When likelihood goes wrong

Nancy Reid
University of Toronto

April 25 2025

Colloque des sciences mathématiques du Québec

CENTRE DE RECHERCHES MATHÉMATIQUES

Examples: a haphazard selection

Clim. Past, 20, 1387–1399, 2024
https://doi.org/10.5194/cp-20-1387-2024

Climate of the Past
Open Access

600 years of wine must quality and April to August temperatures in western Europe 1420–2019

Christian Pfister[1], Stefan Brönnimann[2], Andres Altwegg[3], Rudolf Brázdil[4], Laurent Litzenburger[5], Daniele Lorusso[6], and Thomas Pliemon[7]

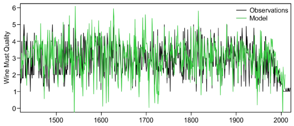


Figure 4. Observed series of wine quality (average; black) from 1420 to 2019 and series obtained with a statistical model calibrated in 1781–1800 (green). The model is explained in Sect. 3.

Scientific question: Can historical records of wine quality be used as temperature proxies?

observational data

Statistical model: “we used a statistical [linear regression] model for wine quality based on local temperature and precipitation”

yes, if used carefully

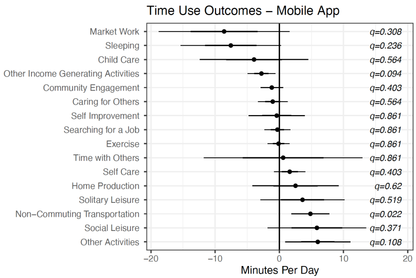
NBER WORKING PAPER SERIES

THE EMPLOYMENT EFFECTS OF A GUARANTEED INCOME:
EXPERIMENTAL EVIDENCE FROM TWO U.S. STATES

Eva Vivalt
Elizabeth Rhodes
Alexander W. Bartik
David E. Broockman
Sarah Miller

Working Paper 32719
http://www.nber.org/papers/w32719

Figure 5: Time Use Results: Mobile App



Scientific question: Does guaranteed income supplement affect labor market measures?

randomized controlled trial

Statistical model: $Y_i = \alpha + \beta Treated_i + \gamma^{\mathrm{T}} X_i + \epsilon_i$

"support for both sides of this debate"

Article https://doi.org/10.1038/s41550-023-01983-1

Variability of extragalactic X-ray jets on kiloparsec scales

Received: 17 May 2022

Accepted: 27 April 2023

Published online: 29 May 2023

Eileen T. Meyer[1], Aamil Shaik[1], Yanbo Tang[2], Nancy Reid[3], Karthik Reddy[1,4], Peter Breiding[5], Markos Georganopoulos[1], Marco Chiaberge[5,6], Eric Perlman[7], Devon Clautice[7], William Sparks[8,9], Nat DeNigris[1,10] & Max Trevor[1,11]

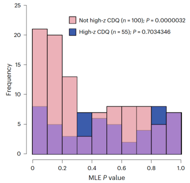


Fig. 3 | Histogram of the single-region *P* values from the directional test, not adjusted for multiple comparison. In pink, the subset of sources that

Yanbo Tang

Scientific question: Are observations of X-ray jets consistent with current theory?

observational data

Statistical model: compare background and sources measurements using Poisson:

$$x_i \sim Po(a_i\beta_i), \quad y_i \sim Po\left(b_i(\beta_i + f_i\mu_i)\right) \qquad H : \mu_i \equiv 0$$

"variability in the X-ray emission is not compatible with proposed mechanism"

The New York Times

Shingles Vaccine Can Decrease Risk of Dementia, Study Finds

A growing body of research suggests that preventing the viral infection can help stave off cognitive decline.

Matins sans frontières | Rattrapage du lundi 7 avril 2025

RADIO-CANADA Ohdio

Le vaccin contre le zona et la réduction de risque de démence

Lundi 7 avril 2025

Lancer l'écoute
12 min

NBC NEWS

Shingles vaccine may protect against dementia, new study suggests

It's been shown that reactivation of the chickenpox virus can lead to the accumulation of aberrant proteins associated with Alzheimer's.

Shingles is awful, but there may be anoth reason to get vacc : may fight dementia

CTV NEWS

CBC

Shingles vaccine tied to fewer dementia diagnoses, study in Wales suggests

2 potential mechanisms might explain how shingles vaccine could reduce risk of dementia


Article

A natural experiment on the effect of herpes zoster vaccination on dementia

https://doi.org/10.1038/s41586-025-08800-x

Received: 4 November 2023

Markus Eyting[1,2,3,9], Min Xie[1,4,9], Felix Michalik[1,4], Simon Heß[5], Seunghun Chung[1] & Pascal Geldsetzer[1,6,7,8]

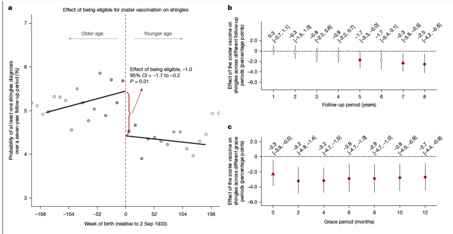


Fig. 2 | The effect of the zoster vaccine on shingles diagnoses. a–c, Effect ... this 10-week increment received in the analysis. For b and c, the MSE-optimal

Scientific question: Does the shingles vaccine reduce the risk of dementia?

natural experiment

Statistical model: "We used regression discontinuity analysis ...
with kernel regression estimates for causal inference"

"receiving the vaccine reduced the probability of a new dementia diagnosis ... by 3.5"%

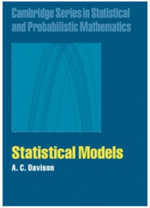
a 20% reduction in relative risk

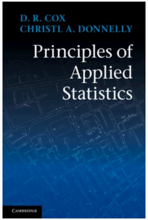
Models and parameters

Why these models?

- motivated by theory: economic, physical, ... *X*-ray jets
- motivated by design: RCT, survey, RDD vaccine
- standard in the literature of that field income
- standard in the publications of that lab breast cancer; world weather attribution
- follow some prescription:
  - binary response — use logistic regression
  - time to event — use PH model
  - time series — use ARMA wine
  - repeated measures — use random effects
  - ...

Some guidance from the experts





- the key feature of a statistical model is that variability is represented using probability distributions
- the art of modelling lies in finding a balance that enables the questions at hand to be answered or new ones posed
- probability models as an aid to the interpretation of data
- perturbations of no intrinsic interest distort an otherwise exact measurement
- substantial natural variability in the phenomenon under study

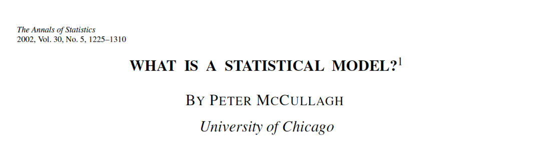
The role of parameters

- probability models very likely be parameterized
- thus defining a class of models $\{f(y;\theta);\theta\in\Theta\}$
- parameters may be finite- or infinite-dimensional parametric vs nonparametric

- ideally one or more parameters represent key aspects of the model for the application at hand

- other parameters complete the specification
- the meaning of various parameters varies with the application

- this sounds simpler than it is


The Annals of Statistics
2002, Vol. 30, No. 5, 1225–1310

WHAT IS A STATISTICAL MODEL?[1]

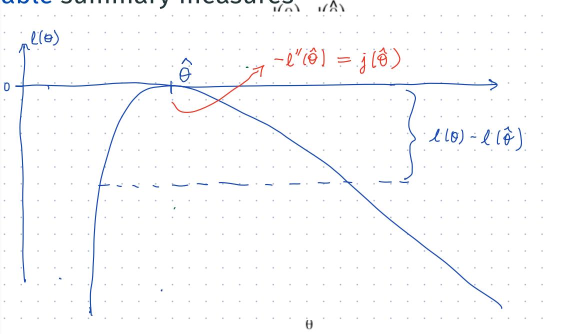
BY PETER MCCULLAGH

University of Chicago

e.g. Box-Cox $y^{\lambda}=x^{\mathrm{T}}\beta+\epsilon$

The likelihood function

- puts the emphasis on the model: $L(\theta; y) \propto f(\boldsymbol{y}; \theta) = \prod_{i=1}^{n} f(y_i; \theta)$ — inverse problem
- provides a convenient way to compare parameter values — e.g. $L(\theta)/L(\hat{\theta})$
- provides reliable summary measures — $\ell(\theta; y) = \log L(\theta; y)$



- can be converted to a probability, given a prior probability for θ — Pfizer vaccine $Bin(162 + 8, \theta)$ via 2 Poissons

Inference and asymptotics

$$(i) \;\; \ell(\theta) = \sum_{i=1}^{n} \log f(y_i; \theta \mid x_i), \quad (ii) \;\; \ell'(\theta) = \sum_{i=1}^{n} \nabla_\theta \log f(y_i; \theta \mid x_i), \quad (iii) \;\; \ell'(\hat\theta) = 0$$

Central Limit Theorem $\quad \frac{1}{\sqrt{n}} \ell'(\theta) \xrightarrow{d} N\{0, I_1(\theta)\}$ $\quad$ observed and expected Fisher information

$\implies$ MLE is approximately normally distributed $\quad J(\theta) = -\ell''(\theta)$

$$\hat\theta \,\dot\sim\, N_p\{\theta, J^{-1}(\hat\theta)\}$$

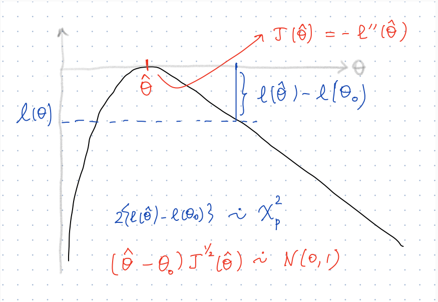
$\implies$ LRT is approximately χ^2 distributed $\quad I(\theta) = \mathbb{E}_\theta\{j(\theta)\}$

$$2\{\ell(\hat\theta) - \ell(\theta)\} \,\dot\sim\, \chi^2_p$$

Large-sample approximation:

$$\hat{\theta} \dot{\sim} N_p\{\theta, J^{-1}(\hat{\theta})\}, \qquad 2\{\ell(\hat{\theta}) - \ell(\theta)\} \dot{\sim} \chi^2_p$$

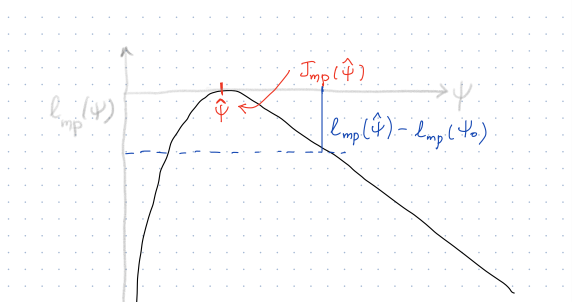
```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   -3.079      0.987   -3.12   0.0018 **
aged1         -0.292      0.754   -0.39   0.6988
stage1         1.373      0.784    1.75   0.0799 .
grade1         0.872      0.816    1.07   0.2850
xray1          1.801      0.810    2.22   0.0263 *
acid1          1.684      0.791    2.13   0.0334 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 40.710  on 22  degrees of freedom
Residual deviance: 18.069  on 17  degrees of freedom
```



A bit too simple

- model $f(y;\theta)$, $\quad \theta \in \mathbb{R}^p$
- $\theta = (\psi, \lambda)$ parameters of interest nuisance parameters
- results above used modified profile log-likelihood function

$$\ell_{\text{mp}}(\psi) = \ell(\psi, \hat\lambda_\psi) - \frac{1}{2}\log|J_{\lambda\lambda}(\psi, \hat\lambda_\psi)|$$



What can go wrong?

- distribution approximations might be poor — likelihood skewed; extremes more relevant
- too many parameters — $p \sim n^{\alpha}, \quad p/n \to C, \quad p/n \to \infty$
- irregular parameter space — $pf(y;\theta_1) + (1-p)f(y;\theta_2), \quad 0 \le p \le 1$
- computational intractability — $L(\theta, \tau; y) = \int_{\mathbb{R}^k} f(y \mid z; \theta) f(z; \tau) dz$
- model is misspecified — true $Y \sim m(y), \quad f(\cdot;\theta) \neq m(\cdot) \quad \forall\theta$

Some approaches to misspecification

1. Classical

Cox 1961,2; Huber 1967; White 1982

- true model $m(\boldsymbol{y})$ fitted model $f(\boldsymbol{y};\theta)$

$\boldsymbol{y} = (y_1, \ldots, y_n)$

$\ell(\theta;\boldsymbol{y}) \equiv \log f(\boldsymbol{y};\theta)$

$\widehat{\theta} \equiv \arg\sup_\theta \ell(\theta;\boldsymbol{y})$

- maximum likelihood estimator $\widehat{\theta}$

- $\widehat{\theta}$ converges to the "closest true value"

KL-divergence

$$\theta^{\mathrm{o}}_m = \arg\min_\theta \int m(\boldsymbol{y}) \log\{\frac{m(\boldsymbol{y})}{f(\boldsymbol{y};\theta)}\} d\boldsymbol{y}$$

- $\widehat{\theta}$ has asymptotic normal distribution, but is not fully efficient

"sandwich variance"

$$\text{a.var.}\,(\widehat{\theta}) = G^{-1}(\theta^{\mathrm{o}}_m), \qquad G(\theta) = J(\theta)I^{-1}(\theta)J(\theta)$$

$I = \mathrm{var}_m(\ell'),\ J = \mathbb{E}_m(-\ell'')$

- change the inference goal, proceed more or less as usual

"we used robust standard errors "

2. More flexible inference functions

Composite likelihood

- true model $m(\boldsymbol{y}_i) = f(\boldsymbol{y}_i;\theta), \boldsymbol{y}_i \in \mathbb{R}^d$ fitted model $\prod_{A\in\mathcal{A}} f(y_{iA};\theta)$ subsets A

- Example: pairwise likelihood $\boldsymbol{y} = (\boldsymbol{y}_1,\ldots,\boldsymbol{y}_n)$

$$L_{pair}(\theta;\boldsymbol{y}) = \prod_{i=1}^n \prod_{s\neq t} f_2(y_{is}, y_{it};\theta)$$

- Example $AR(1)$ likelihood $\boldsymbol{y} = (y_1,\ldots,y_n)$

$$L_{cond}(\theta;\boldsymbol{y}) = \prod_{i=1}^n f(y_i \mid y_{i-1};\theta)$$

interpretation of θ

- Example pseudo-likelihood in spatial models condition on near neighbours; Besag 74

… More flexible inference functions

Quasi-likelihood and generalized estimating equations

$$g\{\mathbb{E}(y_i \mid \boldsymbol{x}_i)\} = g(\mu_i) = \boldsymbol{x}_i^{\mathrm{T}}\beta, \quad \text{var}(y_i \mid \boldsymbol{x}_i) = \sigma^2 V(\mu_i)$$

- estimating equation for β

full distribution unspecified

$$\sum_{i=1}^{n} \frac{\partial \mu_i(\beta)}{\partial \beta} \frac{(y_i - \mu_i)}{V(\mu_i)} = \mathbf{0}$$

column vector

Quadratic inference functions

Qu, Lindsay, Li 2000; Hector 2023

- replace $V^{-1}(\mu_i)$ above with an expansion in basis functions
- apply generalized method of moments

3. More flexible models

- identify one or more parameters of interest — here β
- use a highly flexible specification form for other aspects of the model

- Example: proportional hazards regression — instantaneous failure rate

$$h(t; x, \beta) = h_0(t)\exp(x^{\mathrm{T}}\beta)$$

- Example: empirical likelihood — $T(F)$ to be specified; e.g. $\mathbb{E}_F(Y_i)$

$$\max_F L(F; \boldsymbol{y}), \text{ subject to } T(F) = \theta$$

$L(F; \boldsymbol{y}) = \prod_{i=1}^n F(y_i)$

- Example: semi-parametric regression

$$\mathbb{E}(y \mid T, x) = \psi T + \omega(x)$$

- when does parameter of interest have a stable interpretation — model assumption

Example: exponential matched pairs

- independent exponential pairs (y_{1i}, y_{2i}), $i = 1, \ldots, n$ — $n + 1$ parameters
- rate parameters γ_i/ψ and $\gamma_i\psi$, respectively
- ψ common parameter of interest γ_i pair-specific nuisance parameters

- likelihood function

$$L(\psi, \gamma; \boldsymbol{y}) \propto \prod_{i=1}^{n} \gamma_i^2 \exp\{-\gamma_i(\frac{y_{1i}}{\psi} + \psi y_{2i})\}$$

- possibilities for eliminating nuisance parameters
  - profile (concentrated) likelihood — maximize over γ
  - marginal likelihood: $f(\boldsymbol{t}; \psi) = \prod_{i=1}^{n} f(t_i; \psi)$ — $t_i = y_{1i}/y_{2i}$
  - random effects $\gamma_i \sim g(\cdot; \lambda)$ — more efficient, if ...

... Example: exponential matched pairs

- independent exponential pairs (y_{1i}, y_{2i}), $i = 1, \dots, n$ — $n+1$ parameters
- rate parameters γ_i/ψ and $\gamma_i\psi$, respectively
- random effects: $\gamma_i \sim \text{Gamma}(\alpha, \beta)$ — λ = (shape, rate)
- likelihood function

$$L(\psi, \alpha, \beta; \boldsymbol{y}) \propto \prod_{i=1}^{n} \int \gamma_i^2 \exp\{-\gamma_i(\frac{y_{1i}}{\psi} + \psi y_{2i})\} g(\gamma_i; \alpha, \beta) d\gamma_i$$

- orthogonality:

$$\mathbb{E}_{gamma}\left\{-\frac{\partial^2 \log L(\psi, \alpha, \beta)}{\partial\psi\partial\alpha}\right\} = 0, \quad \mathbb{E}_{gamma}\left\{-\frac{\partial^2 \log L(\psi, \alpha, \beta)}{\partial\psi\partial\beta}\right\} = 0$$

- even better

$$\mathbb{E}_{m}\left\{-\frac{\partial^2 \log L(\psi, \alpha, \beta)}{\partial\psi\partial\alpha}\right\} = 0, \quad \mathbb{E}_{m}\left\{-\frac{\partial^2 \log L(\psi, \alpha, \beta)}{\partial\psi\partial\beta}\right\} = 0$$

any random effects distribution

... Example: exponential matched pairs

- independent exponential pairs $(y_{1i}, y_{2i}), \quad i = 1, \ldots, n$ — $n + 1$ parameters
- rate parameters γ_i/ψ and $\gamma_i\psi$, respectively
- random effects: $\gamma_i \sim \text{Gamma}(\alpha, \beta)$ — λ = (shape, rate)
- likelihood function

$$L(\psi, \alpha, \beta; \boldsymbol{y}) \propto \prod_{i=1}^{n} \int \gamma_i^2 \exp\{-\gamma_i(\frac{y_{1i}}{\psi} + \psi y_{2i})\} g(\gamma_i; \alpha, \beta) d\gamma_i$$

- even better

$$\mathbb{E}_m\left\{-\frac{\partial^2 \log L(\psi, \alpha, \beta)}{\partial\psi\partial\alpha}\right\} = 0, \quad \mathbb{E}_m\left\{-\frac{\partial^2 \log L(\psi, \alpha, \beta)}{\partial\psi\partial\beta}\right\} = 0$$

- and

$$\hat{\psi} \xrightarrow{p} \psi$$

any random effects distribution

Formalisation

PNAS RESEARCH ARTICLE | STATISTICS OPEN ACCESS

On the role of parameterization in models with a misspecified nuisance component

Heather S. Battey[a,2,1] and Nancy Reid[b,2,1]

Contributed by Nancy Reid; received February 8, 2024; accepted July 23, 2024; reviewed by Emmanuel J. Candès and Edward I. George

August 30, 2024 | 121 (36) e2402736121

Significance

Towards a formalisation

- true model $m(\boldsymbol{y})$ with parameter ψ and true value ψ_*
- fitted model $f(\boldsymbol{y}; \psi, \lambda)$ same parameter of interest, (many) nuisance parameters

  interpretation of ψ is stable
- we know maximum likelihood estimates $(\hat{\psi}, \hat{\lambda}) \xrightarrow{p} (\psi_m^0, \lambda_m^0)$ KL divergence
- assume no value of $\lambda \in \Lambda$ gives back $m(\cdot)$ 'truly' misspecified
- Does $\psi_m^0 = \psi_*$? need $\mathbb{E}_m\{\partial\ell(\psi_*, \lambda_m^0)/\partial\psi\} = 0$ (1) λ_m^0 unknown
- can be easier to show $\mathbb{E}_m\{\partial\ell(\psi_*, \lambda)/\partial\psi\} = 0 \quad \forall\lambda$ (2)
- Result 1: (1) $\equiv$ (2) $\iff$ ψ_* is m-orthogonal to Λ: $\forall\lambda \quad \mathbb{E}_m\left\{\frac{\partial^2\ell(\psi,\lambda)}{\partial\psi\partial\lambda}\right\} = 0$

BR: Prop 1.1

... towards a formalisation

- true model $m(\boldsymbol{y})$ with parameter ψ and true value ψ_*
- fitted model $f(\boldsymbol{y};\psi,\lambda)$, maximum likelihood estimate $\widehat{\psi}$
- Result 1: m-orthogonal parameters lead to consistent MLE $\psi_m^0 = \psi_*$

- But, $\widehat{\psi}$ can be consistent without this requirement

- Result 2: A weaker requirement: if BR: Prop 1.2

$$I^{\psi\psi}\mathbb{E}_m\left\{\frac{\partial\ell(\psi_*,\lambda)}{\partial\psi}\right\} + I^{\psi\lambda}\mathbb{E}_m\left\{\frac{\partial\ell(\psi_*,\lambda)}{\partial\lambda}\right\} = 0, \quad \forall\,\lambda,\ \text{then} \quad \psi_m^0 = \psi_*$$

$$I = I(\psi_*,\lambda) = \mathbb{E}_m\left\{-\frac{\partial^2\ell(\psi_*,\lambda)}{\partial\theta\partial\theta^{\mathrm{T}}}\right\}, \qquad I^{-1} = \begin{pmatrix} I^{\psi\psi} & I^{\psi\lambda} \\ I^{\lambda\psi} & I^{\lambda\lambda}\end{pmatrix},$$

still too strong

Parameter orthogonality

- we can often establish parameter orthogonality in the assumed model $f(\boldsymbol{y}; \psi, \lambda)$
- all expectations with respect to this assumed model
- this is not usually the same as parameter orthogonality in the true model $m(\boldsymbol{y}; \psi)$
- Result 3: a special case BR: Prop 1.3

  If $\nabla^2_{\psi\lambda}\ell(\psi, \lambda; \boldsymbol{y})$ a function of $S = (S_1, \dots, S_k)$, and is additive in S, and

  $$E_m(S_j) = \mathbb{E}_{(\psi,\lambda)}(S_j)$$

  then assumed-model orthogonality $\implies$ true-model orthogonality
- easier: information calculations under assumed model

Parameter Symmetry

- matched exponential pairs is a scale model: $\mathbb{E}(Y_{1i}) = \psi/\gamma_i; \quad \mathbb{E}(y_{2i}) = 1/(\psi\gamma_i)$
- the parameter of interest enters symmetrically
- the proof of consistency repeatedly uses the change of variables to y_{1i}/y_{2i} and $y_{1i}y_{2i}$
- how to generalize this observation?

... Parameter Symmetry

- from earlier results, want extended orthogonal parametrization $\ell = \log L$

$$\mathbb{E}_m\{-\partial^2\ell(\psi,\lambda)/\partial\psi\partial\lambda^{\mathrm{T}}\} = 0$$

or at least at ψ_*

- we don't know the true model m, so can't check this
- exponential matched pairs is a group models
- parametrization ensures cancellation of terms

- Result 4: If the joint distribution of Y_1, Y_2 is parametrized ψ-symmetrically, and this parametrization induces anti-symmetry on the ψ-score function, then

$$\psi^* \perp_m \Lambda, \quad \mathbb{E}_m\{\partial\ell(\psi_*,\lambda)/\partial\psi\} = 0, \text{ which implies}$$

$$\widehat{\psi} \text{ is consistent} \qquad \psi_m^0 = \psi_*.$$

- Result 5: a version of Result 4 for two-group problems — stratified not matched

BR: Props 2.1, 2.2

… Formalization and Parameter Symmetry

- Result 4: If the joint distribution of Y_1, Y_2 is parametrized ψ-symmetrically, and this parametrization induces anti-symmetry on the ψ-score function, then

$$\psi^* \perp_m \Lambda, \quad \mathbb{E}_m\{\partial\ell(\psi_*, \lambda)/\partial\psi\} = 0, \text{ which implies}$$

$$\widehat{\psi} \text{ is consistent} \qquad \psi_m^0 = \psi_*.$$

- Example: Location family $\qquad g_\psi \in$ location group

$$f_{Y_1}(y_1; \lambda + \psi) = f_U(y_1 - \psi; \lambda),$$
$$f_{Y_2}(y_2; \lambda - \psi) = f_U(y_2 + \psi; \lambda)$$

- Example: Scale family $\qquad g_\psi \in$ scale group

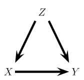
$$f_{Y_1}(y_1; \lambda\psi) = f_U(y_1/\psi; \lambda)(1/\psi),$$
$$f_{Y_2}(y_2; \lambda/\psi) = f_U(y_2\psi; \lambda)\psi,$$
$$U, \stackrel{d}{=} g_\psi^{-1} Y_1 \stackrel{d}{=} g_\psi Y_2$$

Overview

- parameter of interest ψ is well-defined
- model with nuisance parameters may be misspecified — random effects
- when can we recover the true value of ψ
- does parameter orthogonality play a role?

- yes, it does, but may be difficult to verify directly — $\mathbb{E}_m$
- models based on groups satisfy this orthogonality
- with particular parameter structure

- most natural examples seem to involve misspecified random effects — GLM disp
- another example is marginal structural model in a 'frugal parameterization'
- propensity score is the nuisance; other aspects correspond to ψ — Evans & Didelez (2024)
- E&D model has a parameter space cut, hence orthogonal



Tentative conclusions, further work

- Results above only establish consistency
- asymptotic variance is much more difficult — although estimating it might be okay

- in the matched pairs examples, nuisance parameters treated as arbitrary constants can be eliminated by transformation to conditional or marginal distributions
- effectively assuming an arbitrary (nonparametric) mixing distribution
- less efficient when the random effects model is correct

- orthogonality under assumed model $\mathbb{E}_\theta\{-\partial^2\ell(\theta)/\partial\theta\partial\theta^{\mathrm{T}}\} = 0$ — $\theta = (\psi, \lambda)$
- *m*-orthogonality under true model $\mathbb{E}_m\{-\partial^2\ell(\theta)/\partial\theta\partial\theta^{\mathrm{T}}\} = 0$
- connection to Neyman orthogonality? — decorrelated score

$$\partial\ell(\psi,\lambda)/\partial\psi - w^{\mathrm{T}}\partial\ell(\psi,\lambda)/\partial\lambda, \quad w = I_{\psi\lambda}I_{\lambda\lambda}^{-1}$$

- extension to general estimating equations important in 2-debiased ML

Chernozhukov et al 2018, Ning et al 2017, Jorgensen & Knudsen 2004

Conclusion

What can go wrong?

- the distributional approximations might be poor — likelihood skewed; extremes more relevant
- too many parameters — $p \sim n^\alpha, \quad p/n \to C, \quad p/n \to \infty$
- irregular parameter space — $pf(y;\theta_1) + (1-p)f(y;\theta_2), \quad 0 \le p \le 1$
- computational intractability — $L(\theta, \tau; y) = \int_{\mathbb{R}^k} f(y \mid z;\theta) f(z;\tau) dz$
- model is misspecified — true $Y \sim m(y), \quad f(\cdot;\theta) \neq m(\cdot) \ \forall\theta$

Some solutions

• the normal and/or χ^2 approximations might be poor	more accurate approximations different limit theory	HOA extremes
• too many parameters	new asymptotic theory ($p \sim n$) regularization ($p > n$)	Sur & Candès 19; Zhao et al 22 Lasso, SCAD, MCP
• irregular parameter space	different asymptotic theory, e.g. $\chi^2_D \to \Sigma\lambda_j\chi^2_{1j}$	Battey & McCullagh 24
• computational intractability	composite likelihood	Genton et al 15

... and so much more!

Examples

- climate change

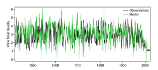
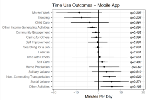


Figure 4. Observed series of wine quality (average; black) from 1420 to 2019 and series obtained with a statistical model calibrated in 1781–1800 (green). The model is explained in Sect. 3.

linear regression; time series

- guaranteed income

Figure 5: Time Use Results: Mobile App



linear regression; treatment effect

- Xray jets

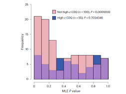


Fig. 3 | Histogram of the single-region *P* values from the directional test, not adjusted for multiple comparison. In pink, the subset of sources that

Poisson distribution

- shingles vaccine

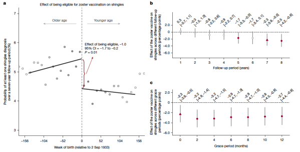


Fig. 2 | The effect of the zoster vaccine on shingles diagnoses. a–c, Effect ... this 10-week increment received in the analysis. For **b** and **c**, the MSE-optimal

nonparametric regression

This just in

World's biggest companies have caused $28-trillion in climate damage: study

SETH BORENSTEIN WASHINGTON

The world's biggest corporations have caused US$28-trillion in climate damage, a new study estimates as part of an effort to make it easier for people and governments to hold companies financially accountable, like the tobacco giants have been.

A Dartmouth College research team came up with the estimated pollution caused by 111 companies, with more than half of the total dollar figure coming from 10 fossil fuel providers: Saudi Aramco, Gazprom, Chevron, ExxonMobil, BP, Shell, National Iranian Oil Co., Pemex, Coal India and the British Coal Corporation.

For comparison, US$28-trillion is a shade less than the sum of all goods and services produced in the United States last year.

At the top of the list, Saudi Aramco and Gazprom have each

and Dr. Mankin said.

The researchers started with known final emissions of the products – such as gasoline or electricity from coal-fired power plants – produced by the 111 biggest carbon-oriented companies going as far back as 137 years, because that's as far back as any of the companies' emissions data go and carbon dioxide stays in the air for much longer than that. They used 1,000 different computer simulations to translate those emissions into changes for Earth's global average surface temperature by comparing it to a world without that company's emissions.

Using this approach, they determined that pollution from Chevron, for example, has raised the Earth's temperature by .025 C.

The researchers also calculated how much each company's pollution contributed to the five hottest days of the year using 80

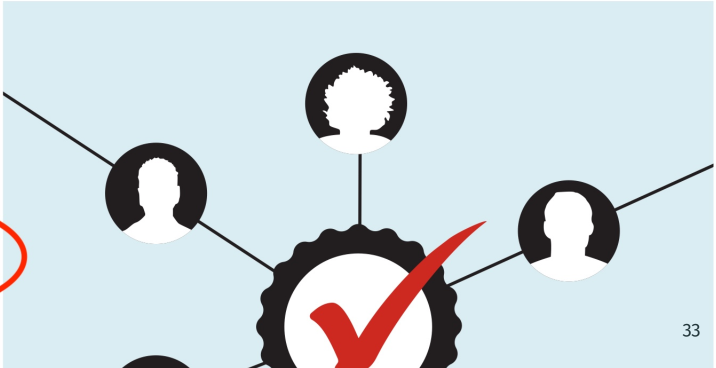
ence gets and the better we know what makes a difference and what does not," Dr. Otto said. So far, no climate liability lawsuit against a major carbon emitter

has been successful, but maybe showing "how overwhelmingly strong the scientific evidence" is can change that, she said.

In the past, damage caused by

individual companies were lost in the noise of data, so it couldn't be calculated, Dr. Callahan said.

ASSOCIATED PRESS



Climate damage

Nature April 24

Perspective

Carbon majors and the scientific case for climate liability

https://doi.org/10.1038/s41586-025-08751-3

Received: 27 March 2023

Accepted: 6 February 2025

Published online: 23 April 2025

Check for updates

Christopher W. Callahan[1,2] ✉ & Justin S. Mankin[1,2,3,4] ✉

Will it ever be possible to sue anyone for damaging the climate? Twenty years after this question was first posed, we argue that the scientific case for climate liability is closed. Here we detail the scientific and legal implications of an 'end-to-end' attribution that links fossil fuel producers to specific damages from warming. Using scope 1 and 3 emissions data from major fossil fuel companies, peer-reviewed attribution methods and advances in empirical climate economics, we illustrate the trillions in economic losses attributable to the extreme heat caused by emissions from individual

August 2023

world weather attribution

Home About Analyses News Peer reviewed research

On average, wildfires burn about 2.5 million hectares in Canada each year. In 2023, wildfires have already burned nearly 14 million hectares. Photo by Audrey Marcoux, SOPFEU.

Home > Wildfire > Climate change more than doubled the likelihood of extreme fire weather conditions in Eastern Canada

Climate change more than doubled the likelihood of extreme fire weather conditions in Eastern Canada

22 August, 2023

During May and June 2023 Canada witnessed exceptionally extreme fire-weather conditions, leading to extensive wildfires that burned over 13 million

Full study

- Download the full study: Climate change more than doubled the likelihood of extreme fire weather conditions in Eastern Canada (26 pages, 1.8MB)

… August 2023

- "As a measure of anthropogenic climate change we use smoothed GMST"

  Global Mean Surface Temperature

- "Methods for observational and model analysis … and synthesis are used according to the World Weather Attribution Protocol" Philip et al. 2020

1. trend using observational data
2. find climate models consistent with 1.
3. compare predictions from 1. and 2.
4. synthesize results in 3. to provide conclusions

… August 2023

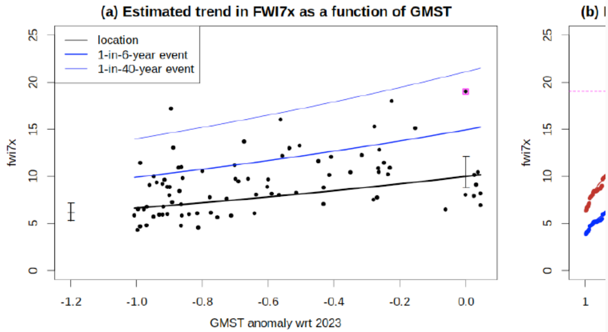


Figure 7: (a) Linear trend in ERA5 FWI7x as a function of GMST.
parameter of the fitted distribution, and the blue lines show estimat
lines show the 95% confidence interval for the location parameter f

Attribution of climate damage

Perspective

Carbon majors and the scientific case for climate liability

https://doi.org/10.1038/s41586-025-08751-3

Received: 27 March 2023

Accepted: 6 February 2025

Published online: 23 April 2025

Check for updates

Christopher W. Callahan[1,2] ✉ & **Justin S. Mankin**[1,2,3,4] ✉

Will it ever be possible to sue anyone for damaging the climate? Twenty years after this question was first posed, we argue that the scientific case for climate liability is closed. Here we detail the scientific and legal implications of an 'end-to-end' attribution that links fossil fuel producers to specific damages from warming. Using scope 1 and 3 emissions data from major fossil fuel companies, peer-reviewed attribution methods and advances in empirical climate economics, we illustrate the trillions in economic losses attributable to the extreme heat caused by emissions from individual

... Attribution of climate damage

- climate model simulations of changes in global mean surface temperature over time
- simulate historical climates and “counter-factual” climates
- counter-factual climates leave out emissions of a given entity

- next convert surface temperature change to likelihood of extreme heat events

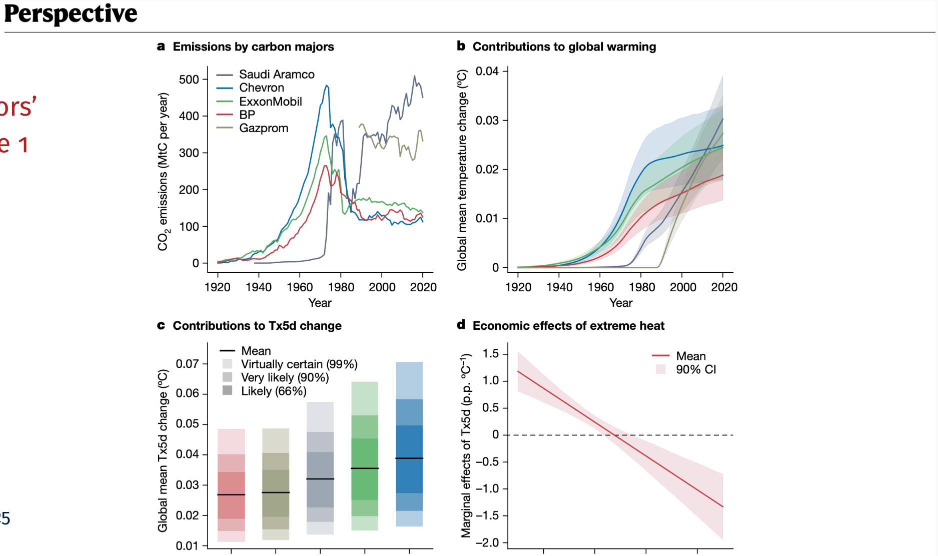
linear regression

- finally convert extreme heat events to economic cost

Specifically, we use the coefficients from the following regression estimated using ordinary least squares:

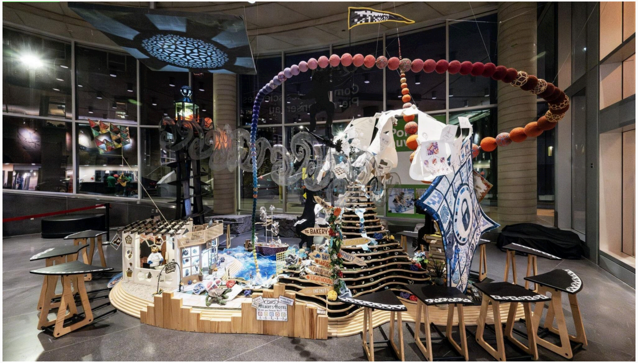
$$g_{it} = \alpha_1 T_{it} + \alpha_2 T_{it}{}^2 + \beta_1 \mathrm{Tx}_{it} + (\beta_2 \mathrm{Tx}_{it} \times T_{it}) + \gamma_1 V_{it} + (\gamma_2 V_{it} \times A_i) + \pi P_{it} + \mu_i + \delta_t + \varepsilon_{it} \quad (1)$$

Authors' Figure 1



Statistics is everywhere!

Thank you



Colloque
des sciences mathématiques du Québec

CENTRE DE RECHERCHES MATHÉMATIQUES

References i

Battey, H.S. & Cox, D.R. (2020). High dimensional nuisance parameters: an example from parametric survival analysis. *Information Geometry* **3** 119–148. matched pairs exp

Battey, H.S. & McCullagh, P. (2024). An anomaly arising in the analysis of processes with more than one source of variability. *Biometrika* **111**, 677–689.

Battey, H.S. & Reid, N. (2024). On the role of parametrization in models with a misspecified nuisance component. arxiv *PNAS*, to appear.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. R. Statist. Soc.* B **36**, 192–236.

Chernozhukov, V. et al. (2018). Double debiased machine learning. *Econometrics Journal* **21**, C1–C68. doi: 10.1111/ectj.12097 Neyman orthogonality

Cox, D.R. (1961). Tests of separate families of hypotheses. *4th Berkeley Symposium* **1**, 105–123.

References ii

Cox, D.R. (1962). Further results on tests of separate families of hypotheses. *J. R. Statist. Soc.* B **24**, 406–424.

Cox, D.R. (1990). Role of models in statistical analysis. *Statist. Sci.* **5**, 169–174.

Cox, D.R. and Donnelly, C.A. (2011). *Principles of Applied Statistics*. Cambridge University Press.

Davison, A.C. (2003). *Statistical Models*. Cambridge University Press.

Genton, M.G., Padoan, S.A. and Sang, H. (2015). Multivariate max-stable processes. *Biometrika* **102**, 215–230.

Hector, E. (2023). Fused mean structure learning in data integration with dependence. *Canad. J. Statist.*

Huber, P.J. (1967). The behaviour of maximum likelihood estimates under non-standard conditions. *5th Berkeley Symposium* **1**, 221–233.

Jorgensen, B. & Knudsen, S.J. (2004). Parameter orthogonality and bias adjustment for estimating functions. *Scand. J. Statist.* **31**, 93–114.

McCullagh, P. (2002). What is a statistical model? (with discussion). *Ann. Statist.* **30**, 1225–1310.

Ning, Y. & Liu, H. (2017). A general theory of hypothesis tests and confidence regions for sparse high-dimensional models. *Annals of Statistics* **45**, 158–195. decorrelated score

Qu, A.m Lindsay, B.G. and Li, B. (2000). Improving generalised estimating equations using quadratic inference functions. *Biometrika* **87**, 823–836.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50**, 1–25.

Zhao, Q., Sur, P. and Candès, E. (2022). The asymptotic distribution of the MLE in high-dimensional logistic models: Arbitrary covariance. *Bernoulli* **28**, 1835–1861.