

# David Catalán Cerezo

PhD Researcher in Machine Learning & Epidemiology

dcatalan@ethz.ch

+41 76 690 0149

linkedin.com/in/david-c

## Education & Training

---

**PhD, Student** | Dept. of Mathematics, ETH Zurich **Feb 2024 – Present**

Research focus: causal inference and multimorbidity trajectories within the SNSF Sinergia project “From Single-Disease Research to Informed Machine Learning: A New Research Paradigm for Multimorbidity”

**MSc in Fundamental Principles of Data Science** | Universitat de Barcelona **2021 – 2023**

Courses: Data Science for Health, Machine Learning, Deep Learning, PGMs, Bayesian Statistics.

Thesis: parametric learning of probabilistic graphical models from multi-sourced data.

**Postgraduate Degree in Data Science & Big Data** | Universitat de Barcelona **2019 – 2020**

Final project: anomaly detection in a lamination train.

**MSc in Economics & Finance** | Barcelona School of Economics (BSE) **2017 – 2018**

Courses: Econometric Methods I-III, Financial Econometrics, Time Series.

**MSc in Economics (M1, Double Degree UPF-TSE)** | Toulouse School of Economics **2015 – 2016**

Modules: Advanced Calculus, R, SAS, Applied Econometrics.

**BSc in Economics** | Universitat Pompeu Fabra **2012 – 2017**

Specialisation in Statistics & Quantitative Methods (8.76/10, top 3).

Thesis: Effects on education of an increase in the minimum working age across Europe: a gender perspective.  
grade (9.6/10)

## Research & Professional Experience

---

**Doctoral Researcher** | Dept. of Mathematics, ETH Zurich **Feb 2024 – Present**

- Developing machine-learning and causal-inference methods for multimorbidity trajectories using large population-based healthcare and electronic health record (EHR) data.
- Co-developing generative-AI methods (causal normalizing flows, LLM-driven imputation) for treatment-effect estimation from incomplete real-world observational data.
- Applying the target trial emulation framework to Swiss and UK primary-care EHRs, including pharmacological effectiveness studies in metabolic disease.

**Researcher** | FEDEA (Applied Economics), Barcelona **Jun 2021 – Jan 2024**

- Econometric modelling of individual participant data, including demographic and health data: marginal-effects analysis, regression, and time-series methods.
- Built machine-learning models (regression and classification); migrated analysis pipelines from Stata to Python.
- Design of public-policy recommendations.

**Lecturer / Teaching Assistant** | Universitat Pompeu Fabra, Barcelona **Jan 2020 – Jan 2024**

- Teaching assistant for Econometrics I (second-year) and Econometrics III (third-year) undergraduate courses.
- Lecturer of labour economics

**Data Analyst** | FocusEconomics, Barcelona **Sep 2019 – Jun 2021**

- Data processing and process automation in Python (APIs, web scraping); ETL pipelines and data-warehouse maintenance; quality controls and client queries in SQL.

**Research Assistant** | CaixaBank Research (Big Data), Barcelona **Sep 2018 – Aug 2019**

- Created and forecast economic indicators from internal data using R and SQL; time-series modelling.

## Grants, Awards & Distinctions

---

- **Fundació La Pedrera Grant**, MSc in Economics & Finance, Barcelona School of Economics (2017–2018).
- **Dean's List in Economics**, Universitat Pompeu Fabra.
- **Double-degree distinction**, Universitat Pompeu Fabra & Université Toulouse 1 Capitole.

## Publications & Presentations

---

### **Preprints & Manuscripts**

- O. Jullian Parra, S. Zoccheddu, D. Catalán-Cerezo, T. Forzy, F. Ulrich, W. Sutcliffe, J. M. Burgstaller, O. Senn, P. Owen, N. Serra. "Joint Treatment Effect Estimation from Incomplete Healthcare Data: Temporal Causal Normalizing Flows with LLM-driven Evolutionary MNAR Imputation", *arXiv:2605.05125* (2026).
- T. Forzy, D. Catalán-Cerezo, et al. "Linking Educational Attainment to Early Marriage and Fertility in West Africa: Findings from a Multi-Country Analysis", *under review* (2025).
- J. Casanova, D. Catalán, F. Felgueroso, M. Jansen. "Firm-Level Employment Dynamics and Minimum Wages: Evidence from Spain", *FEDEA Working Paper* (2025).

### **Conference Presentations**

- F. S. Ulrich, D. Catalán-Cerezo, D. Fatle, A. Martínez-De la Torre, A. M. Burden. "Effect of Orlistat on Bodyweight Among Adults With Overweight or Obesity Without Diabetes: An Emulation of a Target Trial Using UK Electronic Medical Records", *International Conference on Pharmacoepidemiology (ICPE), ISPE* (2025). Poster; second author.

## Technical Skills & Languages

---

**Programming & data:** Python (proficient), R (proficient), Stata (proficient), SQL (intermediate), Matlab (basic); AWS, Tableau, Git.

**Writing:**  $\LaTeX$ , Markdown, MS Office.

**Languages:** Spanish (native), Catalan (native), English (fluent), French (intermediate).