SUPPLEMENTARY MATERIAL TO: Fourie. S Afr J Sci. 2025;121(11/12), Art. #20532

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

---

Data management and augmentation

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

The R code used for this analysis, titled InequalityScript.R, is available upon request to the corresponding author. In future, the complete data set will also be uploaded to a data repository, facilitating broader access and reproducibility for other researchers interested in historical economic inequality. See www.capepanel.org for more information.

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

Tax censuses

The original data are available in the Cape Archives, J-series. The transcribed data were 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 numerical format.

Several transformations were applied to the data set 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 data set 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 data set 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 data set for 1825 was sourced from various district records compiled into a Microsoft Excel file. The data set 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 were extracted from the Excel file, and relevant columns were renamed to create a consistent data set across districts. The extracted data cover 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 were usable for analysis, several cleaning steps were applied:

- The data set'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 visualise the inequality, Lorenz curves were generated for selected districts, comparing the distribution of wheat sown and livestock ownership.

The cleaned and transformed data set 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 data set is available in the Cape Archives. Fourie[1] 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 data set 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 data set captures key measures of inequality within the Cape Colony during the 18th and early-19th centuries, focusing on these specific assets.

Slave emancipation data set

The data set has been transcribed from several original sources. A full explanation is available in Ekama et al.[2] 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 slave 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 data set 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 data set.

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 data set is available in the Cape Archives, J-series.

The transcribed data were 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.

References

1. 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
2. Ekama K, Fourie J, Heese H, Martin LC. When Cape slavery ended: Introducing a new slave emancipation dataset. Explor Econ Hist. 2021;81, Art. #101390. https://doi.org/10.1016/j.eeh.2021.101390