# A journey through the design of (yet another) journal class

# A journey through the design of (yet another) journal class

**Vincent Goulet**

École d'actuariat, Université Laval

TUG 2024
Prague, July 19–21, 2024

**Source code**
View on GitLab

**Cover**
Snowy Owl (*Bubo scandiacus*) easily recognisable due to its mostly white plumage and yellow eyes. One of the largest species of owl, the snowy owl is the avian symbol of Quebec since 1987. It is also present on the coat of arms of the Statistical Society of Canada, where it represents wisdom. Photo credit: © Silver Leapers, CC BY-SA 2.0 Generic, via Wikimedia Commons.

# Preamble

*The Canadian Journal of Statistics* | *La revue canadienne de statistique*
is the official journal of the Statistical Society of Canada

Société statistique du Canada
Statistical Society of Canada

# Context

In late 2022, I was commissioned to develop a new document class for *The CJS*.

- Former class TD-CJS dates *way* back

```
\ProvidesClass{TD-CJS}[1994/07/13 v1.2u Standard LaTeX document class]
```

- Compilation errors
- Evolution of the TeX world
    - LaTeX → pdfLaTeX → XeLaTeX
    - Computer Modern → PostScript fonts → OpenType fonts
    - printed documents → electronic documents

# Yet another journal class

The package **cjs-rcs-article** was first released on CTAN in November, 2023.

- Class cjs-rcs-article
- Bibliographic styles `cjs-rcs-en` and `cjs-rcs-fr`
- Complete documentation (English and French)
- Article templates

# Why not use the Wiley class?

- Distinctive and unique look for *The Canadian Journal of Statistics*
- Independence from large publishers
- Build something with, and for, the community

# Why you?

- Ask the editor Johanna Nešlehová!

# Why you?

- Ask the editor Johanna Nešlehová!
- Maintainer of **ulthese** since 2012

# Why you?

- Ask the editor Johanna Nešlehová!
- Maintainer of **ulthese** since 2012
- Maintainer of a few other packages: **formation-latex-ul**, **actuarialsymbol**, **francais-bst**, ...

# Why you?

- Ask the editor Johanna Nešlehová!
- Maintainer of **ulthese** since 2012
- Maintainer of a few other packages: **formation-latex-ul**, **actuarialsymbol**, **francais-bst**, …

# I am not a designer

I am not a designer

(but I appreciate good design)

**I am not a designer**

**(but I appreciate good design)**

**(and I’m good at replicating stuff)**

# Constraints

**Imposed by *The CJS***

- As close as possible to the published version
- Some mathematical operators (probability, expectation, etc.)
- “French version”

# Constraints

**Imposed by *The CJS***

- As close as possible to the published version
- Some mathematical operators (probability, expectation, etc.)
- “French version”

**Imposed by myself**

- Free software

# Constraints

**Imposed by *The CJS***

- As close as possible to the published version
- Some mathematical operators (probability, expectation, etc.)
- “French version”

**Imposed by myself**

- Free software

**Imposed to myself**

- Sole prerequisite:
  up-to-date modern TeX distribution
- Truly bilingual:
  typography and documentation
- Distinctive high quality fonts
- Compatible with pdfLaTeX and XeLaTeX
- Bibliographic styles

# Demo

The Canadian Journal of Statistics
Vol. 51, No. 3, 2023, Pages 1–27
La revue canadienne de statistique



# Sparse estimation within Pearson's system, with an application to financial market risk

Michelle CAREY[1], Christian GENEST[2] and James O. RAMSAY[2]

[1]*School of Mathematics and Statistics, University College Dublin, Dublin, Ireland*
[2]*Department of Mathematics and Statistics, McGill University, Montréal (Québec) Canada*

*Key words and phrases:* Density estimation; differential regularization; parameter cascading; penalized likelihood; risk measures; S&P 500.

*MSC 2020:* Primary 62G07, 62P05; secondary 62E17, 62R10

*Abstract:* Pearson's system is a rich class of models that includes many classical univariate distributions. It comprises all continuous densities whose logarithmic derivative can be expressed as a ratio of quadratic polynomials governed by a vector $\beta$ of coefficients. The estimation of a Pearson density is challenging as small variations in $\beta$ can induce wild changes in the shape of the corresponding density $f_\beta$. The authors show how to estimate $\beta$ and $f_\beta$ effectively through a penalized likelihood procedure involving differential regularization. The approach combines a penalized regression method and a profiled estimation technique. Simulations and an illustration with S&P 500 data suggest that the proposed method can improve market risk assessment substantially through value-at-risk and expected shortfall estimates that outperform those currently used by financial institutions and regulators.

*Résumé:* La classe de Pearson englobe de très nombreux modèles, dont plusieurs lois univariées classiques. En font partie toutes les densités continues dont la dérivée logarithmique est un rapport de polynômes quadratiques dépendant d'un vecteur $\beta$ de coefficients. L'estimation d'une densité de Pearson est ardue car de petites perturbations de $\beta$ peuvent modifier substantiellement la densité $f_\beta$ correspondante. Les auteurs montrent comment estimer efficacement $\beta$ et $f_\beta$ au moyen d'une vraisemblance pénalisée à facteur de régularisation différentielle. L'approche s'appuie sur la régression pénalisée et l'estimation profilée. Des simulations et une illustration portant sur l'indice boursier S&P 500 suggèrent que l'approche proposée améliore sensiblement l'évaluation du risque de marché grâce à des estimations de valeur à risque et de déficit attendu supérieures à celles couramment utilisées par les institutions financières et les régulateurs.

## 1. INTRODUCTION

The problem of estimating an unknown density $f$ is common in statistics and usually motivated by the need to visualize data, to construct models or to determine the probability of specific, possibly rare, events.

A classical expedient consists of assuming at the outset that $f$ belongs to a parametric class of distributions, thereby reducing the estimation of the density to a finite-dimensional problem. In practice, however, a suitable choice of model can be difficult to make and may even end up occulting data features of critical

* *Author to whom correspondence may be addressed.*
*E-mail: christian.genest@mcgill.ca*



CJS 21-0154



importance, such as a slight asymmetry or heavy-tail behaviour.

As an alternative, nonparametric density estimation methods are often used; see, e.g., Silverman (1986) for an overview. For example, kernel estimation of $f$ from data $X_1, \ldots, X_n$ involves a non-negative function $K$ and a tuning parameter $h \in (0, \infty)$ controlling the regularity of the solution defined, at any $x \in \mathbb{R}$, by

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right). \tag{1}$$

Catalogs of kernels $K$ are available (Granovsky & Müller, 1991) and various methods for selecting an optimal bandwidth $h$ have been proposed; see, e.g., Hall et al. (1992) or Wand & Jones (1994). While many crucial data features can be captured that way, the model is no longer summarized by a few interpretable parameters, and extrapolation beyond the observed range is often poor.

### 1.1. Contribution

This paper describes, studies, and illustrates a highly flexible intermediate solution to the density estimation problem rooted in the maximum penalized likelihood method pioneered by Good & Gaskins (1971); other key early references include Silverman (1982) and O'Sullivan (1988).

The specific approach considered here was briefly mentioned by Ramsay (2000) but never developed. It consists of assuming that $f$ belongs to Pearson's four-parameter system of distributions (Pearson, 1895, 1901). Equivalently, $f$ is taken to be the unique solution $f_\beta$ to the differential equation

$$\frac{d}{dx} \ln\{f(x)\} + g_\beta(x) = 0, \tag{2}$$

where $g_\beta(x) = (x - \beta_1)/(\beta_2 + \beta_3 x + \beta_4 x^2)$ for some unknown parameter $\beta = (\beta_1, \beta_2, \beta_3, \beta_4) \in \mathbb{R}^4$ and values of $x$ in a subset of the real line depending on $\beta$.

This approach is appealing on several accounts. First, it leads to a parametric solution and hence allows for simple and accurate extrapolation beyond the observed range of available observations, which is crucial for estimating the probability of rare events. Second, it is highly flexible in that it can accommodate a diverse range of skewness and kurtosis, as evidenced by Table A in the Appendix, which lists some of the classical models that are encompassed as special cases. Third, the solution is directly interpretable given the direct correspondence that exists between the components of the vector $\beta$ in Pearson's system and the central moments of the distribution (Stuart & Kendall, 1963).

There are, however, challenges in the search for a suitable Pearson distribution. First, the solution $f_\beta$ to Equation (2) is generally not explicit, and hence direct maximization of the likelihood for $\beta$ is excluded unless restrictions are imposed a priori which reduce $f_\beta$ to a specific form. Second, delicate numerical

The Canadian Journal of Statistics | La revue canadienne de statistique
**Vol. 51, No. 3, 2023, 1–20**

# Sparse estimation within Pearson's system, with an application to financial market risk

**Michelle CAREY**
School of Mathematics and Statistics, University College Dublin, Dublin, Ireland
**Christian GENEST** @
McGill University, Montréal (Québec) Canada
**James O. RAMSAY**
McGill University, Montréal (Québec) Canada

**Abstract** Pearson's system is a rich class of models that includes many classical univariate distributions. It comprises all continuous densities whose logarithmic derivative can be expressed as a ratio of quadratic polynomials governed by a vector $\beta$ of coefficients. The estimation of a Pearson density is challenging as small variations in $\beta$ can induce wild changes in the shape of the corresponding density $f_\beta$. The authors show how to estimate $\beta$ and $f_\beta$ effectively through a penalized likelihood procedure involving differential regularization. The approach combines a penalized regression method and a profiled estimation technique. Simulations and an illustration with S&P 500 data suggest that the proposed method can improve market risk assessment substantially through value-at-risk and expected shortfall estimates that outperform those currently used by financial institutions and regulators.

**Résumé** La classe de Pearson englobe de très nombreux modèles, dont plusieurs lois univariées classiques. En font partie toutes les densités continues dont la dérivée logarithmique est un rapport de polynômes quadratiques dépendant d'un vecteur $\beta$ de coefficients. L'estimation d'une densité de Pearson est ardue car de petites perturbations de $\beta$ peuvent modifier substantiellement la densité $f_\beta$ correspondante. Les auteurs montrent comment estimer efficacement $\beta$ et $f_\beta$ au moyen d'une vraisemblance pénalisée à facteur de régularisation différentielle. L'approche s'appuie sur la régression pénalisée et l'estimation profilée. Des simulations et une illustration portant sur l'indice boursier S&P 500 suggèrent que l'approche proposée améliore sensiblement l'évaluation du risque de marché grâce à des estimations de valeur à risque et de déficit attendu supérieures à celles couramment utilisées par les institutions financières et les régulateurs.

**Keywords** Density estimation; differential regularization; parameter cascading; penalized likelihood; risk measures; S&P 500.
**MSC2020** Primary 62G07, 62P05; Secondary 62E17, 62R10.
@ **Corresponding author** christian.genest@mcgill.ca

## 1 Introduction

The problem of estimating an unknown density $f$ is common in statistics and usually motivated by the need to visualize data, $\lambda_{ij}^{1/2}$ to construct models or to determine the probability of specific, possibly rare, events. A classical expedient consists of assuming at the outset that $f$ belongs to a parametric class of distributions, thereby reducing the estimation of the density to a finite-dimensional problem. In practice, however, a suitable choice of model can be difficult to make and may even end up occulting data features of critical importance, such as a slight asymmetry or heavy-tail behaviour.

As an alternative, nonparametric density estimation methods are often used; see, e.g., Silverman (1986) for an overview. For example, kernel estimation of $f$ from data $X_1, \ldots, X_n$ involves a non-negative function $K$ and a





tuning parameter $h \in (0, \infty)$ controlling the regularity of the solution defined, at any $x \in \mathbb{R}$, by

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right). \tag{1}$$

Catalogs of kernels $K$ are available (Granovsky and Müller, 1991) and various methods for selecting an optimal bandwidth $h$ have been proposed; see, e.g., Hall et al. (1992) or Wand and Jones (1995). While many crucial data features can be captured that way, the model is no longer summarized by a few interpretable parameters, and extrapolation beyond the observed range is often poor.

### 1.1 Contribution

This paper describes, studies, and illustrates a highly flexible intermediate solution to the density estimation problem rooted in the maximum penalized likelihood method pioneered by Good and Gaskins (1971); other key early references include Silverman (1982) and O'Sullivan (1988).

The specific approach considered here was briefly mentioned by Ramsay (2000) but never developed. It consists of assuming that $f$ belongs to Pearson's four-parameter system of distributions (Pearson, 1895, 1901). Equivalently, $f$ is taken to be the unique solution $f_\beta$ to the differential equation

$$\frac{d}{dx} \ln\{f(x)\} + g_\beta(x) = 0, \tag{2}$$

where $g_\beta(x) = (x - \beta_1)/(\beta_2 + \beta_3 x + \beta_4 x^2)$ for some unknown parameter $\beta = (\beta_1, \beta_2, \beta_3, \beta_4) \in \mathbb{R}^4$ and values of $x$ in a subset of the real line depending on $\beta$.

This approach is appealing on several accounts. First, it leads to a parametric solution and hence allows for simple and accurate extrapolation beyond the observed range of available observations, which is crucial for estimating the probability of rare events. Second, it is highly flexible in that it can accommodate a diverse range of skewness and kurtosis, as evidenced by Table A in the Appendix, which lists some of the classical models that are encompassed as special cases. Third, the solution is directly interpretable given the direct correspondence that exists between the components of the vector $\beta$ in Pearson's system and the central moments of the distribution (Stuart and Kendall, 1963).

There are, however, challenges in the search for a suitable Pearson distribution. First, the solution $f_\beta$ to equation (2) is generally not explicit, and hence direct maximization of the likelihood for $\beta$ is excluded unless restrictions are imposed a priori which reduce $f_\beta$ to a specific form. Second, delicate numerical issues arise because small variations in $\beta$ can induce large variations in the shape of $f_\beta$, which may live on a bounded interval, a half-line, or the whole real line. Third, as can be seen from Table A, the vector $\beta$ of parameters is often sparse, so parsimony needs to be taken into account in the search for a solution.

### 1.2 Outline

Section 2 describes how estimation can be carried out efficiently and parsimoniously within Pearson's broad class of distributions by relying on a penalized likelihood procedure involving differential regularization and a parameter cascading estimation technique adapted from the functional data analysis literature.

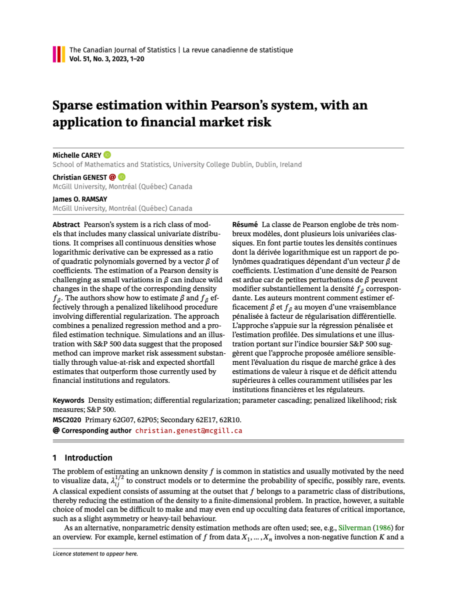
The key step, outlined by Ramsay (2000), consists in identifying, for any vector $\beta$ and any given value of a smoothing parameter $\lambda \in (0, \infty)$, the density $f$ that minimizes the penalized negative log-likelihood score

$$-\frac{1}{n} \sum_{i=1}^{n} \ln\{f(X_i)\} + \lambda \int \left[\frac{d}{dx} \ln\{f(x)\} + g_\beta(x)\right]^2 dx. \tag{3}$$

The solution $f_{\beta,\lambda}$ does not exist in closed form but can be approximated by a linear combination of functions $B_1, \ldots, B_Q$ forming a rich basis, e.g., B-splines.

Each approximation $\hat{f}_{\beta,\lambda}$ induces a likelihood that is easy to compute. Proceeding as in Cao and Ramsay (2007) and Ramsay et al. (2007), one can then construct a profiled likelihood by varying $\beta$ while keeping $\lambda$ fixed. An estimate $\hat{\beta}_\lambda$ of the structural parameter $\beta$ is then found by maximizing this likelihood. Care must be exerted to ensure that the estimate is sparse, however.

Finally, the dependence of $\hat{\beta}_\lambda$ on $\lambda$ can be removed by increasing the value of the smoothing parameter gradually. As $\lambda$ increases, more and more weight is given in expression (3) to the second summand, which


The Canadian Journal of Statistics | La revue canadienne de statistique
Vol. 51, No. 3, 2023, 1–20

# Sparse estimation within Pearson's system, with an application to financial market risk

**Michelle CAREY**
School of Mathematics and Statistics, University College Dublin, Dublin, Ireland

**Christian GENEST** @
McGill University, Montréal (Québec) Canada

**James O. RAMSAY**
McGill University, Montréal (Québec) Canada

**Abstract** Pearson's system is a rich class of models that includes many classical univariate distributions. It comprises all continuous densities whose logarithmic derivative can be expressed as a ratio of quadratic polynomials governed by a vector $\beta$ of coefficients. The estimation of a Pearson density is challenging as small variations in $\beta$ can induce wild changes in the shape of the corresponding density $f_\beta$. The authors show how to estimate $\beta$ and $f_\beta$ effectively through a penalized likelihood procedure involving differential regularization. The approach combines a penalized regression method and a profiled estimation technique. Simulations and an illustration with S&P 500 data suggest that the proposed method can improve market risk assessment substantially through value-at-risk and expected shortfall estimates that outperform those currently used by financial institutions and regulators.

**Résumé** La classe de Pearson englobe de très nombreux modèles, dont plusieurs lois univariées classiques. En font partie toutes les densités continues dont la dérivée logarithmique est un rapport de polynômes quadratiques dépendant d'un vecteur $\beta$ de coefficients. L'estimation d'une densité de Pearson est ardue car de petites perturbations de $\beta$ peuvent modifier substantiellement la densité $f_\beta$ correspondante. Les auteurs montrent comment estimer efficacement $\beta$ et $f_\beta$ au moyen d'une vraisemblance pénalisée à facteur de régularisation différentielle. L'approche s'appuie sur la régression pénalisée et l'estimation profilée. Des simulations et une illustration portant sur l'indice boursier S&P 500 suggèrent que l'approche proposée améliore sensiblement l'évaluation du risque de marché grâce à des estimations de valeur à risque et de déficit attendu supérieures à celles couramment utilisées par les institutions financières et les régulateurs.

**Keywords** Density estimation; differential regularization; parameter cascading; penalized likelihood; risk measures; S&P 500.

**MSC2020** Primary 62G07, 62P05; Secondary 62E17, 62R10.

@ **Corresponding author** `christian.genest@mcgill.ca`

## 1 Introduction

The problem of estimating an unknown density $f$ is common in statistics and usually motivated by the need to visualize data, $\lambda_{ij}^{1/2}$ to construct models or to determine the probability of specific, possibly rare, events. A classical expedient consists of assuming at the outset that $f$ belongs to a parametric class of distributions, thereby reducing the estimation of the density to a finite-dimensional problem. In practice, however, a suitable choice of model can be difficult to make and may even end up occulting data features of critical importance, such as a slight asymmetry or heavy-tail behaviour.

As an alternative, nonparametric density estimation methods are often used; see, e.g., Silverman (1986) for an overview. For example, kernel estimation of $f$ from data $X_1, \ldots, X_n$ involves a non-negative function $K$ and a



text and main title

STIX Two

headings

Fira Sans

computer code

Fira Mono

The Canadian Journal of Statistics | La revue canadienne de statistique
**Vol. 51, No. 3, 2023, 1–20**

# Sparse estimation within Pearson's system, with an application to financial market risk

**Abstract** Pearson's system is a rich class of models that includes many classical univariate distributions. It comprises all continuous densities whose logarithmic derivative can be expressed as a ratio of quadratic polynomials governed by a vector $\beta$ of coefficients. The estimation of a Pearson density is challenging as small variations in $\beta$ can induce wild changes in the shape of the corresponding density $f_\beta$. The authors show how to estimate $\beta$ and $f_\beta$ effectively through a penalized likelihood procedure involving differential regularization. The approach combines a penalized regression method and a profiled estimation technique. Simulations and an illustration with S&P 500 data suggest that the proposed method can improve market risk assessment substantially through value-at-risk and expected shortfall estimates that outperform those currently used by financial institutions and regulators.

**Résumé** La classe de Pearson englobe de très nombreux modèles, dont plusieurs lois univariées classiques. En font partie toutes les densités continues dont la dérivée logarithmique est un rapport de polynômes quadratiques dépendant d'un vecteur $\beta$ de coefficients. L'estimation d'une densité de Pearson est ardue car de petites perturbations de $\beta$ peuvent modifier substantiellement la densité $f_\beta$ correspondante. Les auteurs montrent comment estimer efficacement $\beta$ et $f_\beta$ au moyen d'une vraisemblance pénalisée à facteur de régularisation différentielle. L'approche s'appuie sur la régression pénalisée et l'estimation profilée. Des simulations et une illustration portant sur l'indice boursier S&P 500 suggèrent que l'approche proposée améliore sensiblement l'évaluation du risque de marché grâce à des estimations de valeur à risque et de déficit attendu supérieures à celles couramment utilisées par les institutions financières et les régulateurs.

**Keywords** Density estimation; differential regularization; parameter cascading; penalized likelihood; risk measures; S&P 500.

**MSC2020** Primary 62G07, 62P05; Secondary 62E17, 62R10.

## 1 Introduction

The problem of estimating an unknown density $f$ is common in statistics and usually motivated by the need to visualize data, $\lambda_{ij}^{1/2}$ to construct models or to determine the probability of specific, possibly rare, events. A classical expedient consists of assuming at the outset that $f$ belongs to a parametric class of distributions, thereby reducing the estimation of the density to a finite-dimensional problem. In practice, however, a suitable choice of model can be difficult to make and may even end up occulting data features of critical importance, such as a slight asymmetry or heavy-tail behaviour.

As an alternative, nonparametric density estimation methods are often used; see, e.g., Silverman (1986) for an overview. For example, kernel estimation of $f$ from data $X_1, \ldots, X_n$ involves a non-negative function $K$ and a tuning parameter $h \in (0, \infty)$ controlling the regularity of the solution defined, at any $x \in \mathbb{R}$, by

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right). \tag{1}$$

Catalogs of kernels $K$ are available (Granovsky and Müller, 1991) and various methods for selecting an optimal bandwidth $h$ have been proposed; see, e.g., Hall et al. (1992) or Wand and Jones (1995). While many



option `review`
(using **lineno**)

The Canadian Journal of Statistics | La revue canadienne de statistique
**Vol. 51, No. 3, 2023, 1–20**

# Sparse estimation within Pearson's system, with an application to financial market risk

**Michelle CAREY**
School of Mathematics and Statistics, University College Dublin, Dublin, Ireland

**Christian GENEST** @
McGill University, Montréal (Québec) Canada

**James O. RAMSAY**
McGill University, Montréal (Québec) Canada

**Abstract** Pearson's system is a rich class of models that includes many classical univariate distributions. It comprises all continuous densities whose logarithmic derivative can be expressed as a ratio of quadratic polynomials governed by a vector $\beta$ of coefficients. The estimation of a Pearson density is challenging as small variations in $\beta$ can induce wild changes in the shape of the corresponding density $f_\beta$. The authors show how to estimate $\beta$ and $f_\beta$ effectively through a penalized likelihood procedure involving differential regularization. The approach combines a penalized regression method and a profiled estimation technique. Simulations and an illustration with S&P 500 data suggest that the proposed method can improve market risk assessment substantially through value-at-risk and expected shortfall estimates that outperform those currently used by financial institutions and regulators.

**Résumé** La classe de Pearson englobe de très nombreux modèles, dont plusieurs lois univariées classiques. En font partie toutes les densités continues dont la dérivée logarithmique est un rapport de polynômes quadratiques dépendant d'un vecteur $\beta$ de coefficients. L'estimation d'une densité de Pearson est ardue car de petites perturbations de $\beta$ peuvent modifier substantiellement la densité $f_\beta$ correspondante. Les auteurs montrent comment estimer efficacement $\beta$ et $f_\beta$ au moyen d'une vraisemblance pénalisée à facteur de régularisation différentielle. L'approche s'appuie sur la régression pénalisée et l'estimation profilée. Des simulations et une illustration portant sur l'indice boursier S&P 500 suggèrent que l'approche proposée améliore sensiblement l'évaluation du risque de marché grâce à des estimations de valeur à risque et de déficit attendu supérieures à celles couramment utilisées par les institutions financières et les régulateurs.

**Keywords** Density estimation; differential regularization; parameter cascading; penalized likelihood; risk measures; S&P 500.

**MSC2020** Primary 62G07, 62P05; Secondary 62E17, 62R10.

@ **Corresponding author** christian.genest@mcgill.ca

## 1 Introduction

The problem of estimating an unknown density $f$ is common in statistics and usually motivated by the need to visualize data, $\lambda_{ij}^{1/2}$ to construct models or to determine the probability of specific, possibly rare, events. A classical expedient consists of assuming at the outset that $f$ belongs to a parametric class of distributions, thereby reducing the estimation of the density to a finite-dimensional problem. In practice, however, a suitable choice of model can be difficult to make and may even end up occulting data features of critical importance, such as a slight asymmetry or heavy-tail behaviour.

As an alternative, nonparametric density estimation methods are often used; see, e.g., Silverman (1986) for an overview. For example, kernel estimation of $f$ from data $X_1, \ldots, X_n$ involves a non-negative function $K$ and a



option final

The Canadian Journal of Statistics | La revue canadienne de statistique
Vol. 51, No. 3, 2023, 1–20

SUPPLEMENTARY MATERIAL FOR THE ARTICLE

# Sparse estimation within Pearson's system, with an application to financial market risk

**Michelle CAREY**
School of Mathematics and Statistics, University College Dublin, Dublin, Ireland
**Christian GENEST** @
McGill University, Montréal (Québec) Canada
**James O. RAMSAY**
McGill University, Montréal (Québec) Canada

@ **Corresponding author** christian.genest@mcgill.ca

## S1 Introduction

The problem of estimating an unknown density $f$ is common in statistics and usually motivated by the need to visualize data, $\lambda_{ij}^{1/2}$ to construct models or to determine the probability of specific, possibly rare, events. A classical expedient consists of assuming at the outset that $f$ belongs to a parametric class of distributions, thereby reducing the estimation of the density to a finite-dimensional problem. In practice, however, a suitable choice of model can be difficult to make and may even end up occulting data features of critical importance, such as a slight asymmetry or heavy-tail behaviour.

As an alternative, nonparametric density estimation methods are often used; see, e.g., Silverman (1986) for an overview. For example, kernel estimation of $f$ from data $X_1, \ldots, X_n$ involves a non-negative function $K$ and a tuning parameter $h \in (0, \infty)$ controlling the regularity of the solution defined, at any $x \in \mathbb{R}$, by

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right). \tag{S1}$$

Catalogs of kernels $K$ are available (Granovsky and Müller, 1991) and various methods for selecting an optimal bandwidth $h$ have been proposed; see, e.g., Hall et al. (1992) or Wand and Jones (1995). While many crucial data features can be captured that way, the model is no longer summarized by a few interpretable parameters, and extrapolation beyond the observed range is often poor.

### S1.1 Contribution

This paper describes, studies, and illustrates a highly flexible intermediate solution to the density estimation problem rooted in the maximum penalized likelihood method pioneered by Good and Gaskins (1971); other key early references include Silverman (1982) and O'Sullivan (1988).

The specific approach considered here was briefly mentioned by Ramsay (2000) but never developed. It consists of assuming that $f$ belongs to Pearson's four-parameter system of distributions (Pearson, 1895, 1901). Equivalently, $f$ is taken to be the unique solution $f_\beta$ to the differential equation

$$\frac{d}{dx} \ln\{f(x)\} + g_\beta(x) = 0, \tag{S2}$$

where $g_\beta(x) = (x - \beta_1)/(\beta_2 + \beta_3 x + \beta_4 x^2)$ for some unknown parameter $\beta = (\beta_1, \beta_2, \beta_3, \beta_4) \in \mathbb{R}^4$ and values of $x$ in a subset of the real line depending on $\beta$.

This approach is appealing on several accounts. First, it leads to a parametric solution and hence allows for simple and accurate extrapolation beyond the observed range of available observations, which is crucial



option supplement

# Sparse estimation within Pearson's system, with an application to financial market risk

**Michelle CAREY**
School of Mathematics and Statistics, University College Dublin, Dublin, Ireland

**Christian GENEST** @
McGill University, Montréal (Québec) Canada

**James O. RAMSAY**
McGill University, Montréal (Québec) Canada

**Abstract** Pearson's system is a rich class of models that includes many classical univariate distributions. It comprises all continuous densities whose logarithmic derivative can be expressed as a ratio of quadratic polynomials governed by a vector $\beta$ of coefficients. The estimation of a Pearson density is challenging as small variations in $\beta$ can induce wild changes in the shape of the corresponding density $f_\beta$. The authors show how to estimate $\beta$ and $f_\beta$ effectively through a penalized likelihood procedure involving differential regularization. The approach combines a penalized regression method and a profiled estimation technique. Simulations and an illustration with S&P 500 data suggest that the proposed method can improve market risk assessment substantially through value-at-risk and expected shortfall estimates that outperform those currently used by financial institutions and regulators.

**Résumé** La classe de Pearson englobe de très nombreux modèles, dont plusieurs lois univariées classiques. En font partie toutes les densités continues dont la dérivée logarithmique est un rapport de polynômes quadratiques dépendant d'un vecteur $\beta$ de coefficients. L'estimation d'une densité de Pearson est ardue car de petites perturbations de $\beta$ peuvent modifier substantiellement la densité $f_\beta$ correspondante. Les auteurs montrent comment estimer efficacement $\beta$ et $f_\beta$ au moyen d'une vraisemblance pénalisée à facteur de régularisation différentielle. L'approche s'appuie sur la régression pénalisée et l'estimation profilée. Des simulations et une illustration portant sur l'indice boursier S&P 500 suggèrent que l'approche proposée améliore sensiblement l'évaluation du risque de marché grâce à des estimations de valeur à risque et de déficit attendu supérieures à celles couramment utilisées par les institutions financières et les régulateurs.

**Keywords** Density estimation; differential regularization; parameter cascading; penalized likelihood; risk measures; S&P 500.

**MSC2020** Primary 62G07, 62P05; Secondary 62E17, 62R10.

@ **Corresponding author** christian.genest@mcgill.ca

## 1 Introduction

The problem of estimating an unknown density $f$ is common in statistics and usually motivated by the need to visualize data, $\lambda_{ij}^{1/2}$ to construct models or to determine the probability of specific, possibly rare, events. A classical expedient consists of assuming at the outset that $f$ belongs to a parametric class of distributions, thereby reducing the estimation of the density to a finite-dimensional problem. In practice, however, a suitable choice of model can be difficult to make and may even end up occulting data features of critical importance, such as a slight asymmetry or heavy-tail behaviour.

As an alternative, nonparametric density estimation methods are often used; see, e.g., Silverman (1986) for an overview. For example, kernel estimation of $f$ from data $X_1, \ldots, X_n$ involves a non-negative function $K$ and a




option nocjs

approach is the only one that consistently quantified the market risk in both periods.

In future work, the proposed approach could be extended to the estimation of operational risk (Tursunalieva and Silvapulle, 2014, 2016), as well as to hazard rate estimation (Silverman, 1982). The penalized negative-likelihood could also be broadened to include additional covariate information and handle mixtures of distributions. Finally, Data2Density could eventually be extended to a multivariate context.

## Data sharing

A Matlab package with source code and the datasets used in the examples presented herein are available from `https://data2dynamics.ucd.ie/`


## Acknowledgements

The authors are grateful to the Editor, the Associate Editor and two anonymous referees for their helpful comments and suggestions.

## Funding information

Funding in partial support of this work was provided by the Canada Research Chairs Program, the Natural Sciences and Engineering Research Council of Canada, the Canadian Statistical Sciences Institute, and the Trottier Institute for Science and Public Policy.



## ORCID

Michelle Carey: `https://orcid.org/0000-0002-5603-4264`
Christian Genest: `https://orcid.org/0000-0002-1764-0202`


## A Distributions in Pearson's system

Table A lists some common models within Pearson's class of distributions. For more information, see Elderton and Johnson (1969) or Wolfram Research (2010).

**Table A:** Partial list of distributions in Pearson's system of distributions.

| Type | Distribution | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ |
|---|---|---|---|---|---|
| I | Beta | $-\alpha/(\alpha+\beta)$ | 0 | $1/(\alpha+\beta)$ | $-1/(\alpha+\beta)$ |
| I | Power | $(\alpha-1)/\{k(1-\alpha)\}$ | 0 | $-1/\{k(1-\alpha)\}$ | $1/(1-\alpha)$ |
| III | Chi-Square | $2-\upsilon$ | 0 | 2 | 0 |
| III | Exponential | 0 | 0 | $1/\lambda$ | 0 |
| III | Gamma | $-\alpha\beta+\beta$ | 0 | $\beta$ | 0 |
| III | Normal | $-\mu$ | $\sigma^2$ | 0 | 0 |
| IV | Cauchy (as $\epsilon \to 0$) | $-a$ | $(a^2+b^2)/2$ | $-a$ | $(1+\epsilon)/2$ |
| IV | Student $t_\nu$ | 0 | $\nu/(1+\nu)$ | 0 | $1/(1+\nu)$ |
| V | Lévy | $-\mu-\sigma/3$ | $2\mu^2/3$ | $-4\mu/3$ | $2/3$ |
| VI | $F$-Ratio | $-\dfrac{\upsilon_2-2\upsilon_2/\upsilon_1}{2+\upsilon_2}$ | 0 | $\dfrac{2\upsilon_2/\upsilon_1}{2+\upsilon_2}$ | $\dfrac{2}{2+\upsilon_2}$ |
| VI | Pareto | $-\mu$ | $\dfrac{\mu(\mu-k)}{1+\alpha}$ | $\dfrac{(-2\mu+k)/\alpha}{1+1/\alpha}$ | $\dfrac{1/\alpha}{1+1/\alpha}$ |

## B Auxiliary result

The following result is used in Section 2.1.

**Proposition 1.** *Let $\mathcal{D}$ be the set of functions $w : \mathbb{R} \to \mathbb{R}$ such that*

*(i) the derivative $w'$ exists everywhere and is piece-wise differentiable,*

*(ii) $\int e^{w(x)} dx < \infty$, and*

*(iii) $\int \{w'(x) + g_\beta(x)\}^2 dx < \infty$ for all vectors $\beta \in \mathbb{R}^4$.*

backmatter material

# Interface (select pieces)

# Foundation

cjs-rcs-article is based on memoir.

# Foundation

cjs-rcs-article is based on memoir.

**Pros**

- Experience working with the class
- Page layout, line spacing, nice tables, decorative text, etc.
- Excellent documentation

# Foundation

cjs-rcs-article is based on memoir.

**Pros**

- Experience working with the class
- Page layout, line spacing, nice tables, decorative text, etc.
- Excellent documentation

**Cons**

- Chapter level
- Interaction with **doc** (documentation)

# Authors and affiliations

Information entry system inspired by **authblk** and Bos and McCurley (2023).

```
\author[orcid=0000-0002-5603-4264]
       {Michelle \surname{Carey}}
\affil{School of Mathematics, ...}

\author[orcid=0000-0002-1764-0202,
        email=christian.genest@mcgill.ca,
        corresponding]
       {Christian \surname{Genest}}
\affil{McGill University, ...}

\author{James O. \surname{Ramsay}}
\affil{McGill University, ...}
```

# Authors and affiliations

Information entry system inspired by **authblk** and Bos and McCurley (2023).

```
\author[orcid=0000-0002-5603-4264]
       {Michelle \surname{Carey}}
\affil{School of Mathematics, ...}

\author[orcid=0000-0002-1764-0202,
        email=christian.genest@mcgill.ca,
        corresponding]
       {Christian \surname{Genest}}
\affil{McGill University, ...}

\author{James O. \surname{Ramsay}}
\affil{McGill University, ...}
```

`\author` and `\affil` accumulate data in `\@author`

- text, styling, positioning
- mostly rewritten from scratch

# Authors and affiliations

Information entry system inspired by **authblk** and Bos and McCurley (2023).

```
\author[orcid=0000-0002-5603-4264]
       {Michelle \surname{Carey}}
\affil{School of Mathematics, ...}

\author[orcid=0000-0002-1764-0202,
        email=christian.genest@mcgill.ca,
        corresponding]
       {Christian \surname{Genest}}
\affil{McGill University, ...}

\author{James O. \surname{Ramsay}}
\affil{McGill University, ...}
```

`\surname` typesets and collects surnames

- styling for title page
- collection for the running head (`\runningauthor` to override)
- metadata (not implemented)

# Authors and affiliations

Information entry system inspired by **authblk** and Bos and McCurley (2023).

```
\author[orcid=0000-0002-5603-4264]
       {Michelle \surname{Carey}}
\affil{School of Mathematics, ...}

\author[orcid=0000-0002-1764-0202,
        email=christian.genest@mcgill.ca,
        corresponding]
       {Christian \surname{Genest}}
\affil{McGill University, ...}

\author{James O. \surname{Ramsay}}
\affil{McGill University, ...}
```

key-value interface for metadata

- ORCID iD of authors
- email of authors
- identification of the corresponding author

# Abstracts

Special environments created with **environ** to enter abstracts.

```
\begin{englishabstract}
  Pearson's system is a rich
  class of models that...
\end{englishabstract}

\begin{frenchabstract}
  La classe de Pearson englobe
  de très nombreux modèles...
\end{frenchabstract}
```

# Abstracts

Special environments created with **environ** to enter abstracts.

```
\begin{englishabstract}
  Pearson's system is a rich
  class of models that...
\end{englishabstract}

\begin{frenchabstract}
  La classe de Pearson englobe
  de très nombreux modèles...
\end{frenchabstract}
```

automatic positioning depending on the main language of the article

| | | |
|---|---|---|
| English | Abstract | Résumé |
| French | Résumé | Abstract |

# Keywords and subject classification

Entry using list-like environments.

```
\begin{keywords}
\item Density estimation
\item differential regularization
\item parameter cascading
...
\end{keywords}

\begin{classification}
\item[Primary] 62G07, 62P05
\item[Secondary] 62E17, 62R10
\end{classification}
```

# Keywords and subject classification

Entry using list-like environments.

```
\begin{keywords}
\item Density estimation
\item differential regularization
\item parameter cascading
...
\end{keywords}

\begin{classification}
\item[Primary] 62G07, 62P05
\item[Secondary] 62E17, 62R10
\end{classification}
```

- collection using **environ**
- display using **enumitem**

# Back matter information

Entry as free form text in environments created with **environ**.

```
\begin{supplement}
  The computer code to...
\end{supplement}
```

```
\begin{sharing}
  A Matlab package with...
\end{sharing}
```

```
\begin{acknowledgements}
  The authors are grateful to...
\end{acknowledgements}
```

```
\begin{funding}
  Funding in partial support of...
\end{funding}
```

# Back matter information

Entry as free form text in environments created with **environ**.

```
\begin{supplement}
  The computer code to...
\end{supplement}
```

```
\begin{sharing}
  A Matlab package with...
\end{sharing}
```

```
\begin{acknowledgements}
  The authors are grateful to...
\end{acknowledgements}
```

```
\begin{funding}
  Funding in partial support of...
\end{funding}
```

Display in the correct order (along with ORCID information) using a command.

```
\makebackmatter
```

# Some other features, in no particular order

- Licensing information (`\licence`, `\ccby`, `\ccbysa`, ...)
- Mathematical commands (`\E`, `\Var`, `\prdist`, `\Nset`, ...) and environments (`theorem`, `lemma`, ...)
- Computer code and software (`\code`, `\Rlang`, `\proglang`, ...)
- Compatibility with XeLaTeX and pdfLaTeX
- Exhaustive documentation in English and French
- Templates

# One struggle

I wanted the documentation to use the class (with the `nocjs` option), but memoir and **doc** do not play well together out of the box.

- Use a copy of `\maketitle` from cjs-rcs-article to typeset the title page
- Undefine the environment `glossary` created by memoir
- Restore the standard LaTeX commands for the index and glossary that **doc** relies upon

# Bibliographic styles

*Then the map $\hat{f} = e^{\hat{w}}$ minimizes the expression given in Equation (3) subject to $\int f\,dx = 1$ over $\mathcal{L}$ if and only if the map $\hat{f} = e^{\hat{\bar{w}}}$ minimizes the expression given in Equation (5) over $\mathcal{L}$.*

Proposition 1 holds by the same argument as the proof of Theorem 3.1 of Silverman (1982). It suffices to set $\bar{w} = w - \ln\{\int e^{w(x)}dx\}$, so that $\int e^{\bar{w}(x)}dx = 1$, and to check that, in Silverman's notation, $[\bar{w}, \bar{w}] = [w, w]$, i.e.,

$$\int \{w'(x) + g_\beta(x)\}^2\,dx = \int \{\bar{w}'(x) + g_\beta(x)\}^2\,dx.$$

This identity is easily checked upon substituting $\bar{w}$ into the right-hand side and calling upon the fact that $w - \bar{w}$ is constant.

## References


Barone-Adesi, G., Bourgoin, F., and Giannopoulos, K. (1998). Don't look back. *Risk Magazine*, 11:100–104.

Barone-Adesi, G., Giannopoulos, K., and Vosper, L. (1999). Var without correlations for portfolios of derivative securities. *Journal of Futures Markets*, 19(5):583–602.

Basel Committee on Banking Supervision (1996). *Supervisory Framework for the Use of "Backtesting" in Conjunction With the Internal Models Approach to Market Risk Capital Requirements*.

Becker, M., Klößner, S., and Heinrich, J. (2017). *PearsonDS: Pearson Distribution System*. `https://CRAN.R-project.org/package=PearsonDS`. R package version 1.1.

Bowman, A. W. and Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis: the Kernel Approach with S-Plus Illustrations*. Oxford University Press, Oxford.

Calmon, W., Ferioli, E., Lettieri, D., Soares, J., and Pizzinga, A. (2021). An extensive comparison of some well-established value at risk methods. *International Statistical Review*, 89(1):148–166.

Cao, J. and Ramsay, J. O. (2007). Parameter cascades and profiling in functional data analysis. *Computational Statistics*, 22:335–351.

Carey, M. and Ramsay, J. O. (2021). Fast stable parameter estimation for linear dynamical systems. *Computational Statistics & Data Analysis*, 156:107–124.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International economic review*, pages 841–862.

De Boor, C. (2001). *A Practical Guide to Splines*. 2nd ed. Springer, New York.

Elderton, W. P. and Johnson, N. L. (1969). *Systems of Frequency Curves*. Cambridge University Press, London.

Engle, R. F. and Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *The Journal of Finance*, 48(5):1749–1778.

Favre, L. and Galeano, J.-A. (2002). Mean-modified value-at-risk optimization with hedge funds. *Journal of Alternative Investments*, 5(2):21–25.

Good, I. J. and Gaskins, R. A. (1971). Nonparametric roughness penalties for probability densities. *Biometrika*, 58:255–277. doi: 10.2307/2334515.

Granovsky, B. L. and Müller, H.-G. (1991). Optimizing kernel methods: a unifying variational principle. *International Statistical Review*, 59:373–388.

Haas, M. (2001). *New Methods in Backtesting*. Tech. rep., Financial Engineering Research Center, Bonn, Germany.

Hall, P., Marron, J. S., and Park, B. U. (1992). Smoothed cross-validation. *Probab. Theory Related Fields*, 92(1):1–20. doi: 10.1007/BF01205233.

Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk management models. *Journal of Derivatives*, 3(2):73–84.

McCullough, B. D. and Vinod, H. D. (2003). Verifying the solution from a nonlinear solver: A case study. *American Economic Review*, 93(3):873–892.

McNeil, A. J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, 7(3-4):271–300.


cjs-rcs-en

(not unlike apalike)

Basel Committee on Banking Supervision. 1996, *Supervisory Framework for the Use of "Backtesting" in Conjunction With the Internal Models Approach to Market Risk Capital Requirements*.

Becker, M., S. Klößner et J. Heinrich. 2017, *PearsonDS : Pearson Distribution System*, https://CRAN.R-project.org/package=PearsonDS. R package version 1.1.

Bowman, A. W. et A. Azzalini. 1997, *Applied Smoothing Techniques for Data Analysis : the Kernel Approach with S-Plus Illustrations*, Oxford University Press, Oxford.

Calmon, W., E. Ferioli, D. Lettieri, J. Soares et A. Pizzinga. 2021, « An extensive comparison of some well-established value at risk methods », *International Statistical Review*, vol. 89, nº 1, p. 148–166.

Cao, J. et J. O. Ramsay. 2007, « Parameter cascades and profiling in functional data analysis », *Computational Statistics*, vol. 22, p. 335–351.

Carey, M. et J. O. Ramsay. 2021, « Fast stable parameter estimation for linear dynamical systems », *Computational Statistics & Data Analysis*, vol. 156, p. 107–124.

Christoffersen, P. F. 1998, « Evaluating interval forecasts », *International economic review*, p. 841–862.

De Boor, C. 2001, *A Practical Guide to Splines*, 2e éd., Springer, New York.

Elderton, W. P. et N. L. Johnson. 1969, *Systems of Frequency Curves*, Cambridge University Press, London.

Engle, R. F. et V. K. Ng. 1993, « Measuring and testing the impact of news on volatility », *The Journal of Finance*, vol. 48, nº 5, p. 1749–1778.

Favre, L. et J.-A. Galeano. 2002, « Mean-modified value-at-risk optimization with hedge funds », *Journal of Alternative Investments*, vol. 5, nº 2, p. 21–25.

Good, I. J. et R. A. Gaskins. 1971, « Nonparametric roughness penalties for probability densities », *Biometrika*, vol. 58, p. 255–277, doi : 10.2307/2334515.

Granovsky, B. L. et H.-G. Müller. 1991, « Optimizing kernel methods : a unifying variational principle », *International Statistical Review*, vol. 59, p. 373–388.

Haas, M. 2001, *New Methods in Backtesting*, rapport technique, Financial Engineering Research Center, Bonn, Germany.

Hall, P., J. S. Marron et B. U. Park. 1992, « Smoothed cross-validation », *Probab. Theory Related Fields*, vol. 92, nº 1, p. 1–20, doi : 10.1007/BF01205233.

Kupiec, P. H. 1995, « Techniques for verifying the accuracy of risk management models », *Journal of Derivatives*, vol. 3, nº 2, p. 73–84.

McCullough, B. D. et H. D. Vinod. 2003, « Verifying the solution from a nonlinear solver : A case study », *American Economic Review*, vol. 93, nº 3, p. 873–892.

McNeil, A. J. et R. Frey. 2000, « Estimation of tail-related risk measures for heteroscedastic financial time series : an extreme value approach », *Journal of Empirical Finance*, vol. 7, nº 3-4, p. 271–300.

McNeil, A. J., R. Frey et P. Embrechts. 2015, *Quantitative risk Management : Concepts, Techniques and Tools*, 2e éd., Princeton University Press, Princeton, NJ.

Miao, H., X. Xia, A. S. Perelson et H. Wu. 2011, « On identifiability of nonlinear ODE models and applications in viral dynamics », *SIAM Rev.*, vol. 53, nº 1, p. 3–39, doi : 10.1137/090757009.

Moré, J. J. et D. C. Sorensen. 1983, « Computing a trust region step », *SIAM J. Sci. Statist. Comput.*, vol. 4, nº 3, p. 553–572, doi : 10.1137/0904038.

O'Sullivan, F. 1988, « Fast computation of fully automated log-density and log-hazard estimators », *SIAM Journal on Scientific and Statistical Computing*, vol. 9, nº 2, p. 363–379, doi : 10.1137/0909024.

Pearson, K. 1895, « Contributions to the mathematical theory of evolution. ii. skew variation in homogeneous material », *Philosophical Transactions of the Royal Society of London (A)*, vol. 186, p. 343–414.

Pearson, K. 1901, « Xi. mathematical contributions to the theory of evolution.—x. supplement to a memoir on skew variation », *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, vol. 197, p. 443–459.

cjs-rcs-fr

(very much like francais from **francais-bst**)

# How it's done

Bibliographic styles developed using `makebst` by Patrick W. Daly.

- Original run using `merlin.mbs` from **custom-bib**
- Modification by hand of the driver files
- Creation of the language definition files
- Changes to `merlin.mbs`
  1. function `format.url` reworked to provide a cleaner URL for DOIs

     doi: ⟨*doi*⟩ vs URL `http://dx.doi.org/`⟨*doi*⟩
  2. `\urlprefix` empty
  3. quotes « » typeset by `\frquote` of **babel-french**

# Usage

# Simple as 1, 2, 3

The package **cjs-rcs-article** is part of the standard TeX distributions.

**Standard**

1. Make sure the TeX distribution is up-to-date
2. Grab a template*
3. Start writing

**Alternative**

1. Download and uncompress `cjs-rcs-article-project-install.zip`
2. Grab a template
3. Start writing

*: Not as simple as it should

# Final thoughts

*This was all great fun!*

This document was typeset with the XeLaTeX document preparation system using the **beamer** class and the Metropolis theme. The main text is in Fira Sans, and the computer code in Fira Mono. Icons come from Font Awesome.