<u>Assessment of thesis for WONDATIR ATINAFU GEBRE</u>

Chapter 1. The impact of droughts in low-income countries: Firm-level analysis on Ethiopian manufacturing

1) **Motivation**: You motivate a lot the impact of climate change in paragraph 2. But to me, what matters is to say that the literature on the economic impact of climate in SSA has mainly focused on the agricultural sector. It seems to me you should explain why neglecting firms might be misleading in our general assessment of the climatic threat for SSA. The issue is far from trivial, if the sectors would be considered as independent and there was inter-sectoral mobility of labor (or K), it would help to cope with climate change (the same is true for geographical mobility). If the non-agricultural sector is also affected, it would not be able to absorb excessive labor in Arthur Lewis term. I am not a theorist but maybe more thought there would help to better define the contribution. Actually, in the results section you interpret your results in light of some theoretical framework, but to me, it should be articulated much earlier in the paper.

Thank you for your insightful comment. Neglecting firms in climate change assessments, primarily focused on agriculture, may lead to a limited understanding. The corrected paper, particularly in the Introduction's third paragraph (page 5), now emphasizes the significance of including non-agricultural sectors for a more comprehensive assessment of the climatic threat in SSA.

"Remarkably, the existing literature on the economic impact of climate in Africa has primarily concentrated on the agricultural sector2. However, [Rodrik, 2016] suggests that "Premature de-industrialization" is prevalent among SSA (Sub-Saharan Africa), manifesting a decline in manufacturing at the lower peak of industrialization. Given the acknowledged significance of the manufacturing sector as a tool for structural transformation, neglecting the sector might be misleading in our general assessment of the climatic threat for SSA…………"

2) **Theoretical framework**: On page 5, you describe the possible impact of climatic shocks on firms. You stress the input channels (energy, agricultural products for food-processing firms), but I remember that there was also a large literature on the impact on transport costs, the impact on labor productivity, … It is a bit old but the review paper by Carleton and Hsiang (2016) in sciences is a good starting point: https://escholarship.org/content/qt2vz2d2zz/qt2vz2d2zz.pdf I remember that there was a whole

literature on labor productivity e.g. in China. Also, it seems to me that you present the issue as too trivial. For example, in a Lewis framework, a disproportionate shock on agriculture should free some labor for the non-agricultural sector, no? I think it is not only important to present the issue as far from trivial but also to push the analysis as far as you can to delineate these channels (see comment 7). In that theoretical framework, how do we expect firms of different sizes to be impacted. It would make the results of Section 1.5.3 more interesting if it helps us to understand the theoretical mechanisms behind the average effect?

This is a valuable comment and thank you for providing the reference. I have carefully reviewed the paper and incorporated the theoretical framework into the Introduction, specifically in the second paragraph on page 4. This addition aims to address the complexity of the impact channels, including those related to labor productivity and the agricultural yield.

"Droughts pose a significant threat to local economies and manufacturing firms, particularly in Africa, due to the countries' heavy reliance on hydro-electric power, which is vulnerable to rainfall variations. Low rainfall leads to electricity shortages, forcing firms to resort to manual processes, increasing labor costs, and altering production factors [Abdisa, 2018] and [Traore and Foltz, 2018]. Power outages result in increased labor and material inputs and decreased capital and electricity usage. Also, drought-induced health risks, such as diarrhea and infections, impact labor supply and productivity, affecting manufacturing firms located in densely populated urban areas [Park, 2016, Carleton and Hsiang, 2016a, Zhang et al.]. Furthermore, rising input prices due to agricultural inefficiencies caused by drought affect manufacturing firms relying on agricultural products [Carleton and Hsiang, 2016a]. Overall, droughts create multifaceted challenges for the performance of local manufacturing firms......."

3) **Data**: I think you should spend some time describing the main firm-level outcome variables and control variables. For example, how is firm productivity measured in the literature (one of the first paper on Africa- I think – is the one by Jo Van Biesebroeck: https://econweb2012.econ.kuleuven.ac.be/public/u0044468/CV/vb05edcc.pdf) Maybe also a few descriptive statistics would be welcomed to understand which firms we are talking about (services, which sectors, average number of employees, distribution of firms according to firm size …). You have a nice dataset but you rush into the empirical analysis. Somehow it will help us to understand the data but also what is specific or not in a low-income country like Ethiopia (for example, the linkages with agriculture? Labor-intensive?) compared to high-income countries found in the literature. It is also not clear to me which controls you are using and why (I cannot assess if they are likely pre-determined variables or bad controls)?

I appreciate your helpful comments on providing a detailed description of firm-level outcome variables and control variables in my dataset. In response, I have included a comprehensive descriptive statistics table 1.1 on page 8. This table outlines key summary information about the main variables, their measurement, and the nature of the dataset, which encompasses a census of all manufacturing firms in Ethiopia spanning from 2000 to 2020. This addition aims to enhance clarity and understanding of the dataset.

4) **Data**: Usually the mean in the construction of anomalies are restricted from. Because rainfall data are very poor in terms of quality prior to 1985 (extrapolated from a few weather stations in Africa). I suggest the review paper by Auffhammer et al.(https://www.journals.uchicago.edu/doi/10.1093/reep/ret016)

Thank you for your insightful suggestion regarding the construction of anomalies and the limitations of rainfall data. I have carefully reviewed the paper by Auffhammer et al., and I agree that it provides valuable insights into the challenges associated with climate data quality. Considering the limitations mentioned, I thoroughly assessed various options for analyzing climate impact and found that the Standardized Precipitation Evapotranspiration Index (SPEI) best suits our study objectives. The use of SPEI is supported by its superior performance based on specific criteria. Moreover, for our analysis, we employed the CRU TS 4.03 dataset, known for being the most complete and updated dataset of gridded precipitation and potential evaporation on a global scale. The dataset's better spatial resolution enhances the accuracy and reliability of our climate impact analysis.

5) **Data:** there is no discussion about the sampling frame. I guess it is not a census of firms but a sample? How can you make that sample representative at the country level? Do you have sampling weights ? It is also not clear when firms enter or exit the sample and what are the consequences for your estimations. At this stage, it can be somehow discussed.
I appreciate your attention to the sampling frame. I want to clarify that the dataset is a census of manufacturing firms, not a sample. I've rephrased this information on page 8 for better clarity.

6) **Level of treatment and standard errors**: The treatment variable Trt_i should have the subscript Trt_ij. It seems like a detail but it is important to acknowledge that the treatment is at the district level. Said differently, two firms within the same district will be affected in the same way. Is it correct? To me, it means that there is likely dependency in the error terms within districts (see Moulton problem in chapter 8 of Mostly Harmless Econometrics). Standard errors need to be clustered at the district level (how many districts do you have? You need to check if you have enough, the rule of thumb is above 40,

see again chapter 8 of MHE).

Thank you for the reference and valuable comments. I've made the necessary edits to the treatment variable. Furthermore, in recognition of the potential dependency in error terms within districts due to the district-level treatment, we have proactively clustered the standard errors at the wereda level, the third administrative layer of the country, throughout the regression. This adjustment involves clustering with approximately 504 weredas, addressing concerns related to the Moulton problem.

7) **Channels:** In my opinion, the paper would benefit in separating the estimated effect for firm productivity and the possible channels, e.g. in a discussion section. You stressed a lot the impact on electricity as an input. The effect on labor productivity is interesting. There is an emerging literature showing the effect on heat on productivity. An alternative channel is the effect on the trade of goods and inputs. At the moment, it gives the impression you test some alternative outcomes and then interpret ex-post what it could mean. It would give less an impression of "fishing" if you were defining more clearly a theoretical framework and then try to test (or acknowledge when you cannot test these channels). Again for example, you come with some theoretical propositions on page 26 but it could come earlier and in a more structured way. At the end, e.g. in a "Discussion section", you would like to distinguish different channels : agricultural input, labor productivity due to heat, labor productivity due to power cut, or effect on value chain (transport): A good review paper to frame the issue is the one by Carleton and Hsiang (2016) in sciences : https://escholarship.org/content/qt2vz2d2zz/qt2vz2d2zz.pdf

I have carefully reviewed the suggested paper by Carleton and Hsiang (2016), and it has been immensely helpful in clarifying the possible channels through which climate affects economic variables. In response to your suggestion, I have created a separate Discussion section (pages 24-26) to provide an in-depth explanation of the channels, including labor supply and productivity, power demand and supply, drop in agricultural yield, and the rise in local input prices. This restructuring aims to better distinguish and elaborate on the various channels, addressing concerns about potential post-hoc interpretation.

Minor Suggestions
- Define much earlier than paragraph 5 what you do in the introduction (paragraph 2 or 3).
  Thank you for bringing this to my attention. I have clarified the primary objective of the paper in the third paragraph of the Introduction section.
- Page 6 "this figure is about 18%" . Would ""number" or "estimated impact" would be better?
  Done. Thank you.

4

- Page 6: "the first attempt' instead of "the very first attempt"

Done. Thank you.

- Maybe a matter of taste but do not like so much to have a small section 1.1. lost. Personally, I would create a section "Background", maybe moving part of paragraphs 1 and 2 of the intro?
- Same for section 1.2. I would rather use the related literature to better define your contributions in the introduction.

Thank you for raising points to enhance the paper's structure. I have addressed your suggestion by eliminating small sections and summarizing relevant points within the Introduction. Specifically, I've integrated background information into paragraphs 1 and 4 on pages 4 and 5. Additionally, I used the related literature section to better define the paper's contributions (page 6 and 7).

- In the intro, you stress a lot the years 2011 and 2016. But if people read too quickly (as I did at the beginning), they may not realize that you are working on a long time period between 2000 and 2020, exploiting the occurrence of droughts in 2011 and 2016. I would make that clear in the abstract and in the introduction.

Thank you for highlighting the potential oversight regarding the time period in the introduction. Your observation was insightful, and I have addressed it by explicitly mentioning in the abstract and introduction that the study covers the years 2000 to 2020, with a focus on the occurrences of droughts in 2011 and 2016. "This paper examines the impact of droughts on manufacturing firms in low-income countries, utilizing a unique dataset from Ethiopian firms spanning 2000 to 2020, with a focused examination of the drought events in 2011 and 2016."

- Figure 1.6e and 1.6f, that would be convincing if you could zoom in between 2000 and 2020 and show how we could observe the 2010 and 2016 droughts (personally, I would then show it in the manuscript. On the contrary, If you cannot see it in the raw data, it would not be a good sign, no?

Your feedback is invaluable, and I appreciate the opportunity to clarify the presentation of these figures in the manuscript. For Figure 1.6e, it is important to note that the droughts in 2011 and 2016 were registered as a single disaster in each respective year. The primary intention behind this representation is to showcase the growing number of disasters in the country, particularly since 1980. While the focus is on illustrating the overall trend, I acknowledge your point about the significance of the 2011 and 2016 droughts. Regarding Figure 1.6f, a careful inspection reveals that the SPEI values drop below -1 in both 2011 and 2016. This alignment with our records from the emergency disaster database reinforces the reliability of the SPEI values in capturing and reflecting the severity of drought conditions during those years.

- With climatic data, it is standard to also show Conley (1999) standard errors to deal with spatial dependency between districts. However, I do not think it is possible with the new technique. That should be acknowledge. Eventually, if standard DID gives similar results, show Conley SE as a robustness check?

  Thank you for your suggestion regarding the use of Conley (1999) standard errors to address spatial dependency in climatic data. While I appreciate the recommendation, the nature of the new technique employed in the analysis makes it challenging to incorporate Conley standard errors. However, I will certainly consider your suggestion to conduct robustness checks with Conley standard errors if standard DID results align with the current findings.

- Tables 1.5 and 1.6 are very small for "old eyes" like me. Maybe worth presenting them in landscape?

  Done, thank you.

- Figure 1.3a. I do not see how it is up to 4 years. It seems at best up to 3 years, no? Visually, I would even say 2 years?

  I have revised the figure accordingly to better reflect the correct time span, now up to 3 years.

- Results section: You can help the reader a bit more by showing where to find the results (e.g. 15.1%, 9.2% and 8.9%).

  Done, thank you.

- Figure 1.3. Repeating graph titles is unnecessary.

  Corrected.

Chapter 2 Historical drought and FDI in low-income countries: Community level analysis

1. **Contribution.** The paper can be seen as a paper on education and FDI, exploiting the exogeneous variation of the drought for identification. I am not asking to re-write the paper, but I would better define that contribution in the introduction. Also in the literature review, you mention that literature but I would try to discuss the methodological challenge of the literature seeking to assess the importance of human capital for FDI. That would strengthen your contribution.

   Thank you for your valuable feedback. In the literature review, I have expanded the discussion to highlight the methodological challenges prevalent in the existing literature that attempts to assess the significance of human capital for FDI. This enhancement aims to underscore the paper's contribution by addressing and overcoming these methodological challenges. (Page 51 paragraph 2 & 3)

2. **Literature.** I found a bit strange to claim that Carleton and Hsiang show "that children who are exposed to drought in utero score significantly worse on literacy and numeracy tests than their peers." (page 51). Carleton and Hsiang is a review paper. Also, I may have missed it but I have not seen the JEEA paper by Stefan Dercon and Catherine Porter on the long-term effect of droughts in Ethiopia. I may be wrong. I also suggest to make the effort to relate the literature review to your paper. Let me give an example. For instance, you add one paragraph : "Therefore, there is ample evidence of the impact of early drought exposure on adulthood economic outcomes … Abiona, 2017." So what? What does it say about your contribution?

   Thank you for pointing out the clarification regarding Carleton and Hsiang. Regarding Dercon and Porter paper, they indeed have a paper titled "Long-term effects of the 1984 Ethiopian famine on children's educational and health outcomes," published in the Journal of the European Economic Association (JEEA) in 2014, which supports the claim about the impact of drought on literacy and numeracy scores. I've made the necessary correction, and I've also combined the literature review with the introduction section to enhance the paper's coherence and explicitly highlight its contribution. (2nd paragraph of page 51 and 1st paragraph of page 52)

3. **Identification.** Please make equation 2.2 consistent with the explanation of the variables. For instance, Equation 2.1 includes X_ct, but the explanation below mentions X_i. It is very important for the reader (and for me) to understand what is the variation which is exploited (see below).

   Thank you. Corrected.

4. **Identification.** I suggest the candidate to better explain the identification strategy. From Equation (2.1), Wondatir compares communities affected by the drought, compared to regions not affected by the

drought (I do not know what is the subscript e). So, it is similar to a cross-section (there is no time variation in the treatment variable). My guess is that if you had aggregated the data at the community level, it would have given similar results because of that lack of time variation. From what I understand, Table 2 tells us that there is less FDI in drought-affected communities. But once again I do not see any time variation. That is maybe why you include region fixed effects and not community fixed effect (the treatment would drop if you were including community fixed effects)?

Thank you for your feedback, and I have made the necessary corrections to clarify the notation and address your concerns. You are correct in noting that the identification strategy involves comparing communities affected by drought to regions not affected by drought. I included regional fixed effects, such as ethnic dominance, as community fixed effects would lead to the treatment dropping due to the lack of time variation. The remaining controls are indeed community specific.

5. **Identification & contribution:** I May have missed something. But if Equation (2.1) is identified from cross-sectional variation, it is still fine to me to show a negative correlation between drought and FDI as a starting point. But you need to be clear about what you are identifying. Then, you can use that "puzzling" fact (interesting to see that those communities affected in 1984, are still those with less FDI in 2011-2016?). Then, I would suggest to have a second part of the paper trying to explain that. The conceptual framework on page 56 seems to be a very restricted view. The effect on human capital is only one possible driver? The economic geography or historical economics is full of hysteresis effects showing that an historical shocks may affect long-term economic development, through human capital but also agglomeration effects, cultural norms, local politics, … I do not know enough about Ethiopia but it seems to me, you need to conjecture alternative explanations on how drought affects FDI.

Thank you for your thoughtful review. I have addressed your suggestion by expanding the conceptual framework to include the economic geography channel in addition to the health channel. The revised version, detailed on page 55, provides a more comprehensive understanding of alternative explanations for how drought affects FDI. Furthermore, I have utilized this enhanced framework to reinforce the argument and interpretation of the results obtained from Equation 2.1, as discussed on page 63.

6. **Identification for mediation analysis.** Same comments apply here (subscripts mismatch, which variation, …) Table 2.8 needs to be explained. Without explanations, I cannot see how "The identification design used for identifying these individuals is outlined in Table 2.8." (page 54). I understood that Wondatir selects individuals born between 1978 and 1985 from the LSMS and compute average level of education.

8

But if you want to capture the importance of the drought- induced changes in human capital (in contrast with the fact drought-affected areas have different levels of education for many different reasons), shall you use the ratio between the mean value of education for affected cohorts compared to the average level of education of the rest of the community population? It is actually a good placebo test to see whether you would have an effect for cohorts not directly affected by the drought (if so, you are capturing something else)?

Thank you for the comment and I added the explanation on revised version of the paper on page 54, 2nd paragraph, "This table presents the age cohort and the corresponding number of individuals included in the calculation of the average years of schooling at the community level across three survey waves (2011, 2013, and 2015)"

7. **Mediation analysis:** For the reason above, I do not think you can claim that: **"the findings presented in table 2.3 from equation 2.2 demonstrate that early exposure to drought has an adverse impact on educational attainment within communities"**. I could simply conclude that affected communities have lower level of education, but that might be for many other reasons. For instance, we know from Amartya Sen works that the damages from the drought were particularly strong in disadvantaged areas (due to the political economy of relief). That would begood to look into his seminal work.

In response to the this feedback, it's recognized that several factors could contribute to the lower education levels observed in affected communities. The study acknowledges limitations in the data, such as the inability to consider certain factors like the political economy of relief, which might influence educational disparities. However, it's emphasized that the paper has taken steps to address this by controlling for various relevant variables within the available dataset. These controls encompass factors like access to basic infrastructure and markets, conflict, migration and political dominance, providing a better understanding of how early drought exposure relates to educational outcomes at the community level.

8. **Standard errors.** The treatment is defined at the community level. Standard errors need to be clustered at that level for Equations 2.1 and 2.2. (unless the sampling design requires another approach, see e.g. the work by Abadie et al. 2022: When should you adjust standard errors for clustering?).

Standard error is clustered at community level

9. **Internal migration.** The major drought took place in 1984, therefore you assume that the level of education born during that period can be attributed to the region where they live in. Well, you rightly point to the displacement that occur at the time of the drought. How does it affect your results? Similarly, there is a whole literature investigating the role of selective mortality following events like

drought. How does it affect your results with respect to education?

Thank you for your insightful comments. Regarding migration, the paper has controlled for mobility to and from the community using the available dataset. However, it's important to note that the dataset doesn't explicitly provide information on mortality related to the indicated drought.

10. **Proof-reading.** The paper needs to be proof-read. I give a few example below

- There are quite a lot of references with parentheses (e.g. (Carleton and Hsiang, 2016) instead ofCarleton and Hsiang, (2016)) when parentheses should not be included.

  Corrected

- Page 53 "officially recorded , 2021." ???

  Corrected

- Reorder tables in appendix, the way they appear in the text.

  Corrected

- Page 62: "distance from the main road and distance from the market **was** found statistically significant **in** negatively affecting the probability of FDI arrival in the particular locality."

  Corrected