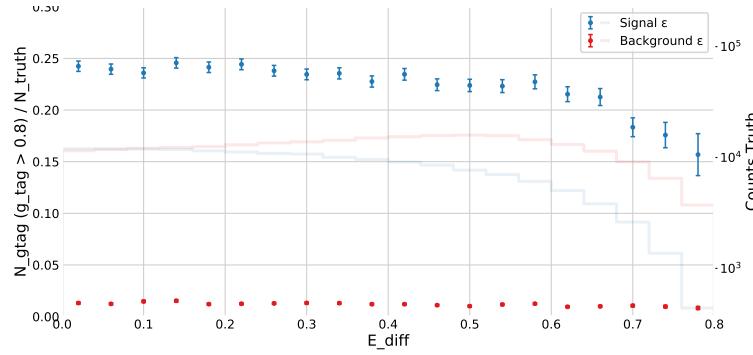


Figure 5.20: Efficiency of the g-tags for 4-jet events as a function of normalized gluon gluon jet energy difference in Monte Carlo. The efficiency is measured as the number of events with a g-tag higher than 0.8 ( $\gamma > 0.8$ ) out of the total number and the normalized gluon gluon jet energy difference  $A$  is  $A = \frac{E_{g\max} - E_{g\min}}{E_{g\max} + E_{g\min}}$  where  $E_{g\max}$  ( $E_{g\min}$ ) refers to the energy of the gluon with the highest (lowest) energy. The efficiency is plotted for **signal events** according to MC Truth in blue and **background events** according to MC Truth in red.

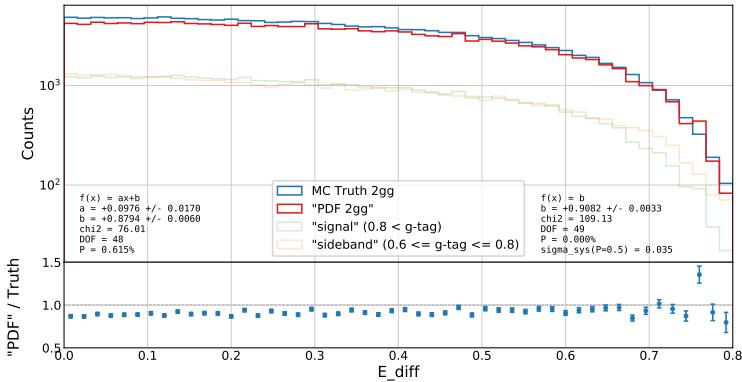


Figure 5.21: Closure plot between MC Truth and the corrected g-tagging model in 4-jet events for the normalized gluon gluon jet energy difference. The corrected g-tagging model is described in further detail in section XXX **TODO!**. In the top part of the plot the **MC Truth** is shown in blue, the **corrected g-tagging model "PDF 2gg"** in red, the **g-signal distribution** in semi-transparent green and the **g-sideband distribution** in semi-transparent orange. In the bottom part of the plot the ratio between MC Truth and the output of the corrected g-tagging model is shown. The normalized gluon gluon jet energy difference  $A$  is  $A = \frac{E_{g\max} - E_{g\min}}{E_{g\max} + E_{g\min}}$ , where  $E_{g\max}$  ( $E_{g\min}$ ) refers to the energy of the gluon with the highest (lowest) energy. Figure 5.22: R CA overview XXX **TODO!**

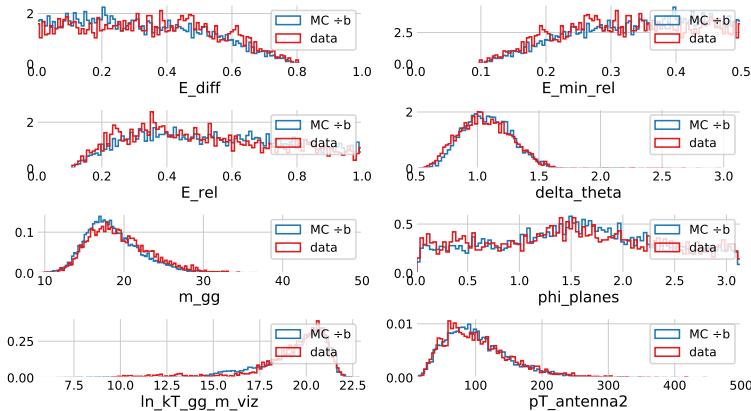
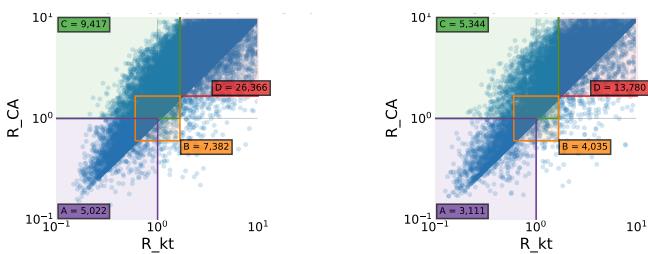


Figure 5.23: R kt CA cut region A XXX **TODO!**

```
\sidenote[][\langle offset\rangle]{Sidenote text.}
```

The empty brackets tell the `\sidenote` command to use the default sidenote number.

If you *only* want to change the sidenote number, however, you may completely omit the `\langle offset\rangle` argument:

```
\sidenote[<number>]{Sidenote text.}
```

The `\marginnote` command has a similar `offset` argument:

```
\marginnote[\langle offset\rangle]{Margin note text.}
```

## 5.2 References

References are placed alongside their citations as sidenotes, as well. This can be accomplished using the normal `\cite` command.<sup>2</sup>

The complete list of references may also be printed automatically by using the `\bibliography` command. (See the end of this document for an example.) If you do not want to print a bibliography at the end of your document, use the `\nobibliography` command in its place.

<sup>2</sup> The first paragraph of this document includes a citation.

## 5.3 Figures and Tables

Images and graphics play an integral role in Tufte's work. In addition to the standard `figure` and `tabular` environments, this style provides special figure and table environments for full-width floats.

Full page-width figures and tables may be placed in `figure*` or `table*` environments. To place figures or tables in the margin, use the `marginfigure` or `margitable` environments as follows (see figure 5.24):

```
\begin{marginfigure}
\includegraphics{helix}
\caption{This is a margin figure.}
\label{fig:marginfig}
\end{marginfigure}
```

The `marginfigure` and `margitable` environments accept an optional parameter `\langle offset\rangle` that adjusts the vertical position of the figure or table. See the “Sidenotes” section above for examples. The specifications are:

```
\begin{marginfigure}[\langle offset\rangle]
...
\end{marginfigure}

\begin{margitable}[\langle offset\rangle]
...
\end{margitable}
```

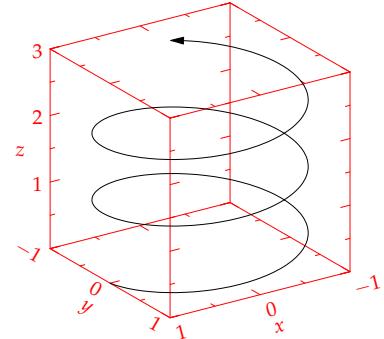


Figure 5.24: This is a margin figure. The helix is defined by  $x = \cos(2\pi z)$ ,  $y = \sin(2\pi z)$ , and  $z = [0, 2.7]$ . The figure was drawn using Asymptote (<http://asymptote.sourceforge.net/>).

Figure 5.25 is an example of the `figure*` environment and figure 5.26 is an example of the normal `figure` environment.

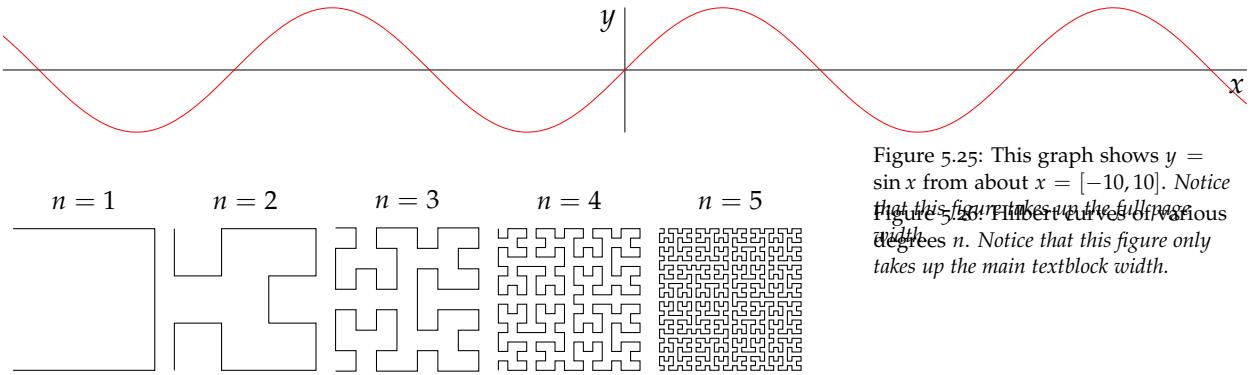


Figure 5.25: This graph shows  $y = \sin x$  from about  $x = [-10, 10]$ . Notice that this figure ranks up the full page width. Figure 5.26: Hilbert curves of various degrees  $n$ . Notice that this figure only takes up the main textblock width.

As with sidenotes and marginnotes, a caption may sometimes require vertical adjustment. The `\caption` command now takes a second optional argument that enables you to do this by providing a dimension  $\langle offset \rangle$ . You may specify the caption in any one of the following forms:

```
\caption{long caption}
\caption[short caption]{long caption}
\caption[][\langle offset \rangle]{long caption}
\caption[short caption][\langle offset \rangle]{long caption}
```

A positive  $\langle offset \rangle$  will push the caption down the page. The short caption, if provided, is what appears in the list of figures/tables, otherwise the “long” caption appears there. Note that although the arguments  $\langle short\ caption \rangle$  and  $\langle offset \rangle$  are both optional, they must be provided in order. Thus, to specify an  $\langle offset \rangle$  without specifying a  $\langle short\ caption \rangle$ , you must include the first set of empty brackets `[]`, which tell `\caption` to use the default “long” caption. As an example, the caption to figure 5.26 above was given in the form

```
\caption[Hilbert curves...][6pt]{Hilbert curves...}
```

Table 5.1 shows table created with the `booktabs` package. Notice the lack of vertical rules—they serve only to clutter the table’s data.

Margin	Length
Paper width	8½ inches
Paper height	11 inches
Textblock width	6½ inches
Textblock/sidenote gutter	¾ inches
Sidenote width	2 inches

Table 5.1: Here are the dimensions of the various margins used in the Tufte-handout class.

OCCASIONALLY  $\text{\LaTeX}$  will generate an error message:

```
Error: Too many unprocessed floats
```

$\text{\LaTeX}$  tries to place floats in the best position on the page. Until it’s finished composing the page, however, it won’t know where those positions are. If you have a lot of floats on a page (including

sidenotes, margin notes, figures, tables, etc.), L<sup>A</sup>T<sub>E</sub>X may run out of “slots” to keep track of them and will generate the above error.

L<sup>A</sup>T<sub>E</sub>X initially allocates 18 slots for storing floats. To work around this limitation, the Tufte-L<sup>A</sup>T<sub>E</sub>X document classes provide a `\morefloats` command that will reserve more slots.

The first time `\morefloats` is called, it allocates an additional 34 slots. The second time `\morefloats` is called, it allocates another 26 slots.

The `\morefloats` command may only be used two times. Calling it a third time will generate an error message. (This is because we can't safely allocate many more floats or L<sup>A</sup>T<sub>E</sub>X will run out of memory.)

If, after using the `\morefloats` command twice, you continue to get the `Too many unprocessed floats` error, there are a couple things you can do.

The `\FloatBarrier` command will immediately process all the floats before typesetting more material. Since `\FloatBarrier` will start a new paragraph, you should place this command at the beginning or end of a paragraph.

The `\clearpage` command will also process the floats before continuing, but instead of starting a new paragraph, it will start a new page.

You can also try moving your floats around a bit: move a figure or table to the next page or reduce the number of sidenotes. (Each sidenote actually uses *two* slots.)

After the floats have placed, L<sup>A</sup>T<sub>E</sub>X will mark those slots as unused so they are available for the next page to be composed.

#### 5.4 Captions

You may notice that the captions are sometimes misaligned. Due to the way L<sup>A</sup>T<sub>E</sub>X's float mechanism works, we can't know for sure where it decided to put a float. Therefore, the Tufte-L<sup>A</sup>T<sub>E</sub>X document classes provide commands to override the caption position.

*Vertical alignment* To override the vertical alignment, use the `\setfloatalignment` command inside the float environment. For example:

```
\begin{figure}[btp]
  \includegraphics{sinewave}
  \caption{This is an example of a sine wave.}
  \label{fig:sinewave}
  \setfloatalignment{b}% forces caption to be bottom-aligned
\end{figure}
```

The syntax of the `\setfloatalignment` command is:

```
\setfloatalignment{\langle pos\rangle}
```

where  $\langle pos \rangle$  can be either `b` for bottom-aligned captions, or `t` for top-aligned captions.

*Horizontal alignment* To override the horizontal alignment, use either the `\forceversofloat` or the `\forcerectofloat` command inside of the float environment. For example:

```
\begin{figure}[btp]
\includegraphics{sinewave}
\caption{This is an example of a sine wave.}
\label{fig:sinewave}
\forceversofloat% forces caption to be set to the left of the float
\end{figure}
```

The `\forceversofloat` command causes the algorithm to assume the float has been placed on a verso page—that is, a page on the left side of a two-page spread. Conversely, the `\forcerectofloat` command causes the algorithm to assume the float has been placed on a recto page—that is, a page on the right side of a two-page spread.

## 5.5 Full-width text blocks

In addition to the new float types, there is a `fullwidth` environment that stretches across the main text block and the sidenotes area.

```
\begin{fullwidth}
Lorem ipsum dolor sit amet...
\end{fullwidth}
```

*Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.*

## 5.6 Typography

### 5.6.1 Typefaces

If the Palatino, Helvetica, and Bera Mono typefaces are installed, this style will use them automatically. Otherwise, we'll fall back on the Computer Modern typefaces.

### 5.6.2 Letterspacing

This document class includes two new commands and some improvements on existing commands for letterspacing.

When setting strings of ALL CAPS or SMALL CAPS, the letter-spacing—that is, the spacing between the letters—should be strings of FULL CAPITAL LETTERS, and the `\smallcaps` command has letterspacing for SMALL CAPITAL LETTERS. These commands will also automatically convert the case of the text to uppercase or lowercase, respectively.

The `\textsc` command has also been redefined to include letterspacing. The case of the `\textsc` argument is left as is, however. This allows one to use both uppercase and lowercase letters: THE INITIAL LETTERS OF THE WORDS IN THIS SENTENCE ARE CAPITALIZED.

## 5.7 Document Class Options

The `tufte-book` class is based on the `LATEX` book document class. Therefore, you can pass any of the typical book options. There are a few options that are specific to the `tufte-book` document class, however.

The `a4paper` option will set the paper size to `A4` instead of the default us letter size.

The `sfsidenotes` option will set the sidenotes and title block in a sans serif typeface instead of the default roman.

The `twoside` option will modify the running heads so that the page number is printed on the outside edge (as opposed to always printing the page number on the right-side edge in `oneside` mode).

The `symmetric` option typesets the sidenotes on the outside edge of the page. This is how books are traditionally printed, but is contrary to Tufte's book design which sets the sidenotes on the right side of the page. This option implicitly sets the `twoside` option.

The `justified` option sets all the text fully justified (flush left and right). The default is to set the text ragged right. The body text of Tufte's books are set ragged right. This prevents needless hyphenation and makes it easier to read the text in the slightly narrower column.

The `bidi` option loads the `bidi` package which is used with `XELATEX` to typeset bi-directional text. Since the `bidi` package needs to be loaded before the `sidenotes` and `cite` commands are defined, it can't be loaded in the document preamble.

The `debug` option causes the Tufte-`LATEX` classes to output debug information to the log file which is useful in troubleshooting bugs. It will also cause the graphics to be replaced by outlines.

The `nofonts` option prevents the Tufte-`LATEX` classes from automatically loading the Palatino and Helvetica typefaces. You should use this option if you wish to load your own fonts. If you're using `XELATEX`, this option is implied (*i.e.*, the Palatino and Helvetica fonts aren't loaded if you use `XELATEX`).

The `nols` option inhibits the letterspacing code. The Tufte-`LATEX` classes try to load the appropriate letterspacing package (either `pdfLATEX`'s `letterspace` package or the `soul` package). If you're using `XELATEX` with `fontenc`, however, you should configure your own letterspacing.

The `notitlepage` option causes `\maketitle` to generate a title block instead of a title page. The `book` class defaults to a title page and the `handout` class defaults to the title block. There is an analogous `titlepage` option that forces `\maketitle` to generate a full title

page instead of the title block.

The `notoc` option suppresses Tufte-L<sup>A</sup>T<sub>E</sub>X's custom table of contents (toc) design. The current toc design only shows unnumbered chapter titles; it doesn't show sections or subsections. The `notoc` option will revert to L<sup>A</sup>T<sub>E</sub>X's toc design.

The `nohyper` option prevents the `hyperref` package from being loaded. The default is to load the `hyperref` package and use the `\title` and `\author` contents as metadata for the generated PDF.



# *6. Discussion and Outlook*

The Tufte-L<sup>A</sup>T<sub>E</sub>X document classes are designed to closely emulate Tufte's book design by default. However, each document is different and you may encounter situations where the default settings are insufficient. This chapter explores many of the ways you can adjust the Tufte-L<sup>A</sup>T<sub>E</sub>X document classes to better fit your needs.

## *6.1 File Hooks*

If you create many documents using the Tufte-L<sup>A</sup>T<sub>E</sub>X classes, it's easier to store your customizations in a separate file instead of copying them into the preamble of each document. The Tufte-L<sup>A</sup>T<sub>E</sub>X classes provide three file hooks: `tufte-common-local.tex`, `tufte-book-local.tex`, and `tufte-handout-local.tex`.

`tufte-common-local.tex` If this file exists, it will be loaded by all of the Tufte-L<sup>A</sup>T<sub>E</sub>X document classes just prior to any document-class-specific code. If your customizations or code should be included in both the book and handout classes, use this file hook.

`tufte-book-local.tex` If this file exists, it will be loaded after all of the common and book-specific code has been read. If your customizations apply only to the book class, use this file hook.

`tufte-common-handout.tex` If this file exists, it will be loaded after all of the common and handout-specific code has been read. If your customizations apply only to the handout class, use this file hook.



# 7. Conclusion

## 7.1 Tufte-L<sup>A</sup>T<sub>E</sub>X Website

The website for the Tufte-L<sup>A</sup>T<sub>E</sub>X packages is located at <http://code.google.com/p/tufte-latex/>. On our website, you'll find links to our SVN repository, mailing lists, bug tracker, and documentation.

## 7.2 Tufte-L<sup>A</sup>T<sub>E</sub>X Mailing Lists

There are two mailing lists for the Tufte-L<sup>A</sup>T<sub>E</sub>X project:

*Discussion list* The `tufte-latex` discussion list is for asking questions, getting assistance with problems, and help with troubleshooting. Release announcements are also posted to this list. You can subscribe to the `tufte-latex` discussion list at <http://groups.google.com/group/tufte-latex>.

*Commits list* The `tufte-latex-commits` list is a read-only mailing list. A message is sent to the list any time the Tufte-L<sup>A</sup>T<sub>E</sub>X code has been updated. If you'd like to keep up with the latest code developments, you may subscribe to this list. You can subscribe to the `tufte-latex-commits` mailing list at <http://groups.google.com/group/tufte-latex-commits>.

## 7.3 Getting Help

If you've encountered a problem with one of the Tufte-L<sup>A</sup>T<sub>E</sub>X document classes, have a question, or would like to report a bug, please send an email to our mailing list or visit our website.

To help us troubleshoot the problem more quickly, please try to compile your document using the `debug` class option and send the generated `.log` file to the mailing list with a brief description of the problem.

## 7.4 Errors, Warnings, and Informational Messages

The following is a list of all of the errors, warnings, and other messages generated by the Tufte-L<sup>A</sup>T<sub>E</sub>X classes and a brief description of their meanings.

Error: \subparagraph is undefined by this class.

The \subparagraph command is not defined in the Tufte-L<sup>A</sup>T<sub>E</sub>X document classes. If you'd like to use the \subparagraph command, you'll need to redefine it yourself. See the "Headings" section on page 17 for a description of the heading styles available in the Tufte-L<sup>A</sup>T<sub>E</sub>X document classes.

Error: \subsubsection is undefined by this class.

The \subsubsection command is not defined in the Tufte-L<sup>A</sup>T<sub>E</sub>X document classes. If you'd like to use the \subsubsection command, you'll need to redefine it yourself. See the "Headings" section on page 17 for a description of the heading styles available in the Tufte-L<sup>A</sup>T<sub>E</sub>X document classes.

Error: You may only call \morefloats twice. See the Tufte-LaTeX documentation for other workarounds.

L<sup>A</sup>T<sub>E</sub>X allocates 18 slots for storing floats. The first time \morefloats is called, it allocates an additional 34 slots. The second time \morefloats is called, it allocates another 26 slots.

The \morefloats command may only be called two times. Calling it a third time will generate this error message. See page 27 for more information.

Warning: Option '*<class option>*' is not supported -- ignoring option.

This warning appears when you've tried to use *<class option>* with a Tufte-L<sup>A</sup>T<sub>E</sub>X document class, but *<class option>* isn't supported by the Tufte-L<sup>A</sup>T<sub>E</sub>X document class. In this situation, *<class option>* is ignored.

Info: The 'symmetric' option implies 'twoside'

You specified the symmetric document class option. This option automatically forces the twoside option as well. See page 30 for more information on the symmetric class option.

## 7.5 Package Dependencies

The following is a list of packages that the Tufte-L<sup>A</sup>T<sub>E</sub>X document classes rely upon. Packages marked with an asterisk are optional.

- xifthen
- ifpdf\*
- ifxetex\*
- hyperref
- geometry
- ragged2e
- chngpage or changepage
- paralist
- textcase
- soul\*

- letterspace\*
- setspace
- natbib *and* bibentry
- optparams
- placeins
- mathpazo\*
- helvet\*
- fontenc
- beramono\*
- fancyhdr
- xcolor
- textcomp
- titlesec
- titletoc



## *Bibliography*

Edgar Anderson. The Species Problem in Iris. 23(3):457–509. ISSN 00266493. DOI: 10.2307/2394164. URL [www.jstor.org/stable/2394164](http://www.jstor.org/stable/2394164).

R. A. Fisher. The Use of Multiple Measurements in Taxonomic Problems. 7(2):179–188. ISSN 2050-1439. DOI: 10.1111/j.1469-1809.1936.tb02137.x. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-1809.1936.tb02137.x>.



# *Index*

- a4paper class option, 30
- \author, 31
- \bibliography, 26
- bidi class option, 30
- bidi package, 30
- booktabs package, 27
- \caption, 27
- \cite, 26
- class options, 30–31
  - a4paper, 30
  - bidi, 30
  - debug, 30, 35
  - justified, 30
  - nofonts, 30
  - nohyper, 31
  - nols, 30
  - notitlepage, 30
  - notoc, 31
  - oneside, 30
  - sfsidenotes, 30
  - symmetric, 30, 36
  - titlepage, 30
  - twoside, 30, 36
- \clearpage, 28
- debug class option, 30, 35
- debug messages, 35
- environments
  - figure, 26
  - figure\*, 26
  - fullwidth, 29
  - marginfigure, 26
  - margitable, 26
- \maketitle, 30
  - marginfigure environment, 26
  - \marginnote, 19, 26
  - margitable environment, 26
  - \morefloats, 28, 36
- \newthought, 17
- \nobibliography, 26
- nofonts class option, 30
- nohyper class option, 31
- nols class option, 30
- notitlepage class option, 30
- \table\*, 26
- \tabular, 26
- error messages, 35
- \figure environment, 26
- \figure\* environment, 26
- file hooks, 33
  - book, 33
  - common, 33
  - handout, 33
- \FloatBarrier, 28
  - fontenc package, 30
- \footnote, 19
- \forceonefloat, 29
- \forceversofloat, 29
- \fullwidth environment, 29
- headings, 8, 17
- \hyperref package, 31
- justified class option, 30
- \letterspace package, 30
- license, ii
- \textsc, 30
- \title, 31
- titlepage class option, 30
- tufte-book-local.tex, 33
- tufte-common-handout.tex, 33
- tufte-common-local.tex, 33
- tufte-handout-local.tex, 33
- twoside class option, 30, 36
- typefaces, 29
- \texttt{XeLATEX}, 30
- notoc class option, 31
- oneside class option, 30
- packages
  - bidi, 30
  - booktabs, 27
  - fontenc, 30
  - \hyperref, 31
  - \letterspace, 30
  - soul, 30
- \setfloatalignment, 28
- \sfsidenotes class option, 30
- \sidenote, 19, 26
- \smallcaps, 29
- \soul package, 30
- \subparagraph, 36
- \subsubsection, 36
- symmetric class option, 30, 36
- \table\* environment, 26
- \tabular environment, 26
- \texttt{tufte-book-local.tex}, 33
- \texttt{tufte-common-handout.tex}, 33
- \texttt{tufte-common-local.tex}, 33
- \texttt{tufte-handout-local.tex}, 33
- twoside class option, 30, 36
- typefaces, 29
- warning messages, 35