Peer review history for:

Fourie J. Inequality in the Cape Colony, 1685–1844. S Afr J Sci. 2025;121(11/12), Art. #20532. https://doi.org/10.17159/sajs.2025/20532

HOW TO CITE:
Inequality in the Cape Colony, 1685–1844 [peer review history]. S Afr J Sci. 2025;121(11/12), Art. #20532. https://doi.org/10.17159/sajs.2025/20532/peerreview

The original manuscript for review is appended below.

Reviewer 1: Round 1
Date completed: 06 March 2025
Conflicts of interest: None

Recommendation: Accept / Revisions required / Resubmit for review / Resubmit elsewhere / Decline / See comments
REVIEWER: Revisions required
Does the manuscript fall within the scope of SAJS? Yes/No
REVIEWER: Yes
Is the manuscript written in a style suitable for a non-specialist and is it of wider interest than to specialists alone? Yes/No
REVIEWER: Yes
Does the manuscript contain sufficient novel and significant information to justify publication? Yes/No
REVIEWER: Yes
Do the Title and Abstract clearly and accurately reflect the content of the manuscript? Yes/No
REVIEWER: No
Is the research problem significant and concisely stated? Yes/No
REVIEWER: Yes
Are the methods described comprehensively? Yes/No
REVIEWER: Yes
Is the statistical treatment appropriate? Yes/No/Not applicable/Not qualified to judge
REVIEWER: Yes
Are the interpretations and conclusions justified by the research results? Yes/Partly/No
REVIEWER: Yes
Please rate the manuscript on overall contribution to the field Excellent/Good/Average/Below average/Poor
REVIEWER: Good

Please rate the manuscript on language, grammar and tone
Excellent/Good/Average/Below average/Poor

REVIEWER: Good

Is the manuscript succinct and free of repetition and redundancies?
Yes/No

REVIEWER: Yes

The number of tables in the manuscript is
Too few/Adequate/Too many/Not applicable

REVIEWER: Too few

The number of figures in the manuscript is
Too few/Adequate/Too many/Not applicable

REVIEWER: Too few

Are the results and discussion confined to relevance to the objective(s)?
Yes/No

REVIEWER: Yes

Is the supplementary material relevant and separated appropriately from the main document?
Yes/No/Not applicable

REVIEWER: Yes

Please rate the manuscript on overall quality
Excellent/Good/Average/Below average/Poor

REVIEWER: Good

Is appropriate and adequate reference made to other work in the field?
Yes/No

REVIEWER: No

Is it stated that ethical approval was granted by an institutional ethics committee for studies involving human subjects and non-human vertebrates?
Yes/No/Not applicable

REVIEWER: Not applicable

If accepted, would you recommend that the article receives priority publication?
Yes/No

REVIEWER: No

Are you willing to review a revision of this manuscript?
Yes/No

REVIEWER: No

With regard to our policy on 'Publishing peer review reports', do you give us permission to publish your anonymised peer review report alongside the authors' response, as a supplementary file to the published article? Publication is voluntary and only with permission from both yourself and the author.
Yes/No

REVIEWER: Yes

Comments to the Author:
In the review of the literature, author should more precisely recall what we know about South Africa's trends in inequality during the 20th century. For recent estimates of wealth inequality, you may also want to refer to: Chatterjee, A., Czajka, L., & Gethin, A. (2022). Wealth Inequality in South Africa, 1993–2017. The World Bank Economic Review, 36(1), 19-36.

"the conventional view that inequality has steadily increased over time" – in South African in particular, or in general? Source? If anything, I would expect that, compared to the 17th, 18th and 19th century, inequality is often lower now. The author claim here seems to be based on a misinterpretation of Piketty's book "Capital in the 21st century". In this book Piketty argues that inequality has been increasing since the end of the second world war, not that it was already increasing before (still in line with most recent estimates available for France or the UK for instance). What is true though, as author point themselves, is that even for data rich countries, inequality levels before the late 19th century are very uncertain. The

contribution of the paper is not that it contradicts "conventional view" but that it reduces uncertainty about (some) inequality levels for South Africa. It is important to rephrase the contribution as such and to revise the "Measuring Inequality" part accordingly. Related to that section also, I would suggest the author to remove the paragraph that list potential determinants of inequality according to Van Zanden, Alfani, Piketty, Derenoncourt, and Milanovic. What drives inequality is a very vast question which many more authors have tackled and this paper is mostly about measuring inequality level before the 20th century. The author should focus on this contribution and tighten the link with historical literature on measuring inequality before the 20th century rather than shortly elaborating on the literature about its determinant. In that regard, the author's review miss at least one very important study using similar methodology "Piketty, T., Postel-Vinay, G., & Rosenthal, J. L. (2006). Wealth concentration in a developing economy: Paris and France, 1807–1994. American economic review, 96(1), 236-256." I strongly suggest the author to read that work and enrich the current paper accordingly.

« As Chancel et al. Note… » missing reference.

On the data: the author should elaborate more about why these data were produced at their time. Take the tax censuses for instance: how was the tax due computed? Who were meant to be taxed? Were they any exemptions, below a certain threshold for instance? Could this lead to a certain type of bias? What type of bias should we expect? For the probate data, you mention that the coverage of the settler population declines after the 1740s: by how much? couldn't you document this with the data you have? Also you say that fixed household assets such as land or property (which you do not measure) "drive wealth inequality" – do you mean "increase"? This is not certain, if housing is more equally distributed than say "horses" then adding it to the picture would decrease inequality.

For non-South-African readers mostly, author could recall that the former Cape Colony and current South Africa do not cover the same area. It would very valuable also to provide maps displaying the boundaries of the different geographical location / region / district within the boundary of the current South Africa.

Producing the evolution gini coefficients is very relevant to the topic but it is not sufficient. For each of the data source they have, the author should produce tables with at least some aggregate statistics such as: a) total population by type (settler, slave and natives) and by region / district, and b) average assets owned, by type of person, by region / district and by asset type, but also average household size, sex ratio etc. In general, it would be great to better document the Cape Colony over that period of time. Any statistics giving a better idea of the size and type of this economy would be interesting. To appreciate your results on the "Stellenbosch-Drakenstein" region or instance, we would need to better understand how it relates to the rest of the Colony, use the available data to give a sense of how important this district is to the entire Colony – (in particular, there are many more things you can do with the 1825 census data where you have the eleven districts).

It is also crucially important to document joint ownership of different assets. Indeed, production may be specialized and assets could be split across specialized groups, in which case (depending on the value of the assets) the colony might not be as unequal as the author claim (in theory it could even be perfectly egalitarian among the non-slave population). It would be interesting for instance to measure the share (in the total population considered) that have none of the assets listed.

"A total of 142,219 observations (at the household level) were recorded across 159 censuses" This figure is not really useful, it is better to the total per year.

Figure 1 – ginis are estimated based on variables measuring counts and not values, but you mention you have combined assets using prices, you should include the gini for this variable into the Figure 1. You also mention you have information on slave ownership – if you can calculate gini of slave ownership as for other types of assets, you should also include it in Figure 1. Having information on the evolution of overall slave to owner ratio would be interesting as well (see my more general comments about documenting other aggregate statistics across observable caracteristics). Be more precise about the statistical unit: is it a)

household ? or b) equal split among censed (free) adults in the household ? or c) equal split among censed (free) individuals (adults + children) in the household ? Ideally you should produce estimates using both b) and c) equivalent scales (decide for one in the core of the paper, leave the other for the appendix). This comment applies to all other figures where the statistical unit is not explicitly mentioned. Also, in the data description, you mention that toward the end of the tax census some information about taxes paid were collected. Even if for the later period, it would be great to provide some statistics as well the distribution of theses taxes.

On Figure 1 and 2: be explicit about the fact that natives are not included here when you present the results. In that regard, it would be interesting to know at least which share of the total population we are missing here.

“relative equality of land ownership” source?

Is there anything you can tell about (and possibly do with) prices? I strongly suggest to produce inequality statistics in total valued wealth whenever possible (even if only for a short period of time).

For comparison with other work, and readability for non-specialists, it would be interesting as well to produce metrics in terms of top shares. Given the levels of inequality –top 1%, top 0.1% or top 0.01% of the population for instance.

Figure 3: in my view, this visualization would be more readable with a bar graphs. Why only 1825? If available data allows you to compare the “Stellenbosch-Drakenstein” region with all, or even only some other region, then plotting the evolution of inequality in Cattle, Horses, Oxen etc. by region, would be interesting.

If the data will be made available, now or later, this is a central contribution of the work that should be mentioned somewhere.

Figure 4: be explicit in the legend about the region covered. Also, you mention that “probates cover the entire period of analysis and all districts” but you do not document variation of probate inequality across regions.

Figure 5: in the legend, “before” and “after” are not clear enough. It would be preferable to use something that indicates what has been done, example: “excluding non-slave owning settlers” “including non-slave owning settlers”.

On the analysis of the Khoe settlements: is there any way to produce inequality statistics combining information about settlers and natives to better document how the within inequality influence overall inequality? In theory at least large within inequality do not necessarily imply similarly large overall inequality.

Please elaborate on the type of economy: what did these household produce? What were the assets you study used for? What were other unmeasured assets and what these used for? The fact that you miss land and housing for that period implies you probably miss (at least one of) the main component of wealth, can you say anything about that type wealth?

I encourage the author to carefully highlight plausible interpretation of some of the temporal variation they observe in the data.

On the overall contribution: most of what you know is about inequality within settlers. Very little can be said about the natives. I would make this much more salient throughout the paper. I suggest the be explicit about it in the title even - “Inequality among settlers in the Cape Colony, 1685-1844” or “Withing group inequality in the Cape Colony, 1685-1844”. Also, it would be great to comment on the link between the

high within group inequality you find and what we know about between group inequality in South Africa over that period of time.

Author response to Reviewer 1: Round 1

In the review of the literature, author should more precisely recall what we know about South Africa's trends in inequality during the 20th century. For recent estimates of wealth inequality, you may also want to refer to: Chatterjee, A., Czajka, L., & Gethin, A. (2022). Wealth Inequality in South Africa, 1993–2017. The World Bank Economic Review, 36(1), 19-36.

AUTHOR: Thank you for the suggestion; I have added a sentence referencing Chatterjee, Czajka, and Gethin (2022) to more precisely situate recent estimates of post-apartheid wealth inequality within the literature review.

"the conventional view that inequality has steadily increased over time" – in South African in particular, or in general? Source? If anything, I would expect that, compared to the 17th, 18th and 19th century, inequality is often lower now. The author claim here seems to be based on a misinterpretation of Piketty's book "Capital in the 21st century". In this book Piketty argues that inequality has been increasing since the end of the second world war, not that it was already increasing before (still in line with most recent estimates available for France or the UK for instance). What is true though, as author point themselves, is that even for data rich countries, inequality levels before the late 19th century are very uncertain. The contribution of the paper is not that it contradicts "conventional view" but that it reduces uncertainty about (some) inequality levels for South Africa. It is important to rephrase the contribution as such and to revise the "Measuring Inequality" part accordingly. Related to that section also, I would suggest the author to remove the paragraph that list potential determinants of inequality according to Van Zanden, Alfani, Piketty, Derenoncourt, and Milanovic. What drives inequality is a very vast question which many more authors have tackled and this paper is mostly about measuring inequality level before the 20th century. The author should focus on this contribution and tighten the link with historical literature on measuring inequality before the 20th century rather than shortly elaborating on the literature about its determinant. In that regard, the author's review miss at least one very important study using similar methodology "Piketty, T., Postel- Vinay, G., & Rosenthal, J. L. (2006). Wealth concentration in a developing economy: Paris and France, 1807–1994. American economic review, 96(1), 236-256." I strongly suggest the author to read that work and enrich the current paper accordingly.

AUTHOR: Thank you for this valuable clarification. I have removed the reference to a "conventional view" of steadily rising inequality and eliminated the earlier discussion of Piketty's Capital in the 21st Century. Instead, I now ground the discussion in the empirical literature on long-run inequality trends, including the referee's suggested reference to Piketty, Postel-Vinay, and Rosenthal (2006), while clearly positioning the contribution of the paper as a reduction of uncertainty about historical inequality levels in South Africa. The "Measuring Inequality" section has been revised accordingly to better reflect the focus on methodological innovation and historical documentation.

« As Chancel et al. Note… » missing reference.

AUTHOR: Thank you. Fixed.

On the data: the author should elaborate more about why these data were produced at their time. Take the tax censuses for instance: how was the tax due computed? Who were meant to be taxed? Were they any exemptions, below a certain threshold for instance? Could this lead to a certain type of bias? What type of bias should we expect? For the probate data, you mention that the coverage of the settler population declines after the 1740s: by how much? couldn't you document this with the data you have? Also you say that fixed household assets such as land or property (which you do not measure) "drive wealth inequality" – do you mean "increase"? This is not certain, if housing is more equally distributed than say "horses" then adding it to the picture would decrease inequality.

AUTHOR: I thank the referee for their detailed and constructive comments on the sources. In response, I have added a new paragraph discussing the opgaafrolle in greater detail, including how taxes were assessed, who was taxed, and the likely implications for selection and reporting bias. This paragraph now addresses questions about enforcement, exemptions, and the direction of possible biases. I have also removed the phrase suggesting that fixed assets "drive wealth inequality"; while the original intent was to

contrast wealth with income inequality, I recognise that the wording was unclear, particularly given the possibility that land and property may have been more evenly distributed than other assets.

For non-South-African readers mostly, author could recall that the former Cape Colony and current South Africa do not cover the same area. It would very valuable also to provide maps displaying the boundaries of the different geographical location / region / district within the boundary of the current South Africa.

AUTHOR: In response to the helpful suggestion regarding historical and geographical context, I have added a brief history of the Cape Colony at the start of the results section. While I have not included a map, this addition is intended to offer non-South African readers a clearer sense of the colony's expansion, regional distinctions, and economic geography.

Producing the evolution gini coefficients is very relevant to the topic but it is not sufficient. For each of the data source they have, the author should produce tables with at least some aggregate statistics such as: a) total population by type (settler, slave and natives) and by region / district, and b) average assets owned, by type of person, by region / district and by asset type, but also average household size, sex ratio etc. In general, it would be great to better document the Cape Colony over that period of time. Any statistics giving a better idea of the size and type of this economy would be interesting. To appreciate your results on the "Stellenbosch-Drakenstein" region or instance, we would need to better understand how it relates to the rest of the Colony, use the available data to give a sense of how important this district is to the entire Colony – (in particular, there are many more things you can do with the 1825 census data where you have the eleven districts).

AUTHOR: "I appreciate the referee's suggestion to include more descriptive statistics on the structure of the Cape Colony's population and economy. Space constraints prevent a fuller tabulation of district- level aggregates across the three datasets, but I have added a new paragraph that uses the 1825 opgaafrolle and 1834 slave valuation rolls to provide contextual figures on household counts, production levels, and the regional distribution of enslaved people. This allows readers to better gauge the importance of the Stellenbosch–Drakenstein region relative to the rest of the colony.

While I agree that further breakdowns – by household size, sex ratios, and asset composition – would be valuable, I have prioritised providing the most relevant comparisons to support the interpretation of inequality trends. I hope future work can build on this foundation to offer a more comprehensive demographic and economic profile of the colony."

It is also crucially important to document joint ownership of different assets. Indeed, production may be specialized and assets could be split across specialized groups, in which case (depending on the value of the assets) the colony might not be as unequal as the author claim (in theory it could even be perfectly egalitarian among the non-slave population). It would be interesting for instance to measure the share (in the total population considered) that have none of the assets listed.

AUTHOR: I thank the referee for raising the important issue of asset complementarity and potential specialisation across settler households. While space constraints prevent me from presenting full cross-tabulations of asset ownership, I have clarified in the text that most settler farms in the Cape combined wheat cultivation, viticulture, and livestock rearing – suggesting a high degree of joint asset ownership rather than specialised production. I also computed Gini coefficients based on total asset value (not shown), which yield similarly high levels of inequality, reinforcing the main results. While the concern that specialisation might artificially inflate asset-based inequality is theoretically valid, the data suggest that this is not a major issue in the Cape context.

"A total of 142,219 observations (at the household level) were recorded across 159 censuses" This figure is not really useful, it is better to the total per year.

AUTHOR: Thank you for the suggestion. I have revised the text to include the number of households recorded per year, reporting averages for the first decade (1685–1694), the final decade (1835–1844), and the full period. This provides a clearer sense of how census coverage expanded over time.

Figure 1 – ginis are estimated based on variables measuring counts and not values, but you mention you have combined assets using prices, you should include the gini for this variable into the Figure 1.

AUTHOR: Thank you for the suggestion. I have now included another line – the total value of all six assets combined – to show what I only stated in words before. I've also updated the relevant paragraph.

You also mention you have information on slave ownership – if you can calculate gini of slave ownership as for other types of assets, you should also include it in Figure 1. Having information on the evolution of

overall slave to owner ratio would be interesting as well (see my more general comments about documenting other aggregate statistics across observable caracteristics). Be more precise about the statistical unit: is it a) household ? or b) equal split among censed (free) adults in the household ? or c) equal split among censed (free) individuals (adults + children) in the household ? Ideally you should produce estimates using both b) and c) equivalent scales (decide for one in the core of the paper, leave the other for the appendix). This comment applies to all other figures where the statistical unit is not explicitly mentioned.

AUTHOR: I thank the referee for raising these important issues. In response, I have made several substantive changes to the manuscript. First, I now present a more detailed analysis of slave ownership and its contribution to inequality in a new section of the paper. This includes a Gini coefficient for slave ownership, which is shown to consistently exceed 0.7 throughout the period. Due to space constraints, I have not added this to Figure 1 (which focuses on asset-based inequality among settler households), but instead expanded Figure 2 to present a range of inequality estimates under different assumptions about slave inclusion. These include scenarios where slaves are treated as assets, and as separate households, both with and without imputed earnings. This approach captures the impact of slavery not only as an economic input but as a structuring force in the colony's broader inequality.

Second, I now explicitly state that the statistical unit throughout the paper is the household, and that inequality is calculated across households as observed in the tax censuses. This is particularly important in a context where assets were typically held and reported at the household level. I agree with the suggestion to explore alternative equivalence scales – for instance, dividing assets among free adults or among all free individuals including children – and I have included a note in the appendix outlining how such adjustments might be implemented in future work. For now, I retain the household as the core unit of analysis, as it aligns most closely with the structure of the sources and the institutional logic of taxation and production in the Cape."

Also, in the data description, you mention that toward the end of the tax census some information about taxes paid were collected. Even if for the later period, it would be great to provide some statistics as well the distribution of theses taxes.

AUTHOR: I agree that the taxation data collected in the later years of the opgaafrolle are of considerable interest, particularly for understanding the relationship between wealth and fiscal extraction in the Cape Colony. However, incorporating this analysis lies beyond the scope of the present paper, which focuses specifically on long-run patterns of asset-based inequality. The taxation records are sparse, inconsistent, and often require different forms of interpretation and validation. I am currently preparing a separate paper that will focus more directly on the evolution of Cape fiscal policy and tax incidence. I hope to engage more fully with the distributional implications of taxation in that context.

On Figure 1 and 2: be explicit about the fact that natives are not included here when you present the results. In that regard, it would be interesting to know at least which share of the total population we are missing here.

AUTHOR: I thank the referee for highlighting this important limitation. I now make explicit in the main text that the inequality estimates presented in Figures 1 and 2 exclude indigenous communities, including the Khoe, who were not systematically recorded in the settler tax censuses. While Khoe labour was vital to the colonial economy, especially on the frontier, population estimates remain too uncertain for meaningful inclusion in colony-wide measures of inequality. For this reason, I have clarified that the results apply to settler and enslaved populations only. In a later section, I return to examine the subset of Khoe households that are recorded in the tax censuses and explore patterns of asset ownership within that group.

"relative equality of land ownership" source?

AUTHOR: Added.

Is there anything you can tell about (and possibly do with) prices? I strongly suggest to produce inequality statistics in total valued wealth whenever possible (even if only for a short period of time).

AUTHOR: Thank you for this helpful suggestion. I have incorporated price data by converting quantities into values (quantity × price) for all relevant goods, and now report inequality in terms of total valued wealth in both Figure 1 and Figure 2. While a separate analysis of price trends would indeed be valuable, I believe it constitutes a distinct research question and falls outside the scope of this paper.

For comparison with other work, and readability for non-specialists, it would be interesting as well to produce metrics in terms of top shares. Given the levels of inequality –top 1%, top 0.1% or top 0.01% of the population for instance.

AUTHOR: Thank you for this helpful suggestion. I have now calculated top wealth shares (top 1%, 0.1%, and 0.01%) using the same value-based asset estimates that underpin Figure 1. These results, which are consistent with the Gini coefficients, suggest that for most of the period, the top 1% of households held between 9% and 12% of total wealth. Only after the 1810s does this share rise sharply – possibly due to inflationary effects. However, because land is not included here, I feel it would be incorrect to report these numbers in the paper, as they will immediately be compared with more recent estimates. I've therefore not included a discussion of these.

Figure 3: in my view, this visualization would be more readable with a bar graphs.

AUTHOR: Thank you for this helpful suggestion. In response, I experimented with several alternative formats, including bar charts, but ultimately chose a bubble plot faceted by asset type. This visualisation retains the Gini coefficient on the y-axis, as before, but now scales the size of each circle to reflect the total amount of that asset produced in each district. To ensure comparability within each facet, circle sizes are normalised relative to the maximum asset quantity in that asset type. This allows the figure to communicate both inequality and economic weight at a glance, while avoiding clutter and making variation across districts more legible than in the original dot plot. I believe this version addresses the clarity concern and adds valuable context.

Why only 1825? If available data allows you to compare the "Stellenbosch-Drakenstein" region with all, or even only some other region, then plotting the evolution of inequality in Cattle, Horses, Oxen etc. by region, would be interesting.

AUTHOR: I agree that regional comparisons over time would be valuable, and this is something I hope to pursue in future work. For the purposes of this paper, however, I focus on 1825 because it is the only year for which we have a complete tax census for all districts in the Colony. This allows for a clean cross-sectional comparison of inequality across regions without the complications introduced by inflation or valuation changes over time, which tend to affect temporal comparisons more than spatial ones. The 1825 dataset thus provides a unique benchmark for contextualising inequality in Stellenbosch and Drakenstein relative to the rest of the Cape.

If the data will be made available, now or later, this is a central contribution of the work that should be mentioned somewhere.

AUTHOR: Thank you for highlighting this important point. I have now added a sentence to clarify that both the data and source code used in this paper will be made publicly available upon completion of the project in 2026. Given the novelty and richness of the archival sources digitised for this study, I agree that this constitutes a central contribution of the work, and have ensured that this is clearly noted in the manuscript. Further details and updates on data availability are provided at www.capepanel.org.

Figure 4: be explicit in the legend about the region covered. Also, you mention that "probates cover the entire period of analysis and all districts" but you do not document variation of probate inequality across regions.

AUTHOR: Thank you for this comment. While these records span the full period of analysis and were filed across multiple districts, they typically only include farm names rather than specific geographic identifiers, and these farm names cannot reliably be mapped to modern district boundaries. As such, it is not possible to document variation in probate inequality across regions. I refer instead to the coverage as representing the broader area of settler settlement in the Colony and will ensure this limitation is clearly stated in the manuscript.

Figure 5: in the legend, "before" and "after" are not clear enough. It would be preferable to use something that indicates what has been done, example: "excluding non-slave owning settlers" "including non-slave owning settlers".

AUTHOR: Thank you for this helpful suggestion. I have updated the legend in Figure 5 to use more descriptive labels: "Excluding non–slave-owning settlers" and "Including non–slave-owning settlers". This more clearly communicates the distinction between the two specifications shown in the figure and aligns with the terminology used elsewhere in the manuscript.

On the analysis of the Khoe settlements: is there any way to produce inequality statistics combining

information about settlers and natives to better document how the within inequality influence overall inequality? In theory at least large within inequality do not necessarily imply similarly large overall inequality.

AUTHOR: Thank you for this helpful suggestion. I have now included new estimates that combine settler households and Khoe settlements within the district of Swellendam. These figures confirm that inequality remained exceptionally high when measured across population groups, with Gini coefficients exceeding 0.77 for horses, oxen, and wheat, and reaching as high as 0.92 for cattle. The text has been updated to reflect these results.

Please elaborate on the type of economy: what did these household produce? What were the assets you study used for? What were other unmeasured assets and what these used for? The fact that you miss land and housing for that period implies you probably miss (at least one of) the main component of wealth, can you say anything about that type wealth?

AUTHOR: Thank you for this thoughtful comment. I have expanded the manuscript to provide additional detail about the type of economy, including the mixed agricultural strategies of most households, the functions of the measured assets (such as livestock for draft, meat, dairy, and market exchange), and the diversified production across wheat, wine, and animal husbandry. I fully agree that land and housing were important components of wealth in this society. However, such assets are not recorded in the tax censuses or probate records used in this study, and are therefore unfortunately not measurable in a consistent or representative way across the period. While this is a limitation (which I do mention in the text), the agricultural assets captured here (including livestock and crop production) formed the core of productive wealth for most households. I have clarified this point in the text and added appropriate caveats where necessary.

I encourage the author to carefully highlight plausible interpretation of some of the temporal variation they observe in the data.

AUTHOR: Thank you for the suggestion. Where appropriate, I now briefly note potential explanations for some of the temporal variation in inequality, such as the possible role of inflation in driving the uptick in the 1810s and 1820s. However, I have chosen not to speculate further, as the primary aim of the paper is to document long-run patterns in inequality rather than to explain their underlying causes. For this reason, I also removed the earlier discussion of possible drivers from the introduction. I do agree that these are important questions and hope they will form the basis of future work building on the dataset presented here.

On the overall contribution: most of what you know is about inequality within settlers. Very little can be said about the natives. I would make this much more salient throughout the paper. I suggest the be explicit about it in the title even - “Inequality among settlers in the Cape Colony, 1685-1844” or “Withing group inequality in the Cape Colony, 1685-1844”. Also, it would be great to comment on the link between the high within group inequality you find and what we know about between group inequality in South Africa over that period of time.

AUTHOR: Thank you for this important suggestion. I agree that the issue of whose inequality is being measured must be handled with care. While the majority of the analysis focuses on settler households, one of the key contributions of the paper is precisely that it goes beyond this group to include enslaved individuals and, where records allow, Khoesan workers and mission residents. Though coverage remains imperfect — particularly for indigenous communities not incorporated into colonial recordkeeping — the inclusion of these groups makes it possible to examine a broader and more representative segment of the colonial population than has previously been possible. I have clarified this contribution more explicitly in the introduction and conclusion. I have also opted to retain the current title, which reflects this broader ambition and avoids reinforcing a narrow settler-focused narrative.

Reviewer 2: Round 1
Date completed: 20 May 2025
Conflicts of interest: None

Recommendation: Accept / Revisions required / Resubmit for review / Resubmit elsewhere / Decline / See comments
REVIEWER: Revisions required
Does the manuscript fall within the scope of SAJS? Yes/No
REVIEWER: Yes
Is the manuscript written in a style suitable for a non-specialist and is it of wider interest than to specialists alone? Yes/No
REVIEWER: Yes
Does the manuscript contain sufficient novel and significant information to justify publication? Yes/No
REVIEWER: Yes
Do the Title and Abstract clearly and accurately reflect the content of the manuscript? Yes/No
REVIEWER: Yes
Is the research problem significant and concisely stated? Yes/No
REVIEWER: Yes
Are the methods described comprehensively? Yes/No
REVIEWER: Yes
Is the statistical treatment appropriate? Yes/No/Not applicable/Not qualified to judge
REVIEWER: Yes
Are the interpretations and conclusions justified by the research results? Yes/Partly/No
REVIEWER: Partly
Please rate the manuscript on overall contribution to the field Excellent/Good/Average/Below average/Poor
REVIEWER: Good
Please rate the manuscript on language, grammar and tone Excellent/Good/Average/Below average/Poor
REVIEWER: Excellent
Is the manuscript succinct and free of repetition and redundancies? Yes/No
REVIEWER: Yes
The number of tables in the manuscript is Too few/Adequate/Too many/Not applicable
REVIEWER: Adequate
The number of figures in the manuscript is Too few/Adequate/Too many/Not applicable
REVIEWER: Adequate
Are the results and discussion confined to relevance to the objective(s)? Yes/No
REVIEWER: Yes
Is the supplementary material relevant and separated appropriately from the main document? Yes/No/Not applicable
REVIEWER: Yes

Please rate the manuscript on overall quality
Excellent/Good/Average/Below average/Poor

REVIEWER: Good

Is appropriate and adequate reference made to other work in the field?
Yes/No

REVIEWER: Yes

Is it stated that ethical approval was granted by an institutional ethics committee for studies involving human subjects and non-human vertebrates?
Yes/No/Not applicable

REVIEWER: Yes

If accepted, would you recommend that the article receives priority publication?
Yes/No

REVIEWER: Yes

Are you willing to review a revision of this manuscript?
Yes/No

REVIEWER: No

With regard to our policy on 'Publishing peer review reports', do you give us permission to publish your anonymised peer review report alongside the authors' response, as a supplementary file to the published article? Publication is voluntary and only with permission from both yourself and the author.
Yes/No

REVIEWER: Yes

Comments to the Author:
This is a very interesting and accessible read which draws on novel historical records to deepen our understanding of the roots of inequality in SA. The authors' are to be congratulated.

Major comments:
A central argument by the authors' is that South Africa's record levels of inequality are not new - in fact inequality has always been extreme.

While I find the evidence convincing, specifically that within-group inequality was very high across a range of communities, I am uncomfortable when the authors take this a step further to imply that 17/18th century wealth/income inequality is higher (or lower) than in the present. For instance consider these statements:

"I find that income and wealth concentration among settlers and within settler society, but also within those Khoe settlements for which suitable data is available, was extremely high, with Gini coefficients exceeding modern levels."

"The most unequal country in the world today may have been even more unequal three centuries ago."

I'm not sure the authors' should make such statements which need more direct comparisons? The bulk of the wealth inequality gini's reported in the paper are based on agricultural related assets (appropriate for the day). However it is not clear to me that these gini scores can be readily compared with post-apartheid related income/wealth gini scores? And most of the gini's reported in the paper are NOT for the whole population but for different groups separately i.e. within group inequality when most modern gini's combine within- and between- gini scores by considering the whole population.

To reiterate, I think the authors' can argue that inequality has always been extreme, but I'm not convinced that they should say it's higher or lower than the present unless we are measuring the same things over time (which is probably not possible or desirable). I think this needs to be toned down in the messaging.

A final more technical issue is the way in which inequality changes when including information on slaves as separate households (which I think comes down to how the authors impute earnings/wealth for slaves) -

which is also important because this approach is more directly comparable with modern gini scores which look at the whole population not groups separately. When income/wealth is assumed to be zero the gini is extreme ("very close to perfect inequality") but when assumed to be positive (1 sheep per month) they report "then the Gini coefficient fits very closely onto the one for settlers only, ranging between 0.75 and 0.8". This is still very high. Does this then imply that the earnings of slaves were essentially the same as the earnings of free settlers (i.e. are you sure your imputation is appropriate)? I find this hard to believe so please explain/clarify.

Minor comments:
"While much of the research attention, for obvious reasons, focuses on post-apartheid trends in inequalities and understanding their determinants, there is little doubt that South Africa's history of racial discrimination and repression explains these persistently high levels of inequality". I think race-based discrimination is part of the explanation for modern day inequality but it not the whole story. For instance, Owen Crankshaw argues that the outcome of inequality might look the same in democratic SA but the causes are fundamentally different - now based on market-based mechanisms of labour market sorting and professionalisation which is class (not race) sorting even if they are correlated.

I would recommend a footnote to explain the term 'VOC mercantilism' when it's first introduced.

I imagine the results might be sensitive to the technical details around how the authors define and measure income/wealth across different sources. I noticed that there is important supplementary materials (to help with transparency and replicability - which is good) BUT I could not find any reference to the supplementary materials in the main text.

"These newly transcribed sources provide a comprehensive basis for analysing the historical roots of inequality in the Cape Colony". It might be necessary to tone this sort of statement down (and in other parts) because there are still gaps and uncertainties in measuring historical records? Do the authors' want to convey that they provide a definitive measure of historical inequality?

"To calculate the actual settler inequality, I add 13033 non-slave owning settlers, and reestimate both Ginis. The results are also reported in Figure 5 (grey dots)." Please note that I think the grey dots referred to here have been changed to grey triangles.

Author response to Reviewer 2: Round 1

This is a very interesting and accessible read which draws on novel historical records to deepen our understanding of the roots of inequality in SA. The authors' are to be congratulated.

Major comments:
A central argument by the authors' is that South Africa's record levels of inequality are not new - in fact inequality has always been extreme.

While I find the evidence convincing, specifically that within-group inequality was very high across a range of communities, I am uncomfortable when the authors take this a step further to imply that 17/18th century wealth/income inequality is higher (or lower) than in the present. For instance consider these statements:

"I find that income and wealth concentration among settlers and within settler society, but also within those Khoe settlements for which suitable data is available, was extremely high, with Gini coefficients exceeding modern levels."

"The most unequal country in the world today may have been even more unequal three centuries ago."

I'm not sure the authors' should make such statements which need more direct comparisons? The bulk of the wealth inequality gini's reported in the paper are based on agricultural related assets (appropriate for the day). However it is not clear to me that these gini scores can be readily compared with post-apartheid related income/wealth gini scores? And most of the gini's reported in the paper are NOT for the whole population but for different groups separately i.e. within group inequality when most modern gini's combine within- and between- gini scores by considering the whole population.

To reiterate, I think the authors' can argue that inequality has always been extreme, but I'm not convinced that they should say it's higher or lower than the present unless we are measuring the same things over time (which is probably not possible or desirable). I think this needs to be toned down in the messaging.

AUTHOR: Thank you for this thoughtful and important suggestion. I have revised both the abstract and conclusion to avoid any direct numerical comparison between historical and modern levels of inequality. In its place, the revised framing now emphasises the historical persistence of extreme inequality as a structural feature of South African society, without implying equivalence in magnitude. The abstract has been updated to reflect this more cautious framing, highlighting within-group inequality and the concentration of productive resources without referring to modern benchmarks. The conclusion similarly refrains from comparative claims and instead underscores the long-run endurance of inequality, culminating in a strengthened final sentence that links the historical findings to their contemporary relevance. I believe these changes improve the clarity, rigour, and precision of the argument and thank the reviewer for prompting this important refinement

A final more technical issue is the way in which inequality changes when including information on slaves as separate households (which I think comes down to how the authors impute earnings/wealth for slaves) - which is also important because this approach is more directly comparable with modern gini scores which look at the whole population not groups separately. When income/wealth is assumed to be zero the gini is extreme ("very close to perfect inequality") but when assumed to be positive (1 sheep per month) they report "then the Gini coefficient fits very closely onto the one for settlers only, ranging between 0.75 and 0.8". This is still very high. Does this then imply that the earnings of slaves were essentially the same as the earnings of free settlers (i.e. are you sure your imputation is appropriate)? I find this hard to believe so please explain/clarify.

AUTHOR: Thank you for this helpful observation. In the course of addressing a related request from Referee 1, I identified and corrected a coding error that had affected the earlier imputed values for slaves treated as separate households. Figure 2 has since been completely revised, and now reflects a more realistic distinction between settler and slave asset distributions under both zero-asset and positive imputation scenarios. This correction also resolves the concern raised here: the revised estimates no longer suggest that slaves earned or owned assets comparable to those of settlers. I have clarified the imputation logic and included a more cautious interpretation in the text.

Minor comments:

"While much of the research attention, for obvious reasons, focuses on post-apartheid trends in inequalities and understanding their determinants, there is little doubt that South Africa's history of racial discrimination and repression explains these persistently high levels of inequality". I think race-based discrimination is part of the explanation for modern day inequality but it not the whole story. For instance, Owen Crankshaw argues that the outcome of inequality might look the same in democratic SA but the causes are fundamentally different - now based on market-based mechanisms of labour market sorting and professionalisation which is class (not race) sorting even if they are correlated.

AUTHOR: Thank you for this important point. I agree that the determinants of inequality in post-apartheid South Africa are complex, and that market-based mechanisms such as class-based labour market sorting now play a significant role alongside the legacies of racial discrimination. However, the aim of this paper is not to explain the causes of inequality in either the colonial or contemporary periods, but rather to document the levels and patterns of inequality in the Cape Colony over time, based on newly available historical data. As such, I have deliberately avoided entering into debates about the drivers of inequality, whether past or present, and have revised the manuscript to clarify this more explicitly. I hope that future work, including by other scholars, will explore the long-run mechanisms in greater depth using the historical estimates established here.

I would recommend a footnote to explain the term 'VOC mercantilism' when it's first introduced.

AUTHOR: Thank you for the suggestion. As I'm not allowed to add footnotes, I have added a few additional explanatory words in the text to make it clear.

I imagine the results might be sensitive to the technical details around how the authors define and measure income/wealth across different sources. I noticed that there is important supplementary materials (to help with transparency and replicability - which is good) BUT I could not find any reference to the supplementary materials in the main text.

AUTHOR: Thank you for this helpful observation. I have now included a reference to the supplementary materials in the main text, clarifying that all data construction procedures, variable definitions, and source code will be made publicly available at the completion of the project in 2026. More information and documentation will be hosted at www.capepanel.org. I agree that transparency is vital for replicability, and I appreciate the opportunity to make this commitment more visible to readers.

"These newly transcribed sources provide a comprehensive basis for analysing the historical roots of inequality in the Cape Colony". It might be necessary to tone this sort of statement down (and in other parts) because there are still gaps and uncertainties in measuring historical records? Do the authors' want to convey that they provide a definitive measure of historical inequality?

AUTHOR: The sentence you refer to has since been removed, and I have taken care throughout the revised manuscript to avoid suggesting that the estimates presented are definitive. While the newly transcribed sources do offer a significant improvement over previous approaches, particularly in moving beyond social tables, I agree that important gaps and uncertainties remain.

"To calculate the actual settler inequality, I add 13033 non-slave owning settlers, and reestimate both Ginis. The results are also reported in Figure 5 (grey dots)." Please note that I think the grey dots referred to here have been changed to grey triangles.

AUTHOR: Thank you for spotting that!

Reviewer 1: Round 2
Date completed: 13 June 2025
Conflicts of interest: None

Recommendation:
Accept / Revisions required / Resubmit for review / Resubmit elsewhere / Decline / See comments

REVIEWER: Accept

Have the authors adequately addressed the concerns raised in the previous review?
Yes/No

REVIEWER: Yes

Is the manuscript publishable in its current revised form?
Yes/No

REVIEWER: Yes

If yes, do you recommend priority publication?
Yes/No

REVIEWER: Yes

Any additional comments for the authors:
Make sure you include the Gini coefficients based on total asset value (p3 of the response, you say it is not shown).

"For now, I retain the household as the core unit of analysis, as it aligns most closely with the structure of the sources and the institutional logic of taxation and production in the Cape." In my understanding, this is not a very convincing argument. If you can, it is better to use household per capita or household per adult as the statistical unit.

"While a separate analysis of price trends would indeed be valuable, I believe it constitutes a distinct research question and falls outside the scope of this paper." To the extend that this paper is about wealth and that wealth valuation require price, my take is that providing information (at least in the text) is

relevant to the paper. In fact in the response to the follow-up comment you explain observed trends by the rise of inflation.

"However, because land is not included here, I feel it would be incorrect to report these numbers in the paper, as they will immediately be compared with more recent estimates." This is not a convincing argument to me. If it is possible to add it, I still believe computing top shares would be valuable.

In two or three instances, you argue that due to space constraints some tables cannot be added. To the editors : I am not aware of the precise rules regarding spaces at the South African Journal of Science, but it case some flexibility exists, it be great to see at least some of the additional exhibit I asked for (which one specifically ? to be decided by the authors/editors).

Author response to Reviewer 1: Round 2

I have made the following targeted revisions:

- Clarified in the text that Figure 1 includes a total-asset Gini series and added a signpost sentence to make this explicit.
- Retained the household as the unit of analysis and briefly set out the main reasons against per-capita or per-adult normalisation at the Cape (household size heterogeneity, treatment of hired and enslaved labour, and incomplete recording of dependants), while noting avenues for future work.
- Added a concise description of the rixdollar price series used to value the six assets, derived from the Cape Orphan Chamber auction rolls (MOOC10, 1701–1825), and summarised their broad stability in the eighteenth century and sharp rise and dispersion from the 1780s, explaining their role in like-for-like valuation.
- Reported top-share metrics for comparability with other work, presenting the top 1%, 0.1% and 0.01% for the six-asset basket only, with a clear caution on non-comparability to later periods that include land and other assets.
- Given the word limit and no specific editorial preference for additional tables, I propose no further tables or text be added.

Inequality in the Cape Colony, 1685-1844

Abstract

South Africa has one of the highest levels of inequality globally. This inequality is not a recent phenomenon. I use several newly transcribed datasets from the seventeenth, eighteenth and nineteenth centuries to calculate wealth inequality in the Cape Colony. The sources – tax censuses, probate inventories and slave valuation rolls – allow comparisons across space and time within the Colony. The Cape was not only a persistently unequal society measured across the entire population of settler, slave and Khoe, but the evidence suggests high levels of within-group inequality. Historically, inequality was the norm.

Significance:

- I challenge the conventional view that inequality has steadily increased over time, revealing that the Cape Colony in the eighteenth and nineteenth centuries already experienced extreme income and wealth concentration.
- By using newly transcribed tax censuses, probate inventories, and slave valuation records, I illustrate how settlers, slaves, and Khoe experienced starkly unequal income and wealth distribution even within their own groups.
- These findings engage with global inequality debates, demonstrating that extreme inequality is not solely a modern phenomenon but has historical foundations.

Keywords: inequality, Gini, wealth, slavery, economic prosperity

---

Introduction

The World Bank's Gini index, a measure of inequality between 0 and 1, ranks South Africa as the most unequal country on earth. Although its most recent observation for South Africa is now a decade old, it calculates a Gini index for income of 0.63, four points above second-placed Namibia.[1] South Africa is even more unequal when considering the distribution of wealth; Von Fintel and Orthofer calculate a wealth Gini between 0.93 and 0.95, noting "wealth is as unequally distributed within South Africa as it is in the world at large".[2]

While much of the research attention, for obvious reasons, focuses on post-apartheid trends in inequalities and understanding their determinants, there is little doubt that South Africa's history of racial discrimination and repression explains these persistently high levels of inequality. What is less clear, however, is when this high level of inequality emerged.

To answer this question, I investigate inequality at the beginning of South Africa's colonial history: the Dutch Cape Colony of the seventeenth and eighteenth centuries and the British Cape Colony of the early nineteenth century. I use several newly transcribed datasets that include information about wealth and income for three population groups, the settler population mostly though not exclusively of European origin, the enslaved population mostly though not exclusively from territories around the Indian Ocean, and the population of indigenous pastoral and hunter-gatherer peoples, the Khoesan.

The Cape Colony, I document, was a severely unequal place. As can be expected, inequality between the three population groups was high; settlers were wealthier than the enslaved and Khoesan. What is surprising, though, is the large inequalities within these groups. The within-group Gini for settlers is as large as the overall Gini; in other words, were the enslaved and Khoesan removed from the population, the Gini coefficient for the Cape would remain extremely high. The same is true of the Khoesan population; Khoesan villages and mission stations exhibited equally high levels of wealth inequality. Severe inequality during precolonial and early colonial eras was the norm.

Measuring Inequality

The study of inequality over the long run has attracted much attention recently. One reason for that is the concern with high levels of inequality today. In his famous magnum opus, Piketty suggests that contemporary inequality follows a trend of rising inequality projecting far back in time.[3] But not all economic historians agree that inequality has worsened.[4] Much of the more recent work on pre-industrial inequality has been built on Milanovic, Lindert, and Williamson.[5] The results suggest that most pre-industrial Western societies were more unequal than those societies today, with mid-nineteenth century Gini coefficients ranging between 0.44 and 0.49 in the United States and rural Palencia in Spain, to 0.52–0.53 in England and Wales, and reaching as high as 0.63 in Holland, for instance.[4]

In new work, Waldenström argues that these pre-industrial estimates of mid-nineteenth-century inequality in European countries may be anomalies, that inequality may have worsened until the early twentieth century, and then declined:

> [W]ealth inequality may not always have been rampant before 1900 but, rather, that the extreme level around that time is a snapshot of inequality when it was approximately at its highest level. Although it is fair to say that, on account of the lack of reliable sources, consensus is not reached regarding pre-industrial wealth inequality…[6]

This idea of an inverted U-relationship echoes earlier work of Van Zanden who, in a seminal work, points to rising capital incomes in premodern Europe, making the case for a "super-Kuznets curve".[7]

Van Zanden was interested not only in documenting inequality trends but also in explaining it. Others have followed. Alfani has emphasised non-economic factors like war, natural disasters and epidemics which reduce population figures, altering the distribution pattern and creating Kuznets-like waves of rising and falling inequality over time.[4,8] Recent work has also approached the question of political and institutional factors. Piketty, for instance, argues that when the rate of return (*r*) exceeds the rate of economic growth (*g*) and wealth remains highly inheritable, inequality increases over time.[3] Working on a global scale, Derenoncourt explores inequality as an outcome of deep institutional-historical differences and factor endowments.[9] Likewise, Milanovic has explored the extent to which a country's colonial status, urbanisation and population density influence its extraction ratios and thus the degree to which elites are able to drive up inequality.[8]

Yet, as Waldenström acknowledges in the above quotation, much of what we know about long-run inequality trends is limited to the nineteenth and twentieth centuries.[6] This is because the sources economists typically use to study such phenomena – national censuses – are mostly a creation of the late eighteenth century; the first US census, for example, was administered in 1790 and the first census of England and Wales in 1801. Economic historians, though, have found ways around these data limitations. One solution is to measure inequality using social tables.

By leveraging group-level data, social tables offer a means for examining long-run inequality in low-resource, "data-poor"[5] settings: i.e., the past. Here, data from transcribed archival records are used to assign aggregate income or wealth values to broadly representative socio-economic groups. In this way, between-group differences are assumed to be a reliable indicator of overall inequality. As Von Fintel, Links and Green note, the use of social tables has provided "new insights on the connections between long-term economic development, institutions, globalization, and inequality."[5]

Because of the method's reliance on data-grouping at the socio-economic-level, however, there is a concern that social tables hide within-group inequality.[8] This can be avoided by including more socio-economic classes. Von Fintel, Links and Green have recently found that how researchers construct social tables is crucial to the credibility of their inequality estimates. A larger "bottom class", for example, can create a downward bias on inequality estimates, as in the case of large subsistence societies.[5]

Another way to address the limitations of social tables is to rely on household-level information. While rare for most pre-industrial societies, the late seventeenth- to early nineteenth-century Dutch and British Cape Colony offers uniquely detailed archival records, enabling an unusually precise investigation of preindustrial inequality. First, though, I offer a brief literature overview of the existing scholarship on the roots of South Africa's high inequality.

Inequality at the Cape

South Africa's high levels of inequality have prompted scholars to seek out its origins. The emphasis has been on the twentieth century, as in Nattrass and Seekings[10], van der Berg[11], and Alvaredo and Atkinson[12]. Apartheid entrenched racially structured income and wealth inequalities through political, social and labour-market mechanisms. All these scholars have, however, recognised that South Africa's twentieth-century inequality has roots stretching back to colonial settlement[2,10], writing that, "a poor pre-colonial society, colonial settlement and then the mineral discoveries laid the basis for a highly dualistic economy that was from the outset highly inegalitarian."[11]

Much of the focus has been on between-group inequality. In the late twentieth century, historians' debates on apartheid, capitalism, and racial inequality in South Africa focused almost exclusively on the disparities between European settlers, enslaved people and native African communities. "Liberal" scholars argued that inequality originated from early European conquest, emphasising dispossession and poverty as consequences of this initial contact.[13] They believed that economic growth would gradually undermine the racial structures of apartheid. In contrast, "radical" scholars used a Marxist lens, viewing colonial contact as a protracted struggle over land, labour and markets, in which capitalist interests were deeply entwined with segregation and apartheid.[14] But, as Nattrass and Seekings note, little was done to study inequality itself by either "liberal" or "radical" scholars.[10]

Economists and economic historians like Wilson[15], Wilson and Ramphele[16], and Feinstein[17] also focused on racial disparities shaped by conquest and control of resources. Settlers established patterns of accumulation based on control of land and water, as Wilson emphasises, and Feinstein highlights how conquest intensified labour scarcity on the frontier, constraining Khoe and African options to wage labour on white farms.[15,17] Robertson and Terreblanche further noted enduring economic ties between Europeans and Africans on the frontier, where Khoe responses to settler encroachment ranged from resistance to cooperation, including shared farming.[18, 20] By the mid-1700s, indentured labour bound Khoe and African workers to settlers through debt, while Bundy observed that slavery and the dispossession of Khoe pastoralists entrenched racially-coded relations of coercion, creating a "population of propertyless labourers".[13] By the mid-1800s, Africans beyond the eastern frontier joined a cash economy, leading to a stratified peasantry, while capitalist agriculture and railway expansion widened disparities among white settlers.[21] Segregation and restrictive land policies further marginalised high-density areas, solidifying poverty under apartheid institutions. [22,23] The implication is that inequality intensified over time as marginalised groups faced mounting economic constraints.

With regards to within-group inequality, Bundy[13], Terreblanche[18], and Feinstein[17] have demonstrated significant stratification in the urban settler society, with 7% of settlers holding over half of the landed

wealth by 1731.[13] This they attribute to mercantilism and slavery. Since VOC mercantilism favoured large-scale producers, landless Europeans were limited to wage labour (for the VOC or landed gentry) or small-scale, subsistence agriculture on the colony's periphery.[18] Using micro-data, Fourie and Von Fintel confirm this finding. They demonstrate significant settler inequality between a rising elite and a progressively impoverished farmer population; a distribution which they find accounts for much of the overall distribution. Their work, though, is limited to a single district and few years.[19]

As Chancel et al. note, "Settler colonialism seems to cast its long shadow on Southern Africa even after the demise of apartheid".[2] While much research has emphasised the high levels of inequality between racial groups, particularly in the twentieth century, relatively little attention has been paid to accurate measures of inequality in the long run and, in particular, attempts to estimate inequality within groups over time. By focusing on the Cape Colony from the seventeenth to the nineteenth centuries and analysing newly transcribed household-level data, this study fills that gap.

New Sources

The Cape Colony has exceptionally rich historical records. Two reasons explain this information abundance: the Cape was governed, first, by a profit-maximising company, the Dutch East India Company, which prioritised recordkeeping to maximise profit and minimise corruption, and, second, by the British, for whom political control meant an accurate assessment of its subjects. The second reason for the information abundance is that Cape Town, where these records have been preserved for more than three centuries, has never had a major war or natural disaster. It is this rich supply of historical records that allows scholars to go beyond simple social table analyses. It is them I use to calculate inequality measures for the Cape over almost two centuries. My analysis begins in 1685, three decades after Jan van Riebeeck and his crew of motley Company workers arrived in Table Bay, and ends in 1844, six years after the enslaved were allowed to leave the farms of the slaveowners. The choice of years depended almost entirely on the availability and completeness of the source material.

I use several new historical sources to calculate and compare wealth inequality at the Cape. The first and largest is the *opgaafrolle* or tax censuses of the Colony. Over the last decade, a team of historians have been meticulously transcribing these records in the Cape Archives in Cape Town's Roeland street archives, or, when possible, from the online collection of the Dutch National Archives in The Hague.[24] The annual censuses of settler households are remarkably rich in their content. (Although "settlers" here refer predominantly to those of European descent, the tax censuses include all *free burghers*, including those of mixed-race residents, freed slaves, and, less frequently, independent Khoesan farmers and mission station residents.) They include the names of the male household head and wife, the number of settler sons and daughters, the number of enslaved workers (by gender and age), the number of Khoesan workers (by gender and age), and then various

agricultural assets, including a count of horses, cattle, sheep, pigs and vines, and quantities of wheat, barley, rye and oats reaped and produced, and wine (and later brandy) produced. Towards the end of the period, the tax censuses also included information about wagon ownership and taxes paid. The comprehensive nature of the records, and the richness of the information for measures of wealth, make this the most reliable source for the purposes of this analysis.[25]

Another source I use is the probate inventories of Cape settlers. The assets of deceased settlers offer another opportunity to measure inequality. Probate inventories were collected by the Master of the Orphan Chamber (MOOC) in Cape Town, an institution responsible for administering estates when individuals died intestate, left heirs under 25 or unmarried or had heirs who were untraceable. In some cases, settlers specifically requested that the Orphan Chamber act as the executor of their estates, and if no such exclusions were made in their wills, the Chamber took responsibility. The series documents the material wealth left behind by deceased settlers, including household goods, livestock, slaves, and other movable assets, allowing researchers to paint a detailed picture of the economic life in the Colony. The MOOC 8-series was first analysed by Fourie a decade ago[26]; for this analysis, however, I use an updated version that includes the full series of probates into the nineteenth century.

Third, I use a recently constructed Slave Emancipation Dataset. The dataset provides detailed insight into slave ownership and wealth distribution in the Cape Colony at the time of abolition in 1834.[27] Compiled from slave valuation rolls, slave registers, and compensation data, it allows for the analysis of inequality among slaveholders.

The foundation of the dataset is the slave valuation rolls, compiled by appraisers appointed by London in 1834. These officials travelled across the Cape Colony's 11 districts, documenting slave owners, the number of slaves they held, and various characteristics of the enslaved, such as their age, gender, and occupation. This dataset provides a near-complete census of slave ownership on the eve of emancipation, but it is crucial to note that it only includes those who still owned slaves in 1834. Valuations were based on historical slave sale averages from 1823 to 1830, adjusted for district conditions, though the Clanwilliam and Swellendam districts lack valuation data.

Following the appraisals, slave owners were compensated by the British government, but the compensation process was controversial. Rather than receiving payment based on the appraised value of each individual slave, slaveholders were compensated according to an average value per slave class, calculated across the entire colony. This method, meant to allocate the limited compensation funds as evenly as possible, often resulted in significant discrepancies between the appraised value of slaves and the actual compensation received. Slave owners, in many cases,

received far less than the estimated value of their slaves, fuelling dissatisfaction, especially among wealthier slaveholders.

Each dataset introduces specific biases due to the limitations inherent in historical records. The annual tax censuses may contain errors from repeated handling and the poor condition of original documents; to reduce this impact, we trimmed outlier values, although this was unnecessary for the smaller probate and slave emancipation datasets. The MOOC 8-series probate inventories also face geographic and demographic bias, as they mostly reflect wealthier, younger estates near Cape Town, likely omitting poorer, remote frontier farmers and older households. Coverage of the settler population declines after the 1740s, potentially increasing sampling bias, yet the series remains invaluable for studying Cape wealth and material culture. Lastly, the Slave Emancipation Dataset undercounts slave ownership by only including owners from 1834, omits data for certain districts, and relies on historical valuation averages rather than current appraisals, which may understate slave values.

Finally, I measure inequality based on agricultural assets recorded in these sources, encompassing elements of both income and wealth inequality. While some assets, like wheat reaped, reflect flow variables akin to income, most assets – such as the numbers of horses, cattle, and sheep – represent stock variables, a form of wealth. However, these are not pure wealth estimates: I do not capture fixed household assets, such as land or property, which typically drive wealth inequality, nor financial assets, except for the compensation settlers received at emancipation. Thus, these Gini estimates should lie between standard income and wealth inequality measures.

Results

These newly transcribed sources provide a comprehensive basis for analysing the historical roots of inequality in the Cape Colony. I explore the patterns of wealth inequality revealed by these datasets, examining how disparities evolved over time and across different groups within the colony.

The section is divided into four main parts. First, I analyse the data from the tax censuses, focusing on patterns of inequality over time and across space. Second, I examine the settler probate inventories, which provide insights into wealth distribution based on the assets of deceased individuals. Third, I present findings from the slave emancipation dataset, highlighting the valuations of enslaved individuals and the compensation received by slave owners following emancipation. Finally, I return to the tax censuses to study inequality within Khoe settlements.

Figure 1: Gini coefficients for six asset types in Stellenbosch-Drakenstein, 1685-1844

Tax censuses

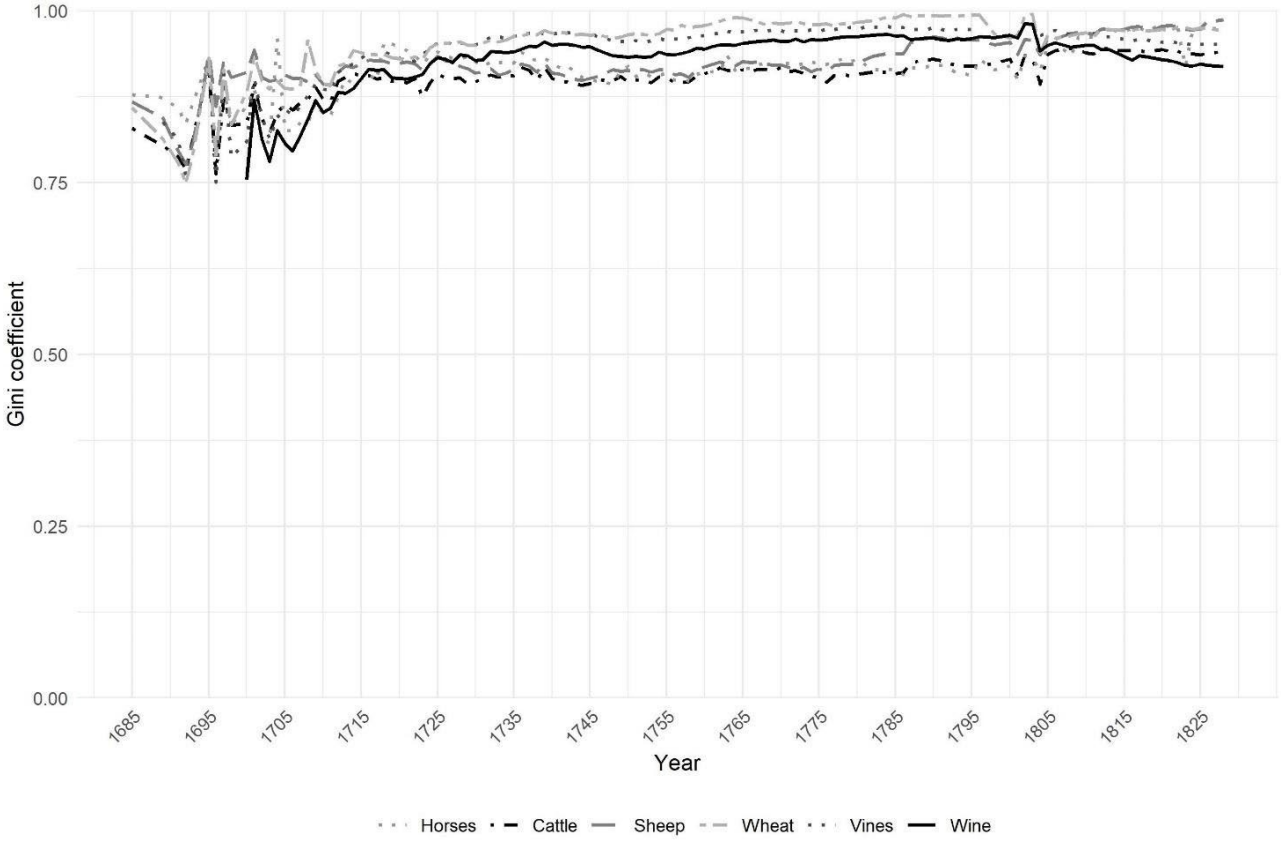
I first plot, in Figure 1, the annual Gini coefficients for six asset types – horses, cattle, sheep, wheat reaped, vines and wine – in the district of Stellenbosch-Drakenstein, one of the oldest and most populous districts within the Colony. A total of 142,219 observations (at the household level) were recorded across 159 censuses conducted between 1685 and 1844. Note that these are not necessarily unique households; the same household could be recorded over multiple years.

Three trends are immediately apparent: First, inequality *within the free population* is remarkably high; not a single asset has a measured Gini of less than 0.5 in any year between 1685 and 1844. Second, though there is some variation over time – the Gini for wine, for example, increases during the eighteenth century and then begins to decline at the turn of the nineteenth century – Gini coefficients are largely persistent. Third, there are moments of volatility, as during the first three decades of the sample, in the period 1795-1805 and again after 1828. These reflect changes to the data generating process – political and institutional change – and should be treated with caution.

I next combine these six assets into a basket and calculate the Gini of the combined assets. To do this, I use prices obtained from auction records. It confirms the above results: a high and persistent Gini coefficient of between 0.6 and 0.75 for the entire eighteenth century, increasing further into the nineteenth century. The settler society of the Stellenbosch-Drakenstein district was clearly highly unequal.

Cape settlers relied heavily on slave labour. Ownership of enslaved people are recorded in the tax censuses too. Although slave ownership declined somewhat during the eighteenth century, it remained more unequally distributed than other asset types throughout most of the period, with a Gini coefficient for slave men above 0.73 across the entire period.

Figure 2: Gini coefficient of Stellenbosch-Drakenstein district, now including enslaved men as separate households, 1685-1844



Slaves were, of course, not only assets in settler households[28,29]; they were part of colonial society. To calculate the inequality within the Cape Colony, as a result, slave men (because we calculate inequality at the household level) should be added as separate households. I do so in Figure 2.

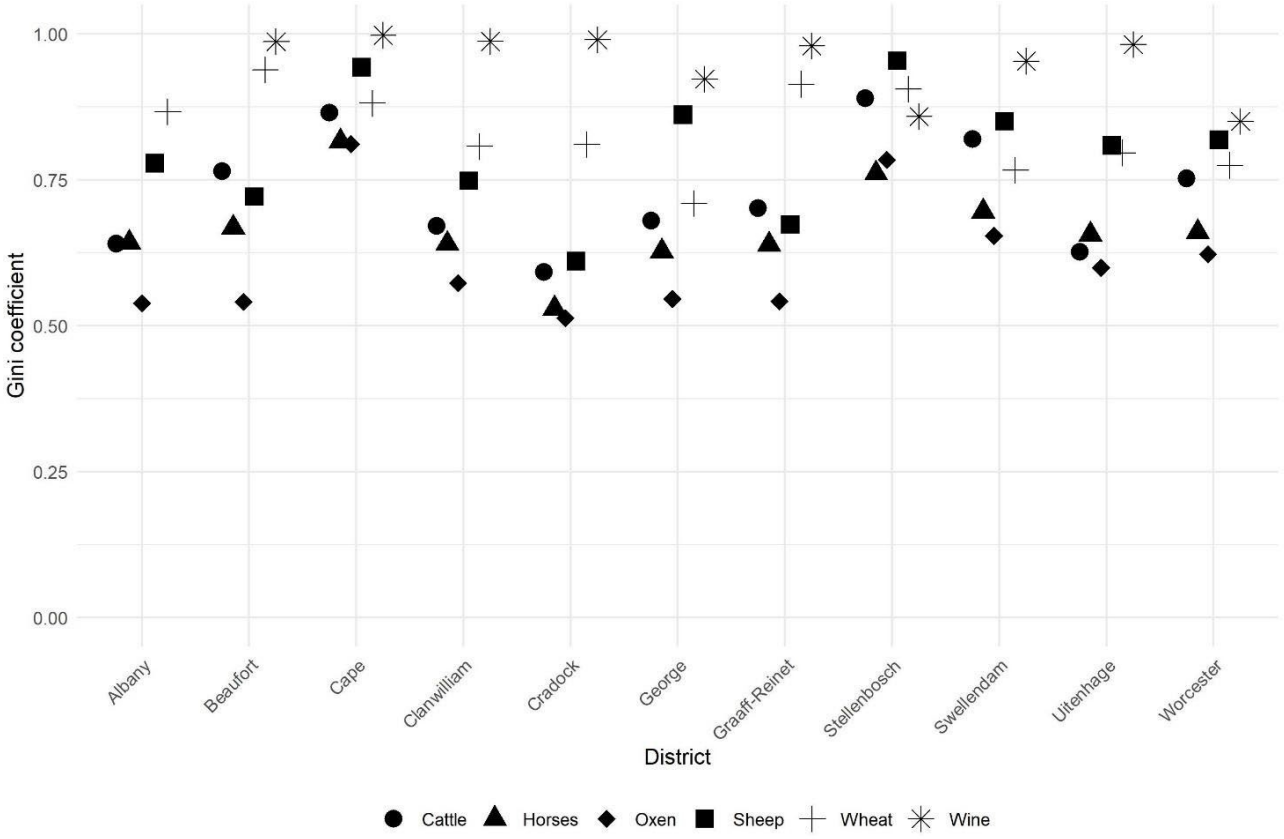
Here I have to make a strong assumption. I assume that slaves owned no assets. Although this is not historically accurate – enslaved people did possess assets and could, and did, save up to manumit themselves or their children[30] – I argue that such savings were rare and limited mostly to the nineteenth century. During the 1820s, laws were introduced that provided greater protection of enslaved people's rights, including protection of their property.[31]

Figure 2 reveals a picture of an extremely unequal society, with Gini coefficients very close to perfect inequality. This is not unheard of for slave societies in the Caribbean[32], but it is surprising given the crop types (wheat, wine and cattle rather than cotton, sugar or tobacco), relative equality of land

ownership and a system of partible inheritance that supposedly ensured a more equitable distribution of wealth across family members compared to primogeniture.

Consider, as a thought experiment, that enslaved people were indeed “compensated”, even if only in in-kind payments of food. Evidence from the early nineteenth century suggests that Khoe labourers received, on average, a sheep a month, or 12 sheep per year.[33] Were we to include enslaved men as household heads, the number of observations expand to 428,780 over the 159 censuses. If we assume enslaved men households earned 12 sheep, then the Gini coefficient fits very closely onto the one for settlers only, ranging between 0.75 and 0.8.

Figure 3: Dot chart of Gini coefficients for eleven districts, by asset type, 1825



These results are for the entire eighteenth century but focus only on the district of Stellenbosch-Drakenstein. Perhaps this region was uniquely unequal. To test whether this is true, I turn to a cross-section of all districts for which tax censuses are available in or around 1825. I use 1825 or the nearest available year for eleven districts. 1825: Cape district, Stellenbosch, Graaff-Reinet, Swellendam, Albany, Beaufort, George, Uitenhage. 1824: Clanwilliam, Worcester. 1823: Cradock. I calculate the Gini coefficient for six agricultural outputs: horses, oxen, cattle, sheep, wheat, and wine. The full dataset includes 10,420 observations. Note that these are unique households, as we only a single year. The results are presented in Figure 3.

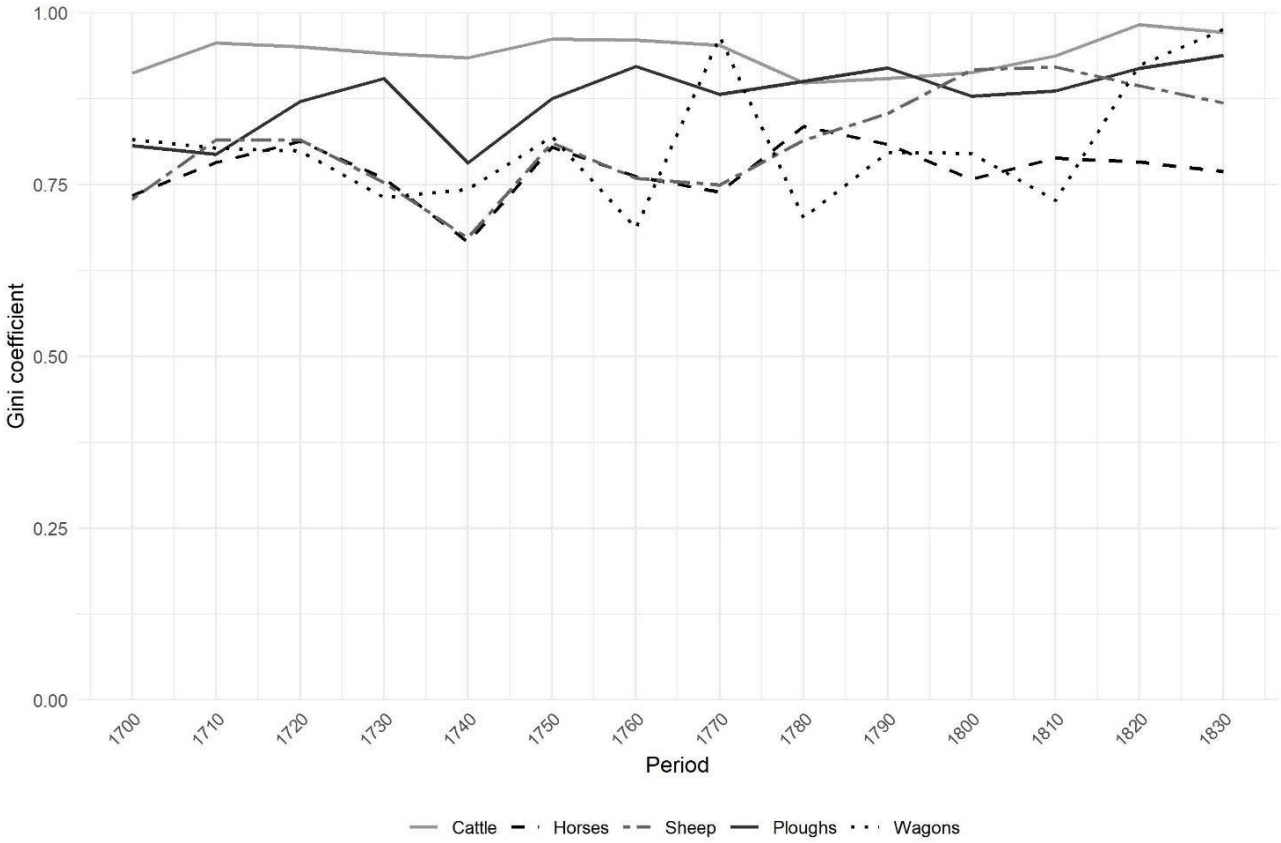
Apart from one case – Albany has no wine production in 1825 – the Gini coefficients are above 0.5 for all output variables in all districts. Some output types, though, are more unequal than others. Wine, for example, is far more unequally distributed than any other crop: all districts record Gini coefficients above 0.8. By contrast, horses and oxen are more equally distributed. This suggests two reasons why Stellenbosch should be seen as the upper end of the inequality distribution: it had high wine output and it had a large concentration of enslaved people.

Probate inventories

I next analyse 4423 probates in total to calculate inequality. Despite concerns about sample selection, there are several benefits from using probates. First, they document a range of household assets, though I do not exploit this feature here. Instead, I only report five asset types to allow for easier comparison with the tax censuses. The levels of inequality for the other assets, however, largely mirror those reported here. Second, probates cover the entire period of analysis and all districts with little change in administrative procedure and reporting. The quality of the recordkeeping and transcription also allows me to include all observations and not trim the top and bottom. Third, they avoid issues of tax avoidance, a serious concern in the tax censuses (and another reason to suspect that inequality is underestimated).

Figure 4 reports the Gini coefficients for cattle, horses, sheep, ploughs and wagons over the eighteenth and early nineteenth centuries. All five Ginis closely mirror those calculated from the tax records. Both measures indicate severe and persistent inequality, with Gini coefficients for all assets exceeding 0.6, and from 1770 onward, surpassing 0.7. There is little doubt that these results reflect a deeply unequal distribution of settler wealth from very early in the colonial period.

Figure 4: Gini coefficients of assets at death, 1700-1830

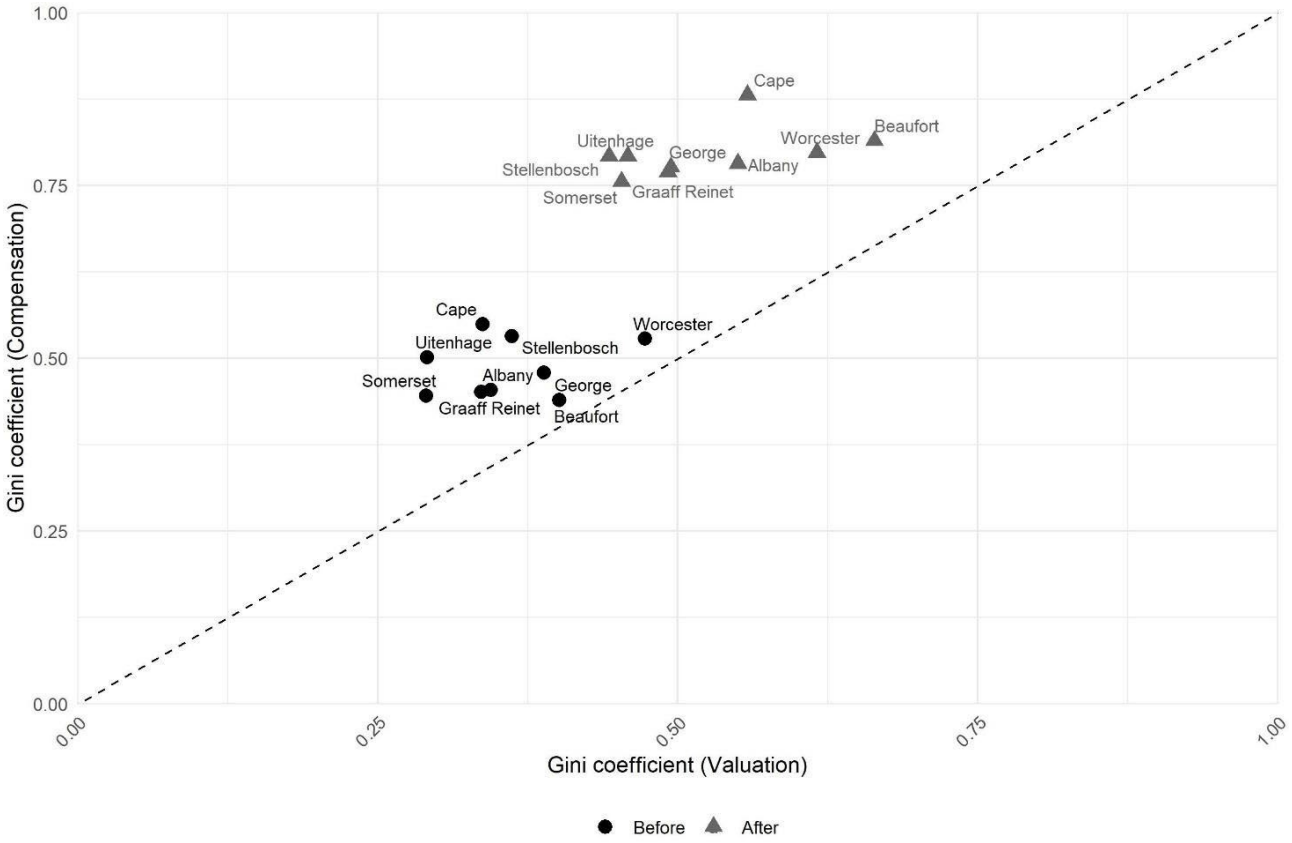


Slave emancipation

The Slave Emancipation Dataset reports the valuation of each enslaved individual as well as the compensation each owner received. The total number of individual-level observations is 36,417. I calculate the Gini coefficient of both. Figure 5 reports the scatter plot of the two, by district (the black dots). What is clear is that slave ownership was unequal in 1834, though not as extreme as the other records analysed: the valuation Gini ranges between a low 0.29 for Somerset and Uitenhage to a relatively high 0.47 for Worcester. Compensation Gini's are more unequal: some slave owners clearly received compensation equivalent to the value of their enslaved while others did not, as also highlighted in the literature. These Gini's range between 0.45 in Somerset and 0.55 in the Cape district.

Although these Ginis suggest that inequality in slave ownership existed – and, therefore, wealth inequality, as enslaved people were considered the most valuable asset at the time – these results hide an important truth: it only includes slave owners, less than 30% of the settler population at the time. To calculate the actual settler inequality, I add 13033 non-slave owning settlers, and reestimate both Ginis. The results are also reported in Figure 5 (grey dots).

Figure 5: Scatter plot of Gini coefficients for valuation of the enslaved and compensation received by slave owners, by district, 1834



The results now align much more closely with the inequality estimates as calculated in the tax censuses and probate inventories. Ginis for the valuation dataset range between 0.49 (Stellenbosch, surprisingly the least unequal district) and 0.75 in Beaufort, though that is largely a consequence of very few slave-owning families in that district. When considering the compensation received, Ginis estimates are extreme, ranging between 0.83 in Somerset to 0.92 in the Cape district. Again, wealth was highly unequally distributed.

Khoe settlements

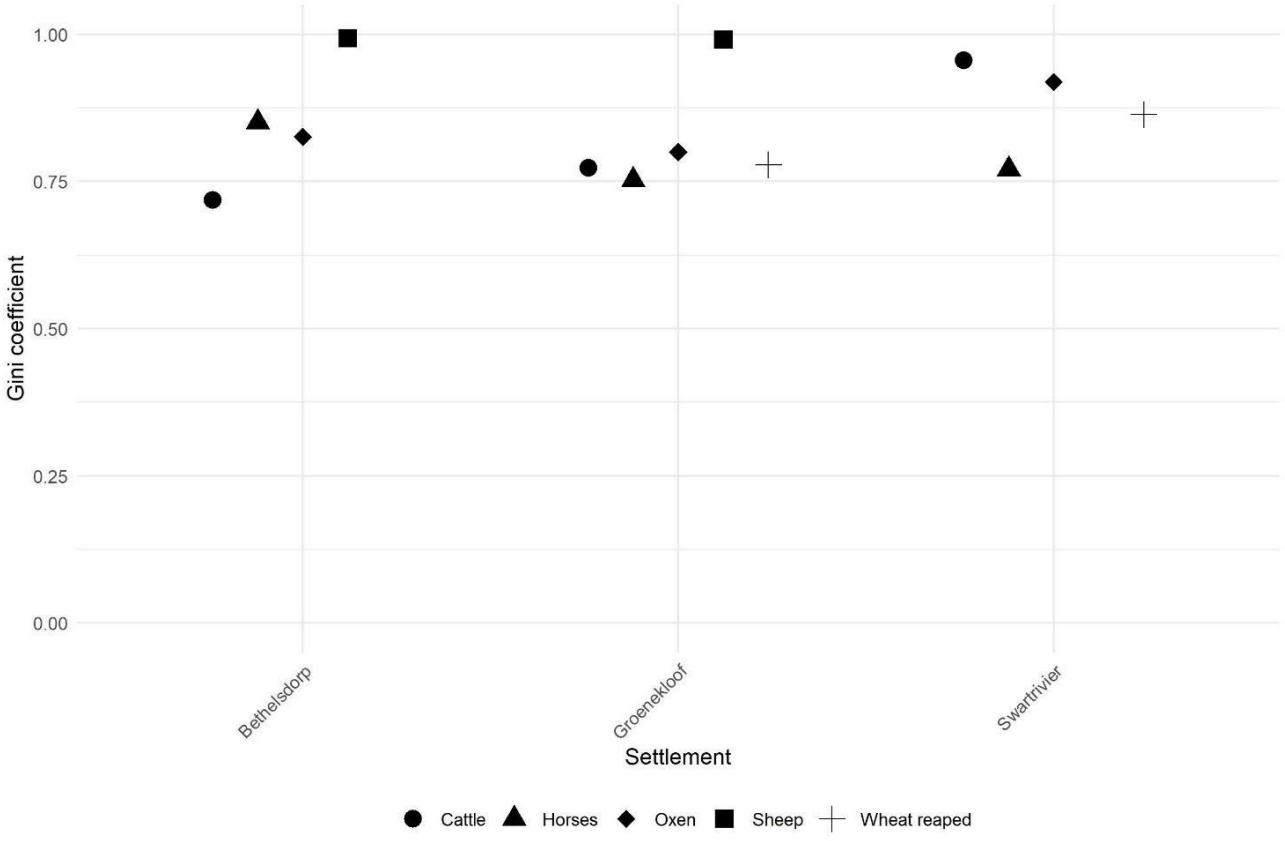
The above results reflect income and wealth inequalities of settler households. Where I have incorporated enslaved people, they appear as zeros. But colonial recordkeeping was not limited to the settler population. By the early nineteenth century, several Khoe settlements were also incorporated into the colonial administrative system, notably mission stations. These records are now being transcribed and analysed.

I consider three of these settlements in 1825: Bethelsdorp, Swartrivier, and Groenekloof. Bethelsdorp, founded in 1803 by the London Missionary Society near Algoa Bay, became an important settlement for the Khoe, offering some protection and autonomy within the colonial

system.[34] By 1814, Bethelsdorp had become a refuge for the 425 Khoe households recorded in the tax census, although it struggled economically due to poor soil and limited resources. Swartrivier, in the Swellendam district, developed as the Khoe were being drawn into the colonial economy through labour contracts and informal patronage.[35] Primarily pastoralists, many Khoe worked for European farmers as herders or labourers, often in exploitative conditions. By 1825, Swartrivier served as a base for 364 Khoe households trying to maintain some independence. Groenekloof (later Mamre), established in the early nineteenth century by the Moravian Missionary Society, became a haven for freed slaves and Khoesan.[36] By 1825, it was a thriving community where 108 households practised subsistence farming and casual labour. Though it offered sanctuary, the settlement's population remained largely reliant on nearby farms for employment.

The tax records for these three settlements are similar in style and content to those collected for settler households, though they typically record fewer variables (on account, presumably, of Khoe households owning fewer assets). For my purposes here, though, I consider five asset types: cattle, horses, oxen, sheep and wheat reaped. I calculate Gini coefficients for each asset type across the three districts and report the results in Figure 6.

Figure 6: Gini coefficients for five asset types in three Khoe settlements, 1814-1825



The Gini coefficients reveal severe inequality within all three Khoe settlements. Aside from cattle in Bethelsdorp with a Gini of 0.72, all other assets have Gini coefficients above 0.75, with some very

close to perfect inequality. For example, only ten families own sheep in Bethelsdorp and only one family owns sheep in Groenekloof. In many ways, the inequality within these Khoe settlements is even more pronounced than that recorded among settlers.

These results provide a clear picture of severe and persistent inequality across the various population groups and time periods in the Cape Colony. Whether examining settler households, slave ownership or Khoe settlements, the data consistently reveals high levels of wealth concentration. Inequality, the household-level data reveal, was a defining feature of the eighteenth and nineteenth century, pre-industrial Cape society.

Conclusions

In their analysis of top incomes in South Africa during the twentieth century, Alvaredo and Atkins conclude that "[t]he share of the top 1% was halved between 1914 and 1993"[12]. The implication: inequality at the start of the twentieth century, long before the apartheid policies were adopted, was much more severe than at its end. This warrants an important but unanswered question: when did these high levels of inequality emerge?

To answer this, I turn to newly transcribed household-level records from the seventeenth, eighteenth, and nineteenth-century Cape Colony. The newly available tax censuses, probate inventories, and slave valuation rolls allow for the first comprehensive analysis of inequality across different groups in the Cape Colony. I find that income and wealth concentration among settlers and within settler society, but also within those Khoe settlements for which suitable data is available, was extremely high, with Gini coefficients exceeding modern levels.

This study not only adds to South Africa's economic history but also engages with broader debates on inequality, showing that extreme inequality is not solely a product of modern capitalism but has deep historical origins. By using detailed data that surpasses the constraints of social tables, it reveals that extreme inequality was the norm in the pre-industrial, colonial Cape Colony, aligning with similar, though more tentative, estimates elsewhere. The most unequal country in the world today may have been even more unequal three centuries ago.

References

1. World Bank. Gini index [Internet]. World Bank Development Indicators; 2024. Available from: https://data.worldbank.org/indicator/SI.POV.GINI
2. Von Fintel D, Orthofer A. Wealth inequality and financial inclusion: Evidence from South African tax and survey records. Economic Modelling. 2020 Sep 1;91:568-78. https://doi.org/10.1016/j.econmod.2020.02.001
3. Piketty T. Capital in the twenty-first century. Cambridge, MA: Belknap Press; 2014. https://doi.org/10.4159/9780674369542
4. Alfani G. Economic inequality in Preindustrial times: Europe and beyond. J Econ Lit. 2021;59(1): 3–44. https://doi.org/10.1257/jel.20191449
5. Von Fintel D, Links C, Green E. Estimating historical inequality from social tables: Towards methodological consistency. Working paper 2023:46. http://dx.doi.org/10.2139/ssrn.4450261
6. Waldenström, D. Richer & more equal: A new history of wealth in the West. Cambridge, MA: Polity Press; 2024.
7. Van Zanden JL. Tracing the beginning of the Kuznets curve: Western Europe during the Early Modern Period. Econ Hist Rev. 1995;48(4):643–64. https://doi.org/10.2307/2598128
8. Milanovic B. Towards an explanation of inequality in premodern societies: The role of colonies, urbanization, and high population density. Econ Hist Rev. 2018;71(4):1029–47. http://doi.org/10.1111/ehr.12613
9. Derenoncourt E. The historical origins of global inequality. In: Boushey H, Bradford DeLong J, Steinbaum M, editors. After Piketty: The agenda for economics and inequality. Cambridge, MA: Harvard University Press; 2017. p. 491–511. https://doi.org/10.4159/9780674978195-021
10. Seekings J, Nattras N. Class, race and inequality in South Africa. New Haven: Yale University Press; 2005.
11. Van der Berg S. Current poverty and income distribution in the context of South African history. Econ Hist Dev Reg. 2011;26(1):120–40. https://doi.org/10.1080/20780389.2011.583018
12. Alvaredo F, Atkinson AB. Top incomes in South Africa in the twentieth century. Cliometrica. 2022;16(3): 477–546. https://doi.org/10.1007/s11698-021-00235-4
13. Bundy C. Development and inequality in historical perspective. In: Schrire R.A, editor. Wealth or poverty?: Critical choices for South Africa. Cape Town: Oxford University Press; 1992. p. 1–17.
14. Murray MJ. The triumph of Marxist approaches in South African social and labour history. In: Gutkind PCW, editor. Third world workers: Comparative international labour studies. Leiden: E.J. Brill;1988. p. 79–101.
15. Wilson F. Historical roots of inequality in South Africa. Econ Hist Dev Reg. 2011;26(1):1–15. https://doi.org/10.1080/20780389.2011.583026
16. Wilson F, Ramphele M. Uprooting poverty: the South African challenge. 2nd edition. Carnegie inquiry into poverty and development in Southern Africa: Cape Town; 1989.
17. Feinstein CH. An economic history of South Africa: Conquest, discrimination and development. Cambridge, UK: Cambridge University Press; 2005.
18. Terreblanche S. A history of inequality in South Africa, 1652–2002. Pietermaritzburg: University of Natal Press; 2002.
19. Fourie J, Von Fintel D. A history with evidence: Income inequality in the Dutch Cape Colony. Econ Hist Dev Reg. 2011;26(1):16–48. https://doi.org/10.1080/20780389.2011.582990
20. Robertson HM. 150 years of economic contact between black and white. S Afr J Econ. 1934;2(4):403–25. https://doi.org/10.1111/j.1813-6982.1935.tb01957.x
21. Bundy C. The rise and fall of the South African peasantry. Berkeley: University of California Press; 1979.
22. Von Fintel DP. Long-run spatial inequality in South Africa: early settlement patterns and separate development. Stud Econ Econom. 2018;42(2):81–102. https://doi.org/10.1080/10800379.2018.12097335
23. Von Fintel D, Fourie J. The great divergence in South Africa: Population and wealth dynamics over two centuries. J Comp Econ. 2019;47(4):759–773. https://doi.org/10.1016/j.jce.2019.08.005
24. Fourie J, Green E. Building the Cape of Good Hope Panel. Hist Fam. 2018;23(3):493–502. https://doi.org/10.1080/1081602X.2018.1509367
25. Rijpma A, Cilliers J, Fourie J. Record linkage in the Cape of Good Hope panel. Hist Methods. 2020;53(2):112–129. https://doi.org/10.1080/01615440.2018.1517030
26. Fourie J. The remarkable wealth of the Dutch Cape Colony: measurements from eighteenth-century probate inventories. Econ Hist Rev. 2013;66(2):419–448. https://doi.org/10.1111/j.1468-0289.2012.00662.x
27. Ekama K, Fourie J, Heese H, Martin LC. When Cape slavery ended: Introducing a new slave emancipation dataset. Explor Econ Hist. 2021;81:101390. https://doi.org/10.1016/j.eeh.2021.101390

28. Fourie J, Swanepoel C. ‘Impending ruin’ or ‘remarkable wealth’? The role of private credit markets in the 18th-century Cape Colony. J South Afr Stud. 2018;44(1):7–25. https://doi.org/10.1080/03057070.2018.1403218
29. Ekama K. Bondsmen: Slave collateral in the 19th-century Cape Colony. J South Afr Stud. 2021;47(3):437–453. https://doi.org/10.1080/03057070.2021.1900467
30. Ekama K. Bound to be free? Manumission in Cape Town, 1825–34. Slavery Abol. 2024;1–22. https://doi.org/10.1080/0144039X.2024.2389549
31. Mason JE. Social death and resurrection: slavery and emancipation in South Africa. Charlottesville: University of Virginia Press; 2003.
32. Rönnbäck K, Galli S, Theodoridis D, Faust Larsen K. The persistence of wealth: Economic inequality in a Caribbean slave colony in the very long run. (In press 2024).
33. Links C, Fourie J, Green E. The substitutability of slaves: evidence from the eastern frontier of the Cape colony. Econ Hist Dev Reg. 2020;35(2):98–122. https://doi.org/10.1080/20780389.2019.1669444
34. Elbourne E. Early Khoisan uses of mission Christianity. Kronos J Cape Hist. 1992;19(1):3-27.
35. Links C. Khoe households in Swellendam, 1825. In: Fourie J, editor. Quantitative history and uncharted people: Case studies from the South African past. London: Bloomsbury Academic; 2023. p. 91–116. https://doi.org/10.5040/9781350335608.ch-4
36. Ludlow H. Missions and emancipation in the South Western Cape: A case study of Groenekloof (Mamre), 1838-1852 [master’s dissertation]. Cape Town: University of Cape Town; 1992.

Supplementary material

Data management and augmentation

The data for this study was sourced from publicly available archival records, all of which are over 100 years old. This age ensures that the data does not fall under POPIA compliance, as it pertains strictly to historical subjects rather than identifiable individuals within modern privacy frameworks. Analyses using most of these datasets have already been published or are under review. However, the data cannot be made publicly available at this time, as it remains central to ongoing research projects. This approach allows for its continued use in current and future studies, ensuring a thorough exploration of its historical insights.

The R code used for this analysis, titled InequalityScript.R, accompanies this submission and will be made publicly available on the project's GitHub repository upon acceptance of the manuscript. This ensures transparency in the analytical methods applied to the data. In future, the complete dataset will also be uploaded to a data repository, facilitating broader access and reproducibility for other researchers interested in historical economic inequality.

Below is a description of the process undertaken for each part of the analysis.

Tax censuses

The original data is available in the Cape Archives, J-series. The transcribed data was sourced from the stellenbosch_clean_july2024.xlsx file, containing detailed records of settlers' and slaves' assets, livestock, agricultural production, and household members. To ensure uniformity and accuracy across all columns, every column was initially read as text. Numeric data were then cleaned by removing non-numeric characters and converting the relevant columns to numeric format.

Several transformations were applied to the dataset to handle missing values. For instance, any missing values in key columns such as livestock or crops were replaced with zeros where appropriate. In cases where multiple related columns existed, such as those representing different categories of cattle or sheep, missing values in the main category (e.g., total cattle) were replaced by summing the related subcategories.

For the analysis of inequality, additional settlers with zero assets were introduced to the dataset to account for individuals who owned no livestock or crops. This adjustment allowed for a more comprehensive representation of the settler population and their economic status.

Outliers in key variables, including livestock numbers and crop production, were trimmed by removing the top and bottom 1% of values within each decade. This helped to focus the analysis on typical patterns of asset ownership and agricultural output, avoiding the distortion caused by extreme values.

The dataset was also expanded to include a calculated value for each household's total assets. This was done by multiplying the quantity of livestock or crops by historical price data for each period, enabling a comparison of total wealth across households and over time.

Finally, Gini coefficients were calculated for various assets and household categories (e.g., settlers, slaves) by decade. The analysis included separate calculations for total household wealth and individual assets such as livestock and crops. These Gini coefficients offer insights into inequality trends over the course of the 18th and early 19th centuries in the Cape Colony.

The original dataset for 1825 was sourced from various district records compiled into an Excel file. The dataset includes detailed information on settler and slave populations, livestock, crop production, and wine production across several districts in the Cape Colony.

Each district's data was extracted from the Excel file, and relevant columns were renamed to create a consistent dataset across districts. The extracted data covers key categories such as the number of men, women, sons, and daughters for settlers, slaves, Khoe (Hottentots), free blacks, and prize negroes. It also

includes details on livestock (horses, oxen, sheep, etc.), agricultural production (wheat, barley, rye, oats), and wine and brandy production.

To ensure the data was usable for analysis, several cleaning steps were applied:

The dataset's column names, which were often spread across multiple rows, were combined into a single set of unique column names.

Non-numeric data were converted into appropriate numeric formats, and missing values were handled by replacing NA values with zeros in numeric columns.

Additional variables were created, such as 'horses,' which sums saddle horses and breeding horses, and 'household_count,' which assigns each row a value of 1 to count households.

After cleaning and transforming the data, district-level summaries were calculated:

- The total number of slaves and households in each district.
- The average number of slave men per household, to understand the extent of slave ownership across districts.

The analysis was extended to inequality measures, using Gini coefficients for various assets like wheat sown, wine production, livestock ownership (horses, oxen, sheep), and household wealth. To visualize the inequality, Lorenz curves were generated for selected districts, comparing the distribution of wheat sown and livestock ownership.

The cleaned and transformed dataset allows for a detailed exploration of demographic and economic structures within the Cape Colony, focusing on the relationships between settlers, slaves, and assets, and the inequality of wealth and resource distribution across districts.

Probates

The original dataset is available in the Cape Archives. Fourie (2013) provides more detail about its origin. The transcribed series was sourced from the MOOC_probates.xlsx file and initially contained probate data covering a wide range of years. For this analysis, I restricted the dataset to the period between 1700 and 1839, excluding any records that fell outside of this range. This was done to maintain a consistent temporal focus relevant to the study.

To facilitate the analysis of trends over time, I grouped the remaining data into decades. Each year was assigned to its corresponding decade, starting from 1700 and continuing in 10-year intervals up to 1839. This allowed for a more structured approach in examining inequality across different time periods.

I calculated inequality, using the Gini coefficient, for several key asset categories: horses, cattle, sheep, ploughs, and wagons. This was done for each decade, enabling a comparison of inequality trends across the different types of assets.

Several adjustments were made to handle missing values appropriately, ensuring that incomplete data did not distort the calculations. The resulting dataset captures key measures of inequality within the Cape Colony during the 18th and early 19th centuries, focusing on these specific assets.

Slave Emancipation Dataset

The dataset has been transcribed from several original sources. A full explanation is available in Ekama et al (2021). The transcribed series is recorded in the Slave Emancipation Dataset.xlsx file, containing information about slave owners, the valuation of their slaves, and the compensation they received upon emancipation.

To facilitate the analysis, the owner's first name and surname were combined to create a unique identifier, 'Owner_fullname', for each individual. This adjustment helped identify the distinct owners in the dataset and enabled the grouping of data by owner for subsequent calculations.

I then calculated inequality, using the Gini coefficient, for both slave valuations and compensations across different districts. These measures provided insights into the distribution of wealth and compensation across slave owners within each district.

To further explore inequality, I added hypothetical settlers who owned zero slaves. This adjustment was made to reflect a broader population of settlers who did not receive compensation. For each district, I calculated the number of zero-asset settlers to add, based on a proportional increase in the number of current settlers. These new settlers were assigned zero values for both valuation and compensation, and they were integrated into the original dataset.

Once the zero-asset settlers were added, I recalculated the Gini coefficients for both slave valuations and compensations by district, to assess how the inclusion of these settlers affected the overall inequality measures.

Finally, I compared the Gini coefficients before and after the addition of zero-asset settlers to evaluate the change in inequality across the districts.

Khoe settlements

The original dataset is available in the Cape Archives, J-series.

The transcribed data was extracted from separate Excel files for each mission station, followed by cleaning and standardising column names. Missing values were replaced with zeroes, and non-numeric entries were converted to numeric values where appropriate. Variables were created to summarise livestock data by combining subcategories, for example, creating a new variable for horses by summing values from categories like horses_saddle and horses_breeding.

Inequality in asset distribution within each mission station was measured by calculating the Gini coefficient for key variables such as horses, oxen, cattle, sheep, and wheat reaped. These Gini coefficients provided insight into the degree of inequality in the ownership or production of these resources.

The results indicated significant disparities in asset ownership among Khoe populations in the studied mission stations. The analysis also allowed for cross-settlement comparisons, highlighting where inequality was most pronounced.