Referee Report for MS-32382

"The Distribution of Capital Gains in the United States"

This paper uses tax return data to describe the distribution of capital gains in the US. The authors use a capitalization method to allocate untaxed (including unrealized) gains to individuals. They find that capital gains are highly concentrated, and, because a large fraction of gains are untaxed, the implied effective tax rate is low.

Understanding how capital gains are distributed and the implications this has for tax policy is an important question. The analysis is carefully done, but the way the paper is written makes it difficult to properly assess its contribution. My concern is that, currently, it comes across as an accounting exercise – I'd like to see the empirical results more closely integrated to the underlying economics (where do gains come from, what does this mean for how we should tax them). This would help to overcome my other concern, which is that the precise quantitative results depend crucially on some key assumptions (e.g. return homogeneity). How much we should worry about this depends on the the conclusions being drawn from the statistics presented.

Below I go into more detail about my major comments, followed by some smaller comments/queries.

Major comments

1. *Conceptual framework.* I think the paper would really benefit from some disciplining framework or discussion that would motivate why the empirical results presented matter. There are glimpses of this throughout the paper, but I think more needs to be done to think about where gains come from & what does that imply about taxation. A recent paper by Aguiar et al. could be a possible starting point for this (see also discussion in the Mirrlees Review about the taxation of normal and excess returns). This doesn't necessary have to be a formal mathematical analysis, but should provide a basis for the empirical analysis carried out.

   One thing I'm particularly worried about the volatility of unrealized gains, and the implications this has for how to tax it. I'd love to see a more in depth discussion of the theory underpinning this, how this connects to the empirical results.

2. *Introduction.* I found the introduction hard going – there's quite a lot of detail (Fig 1 is far too dense & too hard to read so early on in the paper), and a lot of numbers that it's hard to know what to do with. If the paper has a clearer guiding framework and structure, then rewriting the introduction accordingly would likely help a lot with this.

3. *Contribution.* Relatedly, I think the paper needs a clearer statement of the contribution to the existing literature. That capital gains are skewed (and more so for unrealized gains) is reasonably well known. So the question is, what more do we learn from this paper? Is it that the top 1% share goes from 18% to 21%? If so, I'm not really sure what to do with this number, especially given that there is presumably reasonable uncertainty surrounding

it (e.g. because the assumptions made in the capitalization method).

It would also be useful to have a clearer statement of the contribution relative to the DINA and SZZ papers – what does the current paper do that aligns with the previous papers, and what does it do differently? I am reasonably sure that the authors are doing something new and useful, I just don't think it's stated clearly enough at the moment.

See also the recent paper by Msall and Ole-Andreas Naess on "Never Realized Capital Gains" (as an aside, the paper is incredibly US-centric currently, it would be nice to expand the discussion to incorporate the relevant papers/findings from other countries).

4. *Owner-occupied housing.* I didn't understand what the 'SCF methodology' is that is used for the non-itemizers. Should I be worried that non-itemizers are different and live in different properties to itemizers? Can you provide some validation of the estimation of the distribution of property wealth using alternative data sources?

5. *Private business wealth.* This is clearly a hugely important part of the analysis, and so it's important that the reader understands and believes the underlying assumptions. e.g. the business that are sold (and for which there are sale values) are presumably a selected sample - does this matter for the implementation of (3)? How does the estimated value of private business wealth compare with estimates in SZZ (or is this a new estimate)?

6. *Key assumptions.* One thing that might be really helpful is a summary table that describes the different components in the computation of total cap gains, the data used to estimate them, the key assumptions underpinning them (and whether these are new or standard).

7. *Tax rate calculations.* I struggled with what to make of the tax rate calculations, because I wondered (a) *should* we be taxing all unrealized gains? (probably not, but possibly some? if so, which ones?) and (b) this can't really tell us anything about how much revenue raised, because the behavioural responses would be huge. Again, this is a section that would benefit from a much clearer connection to the PF/optimal tax design literature.

Smaller comments/queries

1. Table 1: "current" $ is not very future proofed - should just state the year
2. I am not sure I find the inclusion of the tax form in the main paper necessary
3. A set of asset classes is mentioned in a number of places, but these are not clearly listed anywhere.
4. The colours/lines on a lot of the graphs are hard to read (similarly with legends etc.)
5. The first stat in the first sentence of the intro is a bit misleadingly high (given 2021 is such an odd year) – and unnecessarily so - the topic is sufficiently important that it doesn't need artificially inflating. I also didn't understand/couldn't reconcile this with fn 1.

References

Aguiar, Mark A., Benjamin Moll, and Florian Scheuer. Putting the" Finance" into" Public Finance": A Theory of Capital Gains Taxation. No. w32951. National Bureau of Economic Research, 2024.

https://lucymsall.github.io/research_papers/MsallNaess_stepup_JMP.pdf