

Interactive comment on “A Multiscale Spatial Dataset for Policy-Driven Land Developability across the United States, 2001–2011” by Hung Chak Ho and Guangqing Chi

Anonymous Referee #1

Received and published: 21 April 2020

This paper discusses the creation of a land availability index. For a paper that is meant to discuss the creation of a data set, it is poorly written and incomplete. The text is sloppy and has grammar problems. The introduction is overly generalized and makes references to things like "disasters predicted by Malthusian theory." This is just sloppy writing. I agree with the authors that's important to highlight the potential uses but their suggestions strike me as overly broad and not helpful.

On the index's usefulness: The data may be useful but it's not clear how the index was constructed. Since this is a paper about the construction of the data set much more care needs to be given how the exact details of the index and how, in theory, it could

C1

be replicated.

On Uniqueness: Data sets like this have been created by economists over the last decade or so. So this data set does not strike me as particularly unique. See the following:

Saiz, A., 2010. The geographic determinants of housing supply. *The Quarterly Journal of Economics*, 125(3), pp.1253-1296.

Burchfield, M., Overman, H.G., Puga, D. and Turner, M.A., 2006. Causes of sprawl: A portrait from space. *The Quarterly Journal of Economics*, 121(2), pp.587-633.

Lutz and Sand 2017: paper and data: <https://chandlerlutz.github.io/publication/land-unavailability/>

On its quality: All I can say here is that I opened the shapefile for 2011 and it worked and seemed correct from what I can tell by simply visually inspecting the shapefile. But the authors are quite vague on how they created the index. So there's no way for me to know if what they did is accurate and correct, and how the index values may change with different assumptions or methods to create the index. All indexes should be validated by comparing different versions to see how well they correlate. Also there's nothing in the paper about how they validated their index or shown correlates with steepness or water area or anything to give the reader the confidence that the index is in fact usable data.

Personally I wouldn't use this data for county-level analysis. There's little in the paper that gives me confidence that it's a high-quality index. If the authors want it to be used in data work, they should go through all the elements in fine-grained detail and show the reader how the index was constructed and how its robust to small changes in the assumptions and then show some examples how the index correlates with the components and how the index correlates with other measures of land types and/or other variables that will help the reader have confidence in the index. They mention

C2

they use the index is some regressions for Wisconsin but they ask the reader to take them on faith.

To be honest, the paper feels like it was rushed in order to get another publication.

Interactive comment on Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2020-3>, 2020.