Review of “Flood risks to the financial stability of residential mortgage borrowers: An integrated modeling approach”

General comments

The manuscript "Flood risks to the financial stability of residential mortgage borrowers: An integrated modeling approach" presents an impressive framework for evaluating how flooding impacts mortgage borrowers' financial stability. The authors develop a comprehensive modeling approach that integrates flood damage estimates, property values, mortgage balances, and insurance coverage to identify which borrowers face heightened exposure to mortgage default following flood events.

The study makes a valuable contribution by introducing a model framework that models the relationship between pre-flood financial conditions, flood impacts, and post-flood financial conditions. The integrated modeling approach allows for the assessment of different theorized causal pathways of mortgage default (strategic, cashflow, and double-trigger).

While I appreciate many aspects of the study, I believe it faces several limiting issues that I would like to see the authors address in a revision:

1. **Unclear framing**: The study lacks a clearly articulated research question that aligns with its methodological approach. While the introduction suggests the study addresses "how pre-flood financial conditions affect the relationship between uninsured damage exposure and post-flood risk of mortgage default," the methods and results don't directly answer this question. The study would benefit from explicitly stating what scientific questions it aims to answer and how its modeling approach addresses these questions.
2. **Absence of calibration/validation data for defaults**: Despite citing empirical research on drivers of default, the authors don't calibrate or validate their framework against observed mortgage outcomes. Without these crucial modeling steps, it's difficult to assess whether the modeled financial conditions offer predictive value. The utility of the integrated modeling framework is questionable without demonstrating its predictive accuracy for the outcome of interest, unless the authors pursue a more exploratory approach with more detailed uncertainty quantification and sensitivity analysis.
3. **Unconvincing causal mechanism of flood damage to default**: In light of the comment above, the findings from Kousky et al. (2020), a key reference for the authors, demonstrate significant concerns with using modeled damage estimates to predict mortgage outcomes. Their study shows that catastrophe model damage estimates—even those potentially more accurate than those in the current study because they come from a proprietary catastrophe model—found spurious relationships between predicted flood damage and default compared to results based on actual damage inspections. Specifically, they found that for rare events like default, "predicted damage needs to match better with actual damage at a property level in order to deliver a robust estimated impact." As another example, they found that when the catastrophe model predicted damage of less than 10%, the odds of deep delinquency or default increased, but not when the catastrophe model predicted greater than 10% damage. They wrote, “This counter-

intuitive risk ranking, which we have not seen in other loan performance outcomes, suggests that the inaccurate property-level damage prediction by the catastrophe model can be problematic for a rare outcome, such as deep delinquency or default.” This raises fundamental questions about the reliability of the current study's approach to modeling default risk.

4. **Potential need for reframing study around sensitivity analysis**: Despite the complex integrated modeling approach, the paper doesn't sufficiently explore how uncertainties in model components propagate through to default projections. A more comprehensive uncertainty quantification and sensitivity analysis would strengthen the study by identifying which factors most influence projected outcomes and how robust the findings are to different assumptions. This approach would be particularly valuable given the lack of validation data and would better demonstrate the framework's utility for policy analysis.
5. **Missed opportunity for policy analysis**: The study introduces interesting policy analyses (such as the home repair grant program) but doesn't fully leverage its framework to explore how various policies could influence default rates under different scenarios and assumptions. A more thorough exploration of policy interventions, coupled with comprehensive sensitivity analysis, would significantly enhance the paper's contributions and better justify the development of the integrated modeling approach.
6. **Omission of important contextual factors**: The study takes a real-world framing, which is compelling and raises the stakes of the findings, but the model excludes real-world factors that would influence default outcomes, such as disaster aid programs, employment changes, and the effects of the 2008 financial crisis, without adequate justification for these simplifications.

Specific comments

Abstract

- The writing is generally clear, but the abstract should clarify the degree to which the study is model-based or observational (or both). For instance, is “Here, we evaluate the impact” in L14 an estimated impact or an observed impact (or maybe an estimated impact with methods calibrated to observed outcomes?) This is crucial to clarify because the motivation is about highly consequential (and overlooked in the literature) outcomes, such as mortgage delinquency, default, and foreclosure. The statement I quoted makes it seem like the study observes these outcomes. However, the next line talks about conditions “indicative of default” and then reports somewhat superficial statistics. Lacking insurance or income/collateral to finance repairs does not seem like a predictor of mortgage default in previous empirical studies. It is possible that I (and other readers) am unfamiliar with research showing this, so the authors should add context about the degree to which the conditions are strong indicators of default. Alternatively, if this is an observation-based study, do the authors find that is a different case in North Carolina? If so, this is a major result that the abstract should highlight more prominently.

- Some ambiguous grammar. For example, do the authors mean in L16-17 that they look at default *and* negative equity, or does negative equity refer to one of the financial conditions indicative of default? The ", including" grammar makes the remaining text unclear.

Introduction

- The details are strong and the authors are clearly knowledgeable about this topic, but long paragraphs and sentences make the narrative difficult to follow. Claims like "Given these gaps in existing knowledge" (L131) are ineffective because the authors do not clearly signal knowledge gaps in preceding paragraphs.
- Some of the introduction texts reads like a literature review, which disrupts the flow. It might be helpful to break out a succinct introduction that introduces the paper's focus and contributions, and a separate literature review section. Given that the abstract's focus on default, the paragraph starting on L81 is the most relevant, but we only hear about default after three long preceding paragraphs.
- The text on flood insurance and limited coverage are important, but I wonder if they would be more effective *after* going straight into the main text about financial instability. There is a lot of literature about natural disasters and financial instability – why don't the authors just start on their focus? The issues with the NFIP and flood-risk information are important contributors to these issues since a lack of insurance could be a major driver of bad financial outcomes for households, such as default.
- I don't know that the references in L77 are appropriate. It seems like the point here is that damaged properties are worth less on the market and home equity loan principal will be lower. I'm not sure that the cited studies about post-flood property prices are relevant here because these studies do not control for flood damages. Thus, it is unclear if the change in property prices after floods are due to a damage effect (or similar), and thus concentrated within segments of the housing market (which would be a form of double counting relative to the rest of this paragraph – if a property is structurally comprised, it is worth less and you can't take out a home improvement loan of the same value as pre-flood conditions). There are some studies that control for flood damage and find changes in post-flood prices unrelated to damage. The authors could cite Atreya and Ferreira (2015) - https://doi.org/10.1111/risa.12307 - but this is a small case study and there is conflicting evidence. See Davlasheridze and Fan (2019) - https://doi.org/10.1007/s41885-019-00045-z- or Pollack and Kaufmann (2022) - https://doi.org/10.1016/J.ECOLECON.2022.107350 for evidence on how and why property price changes may not be market-wide after a flood event and specifically indicate double counting in this context.
- There is something difficult about the logic of the key outcome of focus, default. The authors talk about how most of the damage from several of the largest storms in US

history was uninsured, and they also talk about a lack of insurance as a major indicator of mortgage default. So, where is the evidence from these storms that many uninsured households defaulted? The references to literature on this central framing point, starting on L81, would benefit substantially from reporting statistics from the papers to help the reader understand the connection between uninsured damage and default. The most specific statistic in this paragraph is the "50 times higher" one from Kousky, but this is on 90-day delinquency. While an important negative financial outcome, the current study primarily frames a focus on *default* (e.g., L16-17 of the abstract: "Our framework estimates key financial variables to identify borrowers exhibiting financial conditions of default"). The evidence on that from Calabrese et al., (2024) should state what the quantitative evidence is so readers understand how large the effect is.

  - I know the Kousky paper better than the Calabrese one, so I will focus on evidence in there that I believe the authors should pay more attention to in their framing, since they focus on mortgage default as the key outcome. Table 7 of that study seems to suggest that moderate to severe damage is the main predictor of 180 or more days delinquent or default. While the result for moderate to severe damage X in SFHA is not significant, that may be due to the very small sample size of treatment observations. The authors should highlight that damage amount is an important factor and can point out that the study doesn't control for whether a property is insured (which supports the authors' claims either way). However, the study combines 180 or more days delinquency and default into one outcome because the number of defaulted loans is not large enough for identifying a statistical effect. Out of 27,000 loans, there were only 24 defaults as of August 2019 (2 years after the storm). Overall, there are a rather small proportion of observations in the combined outcome. The authors should be transparent that there is not strong empirical evidence on the causal mechanisms their claims rely on.

- I like the Thomson paper a lot, but that study is fully model-based and assumes the causal mechanism of a house being underwater leading to default. As this is the main paragraph on empirical evidence relating uninsured damage and financial preconditions to the main outcome of interest, default, the authors should also review the following related empirical literature (mostly flood, but also related to wildfire risk). These studies generally support the work cited in this study, but some offer surprising insights into household financial resilience in the wake of large natural disasters that this study should reconcile in its framing:
  - Biswas, S., Hossain, M., & Zink, D. (2023). California Wildfires, Property Damage, and Mortgage Repayment. Federal Reserve Board of Philadelphia Working Paper, 23-5.

  - Mota, N., & Palim, M. (2024). Mortgage Performance and Home Sales for Damaged Homes Following Hurricane Harvey. Fannie Mae Working Paper Series.
  - Del Valle, A., Scharlemann, T., & Shore, S. (2024). Household financial decision-making after natural disasters: Evidence from Hurricane Harvey. Journal of Financial and Quantitative Analysis, 1-27.
  - Hopkins, C., Marr, A., & Wilson, N. (2024). How Does Mortgage Performance Vary Across Borrower Demographics Following a Hurricane? Federal Housing Finance Agency Working Paper Series.
  - Issler, P., Stanton, R., Vergara-Alert, C., & Wallace, N. (2020). Mortgage markets with climate-change risk: Evidence from wildfires in california. Available at SSRN 3511843.
  - Rossi, C. V. (2021). Assessing the impact of hurricane frequency and intensity on mortgage delinquency. Journal of Risk Management in Financial Institutions, 14(4), 426-442.
  - Gallagher, Justin, and Daniel Hartley. 2017. "Household Finance after a Natural Disaster: The Case of Hurricane Katrina." American Economic Journal: Economic Policy 9 (3): 199–228.
  - Deryugina, Tatyana, Laura Kawano, and Steven Levitt. 2018. "The Economic Impact of Hurricane Katrina on Its Victims: Evidence from Individual Tax Returns." American Economic Journal: Applied Economics 10 (2): 202–33.
  - Deryugina, Tatyana. 2017. "The Fiscal Cost of Hurricanes: Disaster Aid versus Social Insurance." American Economic Journal: Economic Policy 9 (3): 168–98.
- In particular, Deryugina (2017) and Deryugina et al., (2018) discuss drivers of financial outcomes after natural disasters that are largely missing from the present study's framing. The first study, particularly important for the authors to engage with, investigates the role of non-disaster-based social insurance that can actually improve some households' well-being after disasters. The second examines the role of employment and income, highlighting the role of savings in supplementing households in the aftermath of Katrina (which seems relevant especially here because households outside the SFHA may have lower probability of flooding, so may have higher savings if they don't pay into the insurance program over time; this may not be the case with NFIP because of its risk-rating procedure before Risk Rating 2.0, but seems plausible and is worth mentioning).
- The paragraph on L119 seems too repetitive with previous paragraphs. Can the authors consolidate the presentation of each topic?
- The "gaps in existing knowledge" (L131) are unclear, and it's not clear the authors address them. It seems like the gap (as far as I can tell, the authors state only one knowledge gap) is "Although prior studies such as those by Kousky et al. (2020) and Calabrese et al. (2024) have examined the association between insurance uptake, flood exposure, and mortgage credit risk, there exists a need for additional research into how

the pre-flood financial conditions of a borrower (i.e., equity and liquidity) affect the relationship between uninsured damage exposure and the post-flood risk of mortgage default" (L127-130). However, the objectives of the study do not appear to reconcile the gap. While the authors describe a very impressive analytical workflow for linking flood damage to preconditions, the link between the preconditions and outcome of interest, default, is not addressed in this study. But the authors explicitly claim that there is additional research into how pre-flood financial conditions affect post-flood risk of mortgage default. The last paragraph reads as if the authors address something more like post-flood *exposure* to mortgage default – they link flood damage and pre-flood financial conditions to a post-flood financial state, but do not appear to build evidence on the link between those financial states and mortgage outcomes.

- The Introduction needs to streamline the narrative around the research gaps and what the study focuses on. The current structure does not flow well because it is unfocused. It reads as if the authors developed the analytical workflow and then backed out the research gap the workflow could address, but did not identify a clear science question.
- I think the Introduction needs to soften its claims about what the framework can achieve. The authors need to be more explicit and transparent about what the framework does (from what I have read so far, it is an advanced framework to estimate exposure to bad mortgage outcomes but does not estimate the risk of those outcomes).
- It is now clear that the study is fully model based, not even using data on financial outcomes to calibrate the projections of post-flood financial conditions or vet its performance. This is concerning given the current framing of the study. The main framing of the Intro is "there exists a need for additional research into how the pre-flood financial conditions of a borrower (i.e., equity and liquidity) affect the relationship between uninsured damage exposure and the post-flood risk of mortgage default" (L129-130)." How does this study address this need if there is no data on mortgage default outcomes? At this point, I see a few options to reconcile this. First, the authors might instead frame their study as estimating post-flood exposure to mortgage default. This seems defensible because in the typical risk framing, risk is the potential for adverse consequences, driven by interactions of hazards, exposure, and vulnerability. The workflow the authors describe does not appear to fit this definition. Second, the authors could use the three introduced theories on default causal pathways as their representations of vulnerability. This would enable them to take their exposure estimates and translate them into default outcomes, conditioned on the assumption that one of the theories represents a valid causal pathway. The authors raise nice evidence that these theories have weaknesses, so they would have to be careful in their framing if they take this approach. Third, and complementary to the second, it seems like the authors could take a more exploratory approach by embracing the uncertainty in the system and investigating how different assumptions (e.g., about causal pathways of default) and model uncertainties (e.g., both well-characterized uncertainty around damage projections and deeper structural

uncertainties about the integrated modeling chain) propagate into projections of default risk. I have a preference for the third option because it seems the most appropriate for the question stated on L129-130. Given the model-based approach of the study, with no observations on the outcome of interest, an exploratory modeling approach that identifies drivers of uncertainty in the outcome of interest appears the best way to investigate the relationship between uninsured damage exposure and the post-flood risk of mortgage default. The authors could also clarify their current framing in a revision and explain where I misunderstand a gap between their science question and methods.

Methods

- The first sentence of this section seems like a more modest and appropriate framing than what the Introduction suggests. It also seems to support the need for an exploratory modeling and sensitivity analysis approach. The authors should consider streamlining the Introduction around this.
- Why only use loan-level data for initial financial conditions? If there are new originated loans over the full time period, isn't it possible to identify changes in financial conditions as well? I'm not familiar with the HMDA data. Using only data for initialization leads to very strong assumptions about income and loans over a 23 year period that again seems to support a framing around exploratory modeling and sensitivity analysis.
- I'm confused by the claim that strategic, cashflow, and double-trigger mechanisms are types of defaults (L146-149). The Introduction text specifically frames these as theories of default and talks about limitations in the first two theories. Why are all three theories modeled then? I think it would be helpful to reframe the paper to accommodate the methods. It seems like a worthwhile exercise to map the exposure to three types of default mechanisms, as long as the authors provide more context in the Introduction for why. One reason could be that although there is very insightful empirical research on drivers of default in the context of natural disasters, we don't have a complete understanding of the causal mechanisms (data limitations, a limited number of events, etc.,). There are competing theories about these causal mechanisms, which we can represent with bottom-up models (if taking this approach, please provide evidence that these theories are prominent and inform decisions – which is necessary to support some claims from the Introduction). I think it would be easy to frame the contribution in this way, and helps the authors explain that as more research comes out on the causal mechanisms, their framework could adapt to those pathways and better model mortgage default risk.
- I don't understand why the authors simulate financial conditions at a monthly time step over the 1996-2019 period but only focus on the 7 largest (when they point out that the state faced 14 major disaster declarations over the period). I greatly appreciate the

transparency on this point. Can the authors please justify this modeling choice and explain its potential implications on their results? Why can't the authors simulate just at the storm time steps? What are the consequences of overlooking other major storms (in addition to other flooding events and possibly more important events that affect the outcome of interest such as the great financial recession of '08)?

- The stochastic sampling for certain variables (mortgage loan characteristics – what else?) again suggests the value of an exploratory modeling approach and sensitivity analysis. Is it just one draw from the tract distribution for each household? One draw could lead to spurious projections given the distribution of other model inputs that are correlated with the financial variables (but not sufficiently sampled with one draw).
- In general, the methods and text on cross validation are great and the authors are exceptional related to previous literature. Great job! However, since the point of this study appears to be about the integrated modeling approach to modeling financial stability, it's a first-order concern to investigate how sensitive the modeling framework is to uncertainties in the modeling steps and inputs. The validation does suggest sizable uncertainty in both interpolation & extrapolation, which begs the question: how much does this matter for projections of mortgage default? While I recognize this paper builds on methods under review elsewhere, it would be helpful to contextualize how the sensitivities in the underlying methods are particularly relevant with respect to this study's prediction goals. I think that given the interest in the connection between uninsured losses and mortgage defaults, the most relevant sensitivity is the degree to which the model may overestimate exposure and damage outside of the training data. The authors do a good job of talking about these uncertainties and how well their model does. But the authors probably recognize that there is something different about homes that don't have flood insurance than the homes that do (given the authors' framing around affordability and willingness to pay for insurance and their reference to Bradt et al., 2021, I think they recognize that there are different factors between these populations, including that houses facing higher hazard are more likely to purchase insurance even if they are outside the SFHA). There is also a selection of properties into the claims data, based on factors such as income, deductible, and loss size. I reviewed the Garcia et al., (2025) preprint and sections 4.2 and 4.3 in detail and it's hard to see how their sensitivity checks get at these concerns of selection. I think it's important for the authors, given their current framing and seeming data limitations on "validation" data for default, to incorporate uncertainty in the modeling steps and evaluate how those propagate into sensitivity in the mortgage default projections. I want to emphasize that the cross-validation approach is rigorous, and the authors did a great job writing about it, but in the context of this highly complex modeling workflow, it seems very important to evaluate how the uncertainties in each model component propagate. Again, this especially seems the case because the integrated modeling approach seems to be the study's main contribution (if this is not the case, the study needs to substantially reframe its title and introduction).

- I don't see how the neighborhood-level cross validation results for the damage model are relevant if the outcome of interest gets simulated at the property level. The relevant validation metrics are at the property level, where errors can interact with other uncertain inputs and could propagate into errors in the default projections. The low R^2 is not surprising given previous research but is concerning and it would be very insightful to see how uncertainty in damage at the property-level propagates (especially as it interacts with other uncertain inputs). The later sensitivity analysis that appears to uniformly lower and increase property-level damages by 20% does not justify if 20% is enough. The Wing et al., (2020) study that the authors cite suggest that for most inundation depth, damage at the property-level can vary from 0% to 100% of a structure's value, and that the damage is heteroskedastic with inundation. This suggests that a uniform 20% adjustment is not sufficient for sampling uncertainty and seeing how it propagates.
- I would like to know more about the hedonic model. In particular, can the authors provide more information about the ATTOM dataset? Are the coordinates tax parcels or building footprints? Are the records complete across the state in all observable characteristics? For more context on why it's important to provide more details about this dataset, please see Nolte, Christoph, et al. "Data practices for studying the impacts of environmental amenities and hazards with nationwide property data." Land Economics 100.1 (2024): 200-221.
- How does the property value model account for the existence of flood exposure, events, and damage over the study period? Is there any heterogeneity in property valuation based on exposure, structural defense, or damage? On that note, does the damage model have input features related to structural defense? In addition to the overall sensitivity of the model, it is important to consider sensitivity of the default projections based on heterogeneity in inputs. The results aggregate on geographic and economic factors that may be relevant to flood resilience policy – what about accounting for those factors in modeling the outcome?
- How many samples for each property from the HMDA data? How do the authors account for correlation in income and factors such as property value when sampling from the HMDA data? Accounting for the correlation structure in the HMDA data is great, but it's crucial how those draws are assigned to properties with different hazard, exposure, and vulnerability characteristics.
- Is there anything like Fig. S9 for defaults?
- I don't think Model IV has enough information. How does it account for things like heterogeneous savings for households that are insured or not? With the same income, a household without insurance would accrue more savings by not paying insurance costs. At the property level, there is also a difference in the degree to which a household saves money on purchasing their home based on salient price signals for risk (particularly, in North Carolina Pope (2008) showed how seller disclosures improved price signals to buyers looking to live in the SFHA: https://doi.org/10.3368/le.84.4.551). Seems like the

model assumes everyone's income goes up? There is no job loss? What do recovery trajectories look like after flood events? Is it only possible through loans for uninsured properties? Does income grow for all homeowner types based on county-level annual trends? Do damaged properties recover value over some time period? Does price appreciation/depreciation occur differently for insured and uninsured properties? The authors mentioned some factors like employment and income in some of their mechanisms for default, but I don't see how it's incorporated in the modeling framework in a realistic way. Can the authors talk more about what their assumptions are about exogenous factors over the sample? It would help to have a conceptual model of household finances and indicate what this study includes/excludes and why.

- I recognize that the methods state the study does not look at aid, but why not? How do the results of Deryugina (2017) on social insurance relate to recovery trajectories? In the US, IHP actually requires households to purchase insurance as a condition of the aid – why exclude this type of mechanism from the integrated workflow? It is crucial that the authors justify modeling choices based on their study framing, not based on simplicity. Modeling choices based on simplicity may not always be appropriate for satisfying a study's goals. I'm not sure that the current choices around simplicity are justifiable given the current framing. The study uses real historical flood events and projects defaults based on different theories of how financial conditions produce defaults, but ignores other real-world characteristics that would mediate these outcomes. The "real-world" framing requires very careful justification for departures from reality.
- How does this framework account for the '08 financial recession, which occurs right before Irene? Does the property model adequately project the drop in value during this period? Do the financial models adequately account for the rise in poor financial conditions (and I assume defaults) at the household level? It also begs the question of contextualizing projected defaults from flood stress relative to defaults from the '08 recession. It seems to me that neglecting important exogenous drivers of poor financial conditions could lead to both under estimation in this modeling workflow. The decrease in property values and employment from the great recession, plus Irene in 2011, might mean the model underestimates a large default event based on the double-trigger mechanism?

Results

- What is "flood damage exposure" in Fig 3 and 4a? Is it damage to structures?
- I'm surprised by the amount of attention in the results to describing the damage estimates both overall and stratified across a few groups. I think the authors should restructure their results to focus on their main goal, which is the degree to which default occurs under

different assumptions of causal pathways of damage, financial pre-conditions, insurance, and mortgage debt.

- L576-577, what is the proportion of both located outside SFHA and lacking flood exposure?
- The results on default would be more interpretable if the authors showed how many defaults there would be under the three different causal pathways of default. Most readers will not be able to interpret the 6 metrics of Figure 6 and what it means for default projections. If not this, it could really help to reintroduce the acronyms in the results section and to not use acronyms in the figure caption. But I highly encourage taking the conditions required for certain default mechanisms (e.g., ACLTV > 100% and ADTI > 45% for double-trigger) and showing the proportion of households pre & post flood that met those conditions. I encourage the authors to really focus on their key outcome of interest and focus on interpretability since their topic is extremely important and many readers of this journal come from different disciplines and will need help understanding new concepts.
- I think it could be helpful to compare the default results for both uninsured and insured properties, perhaps to emphasize the additional risk of being uninsured. However, to do this comparison, one would have to account for the additional savings uninsured households might accrue by not paying for flood insurance. While the sample period predates Risk Rating 2.0, that is an especially important consideration as the price of insurance goes up. Also important is that many households, even if insured, need to rely on savings for some time after a large event due to slow timing in damage assessments and payouts. This ties into my questioning about whether the "real-world" framing is appropriate.
- I also think it would be helpful to compare the default projections under the floods versus alternative "normal" rates of defaults (some households go under hard times and may have to default) and "financial crisis" rates of defaults (like '08) to help contextualize the projections.
- The modeling assumptions seem to force the modeling results, particularly that uninsured households have few ways to financially recover between events and have to wait for their income to bounce back. What happens if the main income bringer lose their jobs (insured or uninsured households)? What happens if uninsured households end up with more aid? I'm not suggesting the study has to address all of these questions in its model, but I do think it needs to acknowledge how certain modeling assumptions will lead to certain outcomes when making comparisons across groups. What is left out of the model that might overstate the differences in default outcomes across groups of comparison? This is one reason a conceptual model could help.
- I think comparing the number of at-risk default by causal pathways is great, which is what Figures 7 & 8 do, but I'm confused about the result. Because these are not mutually exclusively, these are likely not the correct visualization types for this result. Separately,

it would help to contextualize these results in terms of the overall residential stock to help readers understand how large this problem is. Further, I think the current framing of the study leads the reader to expect much more results focused on these causal pathways and the default outcome. As I mentioned in my comments on the methods, I think it would be very effective to focus on how uncertainty in inputs & models propagates through the integrated modeling workflow and leads to different projections of default outcomes, conditioned on the different causal pathways one believes.

- For Figure 8, proportions could be helpful to contextualize the results (like the % for the other bar charts). 22,100 damaged homes out of 4.7M is important context when thinking about multisector impacts. Are we talking about financial risks that affect homeowners exclusively, or are there cascading impacts across sectors? It would be helpful to see more about whether 11,000 is a large number or not in this setting.
- The text about liquidity (L594-L604) raises questions about the HMDA sampling. Is it possible that the sampling assigns lower incomes and higher loan amounts to certain high damage households? Looking at results in terms of the sensitivity analysis I suggested above could help readers understand what drives different default outcomes in terms of modeling assumptions. As it currently is, it seems like the authors may be overinterpreting results strongly reliant on plausible but highly uncertain and insufficiently sampled modeling implementation.
- I think it is important not to overinterpret results, which I am worried about for the results shown in Figures 7 and 8. The authors could draw stronger conclusions about liquidity, income, property values, and defaults if they took more of a sensitivity analysis approach and implemented a method like scenario discovery.
- It's great to see the sensitivity analysis about the home repair grant program. It reinforces for me that the strongest insights in this paper would come from a more comprehensive approach to characterizing the drivers of sensitivity in default projections. I think it is the best way for the authors to make an important and reusable contribution in this area. In particular, the current default model results occur under the assumption that there is no grant for home repair, but there are major monetary transfers after large disasters (specifically presidentially declared disasters such as the focus of this study) so it is invalid to ignore this in the "baseline" case. The IHP program specifically requires uninsured households to purchase NFIP insurance, so that requirement seems like "good" policy in terms of reducing default (compliance is a different story, not modeled here but part of the data generating process). This again reinforces why a sensitivity analysis approach would be constructive and insightful. It would be unfortunate to overestimate default risk relative to the current policy environment (though with recent changes to FEMA the authors could frame some of their study around how important these programs are to avoid default risk!).
- It would have been nice to signal earlier to readers that the authors do a sensitivity analysis on damage costs and property values. This sensitivity analysis approach needs a

better description of the methods and a justification for whether it's valid to do a uniform application of the uncertainty instead of randomly sampling from an uncertain distribution for each unit. I'm not sure that it is. For example, the methods show that for the property valuation model, only 54% of observations fall within +/-20%. Given the large uncertainty in the projected outcome based on this under-representation of uncertainty, it seems very important to better contextualize sensitivity of the projected outcome to the actual uncertainty in the property valuation model. In addition, as mentioned earlier, 20% seems like an inadequate representation of uncertainty for the damage model, especially given previous findings on heteroskedasticity with inundation depth.

- It would be better to sample from uncertain factors using an appropriate experimental design at the spatial resolution of the model inputs (i.e., property-level). It would be very insightful for the authors to evaluate at least first-order sensitivities of the default projections to uncertainty in inputs and the authors could interpret results using relatively fast methods such as Method of Morris and scenario discovery. Since there are only ~22k houses under study for the default analysis, it seems like this is feasible. Several of the co-authors are more expert in these methods than I am, so I am interested to hear more about why they did not approach this complex new modeling workflow in a sensitivity analysis approach (e.g., among many works by Saltelli, please see https://www.jstor.org/stable/2676831 given the predominant framing of the integrated modeling workflow).

Discussion

- Nice, tight framing in the first paragraph. I encourage the authors to reflect on the differences in this framing to that of the Introduction, and to better synchronize the two sections to have a consistent story throughout the paper. In addition, the Results should focus on the main story.
- One issue with the first paragraph is the claim on L683-684: "Our results underscore the status of pre-flood home equity and debt-to-income ratio as important determinants of post-flood financial resilience." Is that true, or is that an artifact of implementing theories on causal pathway of default that treat these as determinants of default?
- The discussion is well-written. It also introduces claims that point to some unfocused results on the distribution of damages, as opposed to the most interesting points about the default projections and the uncertain factors surrounding those. For instance, the paragraph on L741 is very interesting. Why didn't the authors model this insurance policy with a deductible equal to 50% of a borrower's equity? Seems like to really illustrate the value of this integrated modeling approach, showing that an end-user can use the approach to stress-test candidate policies against a range of uncertain factors

seems appropriate and very useful to the research and policy community. Just speculating about this when the authors have the tool at their disposal to test their hypotheses is underwhelming.

- I'm confused by the 67,000 exposure estimates on L758. Why make a comparison to this number if the main number of interest is the ~ 22,000 damaged properties with mortgages?
- The discussion does a nice job of mentioning limitations, but generally doesn't justify why the study did not account for some uncertain factors. The idea that the Matthew and Florence grants were small and wouldn't affect the number of mortgages at risk of default is something the authors can test with their framework. Why just speculate?
- What's missing from the discussion is more acknowledgement about the deep uncertainty about whether the outcomes the authors measure actually are strong predictors of defaults. As discussed in the comments on the Introduction, the authors do not present that evidence. I think mortgage default *exposure* framing might be more appropriate than mortgage default risk, unless the authors clarify that the default projections are conditioned on the belief one causal mechanism holds. This calls for more of a deep uncertainty framing to the study, which would work well with the sensitivity analysis approach that I think this complex, prediction-based modeling integration study calls for.

Technical corrections

- L26: "losses from flooding are expected to surpass…" -> not sure this reference helps or is accurate. For one, it doesn't seem like any of the framing points rely on claims about changes in flood risk over time. Second, this is just one study and the reference reads more into the study than the study's findings support. For example, using the passive voice here for "are expected" makes this seem like the estimate is a confident one. It would be more appropriate to say that one study estimates annual losses from flooding may surpass $40B… I also think the claim that losses will surpass $40B as a result of increases in extreme precipitation under climate change is not supported by this reference. The study does not highlight the role of changes in extreme precipitation in changing risk estimates.
- L38 – the text about household ability and willingness to pay and demand could benefit from more specificity. What do the authors mean by "further reduces demand" on L39? Depending on whether they mean the demand curve or quantity demanded, it may be more accurate to say "which reflect low demand."
- The Gourevitch et al., (2023) reference on L125 does not seem to apply here. Where is the event-based specification of Gourevitch? They estimate "overvaluation" based on the degree to which the market capitalized information about properties mapped into the SFHA between sales. The studies that the authors cited on L77 would be more appropriate here, but note the references I mentioned in my comment that suggest the evidence on post-flood price adjustments is mixed (with more detailed specifications on

drivers of risk and household characteristics showing a more heterogeneous market adjustment).