IHEA ECR Pre-Congress Session, 19 July 2025, Bali

Income Inequality and Mental Health in Australia: A Causal Machine Learning Approach

By Han Cheng, PhD Student, Macquarie University

Discussant Comments by:

Sally C. Stearns, PhD

Professor, Department of Health Policy & Management

The University of North Carolina at Chapel Hill

UNC | GILLINGS SCHOOL OF GLOBAL PUBLIC HEALTH

Motivation for the Analysis

- “Income inequality has emerged as a major socioeconomic concern across developed economies, with implications that extend beyond material well-being to population health.”
- Mixed and dated empirical evidence for Australia
  - Existing studies show no effect or negative effects of income inequality on mental health outcomes.
- “Inconsistencies reflect both the complex nature of mental health and limitations in standard empirical approaches, which may not adequately capture non-linear effects, heterogeneous responses, or high-dimensional confounding.”

Overall Contribution of Paper

- Three key contributions:
  - Applies causal machine learning methods to the inequality–mental health relationship, offering more flexible and robust identification than conventional approaches.
  - Uses a range of inequality measures to evaluate how different dimensions of income distribution affect mental health.
  - Conduct gender-specific analyses to uncover potential heterogeneity in response to inequality.

- In total, the paper:
  - Provides new empirical evidence to inform economic and public health policy at the intersection of inequality and mental health;
  - Demonstrates how modern empirical tools can enrich the analysis of complex health determinants.

Hypothesized Mechanisms for Relationship

- Existing frameworks suggest that the inequality–mental health link may operate through multiple individual and contextual channels:
  - **Psychosocial hypothesis**: inequality creates status anxiety and social comparisons that undermine mental health.
  - **Neo-material perspective**: attributes worse health outcomes in unequal societies to reduced investment in social infrastructure.
  - **Social capital theory**: highlights how inequality erodes trust and social cohesion, weakening informal support systems.
  - **Relative deprivation theory**: perceived disadvantage relative to others may be as damaging to mental health as objective deprivation.
- These relationships, which all hypothesize inequality to be negatively associated with better mental health, are likely difficult to detect and distinguish.

Data and Methodology

- Uses Household, Income and Labour Dynamics in Australia (HILDA) survey
  - Nationally representative longitudinal study
  - 22 annual waves covering 2001-2023
- Sample: 24,857 unique individuals aged 15+
  - Average of 17.4 observations per person
  - 433,102 person-year observations in total
- Enough statistical power to detect small effects and examine heterogeneity across subgroups.

Variables and Measurement

- Primary outcome: Mental health component of the Short Form (SF-36) Health Survey. Higher values reflect better mental health.
- Inequality measures capture different distributional dimensions
  - **Gini coefficient** ranges from 0 (perfect equality) to 1 (perfect inequality) for a population
  - **Theil index** is more sensitive to upper-tail changes and decomposable by population subgroups.
  - **Atkinson indices** incorporate explicit inequality aversion parameters, with higher values emphasizing lower-income inequalities.
  - **Palma ratio**, defined as the income share ratio of top 10% to bottom 40%, identifies distributional extremes.
  - **Relative deprivation** captures individual perceptions of economic disadvantage relative to reference groups, measured as the mean income gap between individual i and all higher-income individuals in their neighborhood.
- All inequality measures are calculated at Statistical Area Level 1 (SA1) boundaries, representing local communities of ~ 400-500 people.
  - This geographic level captures the local inequality environment individuals directly experience and is most relevant for social comparison processes.

Other Individual & Contextual Controls

- Person-level
  - age and age-squared, marital status, education level, individual gross income from all sources; number of dependent children aged 0-14;
- Living area level
  - Australian Bureau of Statistics Remoteness Index;
- Wave (time) fixed effects controlling for period-specific shocks; and
- State fixed effects controlling for time-invariant geographic differences.

Empirical Strategy

- Traditional fixed effects regression
  - Concern is omitted time-varying variables
  - So causal identification may not be achieved
- Apply Double Machine Learning (DML) framework (Chernozhukov et al., (2018)) to "estimate the causal effect of income inequality on mental health."
  - DML allows flexible control for high-dimensional covariates while producing consistent estimates of the treatment effect.
  - Technical fine print: The estimation procedure proceeds via sample-splitting and cross-fitting. The data are randomly divided into K=5 folds. For each fold k, we train machine learning models, specifically, Random Forest and LASSO, on the remaining K-1 folds to predict both the treatment variable D and the outcome *Y* using controls *X*.

More Fine Print about Empirical Strategy

- The dual methodological framework combines traditional fixed effects and modern Double ML techniques to strengthen causal identification.
  - Fixed effects models control for time-invariant confounders using within-individual variation;
  - Double ML addresses model misspecification and omitted variable bias via flexible control modeling.
  - Approach mitigates reverse causality by enforcing temporal ordering and excluding individuals with unemployment or health-related work limitations
  - To address selection bias, they exploit the panel structure to track within-individual changes in exposure to inequality.
  - Omitted variables are further addressed through machine learning–based control estimation.
  - Robustness checks include alternative geographic definitions for inequality, different lag structures, sample restrictions, and re-estimation with K=10 folds.
  - Using multiple inequality measures serves both substantive and methodological purposes, capturing diverse distributional dimensions and validating findings through consistent cross-measure patterns.

Key Results: FE Models

- Fixed effects models:
  - The only inequality measure with a statistically significant association is relative deprivation, showing that individual experiences of economic disadvantage relative to local reference groups have substantial adverse effects on mental health outcomes.
  - The gender-specific FE estimates have the same result.

Key Results: DML Models

Table 5 (Dependent variable: Mental Health. 24,857 individuals)						
	Random Forest			LASSO		
Inequality	Estimate	s.e.	z-statistic	Estimate	s.e.	z-statistic
Gini	7.910***	2.170	3.64	3.640**	1.590	-2.28
Theil	-4.490***	0.550	-8.18	-4.040***	0.510	-7.89
Atkinson (ϵ = 0.5)	-3.330	2.270	-1.47	-2.950*	1.770	-1.67
Atkinson (ϵ = 1)	4.370**	1.730	2.52	1.430	1.110	1.29
Atkinson (ϵ = 2)	-8.190***	0.640	-12.87	-6.220***	0.590	-10.63
Palma	0.980***	0.130	7.47	0.390***	0.880	4.40
Relative Deprivation	-11.550***	2.420	-4.77	-14.450***	2.200	-6.58

Additional Estimates and Key Conclusions

- Gender estimates exhibit same patterns across different measures for effect of income inequality on mental health.
  - Han says the estimates for females are larger but does not provide a statistical test.
- Paper concludes with an argument in favor of DML:
  - "The inequality-mental health relationship is more complex than previously understood and may involve non-linear relationships and interactions that traditional methods cannot adequately capture."

My Critique as a Discussant

- I'm a pinch hitter.
  - Paper was originally assigned to Bruce Hollingsworth, who has done work in this area.
  - But I enjoyed learning about this problem, the approach, and Han's results.

- 2025 New York Times article highlighted the importance of this problem in relation to immigration policy in Denmark:
  - Social Democrats in Denmark have a tough/restricted immigration policy because they believe high immigration increases income inequality, and that income inequality is detrimental to society and society's health.
    - Belief that cost of inequality is greatest for lower/poorer classes rather than wealthier individuals.
  - https://www.nytimes.com/2025/02/24/magazine/denmark-immigration-policy-progressives.html

Specific Critiques: Descriptive Stats

- Paper says: "The inequality measures demonstrate meaningful variation that captures different dimensions of income distribution."
  - I wanted more info on the extent of variation in inequality measures over time within areas versus variation across areas.
- Marital status (a six category variable) was incorrectly included as a continuous variable.
  - All model estimates will change; hard to predict how.
  - May be that some results are similar (e.g., gender differences in effect may remain).

Specific Critiques: FE Models

- My main concern for FE models is the extremely limited time-varying area controls.
  - Key area control is rurality, and it's hard to believe there's much within-area variation in rurality over time. (Table 1 could show if there is.)
  - Seems to me like there could be other important time-varying area controls.
    - Political party in power (e.g., low income people may be more optimistic if Labor is in power)? Proportion immigrants? Presence of area-specific social programs? Housing stock? Housing prices?
    - These measures are arguably endogenous, but failure to control for them may be problematic.
    - These measures arguably should be in DML models also.
- Why not use FE at a lower level than state?

Specific Critiques: DML Models

- Is the variation in direction of effect of inequality across the different measures plausible?
- Han describes the different effects, which are very similar for Random Forest and LASSO. For example:
  - "Gini coefficient shows significant positive effects, suggesting overall inequality may have unexpected positive associations with mental health."
  - "Theil index demonstrates strong negative effects, indicating upper-tail inequality concentration has detrimental mental health impacts."

Specific Critiques: DML Models (continued)

- “Atkinson indices progress from negative effects at $\varepsilon=0.5$ to positive effects at $\varepsilon=1$ to strongly negative effects at $\varepsilon=2$, suggesting the relationship between inequality and mental health depends critically on which part of the income distribution drives inequality, with inequalities affecting lower-income groups having the strongest negative mental health implications.”
- “Palma ratio shows significant positive effects, indicating complex relationships between top-bottom inequality and mental health that may reflect aspirational effects or reference group dynamics not captured by traditional approaches.”
- “Relative deprivation maintains strong negative effects, so a robust relationship [of inequality] with mental health.”

Specific Critiques: DML Models (continued)

- Paper argues heavily for value of ML approach, but I want to know more whether the different directional effects for different measures makes sense conceptually.
  - Also, what are the implications of the magnitude of the coefficients?
- While gender estimates exhibit same patterns within gender, the analysis should test whether gender estimates differ significantly from each other.
  - Very hard to eyeball.
  - If they do differ, then what is the explanation?

Specific Critiques: Main Contribution?

- "Our study's primary contribution lies in demonstrating that traditional econometric approaches may systematically underestimate the complexity of the inequality-mental health relationship."
  - Maybe. Don't we already know this as economists?
  - If the model is underspecified, how do we know that the effects in your DML are correct in a case of substantial underspecification?
  - Additionally, if the DML results are correct, then the estimates indicate important relationships that I do not understand.
    - Why are the DML estimates so different in direction?
    - Due to the fact that the indices reflect very different types of inequality?
    - Or due to remaining effects of omitted variable bias?
    - And what does the magnitude of the estimates mean?
- I respect the potential contribution from ML techniques, but I need to better understand what the substantially different estimates tell us about effects of income inequality.

Thoughts from Bruce Hollingsworth (Sally’s paraphrasing from an email)

- A good PhD/ECR paper that is nicely written and easy to understand.
- The motivation could be better
  - The 2012 & 2015 papers on inequality & mental health are dated.
  - Recent work in similar countries (like the UK) could be useful (e.g., Bishop, 2023, and work by Anne Kavanagh).
  - The HILDA data website has more recent info: https://www.unimelb.edu.au/newsroom/news/2025/march/hilda-shows-inequality-rises-to-a-high
- More critically, what is the marginal benefit of this paper?
  - Page 6 says they use ML “t*o address limitations in traditional econometric methods“* Like what? How does ML get over whatever these are?
  - Discussion of what the estimates mean in economic real world terms rather than statistical significance would be useful.
    - Policy makers know there are differences in MH reporting and treatment between genders, and that there are inequalities in MY between rich and poor people.
- In editor/publication terms, I’d ask ’so what?’ question:
  - Fine if updating what goes on in Australia that is interesting.
  - But the contribution for an international audience might be limited.

My Critique as an Editor

- Strengths:
  - Effects of income inequality are very important.
  - Large dataset over a long time frame.
  - Paper is very well written and clearly presented.
- Key Concerns:
  - One empirical estimation illustrates but does not prove value of ML (even with two ML techniques).
  - Hard for me to believe that ML can fully compensate for substantially underspecified model.
    - Maybe omitted variables cause different directional effects of different measures in DML results.
  - More exploration and explanation of conflicting indications from different measures of inequality is needed.
- Overall: Very interesting work that should be continued.

Appendix: Tables from Paper

Table 1			
Descriptive statistics			
Variables	Observations (N)	Mean (s.d.)	
Health			
SF-36 mental health	433,102	45.72 (41.65)	
Inequality			
Gini coefficient*	433,102	0.47 (0.02)	
Theil	433,102	0.42 (0.06)	
Atkinson (ϵ = 0.5)	433,102	0.19 (0.02)	
Atkinson (ϵ = 1)	433,102	0.38 (0.03)	
Atkinson (ϵ = 2)	433,102	0.85 (0.06)	
Palma	433,102	2.89 (0.42)	
Income			
Total income	433,102	38,907.32 (68,170.64)	
Individual controls			
Age	433,102	36.50 (22.8)	
Gender	433,102	1.51 (0.50)	
Marital status	433,102	2.72 (2.09)	
Education (highest)	433,102	5.95 (2.69)	
Number of children	433,102	0.70 (1.07)	
Contextual controls			
Living Area	433,102	30.68 (16.13)	

DML Results for Males (Table 6)

Table 6 Double ML Results Male						
	Dependent Variable: Mental Health			Number of ID: 11,943		
	Random Forest			LASSO		
Inequality	Estimate	s.e.	z-statistic	Estimate	s.e.	z-statistic
Gini	7.380**	3.040	2.420	4.110*	2.260	1.821
Theil	-3.920***	0.761	-5.154	-2.770***	0.722	-3.840
Atkinson (ϵ = 0.5)	-2.440*	3.172	-0.770	-0.080	2.510	-0.030
Atkinson (ϵ = 1)	4.410*	2.440	1.801	3.200**	1.586	2.035
Atkinson (ϵ = 2)	-6.480***	0.890	-7.281	-4.310***	0.840	-5.150
Palma	0.900***	0.181	4.890	0.390***	0.123	3.140
Relative Deprivation	-11.850***	3.440	-3.450	-15.640***	3.120	-5.012

GILLINGS SCHOOL OF GLOBAL PUBLIC HEALTH

DML Results for Females (Table 7)

Table 7 Double ML Results Female						
	Dependent Variable: Mental Health			Number of ID: 12,918		
	Random Forest			LASSO		
Inequality	Estimate	s.e.	z-statistic	Estimate	s.e.	z-statistic
Gini	9.461**	3.062	3.101	3.172	2.241	1.411
Theil	-5.241***	0.771	-6.773	-5.252***	0.722	-7.272
Atkinson (ϵ = 0.5)	-4.462	3.191	-1.404	-5.733**	2.493	-2.302
Atkinson (ϵ = 1)	3.952	2.431	1.631	-0.264	1.554	-0.173
Atkinson (ϵ = 2)	-9.183***	0.891	-10.342	-8.055***	0.822	-9.864
Palma	1.104***	0.180	5.963	0.385***	0.121	3.085
Relative Deprivation	-13.404***	3.401	-3.944	-12.976***	3.082	-4.221

UNC GILLINGS SCHOOL OF GLOBAL PUBLIC HEALTH