

Author Response to the Referees

Manuscript: *The impact of the Canterbury earthquakes on household income and expenditure in the Canterbury region in New Zealand*

Manuscript number: egusphere-2026-1005

We thank both referees for their careful reading and constructive comments. We have substantially revised the manuscript in response. The principal changes include:

1. adding a continuous-MMI specification;
2. adding between-wave compositional-balance and pseudo-outcome tests;
3. clarifying that the Household Economic Survey is analysed as a repeated cross-section rather than a household panel;
4. introducing quadratic and categorical age specifications and richer household-composition controls;
5. re-estimating the event-study models and adding formal joint Wald tests of the pre-earthquake coefficients;
6. adding an ATT-IPW estimator that uses the full eligible comparison pool;
7. expanding the household-level disaster literature and more clearly explaining the paper's ability to distinguish regular income from insurance-related reimbursements and other refund inflows; and
8. clarifying the rationale for the North Island comparison group and revising several variable labels.

Below, we respond point by point.

Response to Anonymous Referee #1

General comment

This is a competent and useful but modest paper.

Response: We thank the referee for this assessment and for the two focused suggestions. We have implemented both requested analyses and have revised the manuscript to explain their implications more clearly.

Comment 1

First, please add a specification with the MMI, rather than its discretization into three classes (zero, low, high).

Response: We agree that it is important to examine whether the results depend on discretising earthquake intensity. We have therefore added a specification using continuous meshblock-level MMI among earthquake-exposed Canterbury households. The results are reported in Appendix H, Table H.1, and discussed in Section 5.5.

In these models, the interaction between continuous MMI and the post-earthquake indicator is not statistically significant for total household income, regular household income, or total household expenditure. The estimated interaction coefficients are:

- total household income: $-1,114.46$, standard error $1,917.09$;
- regular household income: $-1,036.31$, standard error $1,790.38$; and
- total household expenditure: $-1,001.20$, standard error 630.66 .

These estimates provide no evidence of a simple linear relationship between an additional unit of MMI and the post-earthquake change in the outcomes. We therefore retain the grouped specification as the principal specification because the household response may be nonlinear or threshold-dependent, while reporting the continuous-MMI results as a robustness check.

We also clarified an ambiguity in the previous manuscript. The principal model does not treat Canterbury households with MMI equal to zero as the reference category. The omitted category is the matched North Island comparison group. The low- and high-intensity indicators identify treated Canterbury households with MMI in the ranges $[4, 7)$ and ≥ 7 , respectively. We do not assign an arbitrary MMI value of zero to North Island households in the continuous-intensity model, because doing so would combine treatment status and within-Canterbury shaking intensity in a single variable.

Changes in the manuscript: Section 3.3 clarifies the construction of the intensity groups and the omitted comparison category. Section 5.5 discusses the continuous-MMI results, and full estimates are reported in Appendix H, Table H.1.

Comment 2

Second, please add a balancing test between waves. You match treatment and control. Are there changes between waves that could endanger identification?

Response: We agree that conventional post-matching balance statistics within individual waves do not by themselves rule out differential compositional changes between the treated and comparison samples over time.

We have therefore added three complementary sets of diagnostics in Appendix B.

First, Table B.2 reports the four-cell means and standard deviations for the principal matching and household characteristics, separately for treated and comparison households before and after the earthquakes. These variables include age, household size, sex of the reference person, homeownership, and higher educational attainment.

Second, Table B.3 treats each of these characteristics as a pseudo-outcome and estimates a DiD coefficient on $\text{Treat} \times \text{Post}$. The estimated coefficients are small and statistically insignificant:

- age: $p = 0.33$;
- household size: $p = 0.89$;
- female reference person: $p = 0.86$;
- owner/trust tenure: $p = 0.73$; and
- higher education: $p = 0.84$.

Third, Table B.4 repeats this exercise for broad household-composition categories, including one-family, two-family, three-or-more-family, other multi-person, and one-person households. We report both unweighted and weighted specifications. None of the $\text{Treat} \times \text{Post}$ coefficients is statistically significant under the significance convention used in the manuscript.

These results do not prove that all unobserved compositional changes are absent, but they provide no evidence of systematic differential changes in the observable demographic or household-composition characteristics of the treated and comparison samples after the earthquakes. We have added this qualification to the limitations section.

We also now emphasise that propensity-score matching is conducted separately within every HES survey wave. Treated households are therefore compared with contemporaneous controls rather than with a fixed pooled control sample whose composition may change mechanically over time.

Changes in the manuscript: The matching discussion in Section 2.2 now explains the wave-specific matching procedure and summarises the pseudo-outcome results. Appendix B now contains the four-cell means, covariate pseudo-outcome tests, and household-composition pseudo-outcome tests.

Response to Anonymous Referee #2

General comment and literature

The authors are working on an important topic with a reasonable approach and a data set that appears to have some very important advantages. Having read many papers in this area, very few can differentiate between earned income, which can be impacted by a disaster, and insurance payouts. Many studies have reported surprisingly little financial hardship around disaster events, and those authors could only speculate that public or private insurance payments smoothed consumption. The disaggregated expenditure data is also uncommon. Considering how extensive the literature on

disaster impacts on households is at this point, more of the literature should probably be cited, including papers whose limitations you are overcoming.

Response: We thank the referee for highlighting these data advantages. We agree that the ability to distinguish recurring household income from insurance-related reimbursements and other compensatory inflows is one of the paper’s most important contributions, and this point was not sufficiently prominent in the previous manuscript.

We have revised the Introduction to state explicitly that the HES records regular household income separately from “receipts and refunds” in the expenditure accounts. The latter category includes insurance reimbursements and related inflows. This distinction allows us to show that regular household income increases after the earthquakes even when reimbursement-related entries are observed separately. The income estimates are therefore unlikely to be explained solely by insurance settlements or other one-off compensation.

We have also expanded the discussion of the contribution in the Introduction. The revised manuscript explains that many existing datasets either combine insurance settlements with income or do not observe them separately, whereas the HES permits separate examination of:

- regular and total household income;
- insurance reimbursements and related refund inflows;
- general-insurance expenditure; and
- detailed consumption, housing, debt, mobility, health, and service-expenditure categories.

We also expanded the household-level disaster literature, including Carter et al. (2007), who document persistent asset losses and poverty-trap dynamics following environmental shocks, and Gignoux and Menéndez (2016), who find short-run losses followed by medium- and longer-run recovery after earthquakes in Indonesia. The revised literature discussion more clearly situates our findings alongside research on migration, household balance sheets, insurance, credit, and heterogeneous recovery.

Changes in the manuscript: The literature review and contribution paragraphs in the Introduction have been expanded, and the Abstract and Conclusion now more clearly distinguish regular income changes from insurance- and refund-related financial activity.

Comment 1: Persistence of the linked sample

The authors have linked records, which enables them to do analysis on people who stay in place vs. those who migrate. However, I didn’t understand how persistent the sample is. The observation counts in Table A.1 vary by over 66% between high and low years for Group 1 and 100% in Group 4. Do the people in the lowest year persist all the way through while others come and go? Are there people with all combinations of years observable? For example, could someone be observed only in 2006, 2009 and 2016 and another person observed only in 2010, 2011, and 2012? With the

modest sample sizes, it seems like people entering and exiting the sample could have substantial impacts on the estimates and precision. If this is more like a repeated cross section than a balanced individual panel, you would need to approach it with different methods.

Response: We agree that the earlier wording could give the impression that the linked HES sample was a longitudinal household panel. It is not.

The HES observations are analysed as repeated cross-sections. Each annual wave contains a wave-specific sample of households, and the analysis does not assume that the same households or reference persons persist across survey waves. The administrative linkage is used to reconstruct the residential history of each HES reference person and identify whether that person lived in Canterbury at the time of the earthquakes; it does not convert the HES outcome data into a balanced household panel.

Any incidental repeat appearance of an individual in more than one survey wave is not used as a longitudinal observation. The identification strategy compares changes in average outcomes across wave-specific treated and matched comparison samples. The variation in Table A.1 therefore reflects differences in annual HES sample size and the number of households classified into each group, rather than panel attrition from a fixed initial cohort.

We have revised the empirical framework accordingly:

1. Section 2.1 now explicitly defines i as a household observed in survey wave t , rather than the same household followed over time.
2. Section 3.1 states that the HES is not a balanced panel and that observations in different waves should be interpreted as wave-specific samples.
3. PSM is conducted separately within each survey wave.
4. Year fixed effects are included in all main models.
5. The new pseudo-outcome tests assess whether observable sample composition changes differentially between the treated and comparison groups.

The migration analysis is likewise not a household fixed-effects analysis. It compares treated households observed inside Canterbury with treated households later observed outside Canterbury, conditional on age, sex, and household size. We now describe these results as conditional descriptive associations rather than causal effects of migration.

Changes in the manuscript: Sections 2.1 and 3.1 now explicitly describe the repeated-cross-sectional design. Sections 4.2 and 5.4 clarify the interpretation of the relocation sample and the non-causal nature of the migration comparison.

Comment 2: Age and household structure

I was puzzled by the use of covariates in the specification. As the authors discuss in lines 719–729, no one would expect age to have a linear relationship with income or

expenditures. Why put it in the model as linear when you could include it in a more flexible and realistic form? Similarly with the household structure. The relationship between household size, income and expenditures will be very different moving from one to two working adults versus adding a third or fourth child. Why not have indicators for a handful of common household types (single working person, dual income no kids, dual income with kids, single income with kids, retirees, etc.)?

Response: We agree with the referee that a purely linear age specification is unnecessarily restrictive. The principal outcome regressions now include both age and age squared. The estimated coefficients have the expected signs: age is positive and age squared is negative in the main income and expenditure models, producing a concave life-cycle profile.

We also estimated an alternative categorical specification using age groups. Appendix H, Table H.2 compares:

1. the original linear-age specification;
2. the principal specification with age and age squared; and
3. a specification using age categories.

The estimated low- and high-intensity treatment effects remain similar across these specifications, indicating that the main results are not driven by a particular functional form for age.

We also agree that household size alone does not fully characterise household composition. We therefore added several household-composition checks. Appendix B reports pseudo-outcome tests for one-family, two-family, three-or-more-family, one-person, and other multi-person households. Appendix I additionally reports richer PSM and ATT-IPW specifications that include broad household-composition indicators and housing tenure.

We considered the more detailed typology suggested by the referee—combining the number of adults, employment status, children, and retirement status—but judged that it would produce a large number of sparse cells in the HES sample, particularly after dividing the data by survey wave and earthquake intensity. This would reduce precision and create additional confidentiality constraints. We therefore use broader household-composition categories that can be estimated consistently across the repeated cross-sections.

The richer-control specifications do not materially change the direction of the principal treatment coefficients. We now acknowledge that household size remains a parsimonious rather than exhaustive measure of household needs and earning capacity.

Changes in the manuscript: Age squared is now included in the principal models and Tables 2–3. Alternative age-bin specifications are reported in Appendix H, Table H.2. Household-composition pseudo-outcomes are reported in Appendix B, Table B.4, and richer household-composition controls are included in the bridge specifications in Appendix I.

Comment 3: Event-study evidence and timing

The event study graphs in Appendix C do not improve my confidence in the results. It would not be fully transparent to place them in an appendix that most readers wouldn't see. My concern is that all the pre-treatment coefficients are substantially above zero. The wide confidence intervals include zero, but four of the seven post-event coefficients are at the same levels, as if the earthquake had no impact. The three periods of higher income appear 4 to 7 years after the event. Impacts persisting seven years later is plausible, but if these are driven by insurance payouts and rebuilding, shouldn't we see something in post-years 1, 2 and 3?

Response: We agree that visual inspection of the original figures was insufficient, particularly because the pre-earthquake point estimates were positive even though they were imprecisely estimated.

We have re-estimated the event-study models using the same household controls as the principal regressions—age, age squared, sex of the reference person, and household size—together with survey-year and Territorial Authority fixed effects and TA-clustered standard errors. We now report the complete coefficient estimates in Appendix C, Table C.1, and formal joint Wald tests in Table C.2.

For total household income, the joint test fails to reject the null that the four pre-earthquake coefficients for 2007/08–2010/11 are jointly equal to zero ($p = 0.814$). The corresponding result for regular household income is $p = 0.624$. For total household expenditure, only one comparable pre-earthquake full-expenditure wave is available; the 2009/10 coefficient relative to 2006/07 is statistically insignificant, with $p = 0.544$.

We agree that non-rejection does not prove that the parallel-trends assumption is true. The revised manuscript therefore states that the event-study and Wald-test results provide “supportive evidence” rather than confirmation of parallel trends. We also explicitly acknowledge that the evidence is weaker for total expenditure because only one comparable pre-earthquake expenditure coefficient is available.

Regarding timing, the revised event study does not show a uniform immediate response. Regular household income is positive and statistically significant in the first post-earthquake wave, 2011/12, while the largest and most precisely estimated income effects occur in 2014/15 and 2015/16. We have therefore moderated the interpretation. We no longer describe the estimates as a simple immediate insurance-payment effect. Instead, we interpret the later income response as being consistent with the gradual expansion of reconstruction activity, labour demand, wage effects, and regional recovery.

This distinction is also supported by the outcome definitions. Insurance reimbursements and related refund inflows are recorded separately from regular household income. The later increase in regular income is therefore more consistent with recurring earnings and labour-market channels than with one-off insurance settlements alone.

To improve transparency while limiting the length of the main text, we retained the complete figures and coefficient table in Appendix C but moved the substantive interpretation, the exact Wald-test p -values, and the limitations of the evidence into Section 2.3 of the main manuscript.

Changes in the manuscript: Section 2.3 has been substantially expanded. Appendix C now contains the full event-study estimates and the joint Wald tests. The Conclusion also distinguishes the later reconstruction-related income pattern from immediate compensation flows.

Comment 4: Inverse probability weighting and the control pool

Did the authors try inverse probability weighting to see if it is any better than PSM at reducing the pre-treatment coefficients or improving precision (I apologize if I missed it)? Is there any way of expanding the pool of potential controls, which would increase the quality of the matches and give us more confidence in the results?

Response: We had not reported inverse-probability weighting in the previous manuscript. In response to this comment, we added an ATT-IPW estimator and a sequence of bridge specifications.

The ATT-IPW analysis uses the same treated households and eligible North Island control pool as the principal analysis, but it retains the full eligible control pool rather than selecting only the nearest-neighbour matched controls. Control households are weighted by the odds of their estimated propensity scores so that the weighted comparison group resembles the treated group on observed characteristics.

We report:

1. the original PSM specification;
2. PSM with the principal nonlinear age and sex controls;
3. PSM with richer controls;
4. an unweighted regression using the full eligible pool;
5. ATT-IPW using the full pool;
6. ATT-IPW with richer controls;
7. ATT-IPW with household-composition controls; and
8. a trimmed ATT-IPW specification.

The income estimates generally retain the same positive direction under ATT-IPW, but the standard errors are substantially larger and the estimates are less precise. Total-expenditure effects remain statistically imprecise. ATT-IPW therefore does not improve precision relative to nearest-neighbour matching in this application.

We consequently retain the wave-specific PSM-DiD design as the principal estimator. It provides a transparent comparison with close contemporaneous controls and achieves strong within-wave covariate balance. The ATT-IPW results are reported as a complementary robustness exercise showing that the sign of the income response is not solely an artefact of excluding unmatched control households.

The formal assessment of the pre-earthquake coefficients is provided by the revised event-study and joint Wald tests. We do not claim that ATT-IPW itself eliminates the positive pre-earthquake point estimates.

Changes in the manuscript: A new Section 2.4 explains the ATT-IPW estimator and diagnostics. Section 5.5 summarises the results, and Appendix I reports the full bridge specifications for regular income, total income, and total expenditure.

Comment 5: Use of North Island controls

Finally, I'll request a few minor clarifications for readers who have never had the privilege of visiting New Zealand. What are the important differences between the North and South Islands, and why should I be comfortable with the control group being drawn entirely from a different island?

Response: We agree that the previous manuscript did not adequately explain this choice. We do not assume that untreated North Island households are automatically equivalent to Canterbury households simply because they are located elsewhere in New Zealand. The restriction is primarily intended to reduce treatment contamination and spatial spillovers.

Canterbury is located in the South Island. Other South Island regions are geographically closer to the earthquake-affected area and may have been affected through migration, reconstruction-related labour demand, supply chains, housing-market pressures, and other regional spillovers. We also observe households that moved out of Canterbury after the earthquakes. Including these households, or nearby South Island households exposed indirectly to recovery activity, in the comparison group could violate the no-interference component of SUTVA.

We therefore exclude:

1. all households recorded as living in Canterbury;
2. households that moved out of Canterbury; and
3. the remainder of the South Island.

The eligible comparison pool consists of households that consistently resided in the North Island and were not recorded as living in Canterbury in the linked address sources.

We acknowledge that the islands and their regions are not intrinsically identical. For this reason, identification does not rely on an unadjusted comparison between all Canterbury and all North Island households. We:

- estimate propensity scores within each survey wave;
- match on predetermined or slow-moving household characteristics;
- require matches to occur within the same survey wave;
- include survey-year and Territorial Authority fixed effects;

- cluster standard errors by Territorial Authority; and
- verify strong post-matching covariate balance in every wave.

All maximum standardised mean differences are below 0.1, and the variance-ratio diagnostics satisfy the stated balance criteria. The new pseudo-outcome tests also provide no evidence of differential post-earthquake compositional changes.

We have revised the manuscript to make clear that the North Island restriction is a conservative strategy to reduce contamination, not a claim that raw North Island and Canterbury samples are identical.

Changes in the manuscript: Section 2.2 now explains the no-interference rationale for excluding the South Island and Canterbury out-migrants, and Appendix B reports the matching and compositional-balance diagnostics.

Comment 6: Terminology

Consider replacing “Contribution Schemes” with “Retirement Savings.” I didn’t know what specifically was being contributed to until you mentioned it in the text. I expect all readers would understand “retirement.”

Response: We agree that “Contribution Schemes” was not sufficiently informative for readers unfamiliar with the HES classifications. We have replaced the abbreviated label with:

“retirement-savings and contribution schemes”

This wording preserves the broader HES definition while making the principal component immediately understandable. The category may include retirement-savings contributions alongside other institutional contribution payments, so we did not shorten it to “retirement savings” alone.

The revised label is now used consistently in the variable definitions, tables, appendix results, and discussion.

Changes in the manuscript: The category list in Section 3.3 and the relevant appendix tables now use “retirement-savings and contribution schemes.”

Concluding response

We again thank both referees for their comments. The revisions have led us to clarify the repeated-cross-sectional nature of the data, strengthen the assessment of compositional stability and pre-earthquake trends, use more flexible covariate specifications, and add a full-pool weighting estimator. They have also helped us state the paper’s principal data contribution more clearly: the HES allows recurring household income to be examined separately from insurance reimbursements and related refund inflows, while also providing unusually detailed evidence on the composition of post-disaster household expenditure.