

Inferring the causal effect of journals on citations

V.A. Traag

Centre for Science and Technology Studies (CWTS)
Leiden University, the Netherlands

Complexity Science Hub Vienna, 5 February 2021

High-impact journals are more highly cited.

High-impact journals are more highly cited.

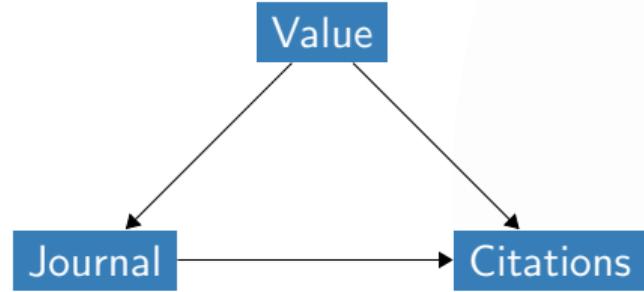
Why?

High-impact journals are more highly cited.

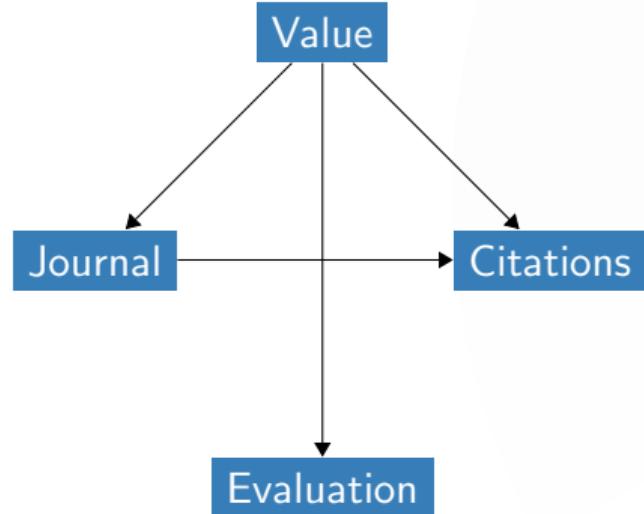
Why?

- ① High-impact journals publish more “citable” articles?
- ② Increase citations because articles appeared in high-impact journal?

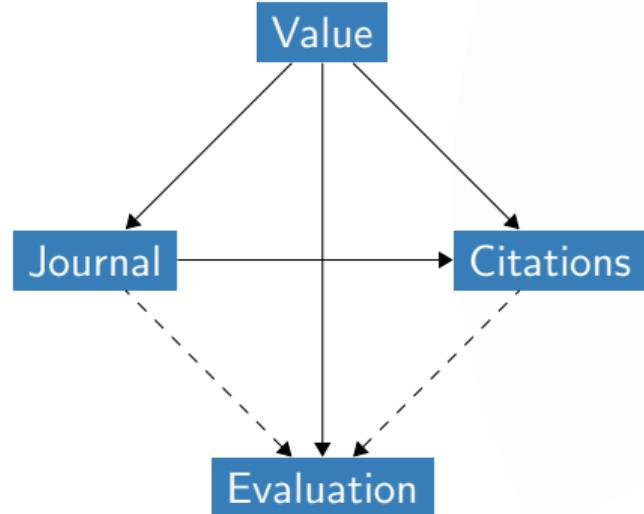
Causal model



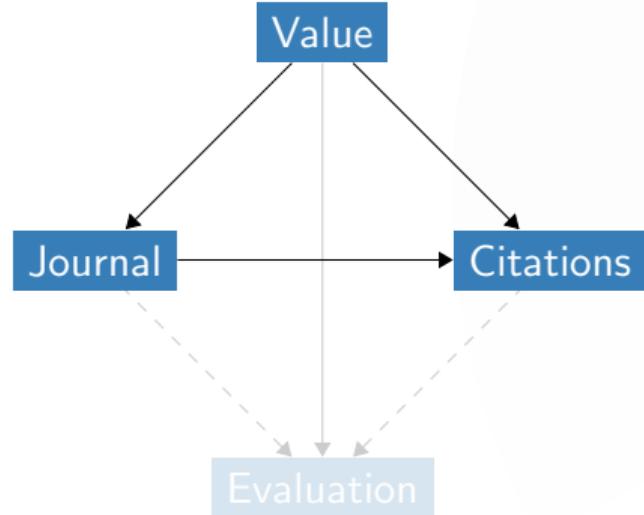
Causal model



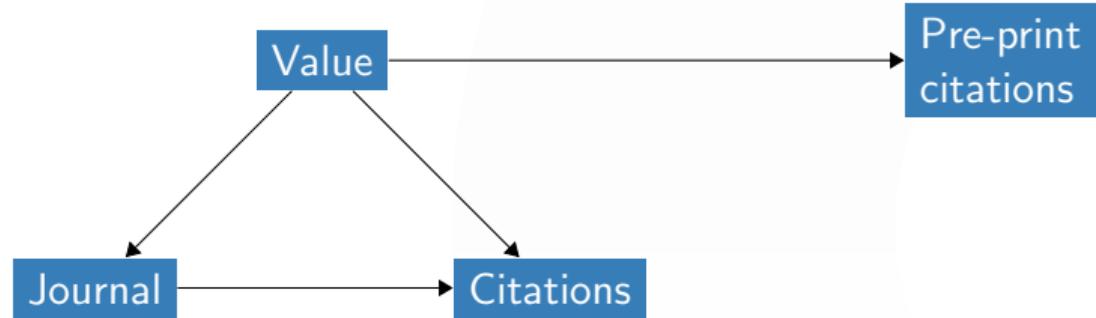
Causal model



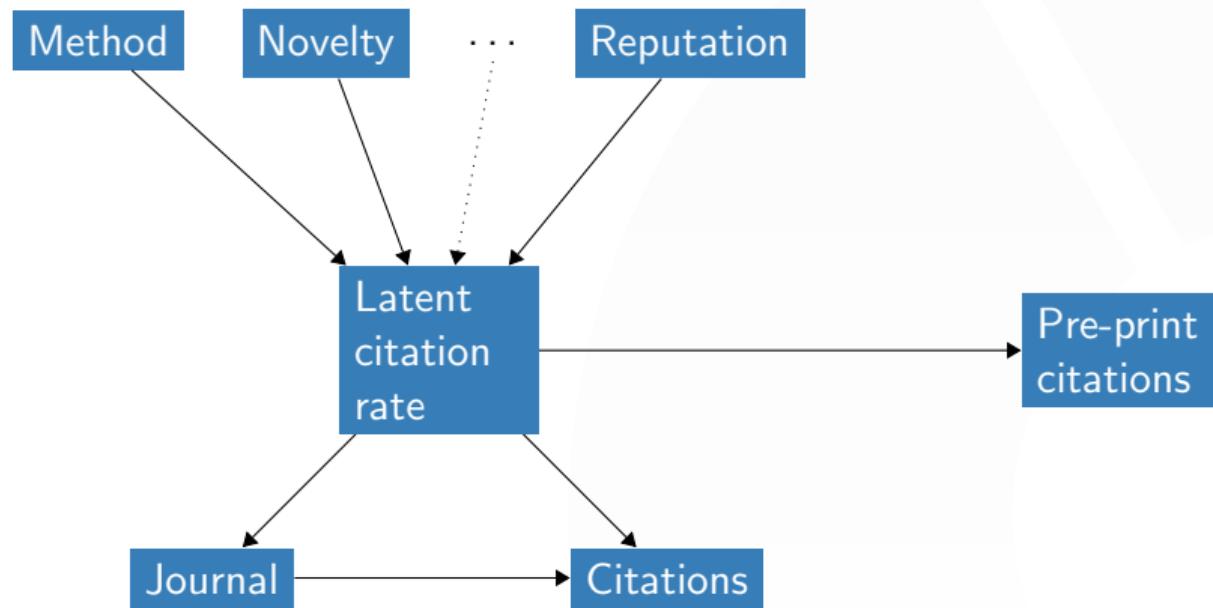
Causal model



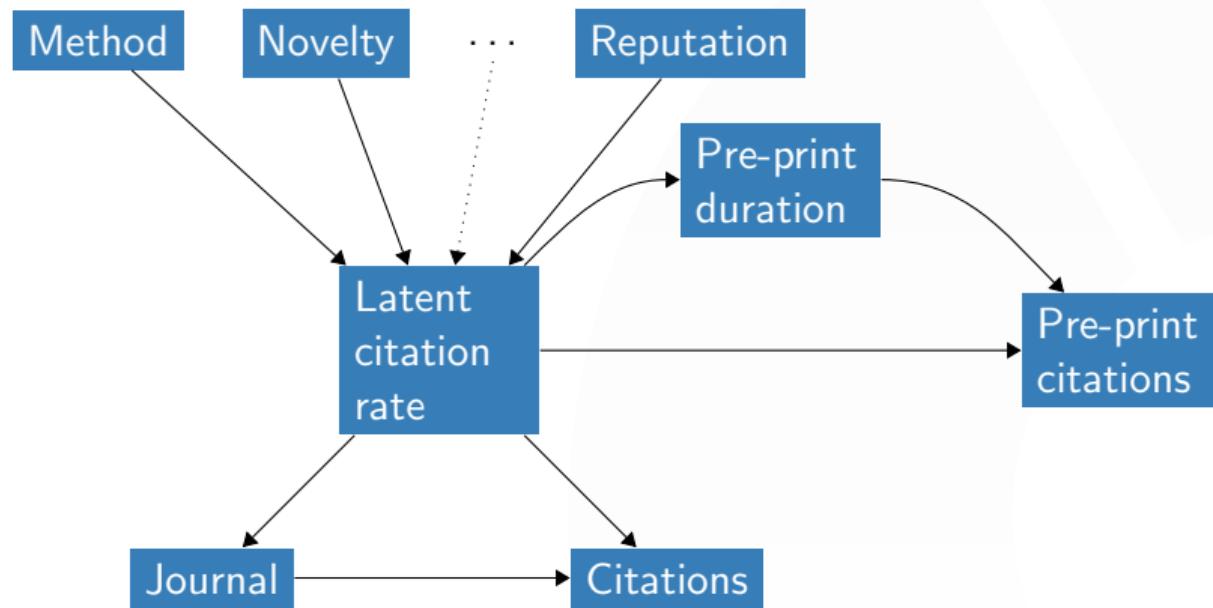
Causal model



Causal model



Causal model



Citations $c_i(t)$ that publication i receives at time t

$$c_i(t) \sim \text{Poisson} \left[\lambda_i(t) \times f_i(t) \times (m + C_i(t-1)) \right]$$

Wang, D, Song, C & Barabási, AL. *Science* 342, 127–132 (2013).

Citations $c_i(t)$ that publication i receives at time t

$$c_i(t) \sim \text{Poisson} \left[\lambda_i(t) \times f_i(t) \times (m + C_i(t-1)) \right]$$

- Effective citation rate

Wang, D, Song, C & Barabási, AL. *Science* 342, 127–132 (2013).

Citations $c_i(t)$ that publication i receives at time t

$$c_i(t) \sim \text{Poisson} \left[\lambda_i(t) \times f_i(t) \times (m + C_i(t - 1)) \right]$$

- Effective citation rate
- Temporal decay

Wang, D, Song, C & Barabási, AL. *Science* 342, 127–132 (2013).

Model

Citations $c_i(t)$ that publication i receives at time t

$$c_i(t) \sim \text{Poisson} \left[\lambda_i(t) \times f_i(t) \times (m + C_i(t - 1)) \right]$$

- Effective citation rate
- Temporal decay
- Cumulative advantage

Wang, D, Song, C & Barabási, AL. *Science* 342, 127–132 (2013).

Model

Citations $c_i(t)$ that publication i receives at time t

$$c_i(t) \sim \text{Poisson} \left[\lambda_i(t) \times f_i(t) \times (m + C_i(t - 1)) \right]$$

- Effective citation rate
- Temporal decay
- Cumulative advantage

Effective citation rate

Wang, D, Song, C & Barabási, AL. *Science* 342, 127–132 (2013).

Model

Citations $c_i(t)$ that publication i receives at time t

$$c_i(t) \sim \text{Poisson} \left[\lambda_i(t) \times f_i(t) \times (m + C_i(t-1)) \right]$$

- Effective citation rate
- Temporal decay
- Cumulative advantage

Effective citation rate

Pre-publication citation rate $\lambda_i(t) = \phi_i$

Wang, D, Song, C & Barabási, AL. *Science* 342, 127–132 (2013).

Model

Citations $c_i(t)$ that publication i receives at time t

$$c_i(t) \sim \text{Poisson} \left[\lambda_i(t) \times f_i(t) \times (m + C_i(t-1)) \right]$$

- Effective citation rate
- Temporal decay
- Cumulative advantage

Effective citation rate

Pre-publication citation rate $\lambda_i(t) = \phi_i \leftarrow$ Latent citation rate

Wang, D, Song, C & Barabási, AL. *Science* 342, 127–132 (2013).

Model

Citations $c_i(t)$ that publication i receives at time t

$$c_i(t) \sim \text{Poisson} \left[\lambda_i(t) \times f_i(t) \times (m + C_i(t-1)) \right]$$

- Effective citation rate
- Temporal decay
- Cumulative advantage

Effective citation rate

$$\begin{aligned} \text{Pre-publication citation rate} \quad \lambda_i(t) &= \phi_i \xleftarrow{\text{Latent citation rate}} \\ \text{Post-publication citation rate} \quad \lambda_i(t) &= \phi_i \times \theta_{J_i} \end{aligned}$$

Wang, D, Song, C & Barabási, AL. *Science* 342, 127–132 (2013).

Model

Citations $c_i(t)$ that publication i receives at time t

$$c_i(t) \sim \text{Poisson} \left[\lambda_i(t) \times f_i(t) \times (m + C_i(t-1)) \right]$$

- Effective citation rate
- Temporal decay
- Cumulative advantage

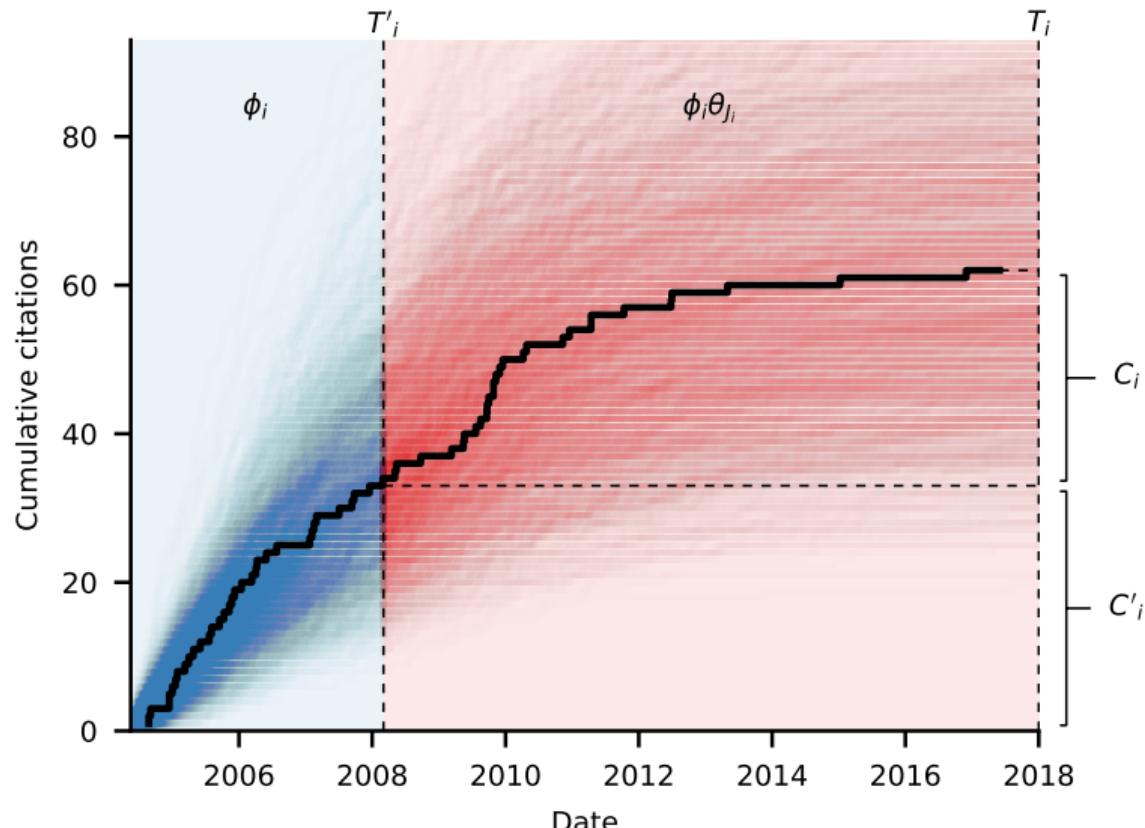
Effective citation rate

$$\text{Pre-publication citation rate} \quad \lambda_i(t) = \phi_i \quad \xleftarrow{\text{Latent citation rate}}$$

$$\text{Post-publication citation rate} \quad \lambda_i(t) = \phi_i \times \theta_{J_i} \quad \xleftarrow{\text{Journal citation multiplier}}$$

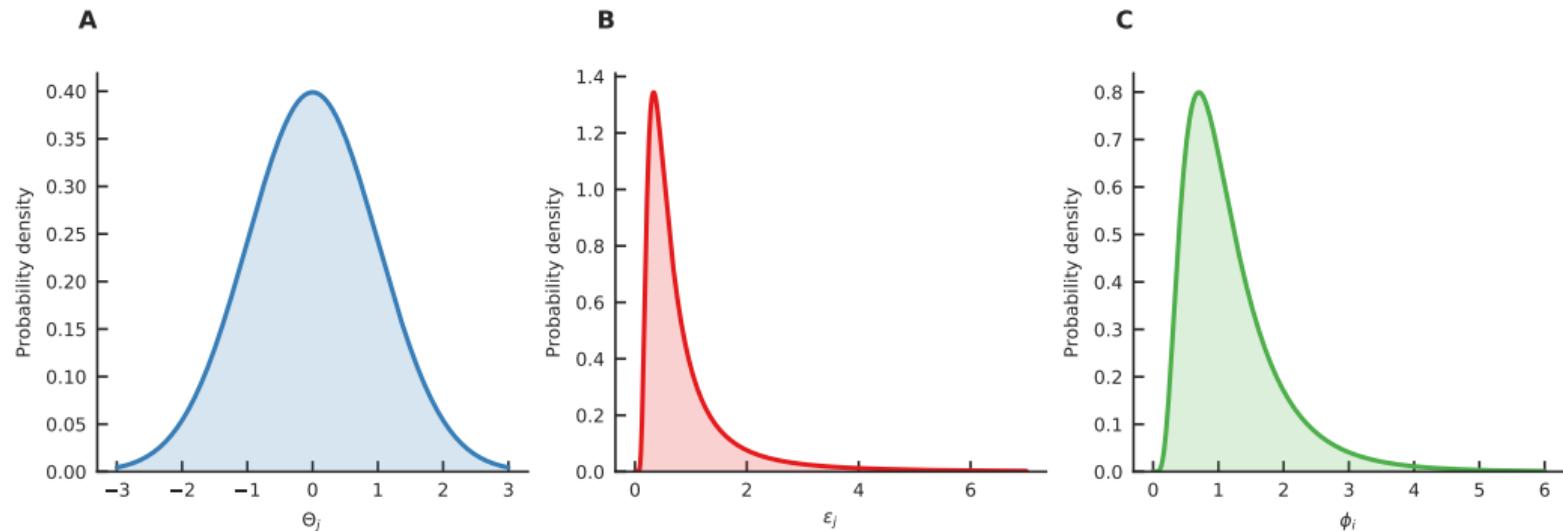
Wang, D, Song, C & Barabási, AL. *Science* 342, 127–132 (2013).

Citation dynamics (astro-ph/0405353)



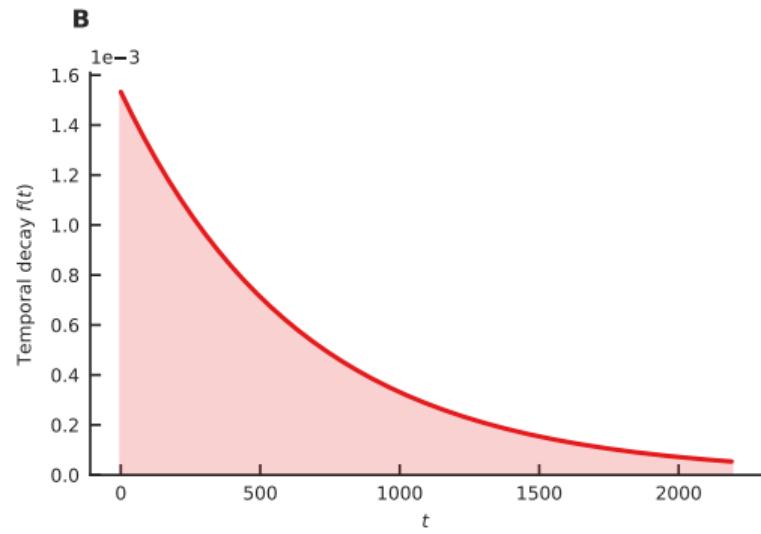
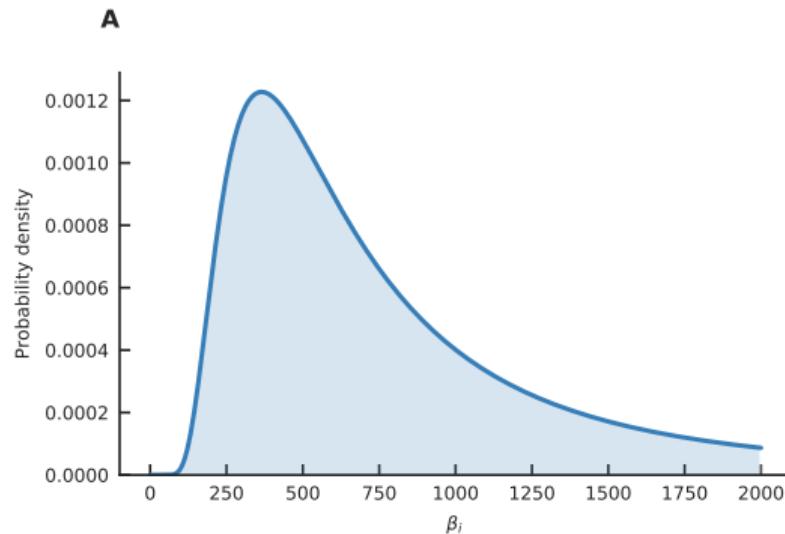
Hierarchical Bayesian model – Latent citation rate prior

Each article i in journal j , $\phi_i \sim \text{LogNormal}(\Phi_j, \varepsilon_j)$.



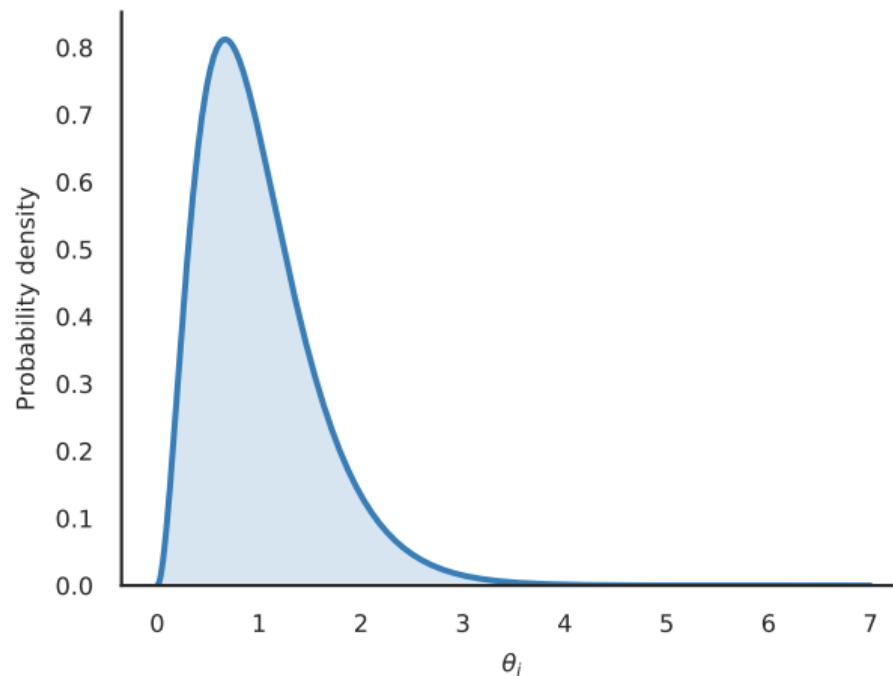
Hierarchical Bayesian model – Temporal decay prior

Decay $f_i(t)$ is exponential, with inverse rate β_i (prior average ≈ 3 years).



Hierarchical Bayesian model – Journal citation multiplier prior

We assume θ_j is centered around 1 (i.e. no increase/decrease).



arXiv

- Collect 727 186 preprints with DOI from arXiv ($\approx 54\%$ of all arXiv preprints).
- Get publication date from CrossRef using DOI.
 - ▶ 521 512 preprints put on arXiv before published.
- Use arXiv subject for scientific fields.

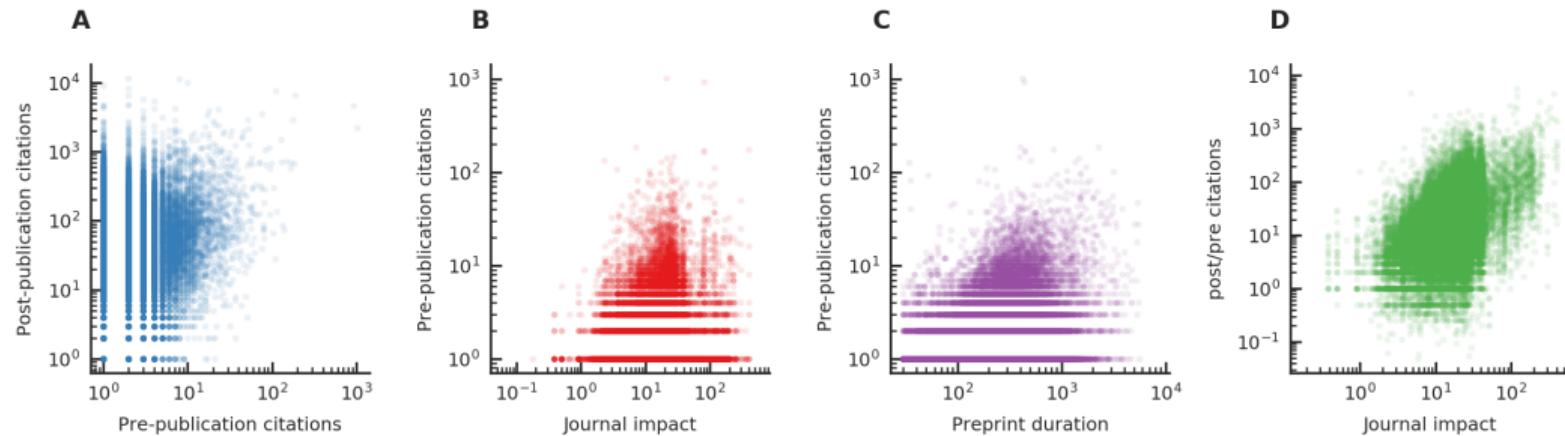
arXiv

- Collect 727 186 preprints with DOI from arXiv ($\approx 54\%$ of all arXiv preprints).
- Get publication date from CrossRef using DOI.
 - ▶ 521 512 preprints put on arXiv before published.
- Use arXiv subject for scientific fields.

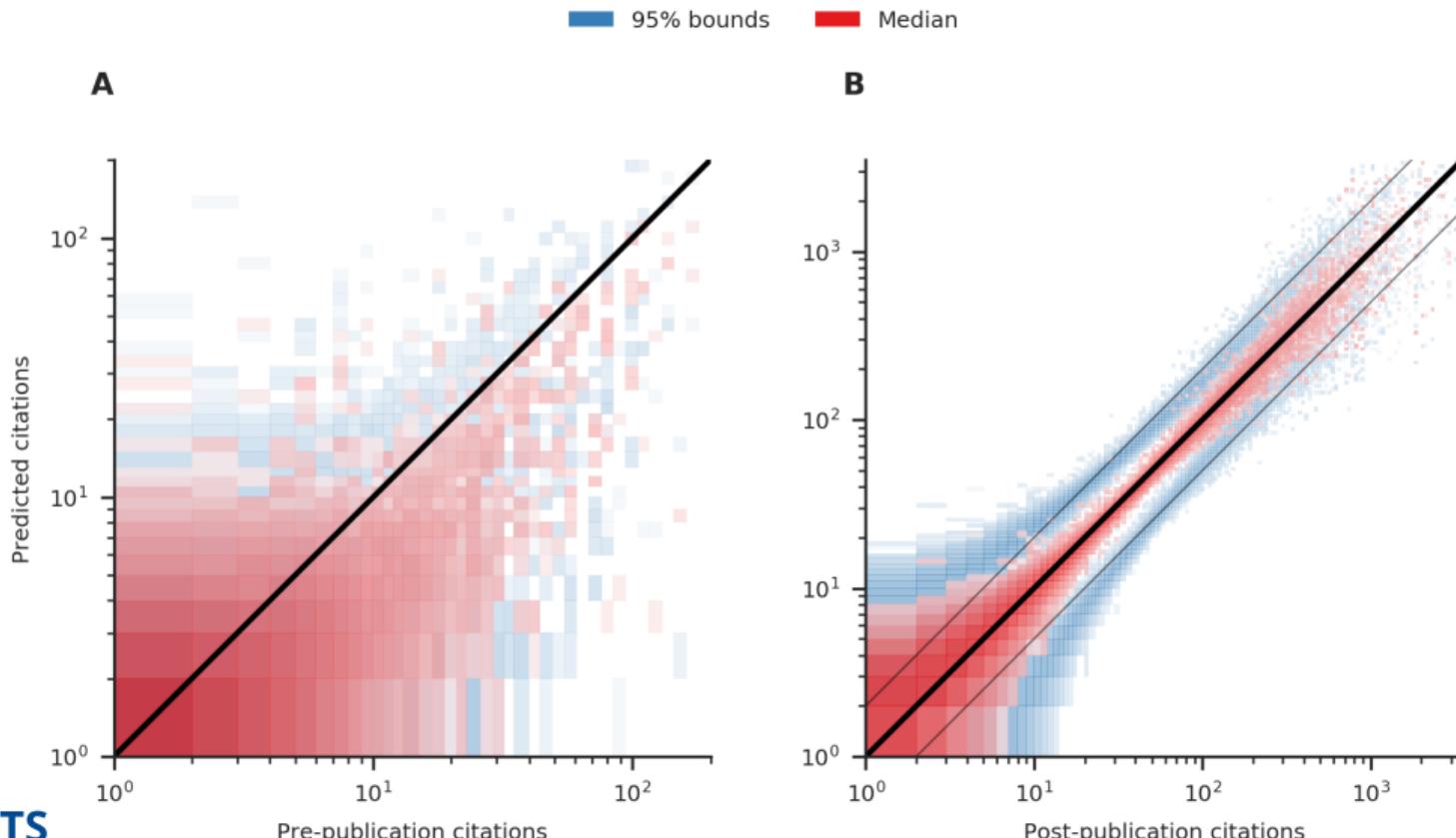
Scopus

- Use Scopus to get citations to preprint (using arXiv identifier) and published version.
- Pre-publication citations: before publication date (156 528).
- Post-publication citations: after publication date (15 939 887).
- Calculate journal impact from Scopus.
 - ▶ Average citations in 5 years for articles & reviews.

Citations and impact



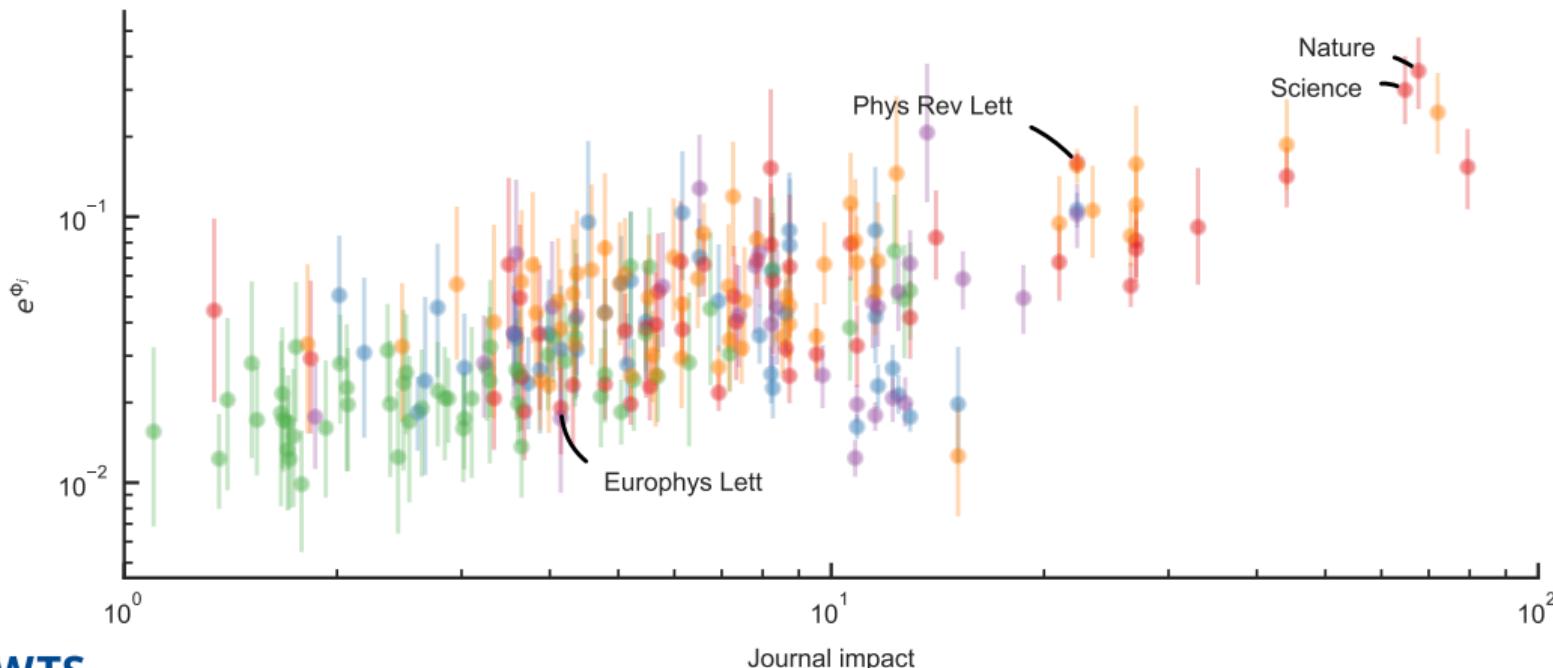
Model fit - article citations



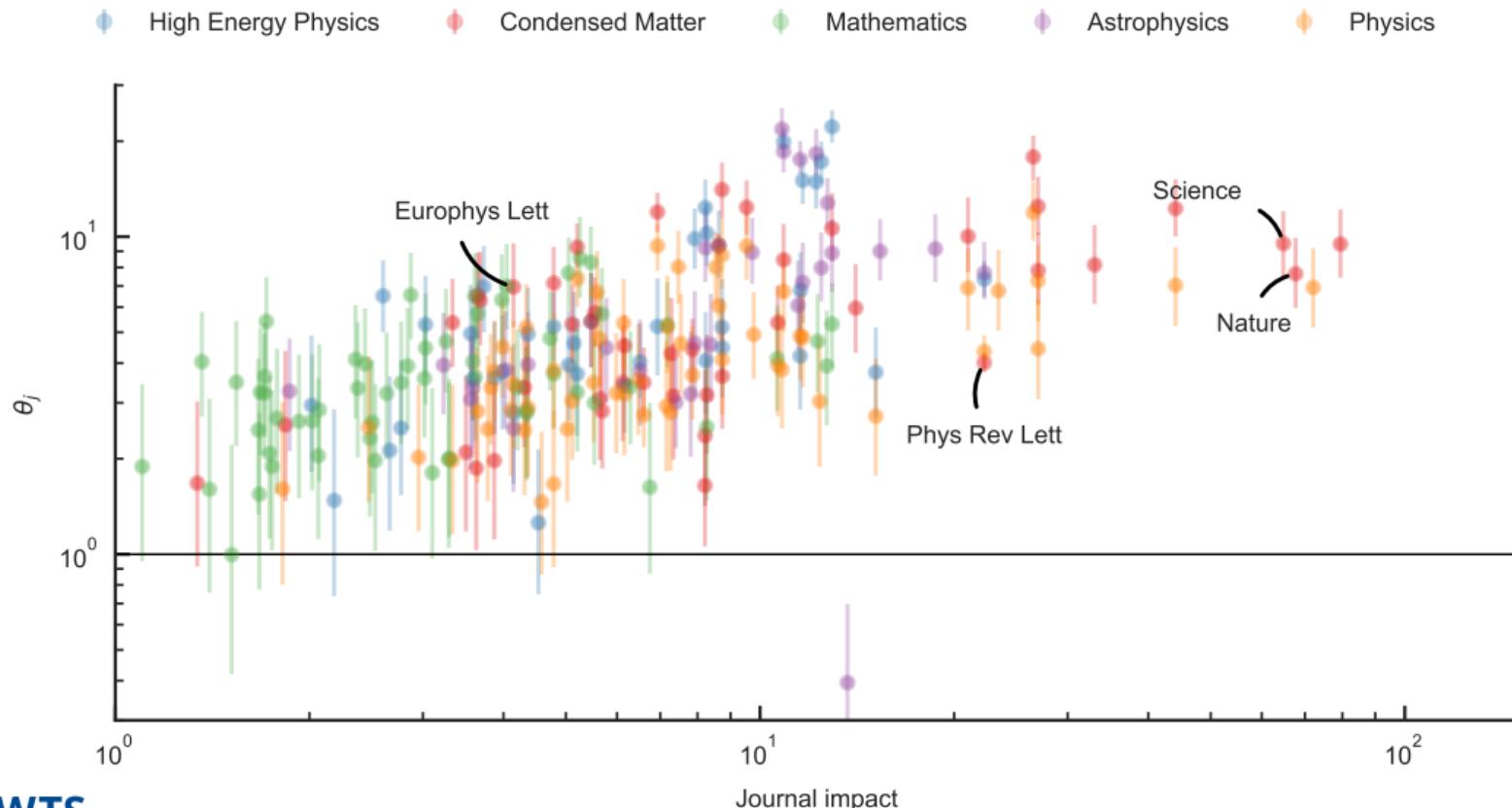
Impact vs median latent citation rate

High Energy Physics Condensed Matter Mathematics Astrophysics Physics

B



Impact vs journal citation multiplier



Conclusion

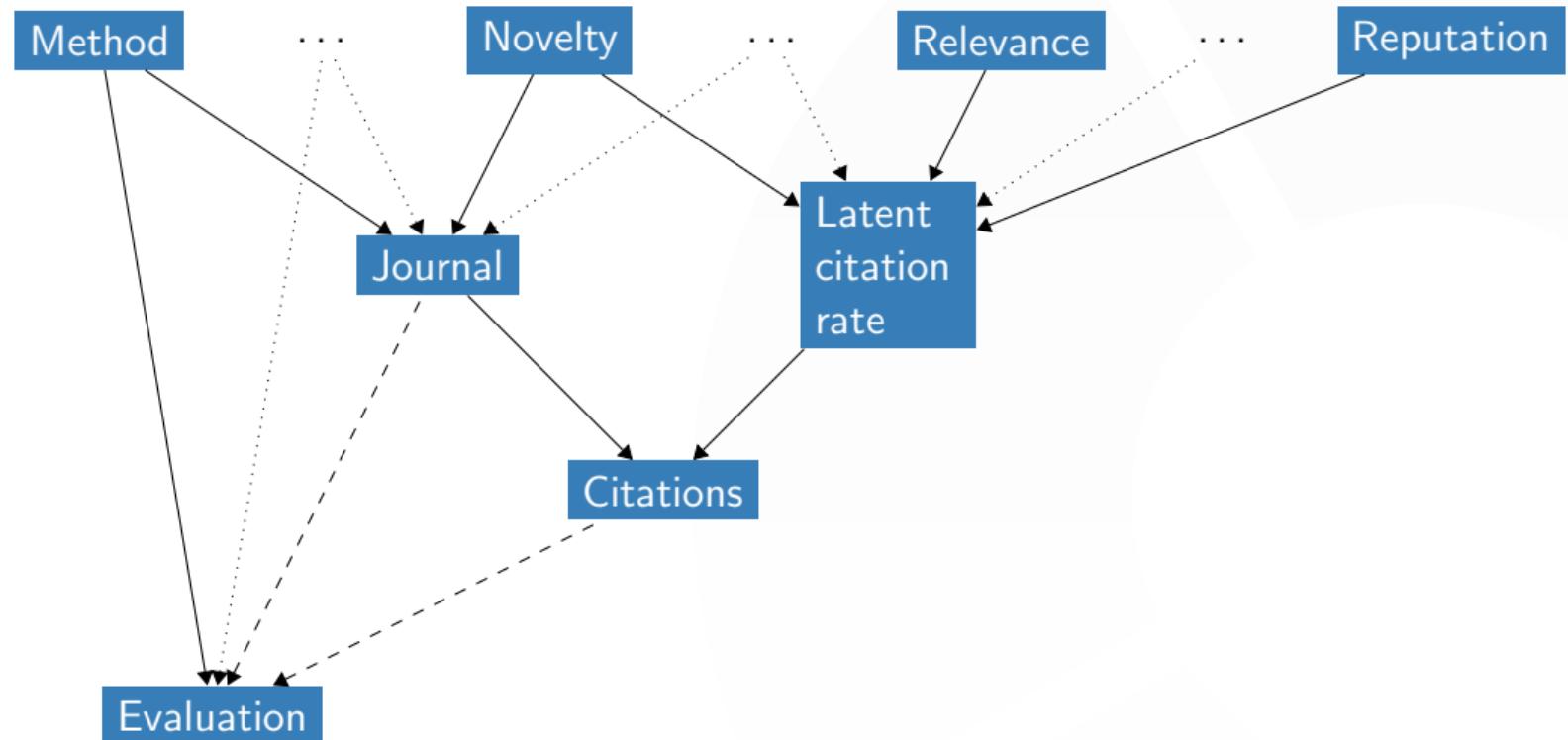
Why are high-impact journals cited more?

- ① High-impact journals publish articles with higher latent citation rates.
- ② High-impact journals increase citations more (cumulative advantage).

Considerations

- Causal effect presumably related to readership/circulation.
- Self-perpetuating journal hierarchy (high-impact because it has high-impact).
- Journal may be useful conduit of information for citations.

Conjectured causal model



Thank you!

Suggestions? Comments? Questions?

✉ v.a.traag@cwts.leidenuniv.nl

 @vtraag