# FORMATTING INSTRUCTIONS FOR ICLR 2025 CONFERENCE SUBMISSIONS

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

The abstract paragraph should be indented 1/2 inch (3 picas) on both left and right-hand margins. Use 10 point type, with a vertical spacing of 11 points. The word ABSTRACT must be centered, in small caps, and in point size 12. Two line spaces precede the abstract. The abstract must be limited to one paragraph.

## 1 SUBMISSION OF CONFERENCE PAPERS TO ICLR 2025

ICLR requires electronic submissions, processed by `https://openreview.net/`. See ICLR's website for more instructions.

If your paper is ultimately accepted, the statement `\iclrfinalcopy` should be inserted to adjust the format to the camera ready requirements.

The format for the submissions is a variant of the NeurIPS format. Please read carefully the instructions below, and follow them faithfully.

### 1.1 STYLE

Papers to be submitted to ICLR 2025 must be prepared according to the instructions presented here.

Authors are required to use the ICLR LaTeX style files obtainable at the ICLR website. Please make sure you use the current files and not previous versions. Tweaking the style files may be grounds for rejection.

### 1.2 RETRIEVAL OF STYLE FILES

The style files for ICLR and other conference information are available online at:

`http://www.iclr.cc/`

The file `iclr2025_conference.pdf` contains these instructions and illustrates the various formatting requirements your ICLR paper must satisfy. Submissions must be made using LaTeX and the style files `iclr2025_conference.sty` and `iclr2025_conference.bst` (to be used with LaTeX2e). The file `iclr2025_conference.tex` may be used as a "shell" for writing your paper. All you have to do is replace the author, title, abstract, and text of the paper with your own.

The formatting instructions contained in these style files are summarized in sections 2, 3, and 4 below.

## 2 GENERAL FORMATTING INSTRUCTIONS

The text must be confined within a rectangle 5.5 inches (33 picas) wide and 9 inches (54 picas) long. The left margin is 1.5 inch (9 picas). Use 10 point type with a vertical spacing of 11 points. Times New Roman is the preferred typeface throughout. Paragraphs are separated by 1/2 line space, with no indentation.

Paper title is 17 point, in small caps and left-aligned. All pages should start at 1 inch (6 picas) from the top of the page.

Authors' names are set in boldface, and each name is placed above its corresponding address. The lead author's name is to be listed first, and the co-authors' names are set to follow. Authors sharing the same address can be on the same line.

Please pay special attention to the instructions in section 4 regarding figures, tables, acknowledgments, and references.

There will be a strict upper limit of 10 pages for the main text of the initial submission, with unlimited additional pages for citations.

# 3 Headings: first level

First level headings are in small caps, flush left and in point size 12. One line space before the first level heading and 1/2 line space after the first level heading.

## 3.1 Headings: second level

Second level headings are in small caps, flush left and in point size 10. One line space before the second level heading and 1/2 line space after the second level heading.

### 3.1.1 Headings: third level

Third level headings are in small caps, flush left and in point size 10. One line space before the third level heading and 1/2 line space after the third level heading.

# 4 Citations, figures, tables, references

These instructions apply to everyone, regardless of the formatter being used.

## 4.1 Citations within the text

Citations within the text should be based on the `natbib` package and include the authors' last names and year (with the "et al." construct for more than two authors). When the authors or the publication are included in the sentence, the citation should not be in parenthesis using `\citet{}` (as in "See Hinton et al. (2006) for more information."). Otherwise, the citation should be in parenthesis using `\citep{}` (as in "Deep learning shows promise to make progress towards AI (Bengio & LeCun, 2007).").

The corresponding references are to be listed in alphabetical order of authors, in the References section. As to the format of the references themselves, any style is acceptable as long as it is used consistently.

## 4.2 Footnotes

Indicate footnotes with a number[1] in the text. Place the footnotes at the bottom of the page on which they appear. Precede the footnote with a horizontal rule of 2 inches (12 picas).[2]

## 4.3 Figures

All artwork must be neat, clean, and legible. Lines should be dark enough for purposes of reproduction; art work should not be hand-drawn. The figure number and caption always appear after the figure. Place one line space before the figure caption, and one line space after the figure. The figure caption is lower case (except for first word and proper nouns); figures are numbered consecutively.

Make sure the figure caption does not get separated from the figure. Leave sufficient space to avoid splitting the figure and figure caption.

[1] Sample of the first footnote
[2] Sample of the second footnote

Table 1: Sample table title

| PART | DESCRIPTION |
|---|---|
| Dendrite | Input terminal |
| Axon | Output terminal |
| Soma | Cell body (contains cell nucleus) |

You may use color figures. However, it is best for the figure captions and the paper body to make sense if the paper is printed either in black/white or in color.

Figure 1: Sample figure caption.

### 4.4 Tables

All tables must be centered, neat, clean and legible. Do not use hand-drawn tables. The table number and title always appear before the table. See Table 1.

Place one line space before the table title, one line space after the table title, and one line space after the table. The table title must be lower case (except for first word and proper nouns); tables are numbered consecutively.

## 5 Methodology

### 5.1 Overview

We study portfolio construction under return uncertainty using a conditional diffusion model trained on firm-level characteristics. The model learns the conditional distribution of future cross-sectional returns given observable characteristics and is used to generate Monte Carlo samples of returns. These samples are then mapped into mean–variance optimal portfolios, optionally accounting for transaction costs. Performance is evaluated in a fully out-of-sample backtest.

### 5.2 Data and Preprocessing

We use a large panel of U.S. common equities observed at a monthly frequency. The sample is restricted to primary common shares listed on major U.S. exchanges. Each observation consists of firm-level characteristics observed at month-end and the subsequent realized return.

Let $X_{i,t} \in \mathbb{R}^K$ denote the vector of characteristics for asset $i$ at month-end $t$, and let $R_{i,t+1}$ denote the realized return from $t$ to $t+1$.

**Return Processing.** Raw returns are winsorized cross-sectionally at the 1st and 99th percentiles within each month to mitigate the impact of extreme outliers arising from microstructure effects, delistings, or corporate actions. The prediction target is defined as the next-period return,

$$Y_{i,t} = R_{i,t+1}. \tag{1}$$

**Feature Processing.** All numeric characteristics are standardized cross-sectionally within each month:

$$\tilde{X}_{i,t,k} = \frac{X_{i,t,k} - \mu_{t,k}}{\sigma_{t,k}}, \tag{2}$$

where $\mu_{t,k}$ and $\sigma_{t,k}$ denote the cross-sectional mean and standard deviation of characteristic $k$ at time $t$. Standardized characteristics are winsorized at $\pm 3$ standard deviations. Missing values are imputed using the cross-sectional mean within each month.

**Universe Construction.** To ensure a fixed input dimension suitable for transformer-based diffusion models, we construct a balanced panel by retaining assets with sufficiently long histories and restricting attention to months with complete cross-sectional coverage. The resulting dataset consists of $T$ months, $N$ assets, and $K$ characteristics, yielding tensors

$$X \in \mathbb{R}^{T \times N \times K}, \qquad Y \in \mathbb{R}^{T \times N}. \tag{3}$$

### 5.3 Conditional Diffusion Model for Returns

We model the conditional distribution of next-period returns using a diffusion probabilistic model. At each time $t$, the model learns

$$p_\theta(R_{t+1} \mid X_t), \tag{4}$$

where $R_{t+1} \in \mathbb{R}^N$ denotes the vector of asset returns and $X_t \in \mathbb{R}^{N \times K}$ is the matrix of standardized characteristics.

The diffusion process incrementally perturbs returns with Gaussian noise and trains a neural network to reverse this process conditional on $X_t$. After training, the model generates Monte Carlo samples

$$\{\tilde{R}_{t+1}^{(m)}\}_{m=1}^{M} \sim p_\theta(\cdot \mid X_t), \tag{5}$$

which provide a flexible, non-Gaussian estimate of the conditional return distribution.

### 5.4 Moment Estimation

From the generated samples, we estimate the conditional mean vector and covariance matrix:

$$\hat{\mu}_t = \frac{1}{M} \sum_{m=1}^{M} \tilde{R}_{t+1}^{(m)}, \tag{6}$$

$$\hat{\Sigma}_t = \frac{1}{M-1} \sum_{m=1}^{M} (\tilde{R}_{t+1}^{(m)} - \hat{\mu}_t)(\tilde{R}_{t+1}^{(m)} - \hat{\mu}_t)^\top. \tag{7}$$

For robustness, we also consider a shrinkage estimator of the covariance matrix that combines the sample covariance with a structured target.

### 5.5 Portfolio Optimization

**Mean–Variance Portfolio.** Given estimated moments $(\hat{\mu}_t, \hat{\Sigma}_t)$, portfolio weights are obtained by solving

$$\max_{\omega} \quad \omega^\top \hat{\mu}_t - \frac{\gamma}{2} \omega^\top \hat{\Sigma}_t \omega, \tag{8}$$

subject to

$$\omega^\top \mathbf{1} = 1, \qquad \omega_i \geq 0 \;\; \forall i, \tag{9}$$

where $\gamma > 0$ denotes the risk-aversion parameter.

**Transaction Costs.** To account for trading frictions, we extend the optimization problem to penalize portfolio turnover. Let $\omega_{t-1}$ denote the previous portfolio weights. The objective becomes

$$\max_{\omega} \quad \omega^\top \hat{\mu}_t - \frac{\gamma}{2}\omega^\top \hat{\Sigma}_t \omega - c_b \sum_i b_i - c_s \sum_i s_i, \tag{10}$$

subject to

$$\omega - \omega_{t-1} = b - s, \tag{11}$$

$$\omega^\top \mathbf{1} = 1, \tag{12}$$

$$\omega_i, b_i, s_i \geq 0, \tag{13}$$

where $b_i$ and $s_i$ denote purchases and sales, and $c_b, c_s$ are proportional transaction cost parameters.

### 5.6 Backtesting and Evaluation

Portfolios are rebalanced monthly using information available at time $t$, and returns are realized over $t \to t+1$. Performance is evaluated out of sample using portfolio returns. We report mean monthly return, monthly volatility, and the annualized Sharpe ratio:

$$\text{Sharpe} = \frac{\mathbb{E}[R]}{\sqrt{\mathrm{Var}(R)}}\sqrt{12}. \tag{14}$$

Results are reported both before and after transaction costs.

### 5.7 Baselines

We compare the diffusion-based portfolio to several benchmarks, including an equal-weight portfolio, a classical empirical mean–variance portfolio, and a shrinkage-based mean–variance portfolio. These comparisons isolate the contribution of distributional modeling via diffusion relative to traditional moment estimators.

## 6 Default Notation

In an attempt to encourage standardized notation, we have included the notation file from the textbook, *Deep Learning* Goodfellow et al. (2016) available at `https://github.com/goodfeli/dlbook_notation/`. Use of this style is not required and can be disabled by commenting out `math_commands.tex`.

**Numbers and Arrays**

| | |
|---|---|
| $a$ | A scalar (integer or real) |
| $\boldsymbol{a}$ | A vector |
| $\boldsymbol{A}$ | A matrix |
| $\mathsf{\mathbf{A}}$ | A tensor |
| $\boldsymbol{I}_n$ | Identity matrix with $n$ rows and $n$ columns |
| $\boldsymbol{I}$ | Identity matrix with dimensionality implied by context |
| $\boldsymbol{e}^{(i)}$ | Standard basis vector $[0, \dots, 0, 1, 0, \dots, 0]$ with a 1 at position $i$ |
| $\text{diag}(\boldsymbol{a})$ | A square, diagonal matrix with diagonal entries given by $\boldsymbol{a}$ |
| $\mathrm{a}$ | A scalar random variable |
| $\mathbf{a}$ | A vector-valued random variable |
| $\mathbf{A}$ | A matrix-valued random variable |

**Sets and Graphs**

| | |
|---|---|
| $\mathbb{A}$ | A set |
| $\mathbb{R}$ | The set of real numbers |
| $\{0, 1\}$ | The set containing 0 and 1 |
| $\{0, 1, \dots, n\}$ | The set of all integers between 0 and $n$ |
| $[a, b]$ | The real interval including $a$ and $b$ |
| $(a, b]$ | The real interval excluding $a$ but including $b$ |
| $\mathbb{A} \backslash \mathbb{B}$ | Set subtraction, i.e., the set containing the elements of $\mathbb{A}$ that are not in $\mathbb{B}$ |
| $\mathcal{G}$ | A graph |
| $Pa_{\mathcal{G}}(\mathrm{x}_i)$ | The parents of $\mathrm{x}_i$ in $\mathcal{G}$ |

**Indexing**

| | |
|---|---|
| $a_i$ | Element $i$ of vector $\boldsymbol{a}$, with indexing starting at 1 |
| $a_{-i}$ | All elements of vector $\boldsymbol{a}$ except for element $i$ |
| $A_{i,j}$ | Element $i, j$ of matrix $\boldsymbol{A}$ |
| $\boldsymbol{A}_{i,:}$ | Row $i$ of matrix $\boldsymbol{A}$ |
| $\boldsymbol{A}_{:,i}$ | Column $i$ of matrix $\boldsymbol{A}$ |
| $\mathsf{A}_{i,j,k}$ | Element $(i, j, k)$ of a 3-D tensor $\mathbf{\mathsf{A}}$ |
| $\mathbf{\mathsf{A}}_{:,:,i}$ | 2-D slice of a 3-D tensor |
| $\mathrm{a}_i$ | Element $i$ of the random vector $\mathbf{a}$ |

**Calculus**

| | |
|---|---|
| $\frac{dy}{dx}$ | Derivative of $y$ with respect to $x$ |
| $\frac{\partial y}{\partial x}$ | Partial derivative of $y$ with respect to $x$ |
| $\nabla_{\boldsymbol{x}} y$ | Gradient of $y$ with respect to $\boldsymbol{x}$ |
| $\nabla_{\boldsymbol{X}} y$ | Matrix derivatives of $y$ with respect to $\boldsymbol{X}$ |
| $\nabla_{\mathbf{\mathsf{X}}} y$ | Tensor containing derivatives of $y$ with respect to $\mathbf{\mathsf{X}}$ |
| $\frac{\partial f}{\partial \boldsymbol{x}}$ | Jacobian matrix $\boldsymbol{J} \in \mathbb{R}^{m \times n}$ of $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ |
| $\nabla_{\boldsymbol{x}}^2 f(\boldsymbol{x})$ or $\boldsymbol{H}(f)(\boldsymbol{x})$ | The Hessian matrix of $f$ at input point $\boldsymbol{x}$ |
| $\int f(\boldsymbol{x}) d\boldsymbol{x}$ | Definite integral over the entire domain of $\boldsymbol{x}$ |
| $\int_{\mathbb{S}} f(\boldsymbol{x}) d\boldsymbol{x}$ | Definite integral with respect to $\boldsymbol{x}$ over the set $\mathbb{S}$ |

**Probability and Information Theory**

| | |
|---|---|
| $P(\mathrm{a})$ | A probability distribution over a discrete variable |
| $p(\mathrm{a})$ | A probability distribution over a continuous variable, or over a variable whose type has not been specified |
| $\mathrm{a} \sim P$ | Random variable a has distribution $P$ |
| $\mathbb{E}_{\mathrm{x}\sim P}[f(x)]$ or $\mathbb{E} f(x)$ | Expectation of $f(x)$ with respect to $P(\mathrm{x})$ |
| $\mathrm{Var}(f(x))$ | Variance of $f(x)$ under $P(\mathrm{x})$ |
| $\mathrm{Cov}(f(x), g(x))$ | Covariance of $f(x)$ and $g(x)$ under $P(\mathrm{x})$ |
| $H(\mathrm{x})$ | Shannon entropy of the random variable x |
| $D_{\mathrm{KL}}(P \Vert Q)$ | Kullback-Leibler divergence of P and Q |
| $\mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ | Gaussian distribution over $\boldsymbol{x}$ with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ |

**Functions**

| | |
|---|---|
| $f : \mathbb{A} \rightarrow \mathbb{B}$ | The function $f$ with domain $\mathbb{A}$ and range $\mathbb{B}$ |
| $f \circ g$ | Composition of the functions $f$ and $g$ |
| $f(\boldsymbol{x}; \boldsymbol{\theta})$ | A function of $\boldsymbol{x}$ parametrized by $\boldsymbol{\theta}$. (Sometimes we write $f(\boldsymbol{x})$ and omit the argument $\boldsymbol{\theta}$ to lighten notation) |
| $\log x$ | Natural logarithm of $x$ |
| $\sigma(x)$ | Logistic sigmoid, $\dfrac{1}{1+\exp(-x)}$ |
| $\zeta(x)$ | Softplus, $\log(1+\exp(x))$ |
| $\|\|\boldsymbol{x}\|\|_p$ | $L^p$ norm of $\boldsymbol{x}$ |
| $\|\|\boldsymbol{x}\|\|$ | $L^2$ norm of $\boldsymbol{x}$ |
| $x^+$ | Positive part of $x$, i.e., $\max(0,x)$ |
| $\mathbf{1}_{\mathrm{condition}}$ | is 1 if the condition is true, 0 otherwise |

## 7 FINAL INSTRUCTIONS

Do not change any aspects of the formatting parameters in the style files. In particular, do not modify the width or length of the rectangle the text should fit into, and do not change font sizes (except perhaps in the REFERENCES section; see below). Please note that pages should be numbered.

## 8 PREPARING POSTSCRIPT OR PDF FILES

Please prepare PostScript or PDF files with paper size “US Letter”, and not, for example, “A4”. The -t letter option on dvips will produce US Letter files.

Consider directly generating PDF files using `pdflatex` (especially if you are a MiKTeX user). PDF figures must be substituted for EPS figures, however.

Otherwise, please generate your PostScript and PDF files with the following commands:

```
dvips mypaper.dvi -t letter -Ppdf -G0 -o mypaper.ps
ps2pdf mypaper.ps mypaper.pdf
```

### 8.1 MARGINS IN LATEX

Most of the margin problems come from figures positioned by hand using `\special` or other commands. We suggest using the command `\includegraphics` from the graphicx package.

Always specify the figure width as a multiple of the line width as in the example below using .eps graphics

```
\usepackage[dvips]{graphicx} ...
\includegraphics[width=0.8\linewidth]{myfile.eps}
```

or

```
\usepackage[pdftex]{graphicx} ...
\includegraphics[width=0.8\linewidth]{myfile.pdf}
```

for .pdf graphics. See section 4.4 in the graphics bundle documentation (`http://www.ctan.org/tex-archive/macros/latex/required/graphics/grfguide.ps`)

A number of width problems arise when LaTeX cannot properly hyphenate a line. Please give LaTeX hyphenation hints using the `\-` command.

### Author Contributions

If you'd like to, you may include a section for author contributions as is done in many journals. This is optional and at the discretion of the authors.

### Acknowledgments

Use unnumbered third level headings for the acknowledgments. All acknowledgments, including those to funding agencies, go at the end of the paper.

## References

Yoshua Bengio and Yann LeCun. Scaling learning algorithms towards AI. In *Large Scale Kernel Machines*. MIT Press, 2007.

Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT Press, 2016.

Geoffrey E. Hinton, Simon Osindero, and Yee Whye Teh. A fast learning algorithm for deep belief nets. *Neural Computation*, 18:1527–1554, 2006.

## A Appendix

You may include other additional sections here.