Referee Responses

Benjamin Bradlow, Stefano Polloni, William Violette

May 31, 2022

1 Response to Report_YJUEC-D-20-00422

South Africa (their setting) is not a typical African country. According to World Bank, it is about 4.8 times richer per capita than the average sub-Sahara African country in 2001, and 4.2 times in 2012. It is important that authors improve the framing of South Africa (their context) in Africa more generally.

Thank you for the helpful observation. In an introduction footnote, we clarify that "compared to its low-income, sub-Saharan neighbors, South Africa shares similar urban slum conditions including a thriving informal housing sector. However, South Africa is unique among its neighbors in its ability to fund large-scale housing programs (Bah et al. 2018)."

I agree with the authors that their question is an important contribution to the literature. However, their discussion of Michaels et al. (2020) and Harari and Wong (2020) is not clear. These papers do not find mixed results. The Michaels paper studies areas that were planned before development, while Harari and Wong study slum upgrading areas. Their results are largely consistent.

We appreciate the clarification and have addressed it in the introduction.

The contribution of this paper, is rather, to study a different policy: public housing, where Michaels studies pre-emptive urban planning, and Harari and Wong study slum upgrading.

We reframe our contribution in the introduction as "our study contributes to this literature by studying public housing projects using spatial panel data."

Related, the authors might also want to look at Franklin's job market paper on public housing projects in Ethiopia.

Thank you for the suggestion of Simon Franklin's job market paper, which we found very interesting. We reference it in the conclusion.

1.1 The counterfactual

A substantive issue with the empirical analysis is the classification of the counterfactual group. The building data are recorded at 2001 and 2012 and census data at 2001 and 2011. The treatment group contains projects described as "current", "under implementation", or "complete" as of 2008. The control group contains projects described as "proposed", "under planning", "future", "investigating", and "uncertain" as of 2008. Since the control group is partly comprised of planned or future projects, development in these areas is likely inhibited by the government. To address this, can results be broken down by proposed, investigating, and uncertain vs. future and under planning projects? It would also be helpful to have more explanation with examples of what each of these statuses mean. Ideally the counterfactual group should contain projects that were approved on paper, but with no action taken on the ground. In a related paper, Franklin (2020) uses housing projects that were planned and approved during his study period but cancelled or delayed in their implementation.

Excellent point to explore heterogeneity in construction categories. It turns out we do not observe substantial heterogeneity as explained below, and unfortunately we do not have better information on categories.

- In Robustness Section 8, "we test whether the results are sensitive to our definition of unconstructed projects by excluding overlap with 89 unconstructed projects labeled "proposed," "investigating," or "uncertain." Table 22 finds results very similar to our main specifications, indicating that our empirical strategy is insensitive to this alternate project definition."

- We do not have more documentation on what the construction categories signify because we received our data from a non-profit research group who received the data from the government. Likely because the projects often become very political, we have been unable to get accurate reports on project status through other sources.

There are also observable baseline differences between treatment and control groups. First of all, Table 1 should statistically test for differences in these characteristics between constructed and unconstructed projects - currently it only provides point estimates. Also, based on Figure 1, it would be good to add differences in connection to the city center (distance to CBD and main highways) to Table 1. All of these observables could influence the change in private housing investment, and so may bias the diff-in-diff estimates. The authors should show that their results are robust to the inclusion of baseline observables that are 'unbalanced' in Table 1.

We share your concern and have concluded that although imperfect, the constructed/uunconstructed comparison is the best available and survives many robustness checks. We try to be transparent about the limitations of our approach throughout.

- Table 1 finds many statistically significant differences and includes many additional variables including distance to CBD and highways.

- Robustness Table 23 controls for distance to CBD, distance to major highways, elevation, and land slope as well as their interactions with the post indicator, finding no change in the main results.

- We augmented our analysis to include all measures available in the pre-period to develop a propensity score and reweight our analysis by this score and results are robust (See discussion in Section 8 and results in Table 18, Figure 19, and Table 19). We also include projects with no housing development at baseline, finding broadly similar results in Table 17 (discussions in Section 8).

1.2 Direct effects

It is unclear what we learn from the direct effects estimates. Since, as far as I can tell, we cannot distinguish between buildings that were built as part of the project and those that were complementary investment. This includes formal buildings, informal buildings, and public service buildings. Consequently, the direct effects are capturing the project itself. Potentially more could be done to distinguish which buildings were built as part of the project and which were not. However this may be prohibitively difficult to do, in which case the paper should focus on the spillover (outside of project boundaries) effects instead, especially for the welfare analysis.

Thank you for the helpful observation. We reframe our analysis given that we cannot separate project from non-project formal houses/infrastructure. We continue to estimate direct effects on informal buildings because they were explicitly not part of the program as described below.

- In the introduction, we more clearly define project goals as building formal houses, moving recipients into houses, and providing basic services/infrastructure, which is consistent with project guidelines.

- While data is not available to precisely identify project structures, we treat informal housing within project footprints as an outcome because qualitative interviews with housing officials indicated that informal housing development is viewed as entirely separate from projects. Also, informal/formal housing results are similar with both aerial building data and census measures (See section 8).

- We reframe our analysis as (1) measuring the projects (formal housing, infrastructure), (2) direct effects on informal housing, and (3) spillover effects.

1.3 Welfare and displacement

 The paper does not address whether the projects provide welfare gains on aggregate or simply displace development from other areas of the city. This is a fundamental concern when evaluating place based policies, and it is very difficult to provide a definitive answer in many cases. I do not expect that it is possible to provide an answer in this context where all analysis is done within one city, but it is important that the paper critically discusses this issue. The welfare results should be heavily caveated in light of this issue.

Helpful observations, and we have removed the analysis. In the conclusion, we included a discussion of why our analysis is not able to quantify welfare effects.

1.4 Measurement

 Census measures are interpolated from census areas to plots. This could confound spillover effects with direct effects because these census areas will straddle the project boundary even if the (interpolated) plot is entirely outside. Therefore, census measures should be analyzed at the spatial unit they are measured in, at least as a robustness check.

Table 24 repeats the census results at the census enumerator area level with roughly 7,500 enumerator areas (EAs). We find similar results in project footprints but zero spillover effects. Zero spillovers from this method are likely due to excluding EAs that straddle any project boundaries leaving only a few EAs that are close to projects but do not straddle boundaries, which is a very conservative approach to detecting spillovers. We detect similar spillover effects using interpolated census measures and precise building measures (Table 25), which gives us some confidence of detecting spillovers in other outcomes with interpolated census data. See Robustness Section 8 for discussion.

1.5 Minor issues

 Since there are no spillovers beyond 500m, it would be interesting to see spillovers broken down over an even finer spatial scale.

Thanks for the interesting suggestion. We tried 100m rings from 0 to 500m and found that results are mostly insignificant without a clear pattern of effects nearest or furthest in this range. See Table 26 and Robustness Section 8 for discussion.

 I very much appreciate the authors' transparency reporting all results even if they are insignificant. Since there are so many outcomes, it would be helpful to correct for multiple hypothesis testing. This could be done by table (i.e. family of outcomes) using, for instance, the False Discovery Rate or Family Wise Error Rate correction.

We added a family-wise error rate correction using the Bonferonni method throughout, which informs our indicators of significance.

 Estimates of theta are very similar for both formal and informal housing. Can the authors explain in more detail why informal housing is only 10% more costly?

We decided to remove the welfare section altogether.

 I assume that there is no data on the footprint sizes of buildings, if so it would be good to see this as an outcome as well.

We agree especially since we likely identify project houses due to their uniform construction. Unfortunately, the data are unavailable to us.

Why use hectometre and not just km? I know that the plots are measured in hectares, but it is much simpler to use km in the text. This would remove the need for the many footnotes on this.

Great suggestion and updated with km's throughout.

The authors continually mention that some housing projects may have already been completed by 2001. Can you find out their dates? How many are there? How long before?

We tried matching strings of project names to government budget reports as well as reaching out to various government housing authorities, both to no avail. Unfortunately, we have been unable to find any quality data on construction dates (possibly because construction completion can be very political).

2 Response to 2021-01-25-jue

This paper contrasts changes in outcomes between 2001 and 2011 for areas where public housing projects were constructed in Gauteng Province, South Africa, compared to areas where planned projects were not constructed. Gauteng contains Pretoria and Johanesburg, with the projects clustered around the latter. The program under which the projects were constructed began in 1994. There are two main novel findings:

Just to note, in Data Section 3, we clarify that "We locate projects using a map of housing projects for the Johannesburg metro area (excluding Pretoria) obtained from the Gauteng City Regional Observatory". We also include a footnote in the introduction "Our data do not cover Pretoria."

1. Along with the expected formal constructions, areas within project boundaries also saw large increases in informal construction, including of "backyard houses" built on formal plots. 2. There is some evidence of small spillover effects of the projects outside their boundaries, which raise the socioeconomic status of the surrounding neighborhoods. On formal house transaction prices, the order of magnitude is similar to those found for low-income areas in Diamond and McQuade (2019) although the results in the present paper are not statistically significant.

2.1 Long-form comments

Let me start by saying that the paper makes a distinct contribution to the literature, which will benefit from a greater understanding of the effects of public housing in a developing country setting. To this end, I would like to see more direct comparison with the results on spillover effects of public housing construction in high-income countries, reviewed for example in Diamond and McQuade (2019). Outside of the effects on informal outcomes, how do the results differ and what does this tell us?

We appreciate the recommendation to better link to the literature. We updated our discussion of price effects in the intro as follows: "We estimate that prices for formal housing transactions increase by 9% between 0 and 0.5 km from project footprints although this result is statistically indistinguishable from zero. This effect is similar to previous price estimates in the literature, which range from 6.5% to 15% in low-income US neighborhoods to 12% in a middle-income neighborhood in Uruguay. Footnote: {See Diamond and McQuade, 2019 and Baum-Snow and Marion, 2009 for US Low Income Housing Tax Credit estimates. See Gonzalez-Pampillon, 2022 for public housing estimates from Uruguay.} Our effect differs from the 2.5% decrease in prices observed in high income US neighborhoods. Footnote: {Diamond and McQuade, 2019 }. Given that South African housing projects are located in relatively low-income neighborhoods, our price effects are largely consistent with the literature."

Because the authors provide regression results with the fine building categories as separate outcomes, I would like to see more validation (or at least description) of the data. I agree that aerial photographs are quite informative for whether a given building is constructed with formal or informal building materials, but I am not convinced it is easy to ascertain whether a building contains firms, how many firms there are, and whether the firms are formal as suggested by the first two columns of Table 8. I would like the authors to provide the details of the validation exercise they conducted correlating the GeoTerraImage data on number

of informal houses with the figures from the census. I'm not sure whether to consider the reported correlation coefficient of 0.85 large or small.

Great idea to better validate the building data. We have success validating the informal/formal housing measures and less success with the business measures.

- We are able to reasonably validate building and census measures for formal and informal housing. Validation exercises are described in the Data Section 3 finding strong correlations. We note that our correlation estimate is different from the previous version because we are now calculating correlations at the level of the enumerator area to make the census and building data as comparable as possible.

- We also find very similar results using both housing measures given by Appendix Table 25 and discussed in Section 8.

- We have less success validating business and utility measures, and include caveats in the result interpretations. Validation exercises (and limitations to the exercises) are described in the Data Section 3.

For the formal/informal contrast in Table 3, the authors could show that results are robust to using census measures as outcome variables. Brueckner et al. (2019) suggests that there is some social data on backyard housing, I don't suppose the distinction is present in the census?

Thank you for the suggestion. Luckily, the distinction is present in the census, and we find similar results as described below.

- The census identifies (1) "Informal dwelling (shack in backyard)," (2) "Informal dwelling (shack not in backyard, e.g. in an informal/squatter settlement or on a farm)," and (3) "House/flat/room in backyard" as given by the new footnote added in Section 3.

- We find very similar results using both housing measures given by Appendix Table 25 and discussed in Section 8.

For the firm variables, I'd like to see evidence that they track with employment numbers from the census. This more-involved validation would be valuable as more researchers use the GeoTerraImage data with Brueckner et al. (2019), at least, having already done so.

Great suggestion, which we tried, but found low correlations possibly because people live and work in different locations. We include a full discussion in Data Section 3 and caveat our results interpretations accordingly.

I'd also like some clarification on what is/isn't known about each housing project. Analysis is at the level of cells of a 100 m 100 m grid imposed on the area of interest, which the authors refer to as plots. The effect of project construction on number of formal houses in a given project area between 2001 and 2011 is presumably known to project administrators, as is the effect on construction of public service buildings. If the authors accessed this information they could, for instance, estimate the average across projects of the number of formal houses built per plot on plots contained within project areas, where () indicates the cardinality of the argument and I've abstracted away from the fact that some plots are not fully inside or outside a project area. As I understand it, this is effectively what () should measure in equation 1 when the outcome variable is number of formal houses. For the net effect of being in a constructed project, the authors could estimate () for unconstructed projects. The only statistical uncertainty comes from estimation of () so this result would be more precise than what the authors report. Can the authors follow this approach? If not, is it because program administrators do not make the information available (why not? Could it be that fewer houses were constructed than should have been?), or because they don't have records?

Thank you for the detailed empirical suggestions, but unfortunately, we face many data limitations described below.

- In the Data Section 3, we clarify that: "With this project definition, we do not know the dates when projects were started or completed. While the majority of projects were likely to have started construction after 2001 consistent with a broad expansion in housing programs in the early 2000's, the data may also include some projects that were already completed by 2001."

- Unfortunately, reliable data on the number of units constructed and dates of construction are not available. We have reached out to administrators to no avail, and we have tried linking project names to government treasury records, but different naming conventions mean that this exercise did not yield any reasonable results. One explanation is that government housing is an intensely political issue in South Africa, which prevents detailed data from being shared.

- We reframed our estimates of the effect of project construction on formal housing in the project footprint to be in terms of measuring the projects themselves (instead of estimating any exogenous effects). See Measuring Housing Project Construction Section 6.1 for a detailed discussion.

In Figure 1, I notice that many unconstructed project areas border constructed project areas so that according to the authors' own analysis the Stable Unit Treatment Value assumption will not hold for the within plot footprint analysis because the unconstructed plots will be subject to spillover effects. As a robustness check, I would like to see the analysis of at least Table 1 repeated when excluding plots where non-constructed and constructed areas are close, recognizing that these areas are somewhat geographically distinct. Since spillovers appear to be small, I do not expect large differences in results.

Excellent suggestion for a robustness test. We find that the main results remain largely unchanged excluding any projects that are within 2 km of any other projects. See new Table 21 and Robustness Section 8

I would also like to see the welfare analysis in section 9 removed altogether. The authors cannot treat the actual construction of the public housing itself as the action of a profit-maximizing developer, which is implied by including direct effects on formal housing in equation 6. The authors imply that there is a first-stage relationship between construction costs and steepness but as far as I can tell, they do not have variation in construction cost so it isn't clear how () in equation 6 is being estimated.

Helpful suggestion, and we have removed it. In the conclusion, we included a discussion of why our analysis is not able to quantify welfare effects.

2.2 Short comments

The authors claim that their measure of plot exposure to the projects is novel since it measures exposure by the fraction of the project area that intersects the plot itself and areas around the plot centroid instead of distance to the nearest project as in Diamond and McQuade (2019). I do not think the measure is novel. At the very least, Autor et al. (2014) have a similarly-continuous measure of exposure to rent controlled buildings in Cambridge, MA, although it is based on number of units rather than area. Gechter and Tsivanidis (2020) contains a similar measure with that paper being developed over the same time period as this one. Maybe the authors mean the measure is novel within the literature on spillover and direct effects of public housing development?

Thank you for the helpful cites. We removed claims that it is novel and added citations to Autor et al. (2014) and Gechter and Tsivanidis (2020) in the measurement section 5.1

I was a bit confused about the area covered by the public housing project data. Am I right that it is all of Gauteng province and therefore, there was no public housing constructed under the program in Pretoria?

Thank you for identifying the lack of context. In the Introduction, we added a footnote clarifying "Our data do not cover Pretoria." In the Data Section 3, we also clarify that the data do not include Pretoria.

I didn't understand the following interpretation of the coefficients in the "Plot foot-print row of columns 4 and 5 in Table 1: Among informal houses, backyard houses replace non-backyard houses consistent with qualitative evidence that housing authorities cleared preexisting slums before constructing new projects". Both coefficients are positive and significant whereas the statement made me expect a negative coefficient on informal houses.

We have revised our description to the following in Section 6.2 and for a revised table 5, "Columns (2) and (3) disaggregate the total effect into effects on informal backyard houses and informal non-backyard houses.

Backyard housing is a common phenomenon in South Africa where households erect informal dwellings within the plots of formal dwellings. We observe a large increase in backyard housing at the same magnitude as the increase in formal housing in Table 2. This finding suggests that almost every formal project house is accompanied by an informal, backyard house constructed on the same parcel of land. We also estimate a relative decline in informal, non-backyard houses in constructed project areas relative to unconstructed project areas. The number of people living in informal housing declines as a result of project construction. These findings are consistent with larger non-backyard houses being replaced by smaller backyard houses."

On page 27, the first instance of the word "which" should be deleted.

Great catch, and we have removed the welfare section that contained it.

There is a good reason why Diamond and McQuade (2019) and Rossi-Hansberg et al. (2010) do not consider effects on public housing project plots: they are not considering informal activity. I wouldn't imagine informal activity is first-order in their settings. Perhaps it was in Hornbeck and Keniston (2017)'s historical setting. As such, I would like the authors to moderate their statements in the conclusion about a general need to consider within-area effects of public housing projects.

Great observation. In conclusion section 9, we clarify "researchers and policymakers have largely focused on changes outside of policy footprints because policies in developed settings leave little scope for informal activity within their footprints" and emphasize that our results are relevant for developing settings.

3 Response to YJUEC-D-20-00422_review

3.1 Framework and welfare

The paper lacks a clear framework to understand what impacts housing projects may have and why the outcomes the authors consider are important. There are references to the literature on place-based policies, but no adequate discussion of key concepts such as amenity capitalization, externalities, or place-based distortions. The mechanisms discussion is not very sharp and some of the language is vague.

We appreciate your suggestions to sharpen our three mechanisms discussion. We reframed the Discussion in Section 7. We also reframed the introduction as considering three potential mechanisms in the following way: (1) "Project infrastructure investments including electricity and water may also benefit nearby houses", (2) "Buildings identified as businesses increase within projects, which may invite employees to locate nearby projects," and "Projects may produce positive housing externalities for surrounding households. For example, neighbors may benefit from social interactions, improved perceptions of safety, and community development."

The welfare analysis is not very convincing. Trying to back out welfare effects without having any reliable price data is going to be very difficult in this context. The authors should probably downplay it to a more tentative "back of the envelope" exercise and discuss caveats and limitations. It is hard to think of welfare effects without having a clear counterfactual in mind. Do we know anything about who moved into the project areas and were they are coming from? Absent the projects, are these people who would have srowded slums in the periphery? Would the schools and health centers have been built somewhere else absent the program? The welfare impacts are also likely to be very different depending on whether the projects are on empty land or are replacing slums, whose residents were evicted and only in part compensated with public housing. How should we think about the displaced residents' welfare? The welfare calculations are not very clearly explained. It is not clear what is a plot-specific endogenous variable and what is a model parameter. I don't think delta is defined anywhere. The welfare analysis assumes that projects only change amenities levels and therefore the demand side. However there is also an effect of expanding formal sector supply. The authors state that developers are competitive, yet they don't have zero profits?

Helpful observations, and we have removed the welfare analysis entirely. In the conclusion, we included a brief discussion of why our analysis is not able to quantify welfare effects.

3.2 Empirics

The authors should do more to justify the "planned but not constructed" empirical strategy. What is the nature of selection bias in this context and why is this approach alleviating it?

To better justify our empirical strategy, the Introduction adds a discussion of how near/far or pre/post comparisons may lead to biased estimates as follows:

- "the empirical strategy addresses two main challenges in measuring the impacts of housing projects. First, projects are often located not only in poor neighborhoods, but also on relatively undeveloped plots of land within these neighborhoods. Comparing projects to neighboring areas would likely capture these preexisting differences. Our strategy controls for these differences because constructed and unconstructed projects are often located on similar types of land plots in similar types of neighborhoods. Second, projects are often located in rapidly growing cities with high demand for housing. Measuring the evolution of project areas would likely also capture urban growth that is unrelated to the projects. Our strategy controls for trends in urban development by comparing relative growth for areas with constructed or unconstructed projects."

Constructed projects are not randomly selected within the list of potential projects considered. They are likely to be positively selected based on their infrastructure access, location (through zoning) and environmental suitability. It is not obvious to me that this selection bias is less pronounced than selection bias that would be encountered by considering areas that are just adjacent to constructed projects.

Thank you for highlighting an important limitation in our strategy which we first examine at baseline and later test with robustness exercises.

- We provide evidence that while construction status is non-random, constructed/unconstructed comparisons are likely to be less problematic than near/far comparisons. See Descriptives Section 4 for discussion and Table 1 for empirical results.

- We show that our results are robust to (1) controlling for baseline characteristics with a propensity score approach, (2) only including projects with no houses at baseline, (3) examining pre-trends, (4) alternate unconstructed definitions, and (5) time-varying geographic controls. See Section 8 for discussion.

The authors should account for the observables that are likely to matter for selection of projects: distance from the central business district or other city amenities, terrain characteristics (elevation, slope, distance to water bodies), distance to preexisting infrastructure etc. At the very least the authors should show a "balance test" table accounting for these sross-sectional characteristics, in addition to the "pre" outcomes reported in Table 1.

Table 1 provides a balance test finding significant differences between many baseline characteristics. This Table includes distance to CBD, distance to major highways, slope, and elevation. Since there are many small water bodies and pre-existing infrastructure buildings, it is not computationally feasible for us to calculate plot distances to all of these objects.

They could include time-varying controls for these characteristics and also include them in the propensity matching exercise.

We include distance to CBD, distance to major highways, elevation, and land slope as well as their interactions with the post indicator, finding no change in the main results in Table 23. We updated the analysis to include all measures in the pre-period to develop a propensity score and reweight our analysis by this score and results are robust (See Table 18, Figure 19, and Table 19). We also include projects with no housing development at baseline, finding broadly similar results in Table 17. See Section 8 for discussions.

Beyond selection, the location of plots and all the "natural advantage" characteristics listed above are likely to be important sources of heterogeneous effects, which the authors at the moment are not exploring.

While we agree there are many interesting dimensions of heterogeneity to explore, we feel underpowered with only around 300 projects to cut the data across too many dimensions. Section 6.4 and Tables 12 and

13 explore heterogeneity by distance to CBD, which is likely an important dimension for policymakers.

An inherent limitation of the data is that the authors can't really observe prices well. While they don't discuss whether their data on housing transactions include formal or informal housing, my sense is that it has very limited coverage.

Thank you for calling attention to price measurement challenges. We clarify in Data Section 3 that "we analyze deeds data covering the universe of formal housing transactions from 2001 to 2011 in "affordable areas," which are defined as census enumeration areas with 2010 mean house prices below R500,000". We also emphasize "Additional considerations may introduce measurement error into our price measure. Buyers and sellers may not have strong incentives to accurately report transaction prices on deeds records. The timing of the transaction may also differ from the date that the deed is recorded."

Unsurprisingly, the price effects are too noisy to be interpretable.

We agree that noisy estimates are difficult to interpret and take the approach of presenting results but including caveats in our interpretations. We also note in the introduction that our point estimates are at least in line with the literature.

- Our interpretation in Section 6.3 is as follows "Positive effects on local prices are consistent with public housing projects providing positive local amenities. Yet, the statistical insignificance of these results means that we cannot exclude the possibility of zero or even negative effects of projects on local prices. One reason for these statistically insignificant results may be that deeds records are relatively rare occurring on only 7.6% of plots. This finding underscores the difficulty of estimating spillover effects using official deeds records in contexts where housing transactions are unlikely to be officially recorded."

For many outcomes (most importantly prices but also demographics and some house quality measures) the study appears underpowered. These results should be probably downplayed or relegated to the appendix, unless the authors are able to uncover some patterns in a heterogeneity analysis. Accounting for differences in market access / market potential of different areas may help.

Thank you for the suggestion. Although our power is low with only ∼300 projects and possible noise in many of the measures, we agree with another reviewer that a strength of the study is that we have many measures and our analysis is transparent about including measures that have zero or noisy effects. We also correct for multiple hypothesis testing throughout. Section 6.4 and Tables 12 and 13 explore heterogeneity by distance to CBD, which is likely to be policy relevant.

I would downplay the measure of project exposure at the plot level, which is presented by the authors as a novel element of their strategy.

Great suggestion. We downplay the novelty of our measure throughout and note similarities between our approach and previous approaches in the literature like Autor et al. 2014 and Gechter and Tsvanidis 2020.

I don't expect the results to be much different if instead of a continuous measure of overlap the authors used a simpler binary indicator for whether a grid cell includes a housing project. This would also streamline the notation of expressions (1), (2), and (3), which is currently very heavy.

Our spatial overlap approach helps deal with the close proximity of some projects to each other as well as their varying sizes, which we discuss in Section 5.1. We streamline the expressions and exposition by estimating direct effects and spillover effects separately. See equations (1) and (2) in Section 5.2. Estimating separately leads to virtually identical coefficient estimates and standard errors.

The authors may want to consider alternative types of standard errors for robustness, such as Conley.

We appreciate the helpful suggestion. After many attempts, we determined that Conley errors are not feasible with a dataset this large. As an alternate robustness approach, we exclude projects within 2 km of each other and find similar results with isolated projects. The results are not substantially less significant despite losing

about half the sample. Since isolated projects are unlikely to suffer from spatial correlation, this approach may at least partially address spatially correlated errors. See Robustness Section 8 and Table 21.

The authors should provide more detail on the aerial photograph dataset they employ. What are the 30 categories provided for buildings?

The 30 categories are divided into agriculture, forestry, conservation, mining, transport, utilities and infrastructure, residential, community services, health care facilities, education, commercial industrial, recreation and leisure, tourism, institutions, and not classified.

How is informal defined in that dataset? How does that dataset distinguish between schools, health centers, etc? How does that dataset detect informal businesses (which are often carried out inside one's house)?

Since the company that provided the dataset uses proprietary methods, we do not have detailed information on how categories are determined. We do know that the company combines aerial photographs with ground-truthing such as surveys. Our approach is to present the following validation exercises and appropriately caveat the interpretations of our results.

- We are able to reasonably validate building and census measures for formal and informal housing. Validation exercises are described in the Data Section 3 finding strong correlations. We note that our correlation estimate is different from the previous version because we are now calculating correlations at the level of the enumerator area to make the census and building data as comparable as possible.

- We also find very similar results using both housing measures given by Appendix Table 25 and discussed in Section 8.

- We have less success validating business and utility measures, and include caveats in the result interpretations. Validation exercises (and limitations to the exercises) are described in the Data Section 3.

I am still not clear on how the "post" dummy is defined. The authors state that their project maps are from 2008 and that they consider "built" projects that show up as completed in those maps. That seems to suggest that "post" is defined as after 2008. Then on p. 15 the authors state that post refers to "after scheduled construction", which sounds like project-specific. Then on p. 17 it is said that construction occurs between 2001 and 2012, which sounds like post should be 2012. Then the post definition changes again in Table 4.

Thank you for pointing to the confusing definitions in the draft. We updated our Estimating Equations Section 5.2 to clarify "POST_t equals one for 2012 and zero for 2001." We also clarify the timing of the data in the introduction and Data Section 3.

In Table 1 it would be helpful to have a formal test of whether the averages are statistically different between the two samples.

Table 1 now includes t-tests clustered at the project level and finds many significant differences.

3.3 Mechanisms

The framework provided to distinguish between mechanisms is not very useful. Perhaps the authors should distinguish between direct program effects (public investments), complementary private investments (house quality), productive amenity spillovers (businesses).

We appreciate the suggested categories which we used to sharpen our mechanism categories in Discussion section 7 into:

- "First, surrounding neighborhoods may benefit directly from public investments in utilities and other services within project footprints."

- "Second, projects may serve as employment centers" that "may invite people to live nearby and work within projects."

10

- "Third, by increasing the quality and quantity of houses, projects may produce positive housing externalities for surrounding neighborhoods (Rossi-Hansberg et al. 2010; Diamond and McQuade 2019). Investment may increase in nearby housing markets because neighbors benefit from projects in many non-market ways such as improved social interactions, perceptions of safety, and aesthetics."

Heterogeneous effects analysis could be used to assess some of the mechanisms. For instance, the authors argue that the spillovers may come from improved access to schools and health centers. Do we see any differential effects at different distances from schools / health centers?

We updated Discussion section 7 to consider how our estimates might help distinguish between mechanisms. While we agree that heterogeneity may be helpful, we feel underpowered to cut the data in many different ways. Instead, we focus on a policy relevant dimension – distance to CBD – documented in Section 6.4 and Tables 12 and 13.

3.4 Minor comments:

The control areas correspond to projects that are marked as "under planning", "future", "proposed" etc., suggesting they may be actually implemented by the time the outcomes are observed. Perhaps it would be more conservative to consider those treated or exclude them.

Excellent point to consider heterogeneity in construction categories. It turns out we do not observe substantial heterogeneity as explained below, and unfortunately we do not have better information on categories.

- "We test whether the results are sensitive to our definition of unconstructed projects by excluding 89 unconstructed projects labeled "proposed," "investigating," or "uncertain." Table 22 finds results very similar to our main specifications, indicating that our empirical strategy is insensitive to this alternate project definition." from Robustness Section 8.

- We do not have more documentation on what the construction categories signify because we received our data from a non-profit research group who received the data from the government. Since the projects often become very political, it can be difficult to get accurate reports on project status through other sources.

I don't quite follow the argument on p. 16 by which the spatial decay in spillover effects would suggest omitted variable bias.

Thanks for pointing to the confusing discussion and we removed the argument.

In Table 3, is "mean pre" referring to the control group?

We clarify in the table notes that "mean pre and post are average outcomes across all projects pre and post construction."

The number of observations sometimes changes from column to column in the tables and it is not always clear why.

We add table notes to clarify changing observations, which is generally due to the fact that census enumerator areas do not cover all of the plots.

Is "informal backyard house" a sub-category of "informal house" or a separate category?

We clarify that informal backyard houses are a subcategory of informal houses, and we focus our main analysis on informal, backyard houses and informal, non-backyard houses separately for clarity.

3.5   Exposition:

The literature review should emphasize how the context studied in this paper differs from the developed country settings where most urban renewal / public housing projects are studied. This is a setting with weak property rights and pervasive informality: how does this change the potential effects?

We appreciate this observation. In the introduction, we clarify that "since we focus on a developing context with pervasive informal housing, our study differs in two main ways. First, it may be important to account for spatial outcomes within project footprints such as backyard housing. Second, to the extent that informal housing grows in response to place-based policies and is itself a negative amenity, informal housing may undermine policy goals of producing positive amenities."

The intro doesn't explain the timeline and time structure of the data well: when was this public housing program implemented? Do the authors have time variation in the outcomes? When are the outcomes measured? Is the goal to estimate short or long term impacts? All this should come asross clearly from the very beginning of the paper.

Thank you for pointing out the lack of clarity on project/data desrriptions. We added the following early in the introduction: "Annually since 1994, the government has acquired parcels of land and constructed neighborhoods of single-story, two-room houses." and "We compile detailed spatial data in 2001 and 2011 for the Johannesburg metro-area including household census data, deeds records of formal housing transactions, and a remote-sensed panel of all buildings. This dataset allows us to measure the medium-term outcomes for housing projects built between 2001 and 2011. The building panel provides a novel measure of slum growth by distinguishing between formal and informal house types. Our main outcomes are population, formal and informal housing, formal house prices, and infrastructure investments."

In order to explain their identification strategy the authors contrast their approach with what they consider the "standard" one (comparing treated areas to nearby untreated ones). The exposition would be sharper if instead they spelled out clearly what the identification challenge is (unobserved characteristics of treated areas), how they propose to solve it, and what the identifying assumptions are.

The introduction adds: "This strategy addresses two main challenges in measuring the impacts of housing projects. First, projects are often located not only in poor neighborhoods, but also on relatively undeveloped plots of land within these neighborhoods. Comparing projects to neighboring areas would likely capture these preexisting differences. Our strategy controls for these differences because constructed and unconstructed projects are often located on similar types of land plots in similar types of neighborhoods. Second, projects are often located in rapidly growing cities with high demand for housing. Measuring the evolution of project areas would likely also capture urban growth that is unrelated to the projects. Our strategy controls for trends in urban development by comparing relative growth for areas with constructed or unconstructed projects."

The introduction should be clearer from the very beginning on what the treatment is. What were the program components? What exactly gets built within project footprints? What were the projects' goals and what does it mean that projects were "successful"?

The introduction adds context including: "We examine these effects in the context of South African housing projects" and "Annually since 1994, the South African government acquires parcels of land and contracted construction of many housing projects. Projects consist of (1) building neighborhoods of new single-story, two-room houses each on their own plot of land, (2) allocating full ownership of houses to recipient households, and (3) servicing houses with basic infrastructure including water, electricity, community centers, and other complementary investments. Housing projects range widely in size from a few dozen to several hundred houses per project."

It may be easier to express all distances as meters (e.g. say the units are 100 m x 100 m plots) rather than using hectars and reminding the reader many times throughout the paper.

Great suggestion and updated with km's throughout.