




RESEARCH ARTICLE

# Income Mobility before Industrialization: Evidence from South Africa's Cape Colony

Johan Fourie<sup>1</sup> , Erik Green<sup>2,3</sup>, Auke Rijpma<sup>4</sup> , and Dieter von Fintel<sup>5</sup> 

<sup>1</sup>Department of Economics, Stellenbosch University, Stellenbosch, South Africa, <sup>2</sup>Department of Economic History, Lund University, Lund, Sweden, <sup>3</sup>Department of Economics, Stellenbosch University, Stellenbosch, South Africa, <sup>4</sup>Department of History and Art History, Utrecht University, Utrecht, the Netherlands and <sup>5</sup>Department of Economics, Stellenbosch University, Stellenbosch, South Africa

**Corresponding author:** Johan Fourie; Email: [johanf@sun.ac.za](mailto:johanf@sun.ac.za)

(Received 4 July 2023; revised 8 May 2024; accepted 10 May 2024)

## Abstract

Attempts to measure social mobility before the twentieth century are frequently hampered by limited data. In this paper, we use a new source – annual, matched tax censuses over more than 70 years – to calculate intragenerational income mobility within a preindustrial, settler society, the Dutch and British Cape Colony at the southern tip of Africa. Our unique source allows us to measure income mobility along several dimensions, helping to disentangle reasons for the high levels of persistence we find.

**Keywords:** income mobility; social mobility; Cape Colony; South Africa; settler economies

## Introduction

From the seventeenth to nineteenth centuries, abundant land and low disease risk in settler colonies enabled a continued inflow of European migrants that opened up new farmland at relative low cost (Acemoglu et al. 2001; Engerman and Sokoloff 2012). Europeans could arrive in the newly colonized territories almost empty handed, it is said, and gradually accumulate wealth by tilling fresh land or as wage workers in the high-wage environment (Carter and Sutch 2013). Mainland North America is often cited as an example of how poor but hard-working Europeans were able to escape poverty in Europe and build a more prosperous future in the “New World” (Abramitzky and Boustan 2017), and evidence from South America suggests similar conclusions (Pérez 2017). The implication is that upward income mobility among settlers must have been high.

This is confirmed by estimates from the United States during the nineteenth century. Kearl and Pope (1984) match household in US censuses in the second half of the nineteenth century and calculate the distribution and mobility of wealth. They find “considerable mobility” in this period, aided by large numbers of new

immigrants that reported no wealth (Kearl and Pope 1984: 194). Steckel (1990: 277), using a nationally representative sample from the censuses, finds less upward mobility among those at the bottom of the distribution but “roughly the same pattern of movement among those who were initially rich.” Steckel also emphasizes that the 1850s, his period of analysis, coincided with America’s industrialization. More recent literature measuring intergenerational mobility, matching fathers to sons across US censuses ten years apart, also reports relatively high levels of mobility (Long and Ferrie 2013; Olivetti and Paserman 2015).

There are a number of concerns with both the earlier intragenerational estimates and recent intergenerational estimates. Data availability is an important limitation. Because US censuses before 1940 do not record income, occupations are often used as a proxy for economic status. Occupations are, at best, an imprecise indicator of income or wealth. Using occupations as a proxy is especially challenging in preindustrial agrarian economies, where the majority of settlers are farmers. Farmers are often treated as one category despite significant differences in wealth and income within the group. Poor matching rates between fathers and sons complicate matters further. Finally, most recent studies refer to the period after 1850, when census data is readily available (Long and Ferrie 2013; Song *et al.* 2020). Yet this period is characterized by the industrial transformation of both the Old and the New World. For those interested in the mobility of a settler society before structural change, census data are too late.

The main contribution of this paper is to provide a detailed study of late-eighteenth and early-nineteenth-century intragenerational income mobility in a preindustrial, colonial society. We use the Cape Colony, today part of South Africa, as a case study. We can do so because, since 2015, we have transcribed annual tax censuses started by the Dutch East India Company (VOC) of the settler population at the Cape. These rich data sources allow us to calculate household-level incomes for settler farmers in the Colony, and then match these households across time at a much higher rate and time resolution than are available for other settings (Rijpma *et al.* 2020). Our baseline results, based on 138,784 observations, suggest that intragenerational income mobility was surprisingly low; in other words, we find high levels of persistence, contradicting the stereotypical view of high upward mobility in settler societies. Our rich sources allow us to investigate to what extent our methodological choices explain this surprisingly low mobility in a way that existing mobility studies based on census data cannot.

Our dataset allows us to calculate mobility in several ways. Instead of only matching households every ten years, we can measure income mobility across one-year, five-year, and ten-year transitions and compare the results. The number of transition years matters; fewer years between starting and finishing periods yield higher persistence estimates. We also find more mobility at the top of the distribution as the number of transition years increase.

The Cape Colony was a frontier society, characterized by a continuous expansion of its borders. The traditional view has been to emphasize that the expanding frontier enabled migrating men and women who were willing to work hard to experience upward income and/or social mobility (Mann 1984; McGregor 1966; Mendolia and Siminski 2016). Because frontier expansion is key to explaining upward income mobility in settler economies, there may be spatial variation in the

opportunities available to settlers at the center and the periphery (Bazzi et al. 2020). Although some scholars distinguish between open and closed frontiers, one assumption, particularly in earlier studies, was that the results of one district (or state) would be externally valid (Lamar 1981; Willebald and Juambeltz 2018). But because land is still in abundance in an open frontier, and the social and economic order is fluid, we would expect upward income mobility to be greater there than in a closed frontier characterized by land shortages and an institutional order that tends to favor those who are already well established. This paper, then, compares income mobility between the Stellenbosch district, considered a closed frontier, and the Graaff-Reinet district, an open frontier. We find evidence of initial differences that support our hypothesis. Surprisingly, however, the mobility estimates soon reverse, with the frontier experiencing lower levels of mobility rapidly after the district became economically established. Frontier and closed districts provide the same opportunities for economic mobility within a short space of time.

We test several additional assumptions. We calculate income mobility elasticities that both include and exclude zero-income households. Zero-income households might include non-farmers, farmers with alternative income streams not recorded in the tax censuses or households that are mismeasured for a variety of reasons. Although including zeros does inflate the bottom of the distribution, resulting in higher persistence, the differences are not as large as might be expected. We compare mobility estimates when capital goods – notably slave holdings – are included and excluded. Our estimates do not change substantially, suggesting that it is not slave ownership that determines the high levels of persistence, though the agricultural outputs produced by enslaved people may well do so. We also compare mobility estimates for households with a single (male) household head versus those that are married. Here singleness may proxy for being young, engaging in other nonagricultural activities, or simply lower-quality matches when constructing the panel. We indeed find evidence of higher persistence when limiting the analysis to single-headed households, suggesting that poor matching rates (and thus higher measurement error) may result in overestimating mobility (Bailey et al. 2020).

Finally, we split income mobility into the three main types of production – crop, viticulture, and stock farming. We would expect that agricultural activities with immovable inputs such as vines, for example, would have substantially higher persistence. We find, by contrast, the highest income mobility amongst viticulturalists and the lowest among crop farmers, suggesting that temporary shocks such as weather changes have little bearing on calculating mobility estimates when using a large sample.

Although our study provides new insights into intragenerational income mobility in a preindustrial, settler society, it has its own limitations. As we use tax censuses as primary source, we primarily consider a population of farmers in our analysis; while that would include more than 90% of settlers in the districts we study, the numbers would be substantially lower for the Cape district, which included Cape Town with its large services sector. Second, just as others derive income from occupational classifications, we impute (gross) income from agricultural output available in tax censuses. We fail to include output not reported in these tax

censuses. Third, although we look at two districts, one at the center and one on the periphery, we do not account for in- and out-migration between districts or, indeed, to other districts.

Yet by utilizing annual, matched tax censuses over a long-time span, we are able to test many assumptions standard in the historical intergenerational literature. Our results reveal surprising levels of immobility in our preindustrial setting. They thus emphasize caution when calculating and interpreting historical social mobility estimates. Measurement error affects income mobility elasticities; a data-generating process affected by poor data recording and transcription practices, low match rates across time, and the use of proxies are likely to overestimate mobility. Incomes in the past were probably more persistent than we currently think, even in settler societies that promised a better life for those settlers at the bottom of the distribution.

### Income mobility in the present and past

Social mobility was measured by calculating the likelihood that an individual – or class of individuals – would move from one occupational category to another (Grusky and Sørensen 1998). Intergenerational social mobility linked parents to children, measuring the extent of social mobility between generations (Erikson and Goldthorpe 2002).

Estimating intergenerational mobility elasticities (IGE) has a long tradition in the social sciences. The intuition behind the IGE is simple: A low IGE indicates that parental earnings have a minimal impact on the earnings of their children, thereby allowing for greater economic mobility across generations. In other words, a lower IGE value implies that individuals have higher chances of changing their economic status independent of their family's financial background. Initial work suggested very high mobility (or a low IGE) for US males; Becker and Tomes (1986), for example, find an IGE of less than 0.2. But later studies that account for possible measurement error or use administrative data tend to find more intergenerational persistence (i.e., low mobility) (Mazumder 2005; Solon 1999).

Measurement error is one reason for the higher rates of persistence. If left unaccounted for, such an error will create the appearance of more mobility than actually exists. There are many reasons why earnings data may not be completely reliable, and these have been well documented in the literature on contemporary intergenerational earnings mobility. Self-reported earnings or wealth data taken from surveys or census data may not exactly match actual earnings (Blattman *et al.* 2016; Bruckmeier *et al.* 2015). When we compare parents' and children's earnings at different points in their lifecycles, this noncomparability will produce inflated estimates of earnings mobility. IGE estimates can also be biased by temporary shocks to household incomes; Mazumder (2005) demonstrates that persistence may be underestimated by 50% if temporary shocks are unaccounted for.

One way to overcome such measurement error is to use instrumental variables; Solon (1999) reports that studies using this method usually find IGE estimates closer to 0.5. To combat the effect of temporary shocks, income is measured by using the average mid-life earnings. Another way to improve mobility estimates is to use more

reliable data. Lee and Solon (2009) use the full Panel Study of Income Dynamics and report IGEs of between 0.43 and 0.56. Chetty et al. (2014) use US tax returns to calculate a surprisingly low IGE of around 0.3. Mazumder (2016: 83) demonstrates, however, that the age structure and limited panel dimension of the data used by Chetty et al. (2014), notably the short panels that cover only certain portions of the lifecycle, lead to “considerable downward bias in estimating the IGE.” He shows that when these data limitations are relaxed, the IGE of family income in the US is likely to be higher than 0.6.

The problem of measurement error extends to estimating intragenerational social mobility. It is more pronounced in high frequency data over individuals' life course, and specifically in developing country contexts where respondents find it hard to report on irregular incomes from informal and self-employment activities (Fields et al. 2003). But it is also true of data created before the twentieth and twenty-first centuries. The first wave of mobility studies on historical data for the US found unrealistically high rates of mobility (Galenson and Pope 1989; Parkerson 1982; Steckel 1989). Armed with stronger computing power, Ferrie (2005) was one of the first to construct more reliable estimates of mobility, calculating that white American men of the mid-nineteenth century were indeed highly socially mobile. In later work, Long and Ferrie (2013), comparing American and British intergenerational occupational mobility, confirm these results by finding higher social mobility in the United States until the 1950s. American exceptionalism – its higher social mobility – seems to disappear during the second half of the twentieth century, although there is some disagreement about the timing of this decline. Olivetti and Paserman (2015) calculate IGE for both sons and daughters in the United States and find a sudden increase in mobility from 1900 to 1920. By contrast, Song et al. (2020) argue that most of the decline happened prior to 1900, largely driven by the movement of children from farm to factory during the era of industrialization.

The field has since extended in many directions. One line of inquiry has been to calculate intergenerational mobility estimates for more regions, time periods, and groups previously unaccounted for, like Black Americans (Alesina et al. 2021; Chen et al. 2015; Collins and Wanamaker 2022; Deroncourt 2022; Pérez 2019). One hypothesis is that land abundance made the transition towards independent farming easier in the Americas compared to Europe, which in turn spurred upward occupational mobility.

A second line of inquiry has been to link more generations. One finding is that grandfathers matter: adding more generations results in lower intergenerational mobility estimates (or higher persistence) (Collado et al. 2023; Long and Ferrie 2018; Solon 2018). This is because adding more generations captures cumulative advantages or disadvantages that are passed down through family networks. Using rich Swedish population data, for example, Adermon et al. (2021) show that traditional two-generation estimates underestimate intergenerational persistence by at least one-third.

A third line has been to measure intergenerational mobility for various subgroups of society. Migrants have been a particular focus (Abramitzky et al. 2021a; Collins and Wanamaker 2015; Ward 2022). A fourth line has been to develop more sophisticated matching techniques (Abramitzky et al. 2021b; Bailey et al. 2020; Helgertz et al. 2022). A fifth line of inquiry has been attempts at causal explanations

for social mobility, including inheritance (Adermon *et al.* 2018), education (Card *et al.* 2022), and land (Collins *et al.* 2022).

Despite these advances in the field, valid concerns remain regarding the study of income mobility. Many of these concerns apply to both intergenerational and intragenerational studies. One major concern is the classification and ranking of occupations, which becomes problematic in a preindustrial agrarian setting. Farmers, who comprise a significant portion of historical occupations, pose a challenge as there is no information available to distinguish between more and less prosperous farmers, leading to inevitable bias in mobility estimates. The categorization of farmers as a single occupational category introduces another potential bias. If a farmer is classified as “farmer” in one census but given a more specific job title, such as “milkman” or “breeder” in the next census while performing the same tasks, social mobility is likely to be overestimated. Additionally, the wages and status associated with occupations are expected to change over time (Ward 2023). In short, measurement error in data generation for farmers may be larger than for other occupations.

A second concern is the issue of matching. Most studies rely on historical censuses conducted a decade apart, and match rates of 30% or lower are considered fair but may contain errors (Bailey *et al.* 2020; Massey 2017). Common names, such as John Smith in the United States, are unlikely to be matched correctly across datasets separated by several years. The selection into being matched may introduce bias in mobility estimates.

Another limitation is the focus on the industrial era. National censuses, which form the basis for many mobility studies, only became part of government bureaucratic functions during the second half of the nineteenth century, coinciding with the expansion of industrial production (Whitby 2020). Therefore, available estimates primarily cover countries during or at the dawn of industrialization, limiting our understanding of mobility in preindustrial societies.

In our paper, we aim to address these limitations by estimating household-level income mobility in a settler colonial setting: the Cape of Good Hope at the southern tip of Africa. Leveraging the wealth of information preserved by the Dutch East India Company and its successors, we combine annual tax censuses spanning over 70 years for two districts – one close to the colonial center and one on the colonial periphery – to calculate measures of social mobility in a preindustrial, colonial context. Our findings reveal low mobility at the Cape, with some variation observed at the frontier. These results confirm the existence of a “colonial landed elite” (Dooling 2005; Guelke and Shell 1983; Ross 1983; Williams 2013). They also raise questions about the assumed high mobility of settler societies during the phase of frontier expansion.

## The Cape and its institutions

The Cape Colony was established when a small group of VOC employees was sent by the Company to establish a refreshment station for ships sailing between Europe and the East Indies in Table Bay. Soon after their arrival, in 1652, the commander of the station, Jan van Riebeeck, recognized the need to expand production to service the more than 6000 sailors and soldiers that arrived in the fledgling community every year (Boshoff and Fourie 2010). He thus released nine Company servants in

1657 to become settlers: farmers that produced wheat and meat to be sold to the fort in Cape Town.

The history of the Dutch Cape Colony is a history of geographical expansion. Twenty-five years after the establishment of the trading post the area of the VOC's territorial control was still small, yet by the end of the eighteenth century, it was about the same size as Great Britain – a truly remarkable expansion (Guelke 1976). Stellenbosch is located in the south-west and was one of the earliest settlements in the Cape Colony, established in 1679, while Graaff-Reinet is located in the interior and was established more than a century later.

Limited knowledge about local climatological conditions made farming difficult for these early farmers, while constant skirmishes with the indigenous Khoesan, a pastoral people that inhabited most of the fertile land in the area for the past 2000 years, made it dangerous. It was only after a series of violent conflicts with the Khoesan in the 1670s that the fertile valleys below the first mountain ranges opened up. The numbers of the settler population increased substantially and so, too, did the crops and vines under cultivation, satisfying the demand of the passing ships. By contrast, Khoesan numbers dwindled as a result of violence or migration deeper into the interior, though some remained, sometimes forcibly, on settler farms as laborers (Fourie and Green 2015).

The abundance of land that could be acquired relatively cheaply through the loan farm system that was applied in the frontier regions – a system where a settler had to pay an annual fee to get a farm registered – gave poorer settlers a relatively fair chance to accumulate wealth by investing in land. The fee remained unchanged for long periods of time. Farmers had to pay 12 rixdollars annually until 1732 when it was doubled to 24 rixdollars. It then remained unchanged throughout the 18<sup>th</sup> century. Land in the eastern and northern frontiers was of lower quality compared to the commercial center of the southwestern Cape and the distance to the market (i.e., Cape Town) was significantly longer. However, land at the frontier was in general suitable for pastoral farming and the demand for livestock from passing ships steadily increased throughout the nineteenth century. It is this demand, argues Neumark (1957), that explains the rapid expansion of the frontier during the second half of the eighteenth century, relying on a system, as Swanepoel and Fourie (2018) shows, of loan farms rather than farms endowed with freehold property rights. Neumark (1957) argues that settlers with limited means were rational profit-maximizers by moving towards the frontier, establishing cattle farms.

Yet despite Neumark's (1957) emphasis on economic profitability, the general perception of the Cape economy was one of economic backwardness, a settler society that, with the exception of a few elite families, struggled for survival. Ross (1983) first suggested a reinterpretation of the Cape economy. Instead of "more a static than progressing community" that "advanced with almost extreme slowness (Fourie 2013: 419)," Duin and Ross (1987) proposed that the Cape had a thriving agricultural sector, one that was expanding over the eighteenth century. His views have received recent support. Using both probate inventories that record household assets and tax censuses, empirical evidence now supports the notion that Cape settlers were, on average, relatively affluent, at least compared to other settler societies of the time (Fourie 2014; Fourie and Garmon 2023).

Not everyone was equally well-off. According to Ross (1983: 206–207), “this new prosperity was not spread evenly throughout the colony.” When the German traveler, Otto Mentzel (1925), visited the Cape in the 1730s, he classified the settlers into four categories: absentee-landlords who lived in Cape Town and employed wage laborers on their farms to oversee productions, resident landlords who lived on large estates in affluence, just “like a gentry” (Mentzel 1925: 102), settlers who worked alongside their workers, producing for a market, and, finally, the poor stock farmers of the interior.

Historians have expanded on Mentzel’s observations, documenting the seemingly deep and entrenched class divisions within Cape society. Guelke and Shell (1983: 265) noted that the early eighteenth century was a period of rationalization in settler society: “Large estates swallowed up unsuccessful small farms and a few were enlarged by dynastic marriages. Large-scale production yielded low returns on capital, but small farms often ceased to make profits.” Indeed, as Fourie and Von Fintel (2011) show using select tax censuses, Cape settler society was a highly unequal society. A “colonial landed elite” emerged, giving rise to severe levels of inequality and, given the static nature of the economy, low social mobility (Dooling 2005: 148).

It is in this setting that we turn to tax censuses – the *opgaafrollen* – to measure the nature of mobility in two districts of the Cape: the wheat and wine farmers of the fertile Stellenbosch district close to Table Bay, and the largely pastoral farmers of the eastern frontier district, Graaff-Reinet. The tax censuses mainly record essential household and production details.<sup>1</sup> They count the number of men, women, and children (sons and daughters), as well as the number of European settler laborers and enslaved people (by age and sex). In Graaff-Reinet, they also include a count of the number of Khoesan laborers by age and sex. Quantities of livestock (horses, cattle, and sheep), crops (wheat, barley, rye, and oats), vines, and wine production represent agricultural production. For some years, the records included counts of weapons like flintlocks, pistols, and swords. Post-1800 censuses sometimes added more production details, such as brandy volume, and occasionally listed other assets like wagons and carts. The censuses also sometimes included summaries, mortality rates, church donations, and taxes paid. In some cases, the household head’s signature was present. The Cape of Good Hope Panel project is currently transcribing these records to cover more years and districts (Rijpma *et al.* 2020).<sup>2</sup>

The tax censuses also have limitations. Apart from district or subdistrict names, they included no information about the location of farms or, importantly, of the size of farms. The Cape had two property right regimes: smaller, freehold farms in the regions closer to Cape Town and larger, loan farms in the interior (Swanepoel and Fourie 2018). We have no way of knowing either the value of farms or ownership type.

The censuses also exclude all nonagricultural production. Despite manufacturing being prohibited by the Dutch East India Company at the Cape, there is some evidence from settler probate inventories of light manufacturing – barrel making, for example (Fourie 2013). Even where information about output was recorded,

<sup>1</sup>See Fourie and Green (2018) for an extensive discussion of the process of transcribing the tax censuses.

<sup>2</sup>For more information, see [www.capepanel.org](http://www.capepanel.org).



such as cattle ownership, for example, we should also worry about underreporting. Fourie (2013), in comparing the probates to the tax censuses, finds substantial underreporting in those goods that were taxed, notably cattle and sheep. In the context of measuring mobility, however, such underreporting is unlikely to vary systematically year-by-year.

Importantly, though, the tax censuses include the names of the household head and, often, his wife. This allows us to link households between years, making it possible to ascertain the likelihood that individuals of different classes were able to move up or down in the late eighteenth and early nineteenth centuries.

### Constructing the panel

The tax censuses for our two districts comprise a total of 138,784 observations, necessitating the use of automatic record linkage to create a panel that traces households over time. We employ an updated version of the supervised machine-learning approach introduced by Rijpma et al. (2020). This approach estimates optimal weights and thresholds using manually labeled data, which tends to outperform manual linking rules and unsupervised machine learning methods in similar contexts (Aiken et al. 2019). Additionally, this approach provides precise measures of the linkage's accuracy.

The first step involves selecting candidate pairs from the cross sections, which are plausible links based on an efficient metric. In this case, we use the cosine similarity between the bigrams of household heads' surnames to identify candidates. The original training data created for linking the Graaff-Reinet dataset, using manual methods to match households to themselves, demonstrated that this metric captured over 99% of all true links, with true links accounting for 9% of all candidates generated by the cosine similarity procedure. Therefore, the dataset is of manageable size.

To address the discrepancy in performance between districts, a separate training dataset was created for the Stellenbosch district. The model that was trained solely on the Graaff-Reinet data underperformed when applied to the Stellenbosch data due to the higher prevalence of single-headed households in Stellenbosch, which weakened some of the heuristics employed for Graaff-Reinet. We ultimately combined the two training datasets to make predictions for both districts.<sup>3</sup>

The manually labeled data was split into 70% training and 30% evaluation sets for training a predictive model. Table 1 lists the features used in the model, which includes Jaro-Winkler string distances to measure differences between string variables across tax census cross sections. Based on the evaluation data's performance, we selected gradient boosting as the predictive model, as it can handle nonlinearities in string distances and missing data (Chen and Guestrin 2016). The most influential features in the final model were the first and last names of the husband and wife.

---

<sup>3</sup>We combine the two datasets to maximize the training data available, allowing the model to learn how to predict difficult cases as well – in particular, more cases of households with no wives. We include district dummies to allow the model to adapt to data differences between the two regions.

**Table 1.** Features used to predict links in gradient boosting model

Variable	Description
mlastdist	Jaro-Winkler (JW) string distance between male surnames
mfirstdist	JW distance between male first names
minitialsdist_osa	Optimal String Alignment (OSA) string distance between male initials
mlastdx	Soundex distance between male surnames
mfirstdx	Soundex distance between male first names
wlastdist	JW distance between wife's surnames
wfirstdist	JW distance between wife's first names
winitialsdist_osa	OSA distance between wife's initials
wlastdx	Soundex distance between wife's surnames
wfirstdx	Soundex distance between wife's first names
namefreq_from	Standardized name frequency "fro" candidate
namefreq_to	Standardized name frequency "t" candidate
spousenamedist_from	Distance between spouse surnames "fro" candidate to detect drop of maiden name
spousenamedist_to	Distance between spouse surnames "fro" candidate
wifepresent_from	Wife present in "fro" candidate
wifepresent_to	Wife present in "t" candidate
wifeinboth	Both wives present
settlerchildrengauss	Difference in number of children, normalized using a Gaussian kernel
nextmfir	JW distance with next best man's first name
mfir_uniqueness_from	Uniqueness (mean JW distance to all other names in the data) of man's first name of "t" candidate
mfir_uniqueness_to	Uniqueness of man's first name of "t" candidate
matches	Number of candidates
husb_wife_surnamedist	JW distance between husband and wife's surname distance
region1	Region dummy

The predictive model's performance is evaluated using two metrics: precision, which measures the share of all true predictions that are accurate links, and recall, which measures the share of all true links correctly predicted. Our procedure achieves a recall of 86% and a precision of 93%, resulting in a combined F1 score of 89% (Table 2).<sup>4</sup> This performance is similar that of logistic regression and random forests models used by Rippma *et al.* (2020): 89% and 86% recall; 90% and 94%

<sup>4</sup>The F1 score is the harmonic mean of the precision and recall. It could be improved by accepting lower predicted link probabilities than the current threshold of 0.5, but we favor a higher precision to minimize false positives.

**Table 2.** Confusion matrices for training and validation data, comparing actual links (rows) from the manually labeled data with the predictions from the model (columns)

		Predicted			
		Training		Test	
		0	1	0	1
Actual	0	19284	5	8408	15
	1	3	506	31	195

Test set precision =  $195/(195 + 15) = 0.93$ ; recall =  $195/(195 + 31) = 0.86$ ;  $F1 = 2/(1/0.93 + 1/0.86) = 0.89$ .

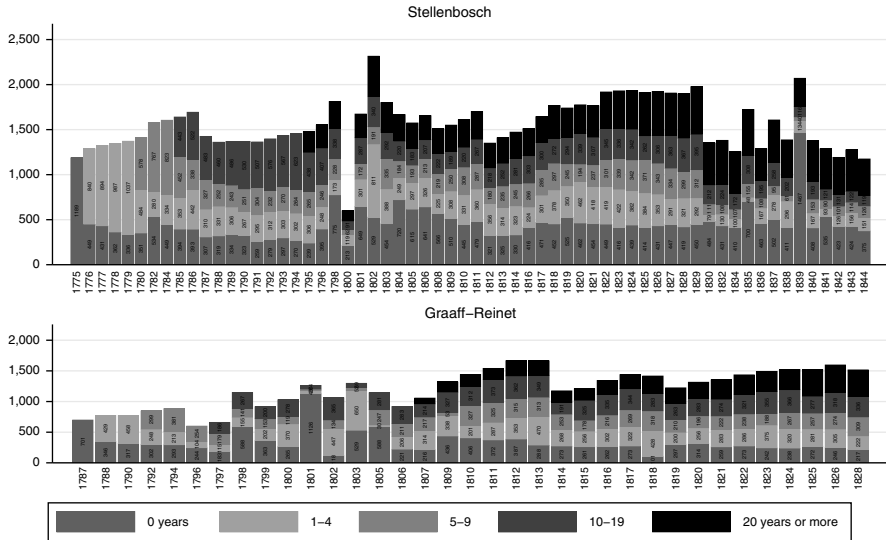
precision respectively. However, gradient boosting was found to scale better, helping us solve several computational bottlenecks, and it allowed us to include missing values such as names in our data, easing data handling and increasing the performance of the model for these records. This performance compares favorably to other historical record linkage studies and instills confidence in the presented results (Abramitzky et al. 2021b; Massey 2017; Price et al. 2021).

The procedure's success can be attributed to several factors. First, the use of annual data reduces issues related to data attrition compared to census linkage studies with longer intervals. Second, the utilization of household-level data, as opposed to individual-level data, provides more information for accurate linking (Akgün et al. 2020; Rijpma et al. 2020). While a single person's name might not uniquely identify them, the combination of names for a couple often does. This linking strategy, however, does imply that marital status affects the probability of being linked. Performance is excellent for individuals married in two cross sections ( $F1: 0.95$ ) and good for single individuals in two cross-sections ( $F1: 0.81$ ), but weaker for individuals married in one year and not in the other ( $F1: 0.55$ , with a recall of 50%, meaning that we recover only 50% of these links) due to the difficulty of determining whether the change in marital status represents a widow or a son succeeding the household. It should be noted that this aspect is inherent to the data rather than an artifact of the procedure. Overall, the record linkage procedure performs well, although households that experience changes in marital status may be underrepresented. Since disrupted households would probably face downward mobility, we potentially overestimate persistence. In online Appendix A we quantify the impact of this bias, finding that the overall results are robust to reweighting for this type of attrition.

The predicted links are utilized to construct a panel by working backward from the most recent cross-section. For each household in each base year, the candidate with the highest predicted probability being a match – and a predicted probability of at least 0.5 – is selected. This approach results in approximately 80% of households being linked to at least one other year in the resulting panel. The linking rates are slightly lower in years where spouse information is not recorded in the tax censuses.<sup>5</sup> Moreover, the resulting series are of sufficient length, with the combined data from the two districts revealing the presence of 10 series with a minimum length of 40 observations, over 1,000 series with a minimum length of 18, and more than

<sup>5</sup>Graaff-Reinet: 1801, 1803; Stellenbosch: 1830, 1832, 1834, 1839, 1841–4.

Evolution of sample by number of years since household entered panel



**Figure 1.** Evolution of sample in Stellenbosch and Graaff-Reinet.

Notes: Linked and unlinked households are shown in this figure. Unlinked households are, by definition counted as having entered the panel 0 years ago. Figures inside the bars indicate the sample size.

12,000 series with a minimum length of 3 (see Figure 1). This panel dataset, known as the Cape of Good Hope Panel, enables the calculation of social mobility across time and in various districts.

With regards to weighting, Bailey *et al.* (2020) emphasize that longitudinal matching generates both errors of inclusion (“false matches” or low precision in our terminology) and exclusion (“missed matches” or low recall). Both errors introduce significant bias in estimating social mobility. The extent of the first problem can be investigated by limiting samples using “validation variables,” auxiliary information which supports when matches are “real” or not, such as agreement over time about where parents were born. The second problem can be addressed by reweighting estimates of social mobility using inverse propensity scores. This process gives greater weight to households possessing characteristics with typically poor match rates, and are, consequently, mostly underrepresented in estimation samples. Because we do not have age and birthplace in our sample, we use “name commonness” in our matching algorithm, similarly to Abramitzky *et al.* (2014). This information would have been our strongest candidate for a validation variable. But because this variable was used to generate matches, it cannot be used as a validation variable for sensitivity testing. However, our data are suited to reweighting using a logit specification. Appendix A provides full details of how our weights were generated. The strongest predictor of matching households over time was whether the tax records included two names of a couple, providing sufficient information to verify that the same household was traced over time. Our results – both estimates of persistence elasticities and transition plots – do not change meaningfully when we reweight our data and when we trim samples to

exclude observations with low propensity scores – those at the greatest risk of not being matched. However, in the case of 10-year transitions, the results are sensitive to trimming. Because most of our analysis focuses on 5-year transitions, our results are highly robust to these methodological variations. Going forward, all results in the main text are reweighted by propensity scores and all observations with propensity scores below 0.3 are omitted from the estimation sample. In the online appendix, we report results using unweighted, weighted, and trimmed estimates.

## Measuring social mobility at the Cape

### Method

The Cape of Good Hope Panel enables the measurement of social mobility within the same household over time. To accomplish this, we construct the total annual income for each household by combining flow and stock variables that generate income. Flow variables involve the difference between the volumes of crops reaped and sowed by households, multiplied by their respective prices to derive crop income. Stock variables, such as slave ownership and livestock (cattle, horses, and sheep), are assessed using prices multiplied by 6% of each household's capital stock. The 6% figure represents the standard interest rate at the Cape. As per Fourie and Garmon (2023), we do not account for depreciation.

We rely on price data based on a limited set of auction records spanning approximately 1798 to 1802. While prices may have varied over the long term, our assumption of fixed prices is unlikely to significantly impact social mobility estimates within districts and over relatively short periods of time, such as five years. One reason for this is that the VOC fixed the prices of agricultural goods for sale in Cape Town; if anything, the Cape experienced deflationary pressures for most of the eighteenth century (Du Plessis and Du Plessis 2012). We further examine the assumption of price stability by presenting separate estimates based on income sources. It is our hope that future research will explore the unaccounted price effects in this analysis.

Empirical studies on income mobility seek to understand the persistence of income for the same household ( $h$ ) over time.<sup>6</sup> Denoting the natural logarithm of

---

<sup>6</sup>We calculate gross income rather than net income. Net income would require us to subtract costs of production. Given that we do not have accurate information at the farm level on wages for European laborers or Khoen laborers and that enslaved workers did not receive a wage (and we have little information about the costs of subsistence), we believe gross income elasticities to be a more accurate reflection of mobility across time and space. We construct our income measure by combining a number of agricultural outputs, multiplied by unit prices: volumes of grains harvested less volumes sown, as well as wine and brandy production. For the latter two outputs, we cannot determine the inputs. We also include a number of stock variables multiplied by their market asset value and assume a return of 6% per annum on each: horses, cattle, sheep, and enslaved workers, an approach followed by Du Plessis et al. (2015). This measure is imperfect in that grain outputs and horses serve as inputs for cattle breeding, for example, and that it is not possible to convert stocks to incomes without further assumptions. To test the robustness of our results to this type of “double counting,” we also introduce robustness checks where we only analyze specific gross income sources separately. The patterns we observe remain stable so that we continue to use our combined measure in our core analysis.

household income as  $y$  in year  $t$ , we estimate the following equation using annual panel data:

$$y_{th} = \mu + \beta y_{t-k:h} + u_{th}$$

where  $k = 1, 5, 10$  and  $u$  is an idiosyncratic error term. The estimated elasticity ( $\beta$ ) reflects the percentage change in current income associated with a 1% change in past income. Elasticities approaching 1 indicate perfect persistence in incomes, implying no mobility. We present these elasticities in brackets in all our figures and provide an estimation summary in Table 3.

While elasticities offer a holistic perspective on mobility/persistence across the entire distribution of households, new analytical tools enable a more disaggregated examination. We utilize transition probability color plots to visualize mobility across the entire income distribution (Van Kerm 2011). Each row in a transition probability color plot represents a marginal distribution, with rows arranged so that low-income groups occupy the top and high-income groups occupy the bottom in the origin period. Different shades indicate the marginal distribution in the destination period, with low-income groups shown in white and high-income groups in black. A plot with horizontal white lines at the top and horizontal black lines at the bottom exemplifies complete immobility: individuals at the bottom of the income distribution (top row) remain at the bottom (white). In the case of complete reversal, the top line would become black, indicating that individuals at the bottom of the income distribution (top row) transition to the top of the distribution (black), while those at the top of the income distribution (bottom row) move to the bottom (white). Conversely, a plot with vertical lines represents complete independence, signifying an equal likelihood of mobility for groups at the bottom and top of the income distribution. The usual situation falls between these extremes, with movements the proportions in each quintile changing over time.

We use these transition probability color plots to provide a more nuanced understanding of mobility patterns. By visualizing the movement of different income groups over time, we can identify trends and variations in mobility across the income distribution.

### **Data and descriptive statistics**

We start by exploring the nature of the analysis sample. Figure 1 displays the number of households in each district by year and classifies them according to how long they have been in the panel. In the first year of our panel, as expected, all households have been in the panel for zero years. Starting from 1776, we consistently observe that approximately two-thirds of the Stellenbosch observations are linked to at least one previous year. By 1780, nearly half of the Stellenbosch panel matches to five or more years, and by 1785, one-fifth of the panel matches to ten or more years. By 1795, around one-tenth of the panel matches to 20 or more years. A similar increase in match rates emerges for Graaff-Reinet.

Figure 1 also reveals the presence of outlier years, such as 1800 in Stellenbosch, where the sample size decreases by over 50% from the previous observation, and 1802, where the sample size increases by more than 10%, only to decline the following year. These discrete sample changes correspond with times when record-

**Table 3.** Summary of elasticity estimates by period, district and assumption

	Gap length		Stellenbosch			Graaff-Reinet		
			Raw	Weight	Trim	Raw	Weight	Trim
All periods	1	$\beta$	0.919	0.915	0.918	0.876	0.871	0.869
(without zeroes)		(se)	(0.002)	(0.002)	(0.002)	(0.003)	(0.004)	(0.004)
		N	25274	25274	25021	17049	17049	16811
	5	$\beta$	0.757	0.756	0.774	0.677	0.665	0.625
		(se)	(0.005)	(0.005)	(0.005)	(0.006)	(0.007)	(0.006)
		N	15240	15240	13636	10200	10200	9137
	10	$\beta$	0.557	0.532	0.418	0.568	0.529	0.553
		(se)	(0.008)	(0.009)	(0.009)	(0.011)	(0.013)	(0.012)
		N	7622	7622	3990	5186	5186	3690
Period 1	5	$\beta$	0.769	0.774	0.796	0.292	0.299	0.227
(without zeroes)		(se)	(0.008)	(0.008)	(0.008)	(0.033)	(0.034)	(0.035)
		N	6416	6416	6225	442	442	367
Period 2	5	$\beta$	0.613	0.633	0.478	0.744	0.701	0.730
(without zeroes)		(se)	(0.015)	(0.016)	(0.017)	(0.013)	(0.016)	(0.013)
		N	1808	1808	1326	1952	1952	1658
Period 3	5	$\beta$	0.695	0.719	0.603	0.701	0.701	0.645
(without zeroes)		(se)	(0.007)	(0.008)	(0.008)	(0.007)	(0.008)	(0.007)
		N	7016	7016	6085	7806	7806	7112

(Continued)

Table 3. (Continued)

	Gap length		Stellenbosch			Graaff-Reinet		
			Raw	Weight	Trim	Raw	Weight	Trim
All periods	5	$\beta$	0.739	0.728	0.723	0.703	0.692	0.650
(with zeroes)		(se)	(0.005)	(0.005)	(0.005)	(0.007)	(0.007)	(0.007)
		N	18883	18883	15150	11019	11019	9542
All periods	5	$\beta$	0.672	0.684	0.703	0.720	0.680	0.720
(without capital,		(se)	(0.006)	(0.006)	(0.006)	(0.007)	(0.008)	(0.008)
without zeroes)		N	13039	13039	11850	9416	9416	8703
All periods	5	$\beta$	0.634	0.629	0.628	0.656	0.619	0.625
(wife present,		(se)	(0.006)	(0.006)	.628.00	(0.006)	(0.006)	(0.006)
without zeroes)		N	10430	10430	10158	9278	9278	9027
All periods	5	$\beta$	0.736	0.694	0.812	0.593	0.522	0.583
(wife not present,		(se)	(0.009)	(0.009)	(0.01)	(0.025)	(0.026)	(0.068)
without zeroes)		N	4810	4810	3478	922	922	110
All periods	5	$\beta$	0.705	0.707	0.707	0.541	0.555	0.550
(Crop income,		(se)	(0.016)	(0.016)	(0.016)	(0.03)	(0.03)	(0.03)
without zeroes)		N	2398	2398	2300	1028	1028	1010
All periods	5	$\beta$	0.645	0.644	0.644	0.449	0.435	0.425
(Alcohol income,		(se)	(0.011)	(0.011)	(0.011)	(0.047)	(0.047)	(0.047)
without zeroes)		N	3598	3598	3523	343	343	337

(Continued)



Table 3. (Continued)

	Gap length		Stellenbosch			Graaff-Reinet		
			Raw	Weight	Trim	Raw	Weight	Trim
All periods	5	$\beta$	0.686	0.679	0.688	0.730	0.682	0.730
(Animal income, without zeroes)		(se)	(0.008)	(0.008)	(0.009)	(0.007)	(0.008)	(0.008)
		N	7660	7660	7316	9370	9370	8664
		Weighted	No	Yes	Yes	No	Yes	Yes
		Trimmed	No	No	Yes	No	No	Yes

Notes: Own calculations from *Opgaafrollen*. Inverse propensity weights were generated from columns (1) and (4) in Table A1. The first column of figures presents unweighted estimates, the second column uses the same sample as in the first column, but reweights, while the third column reweights, but also limits the sample to observations with propensity scores above 0.3. In Stellenbosch, period 1 corresponds to 1780–1798, period 2 to 1803–1811, and period 3 to 1812–1829. In Graaff-Reinet, period 1 corresponds to 1792–1800, period 2 to 1805–1813, and period 3 to 1814–18.

keeping was disrupted by transition from VOC to British rule (such as a drop in sample size in 1800 in Stellenbosch and the high proportion of “new entrants” in Graaff-Reinet in 1801) and in some cases boundary changes (with some of the sample being included in newly established districts – as is evident in 1814 and 1819 in Graaff-Reinet and after 1830 in Stellenbosch). As noted before, these periods also correspond to times when spouses’ names were not reported. We therefore limit our sample appropriately so that individual analyses do not span any of these discrete changes. We therefore focus on three distinct periods: the first spans from 1780 to 1798, coinciding largely with the Dutch East India Company era. The second period, from 1803 to 1811, encompasses the transition from the Batavian Republic to British rule when major legislative changes were introduced, including the cessation of the slave trade. The third period, from 1812 to 1829, covers a phase of stable British rule. Our choice of periods also ensures, to the extent possible, the use of data within common district boundaries for each episode.

The advantage of an annual panel is the flexibility to vary transition lengths. We start by calculating three types of year-transitions: one-year, five-year, and ten-year transitions. One-year transitions are by nature associated with low mobility since a household’s movement up or down the income ladder is likely to be less significant over one compared to ten years. This exercise illustrates that panel data with an annual frequency is well-suited to analyze more precisely the time it takes for households to achieve upward mobility. In subsequent estimation, we utilize five-year transitions, requiring a match to an observation at least five years earlier in the panel. This entails that more than half of all observations are consistently used to estimate mobility, a rarity in the historical mobility literature (Ward 2023). Furthermore, as discussed in online Appendix A, patterns in 5-year transitions are robustly estimated, regardless of whether we reweight for the propensity of being matched across waves, or whether we truncate households that have low propensity scores of being matched.

Table B1 in the online appendix presents descriptive statistics for each period and district, including all families that appear in the tax records, regardless of whether they are linked or not. In both the Stellenbosch and Graaff-Reinet districts the majority of households report positive incomes, indicating the large contribution of agricultural activity to the economy. However, there are notable differences in income sources and trends between the two districts. Stellenbosch gradually diversifies beyond agriculture, with an increasing proportion of households reporting zero agricultural incomes over time. As documented by Fourie and Von Fintel (2011), many households report zeroes on agricultural assets, because they are involved in other sectors of the economy. The increase in zero incomes suggests a transition to alternative sources of income beyond traditional farming activities. On the other hand, Graaff-Reinet shows a rise in positive incomes from animal holdings, indicating the importance of livestock-related activities in the district.

In Stellenbosch, there is a significant increase in incomes across various percentiles of the distribution after 1800, indicating overall economic growth and prosperity. However, it is important to note that these aggregate statistics do not capture changes in individual rankings and relative income mobility. In Graaff-Reinet, the distributional aggregates remain relatively stagnant over time, with incomes starting at lower levels compared to Stellenbosch. Although there is a

modest increase in incomes derived from crops and alcohol in Graaff-Reinet, the overall income dynamics in the district are less pronounced than in Stellenbosch.

These descriptive statistics already demonstrates the dynamic nature of economic mobility over time, highlighting how income mobility can differ significantly between districts and across periods marked by political and administrative transitions. In the next sections, we test how robust these results are to changes in sampling and methodological approaches.

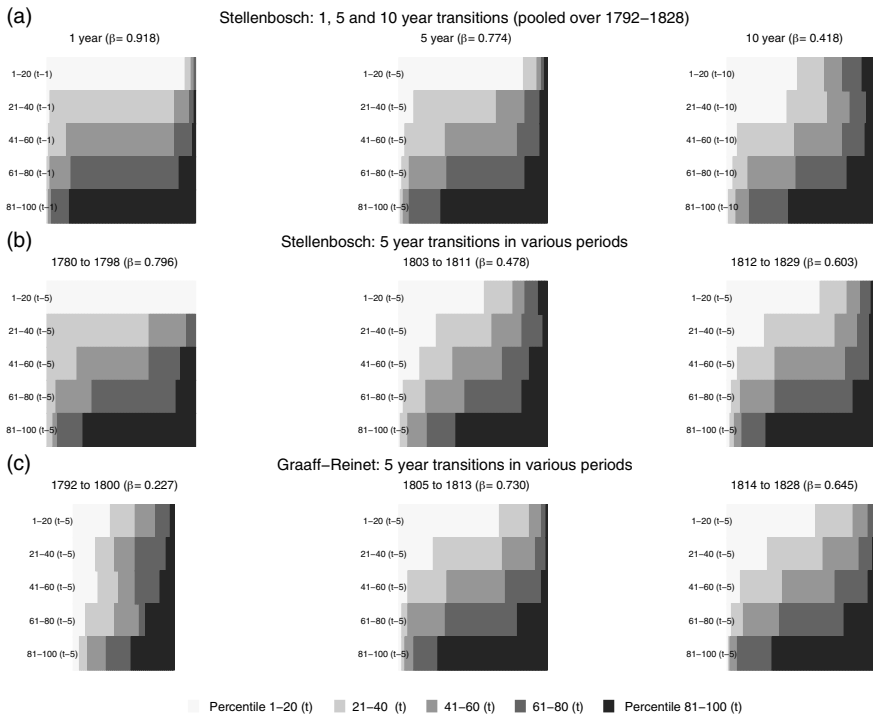
### Baseline results

Our baseline analysis excludes households with zero incomes. Zero-income households are not necessarily independent households. The officials who administered the censuses often included adult children (older than 16 years) who were still living with their parents as separate observations. Additionally, some of these households were involved in nonagricultural production, which is not recorded in the *opgaafrollen*, so that our data falsely reflects zero incomes for these households. To avoid artificially inflating persistence, we remove these zero-income households from the analysis. Below we offer additional robustness tests to examine how this approach affects our results.

Panel (a) of Figure 2 starts by comparing mobility estimates for one-year, five-year, and ten-year transitions using pooled data from 1780 to 1829 in Stellenbosch. Estimates are weighted by inverse propensity scores of being matched and the sample truncated to observations with propensity scores above 0.3. As expected, mobility is lower when measured over one-year transitions, with a persistence elasticity of 0.918, compared to five-year (0.774) and ten-year (0.418) transitions. Table 3 shows these estimates, but also presents them for Graaff-Reinet where similar reductions in persistence occur depending on the transition period. The table also shows that weighting and trimming by propensity scores does not change conclusions significantly.

Examining multiple transition years also allows us to observe mobility differences across the income distribution. The one-year transition plot shows almost symmetric results across rows, indicating that those at the top and bottom of the distribution are equally likely (or unlikely) to experience very short-term mobility. However, as we move to longer transition periods, the plots become less symmetric, with those at the top of the distribution exhibiting more persistence than those at the bottom. This pattern can only be observed with the benefit of an annual panel over an extended period of time.

A further benefit of a very long panel dataset is that it is possible to measure mobility in different periods that correspond to changing political regimes and economic contexts. We compare 5-year transitions in Figure 2(b) for Stellenbosch, an established district at the core of agricultural activity at the time. Mobility is low at the end of the 18<sup>th</sup> century in Stellenbosch – incomes are highly persistent with an elasticity of 0.796. Mobility increases in the second period, with the elasticity declining to 0.478. In the third period, mobility decreases again, although not to the same level as in the first period, with a persistence elasticity of 0.603. The first plot in panel (b), corresponding to Stellenbosch from 1780 to 1790, shows a white horizontal line at the top, indicating that households at the bottom of the income



**Figure 2.** Transition plots by various transition lengths, by historical period and by district.

*Notes:* Estimates are weighted and trimmed using inverse propensity weights that were generated from columns (2) and (5) in Table A1. Only observations with propensity scores above 0.3 are included. The first row in panel (a) presents 1, 5, and 10 year transitions for Stellenbosch, pooled over the entire period covered by the data. Panel (b) shows only five-year transitions for Stellenbosch, but differentiated by period of analysis. Panel (c) shows 5 year transitions for Graaff-Reinet.

distribution remained trapped in the lowest quintile five years later. Some upward mobility is observed for those starting in the middle of the distribution, as well as limited downward mobility for those starting at the top. In contrast, the second and third plots for later periods show that upward social mobility is more likely for those at the bottom of the distribution, and both upward and downward mobility are spread across the distribution. During the third period (1812–1829), greater persistence returns, with those at the very bottom of the distribution being less likely to move up the rankings. Interestingly, those at the top of the distribution also show more persistent outcomes compared to the previous periods. The main takeaway from the baseline results is that calculating income mobility over different periods we find higher persistence in shorter intervals and more mobility at the upper income levels over longer periods.

An important question is whether these elasticities are indeed high and whether Cape settler income was persistent. Comparisons are complicated by two factors: first, most estimates measure inter- rather than intragenerational mobility, and secondly, our unit of analysis is the family farm rather than the individual incomes or occupations followed in most studies (Ward 2023). Nonetheless, we conclude

that our estimates of Cape persistence seem high compared to a range of studies. They are higher than twentieth-century intragenerational persistence estimates ranging from 0.3 to 0.6 (Chetty et al. 2014). They are also substantially higher than those of the four countries analyzed by Fields et al. (2003) in the 1990s, of which South Africa's estimate of 0.44 over 5 transition years was the most persistent. They also seem high compared to the historical intergenerational estimates by Long and Ferrie (2013). They instead seem akin to the high historical intergenerational estimates of Clark and Cummins (2014) who, using surname matches, find much higher intergenerational elasticities of between 0.7 and 0.9 for England over the last 800 years.

Sadly, there are few direct historical comparisons of intragenerational income mobility. Ward (2023), using occupations for the United States, is an exception. He reports the fraction of fathers with the same occupation across two US censuses between 1840 and 1910, ranging from 0.58 in 1850 to 0.51 in 1910. Although these numbers are not directly comparable, they suggest a less mobile society not too dissimilar from what we find in our ten-year transitions.

### **Adding another district**

To examine regional variation in mobility estimates, we compare the Stellenbosch and Graaff-Reinet districts. These two regions had notable differences, of which property rights was the most important (Swanepoel and Fourie 2018). Stellenbosch farms, established in 1679, were smaller in size and predominantly owned by settlers in freehold. In contrast, Graaff-Reinet farms were larger loan farms, mainly dedicated to livestock production. The district was formally established in 1786 and experienced significant geographic mobility in the first two decades as settlers moved to and from the frontier. Factors such as conflicts with the Khoesan and later the amaXhosa, as well as dissatisfaction with VOC colonial rule, contributed to this mobility. Although the announcement of an independent Republic of Graaff-Reinet in 1795 was short-lived due to British intervention, the frontier only stabilized after the second British takeover in 1806. We anticipate a high degree of social mobility in Graaff-Reinet initially, followed by a decline after the turn of the century.

Figure 2(b) and (c) compare the transition probability plots for Stellenbosch and Graaff-Reinet over the three periods. The results support our hypothesis of a significant distinction between the two regions, but only in the first period. Social mobility is low in Stellenbosch, indicated by the white horizontal lines at the top, while Graaff-Reinet exhibits high social mobility. The almost vertical color bars demonstrate that even households lower down the income distribution in the origin year have a significant chance of reaching the top of the income distribution five years later.

The situation changes dramatically in both regions in the second and third periods after the British take-over. Social mobility in Stellenbosch increases, while mobility in Graaff-Reinet declines. In fact, during the second and third periods, social mobility in Stellenbosch (with persistence elasticities of 0.478 and 0.603) surpasses that of Graaff-Reinet (0.730 and 0.645).

This unexpected finding helps us reinterpret the impact of British annexation. First, British arrival in 1795, followed by three years of Batavian rule, contributed to

a period of high mobility in Stellenbosch. By contrast, British rule had little impact on increasing mobility on the frontier. An alternative interpretation is that the higher mobility estimates reflect not an actual increase in mobility, but rather greater measurement error as new officials visited farms or some antagonism towards the British contributed to higher variation in reporting quality. However, considering the results for Graaff-Reinet, where greater antagonism would be expected, this alternative seems unlikely. Second, by the turn of the century, the frontier had closed, and new arrivals to the frontier had similar or even poorer prospects compared to the inhabitants of the more established regions near Cape Town. Third, despite some convergence in mobility estimates, these rankings persisted well into the nineteenth century, with Graaff-Reinet remaining a place of relatively high persistence. This persistence in mobility rankings may explain why, by the early 1830s, groups of frontier farmers began moving further into the South African interior in search of better opportunities. The limited prospects for upward mobility in Graaff-Reinet likely contributed to this outward migration. The comparison of mobility between the Stellenbosch and Graaff-Reinet districts illustrates significant regional differences and shifting patterns over time, highlighting just how important data availability is in shaping our interpretation of historical events on income mobility.

### **Robustness Checks**

#### *Adding zeros*

One potential explanation for the reversal of social mobility estimates between Stellenbosch and Graaff-Reinet is that our results may be influenced by excluding zero-income and zero-asset households. As noted before, young men aged sixteen were recorded as separate households in these censuses, even if they were still living with their parents and did not report separate assets. If there were relatively more of these men on the frontier compared to the interior or specific time periods, it could distort our mobility estimates. To retain zero incomes in the regression samples, they are transformed by the inverse hyperbolic sine transformation instead of the log transformation (Bellemare and Wichman 2020).

The second row of Figure C1 in the online appendix includes zero-income households for both Stellenbosch and Graaff-Reinet to examine the impact. The inclusion of zeros increases the sample size by 23,873 observations in Stellenbosch and 4,263 observations in Graaff-Reinet. Table B1 in the online appendix provides information on the proportion of households reporting no assets or income, ranging from 5% to 30% depending on the period. However, reintroducing the zeroes does not significantly alter our overall observations about household mobility.

The incorporation of zeros leads to a small decrease in persistence elasticities in Stellenbosch. This minor shift could be attributed to young men who initially reported no assets but quickly accumulated their own belongings and incomes as they established separate households from their parents. However, despite this adjustment, the ranking of the two districts remains the same: households in Stellenbosch exhibit slightly less mobility than those in Graaff-Reinet over the entire period.

### *Removing capital goods*

The relatively high persistence we calculate for the Cape Colony may be influenced by the inclusion of returns on capital goods, which are likely to be more persistent across time than income from other streams. As we show in Table 3 and Figure C2 in the online appendix, we find this to be true for Stellenbosch but not for Graaff-Reinet. In Stellenbosch, the income persistence elasticity is 0.774. However, when excluding income from capital goods, specifically the 6% interest earned on the value of horses and slave holdings, the persistence elasticity falls to 0.703. Conversely, for Graaff-Reinet, where incomes are primarily derived from other stock variables such as cattle holdings (contributing around 40% of income), the exclusion of capital goods increases the persistence elasticity.

### *Disaggregating by type of household*

We further analyze the sample by splitting it into households with or without a wife present. It is expected that married households exhibit lower mobility, while single male-headed households, typically starting with limited resources, may experience faster upward mobility. This division also indirectly captures age differences, as more established families are likely to have lower mobility. Household composition and marriage patterns have played a formative role in determining patterns of social mobility at the Cape, particularly for single men (Dooling 2005; von Fintel et al. 2013). Specifically, Roman Dutch inheritance laws favored widows, with sizeable proportions of deceased estates allocated to them. This created significant opportunity for single men to achieve rapid social mobility by marrying wealthy widows.

Our hypothesis is supported in the frontier district of Graaff-Reinet. Single households in Graaff-Reinet demonstrate high income mobility, with an elasticity of 0.583. Married households, on the other hand, exhibit greater persistence, with an elasticity of 0.625. Surprisingly, the opposite pattern is observed for households in Stellenbosch. Figure C3 in the online appendix illustrates that income mobility is higher in Stellenbosch when a wife is present compared to households without a wife. One possible explanation for this anomaly is that young men who were less likely to establish separate farming operations from their fathers were also less likely to marry. Marriage and upward mobility may be closely linked. Additionally, we observe a significant “underclass” of single men with minimal mobility in both the first and last periods. Those seeking better opportunities and facing limited prospects in Stellenbosch may have migrated to the frontier, where single men had greater chances of upward mobility. While we cannot directly match households across districts, this presents a promising avenue for future research.

Another factor to consider is that our matching algorithm may partly contribute to the low mobility observed among single men. As mentioned earlier, our match rates are lowest for single men who subsequently marry. This means we are more likely to select single men who remain single in our sample and underestimate the mobility of single men who marry and likely experience upward mobility. The selective matching process may result in lower estimates of social mobility for single men than what they actually experience. Future improvements to our

matching algorithm, including enhancing match rates for single men and matching across districts, will help clarify the unexpectedly high mobility observed among married households compared to single, male-headed households.

### *Disaggregating by type of production*

We return to our baseline estimates and examine mobility by differentiating between types of production. We categorize them as crop production (wheat, barley, rye, and oats), stock farming (cattle and sheep), and viticulture (wine and brandy). We anticipate higher volatility in the first production type, given the requirement to sow every year and the higher likelihood that returns could be affected by extreme weather. Since most farmers did not specialize in a single type of production, individual households may contribute to all three mobility estimates. Figure C4 in the online appendix presents the results, where we study all families, but limit analyses to incomes associated only with the respective sources.

We find, surprisingly, the lowest income mobility in crop production (with an elasticity of 0.707 in Stellenbosch, as shown in Table 3). This is surprising because crop yields – outputs minus inputs – are dependent on many external factors beyond the farmers' control, including climate shocks. Earlier research has also shown that crop production was most likely to be correlated with ship arrivals in Table Bay (Boshoff and Fourie 2010). Table B1 confirms that around 90% of households in both districts earned no incomes from crops. This emphasizes the unpredictable nature of cultivation and the low rates of engagement in this economic activity.

Stock farming (cattle and sheep) has a lower income persistence elasticity than crop farming. Mobility is highest for viticulture. Again, this is somewhat surprising; many more idiosyncratic factors potentially affect the ownership of cattle and sheep – theft and wild animals, for example – than planting vines. One possible explanation for the high-income persistence of wine and brandy is the low units of account used in reporting wine and brandy output (leaguers, approximately 500 liters), increasing the volatility. This demonstrates yet again the types of measurement error that might contribute to overestimating the extent of mobility in a settler society.

## Conclusions

Measuring social mobility accurately, today and in the past, is necessary for a full understanding of the distributional effects of growth. Historical social mobility estimates – notably those before the twentieth century – have relied almost entirely on occupational classifications found in censuses. Individuals are linked across censuses ten years apart, often at low rates. The measurement error of occupations, and the low matching rates, can potentially bias mobility upwards (Bailey *et al.* 2020).

One way to overcome some of these biases, Ward (2023) shows, is by using an average of multiple observations for the same individual to proxy for their permanent status. Almost all early intergenerational studies use only one observation for the father and one for the son. A single observation may, however, be a noisy signal of an individual's permanent income, especially if social status is measured when



individuals are young and only starting out or old and retiring, rather than in prime age. Such measurement error is likely to increase variation in the incomes of both fathers and sons, variation that will likely decrease persistence elasticities and overstate mobility. Improving measurement, as we have done in this paper, is likely to lower mobility estimates, and reflect more precisely the true persistence of the society.

Estimates of historical intergenerational mobility often still rely on occupational classifications. For the reasons detailed above, this remains a problematic measure. Our use of income, despite its own limitations as detailed above, requires no assumptions about the exact ranking of occupations. Constructing an annual panel allows us to study the life course of a household within a generation, and without the need to link households intergenerationally, where fathers and sons are observed at different ages and eras.

We find high levels of persistence for settler households at the Cape. Although Otto Mentzel visited the Cape about half a century before the time period that our analysis covers, his observation of a highly stratified settler society seems accurate, in particular the existence of a highly persistent colonial elite. Yet our results also question several stylized facts. We find high levels of mobility in the early years of the Cape's Eastern Frontier, but persistence soon set in after British arrival. This contrasts with the more settled regions close to Cape Town where British arrival was associated with greater social mobility, perhaps because of the demise of VOC networks of power and privilege. More work is needed to disentangle who benefited, and why, from this tumultuous period.

The detailed information contained in the tax censuses allows us to test data choices, which is what we believe to be the main contribution of our paper. Exploring income mobility over different periods, our results reveal higher persistence in shorter intervals and more mobility at the upper income levels over longer periods. Further analyses test assumptions around zero-income households, the impact of slave ownership, household head marital status, and type of agricultural production on mobility, indicating high persistence across different conditions and challenging previous understandings of mobility determinants in preindustrial societies.

This is important considering the limitations of current studies that depend largely on US census data, which is only available for every ten years. Our results confirm that using better linkage methods (Bailey et al. 2020; Ghosh et al. 2024) or matching census data to additional sources (Bailey et al. 2023; Buckles et al. 2023) would strengthen the reliability of existing estimates – or perhaps expose inaccuracies. We show that the choice of samples and methods significantly influences the interpretation of historical income mobility. Of course, our results also have limitations too. Another concern Ward (2023) raises is that mobility studies often exclude marginalized groups in society. Most of the intergenerational estimates produced for the United States exclude African Americans. When these subgroups are included, mobility falls further; African Americans, given the various discriminatory laws of the time, were less likely to experience upward mobility, even for those who were geographically mobile (Derenoncourt 2022). Our estimates, sadly, suffer from the same bias; the Khoesan, to a large degree, and the enslaved are excluded from our analysis. There is no doubt that some Khoesan and manumitted slaves are included in

the tax censuses we use, but this is the exception rather than the rule. Our results are therefore likely to be a lower-bound estimate of persistence in the Cape Colony.

There are several extensions that we hope to pursue in future work. We are busy transcribing the tax censuses further back in time; extending the series to the mid-seventeenth century will tell us whether VOC rule was generally characterized by high persistence or whether there was initial mobility that gradually settled into a stratified colony. We would also benefit from adding more Cape districts, and to link households across districts. A third extension would be to add more accurate (and volatile) regional market prices to our calculation of incomes. A final extension is to link the tax censuses intergenerationally. This is possible by linking the tax records to genealogical records, which would provide a more reliable way of matching individuals across generations. All these attempts are likely to reduce measurement error. Given what we have shown in this study, that is likely to further increase our estimates of income persistence. Owing to better data and methods, the Cape – and other settler societies in all likelihood – were far less mobile than previous estimates of social mobility that economic historians have proposed.

**Supplementary material.** The supplementary material for this article can be found at <https://doi.org/10.1017/ssh.2024.24>

**Acknowledgements.** The authors would like to thank Hoyt Bleakley, Rulof Burger, Jeanne Cilliers, Bill Collins, James Fenske, Joseph Ferrie, Igor Martins, Anne McCants, Paul Rhode, Robert Ross, Marianne Wanamaker, Zach Ward, Yiling Zhao, the editors Kris Inwood and Rebecca Jean Emigh, and seminar participants at the University of Michigan, Peking University, University of Southern Denmark, Stellenbosch University, the EHDR seminar and the EHS conference for helpful comments on an earlier version. The authors would like to acknowledge financial support from the Riksbankens Jubileumsfond (Cape of Good Hope Panel project: M20-0041), the Andrew W. Mellon Foundation (Biography of an Uncharted People project) and South Africa's National Research Foundation. For more information about the project, visit [www.capepanel.org](http://www.capepanel.org).

## References

- Abramitzky, Ran, and Leah Boustan** (2017) "Immigration in American economic history." *Journal of Economic Literature* 55 (4): 1311–45.
- Abramitzky, Ran, Leah Boustan, Katherine Eriksson, James Feigenbaum, and Santiago Pérez** (2021b) "Automated linking of historical data." *Journal of Economic Literature* 59 (3): 865–918.
- Abramitzky, Ran, Leah Boustan, Elisa Jácome, and Santiago Pérez** (2021a) "Intergenerational mobility of immigrants in the United States over two centuries." *American Economic Review* 111 (2): 580–608.
- Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson** (2014) "A nation of immigrants: Assimilation and economic outcomes in the age of mass migration." *Journal of Political Economy* 122 (3): 467–506.
- Acemoglu, Daron, Simon Johnson, and James A. Robinson** (2001) "The colonial origins of comparative development: An empirical investigation." *American Economic Review* 91 (5): 1369–1401.
- Adermon, Adrian, Mikael Lindahl, and Márten Palme** (2021) "Dynastic human capital, inequality, and intergenerational mobility." *American Economic Review* 111 (5): 1523–48.
- Adermon, Adrian, Mikael Lindahl, and Daniel Waldenström** (2018) "Intergenerational wealth mobility and the role of inheritance: Evidence from multiple generations." *The Economic Journal* 128 (612): F482–513.
- Aiken, Vera Cardoso Ferreira, João Ricardo Rebouças Dórea, Juliano Sabella Acedo, Fernando Gonçalves de Sousa, Fábio Guerra Dias, and Guilherme Jordão de Magalhães Rosa** (2019) "Record linkage for farm-level data analytics: Comparison of deterministic, stochastic and machine learning methods." *Computers and Electronics in Agriculture* 163: 104857. DOI: [10.1016/j.compag.2019.104857](https://doi.org/10.1016/j.compag.2019.104857).

- Akgün, Özgür, Alan Dearle, Graham Kirby, Eilidh Garrett, Tom Dalton, Peter Christen, Chris Dibben, and Lee Williamson** (2020) "Linking Scottish vital event records using family groups." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 53 (2): 130–46.
- Alesina, Alberto, Sebastian Hohmann, Stelios Michalopoulos, and Elias Papaioannou** (2021) "Intergenerational mobility in Africa." *Econometrica* 89 (1): 1–35.
- Bailey, Martha, Peter Z. Lin, AR Shariq Mohammed, Paul Mohnen, Jared Murray, Mengying Zhang, and Alexa Prettyman** (2023) "The creation of LIFE-M: The Longitudinal, Intergenerational Family Electronic Micro-Database project." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 56 (3): 138–59.
- Bailey, Martha J., Connor Cole, Morgan Henderson, and Catherine Massey** (2020) "How well do automated linking methods perform? Lessons from US historical data." *Journal of Economic Literature* 58 (4): 997–1044.
- Bazzi, Samuel, Martin Fiszbein, and Mesay Gebresilashe** (2020) "Frontier culture: The roots and persistence of 'rugged individualism' in the United States." *Econometrica* 88 (6): 2329–68.
- Becker, Gary S., and Nigel Tomes** (1986) "Human capital and the rise and fall of families." *Journal of Labor Economics* 4 (3, Part 2): S1–39.
- Bellemare, Marc F., and Casey J. Wichman** (2020) "Elasticities and the inverse hyperbolic sine transformation." *Oxford Bulletin of Economics and Statistics* 82 (1): 50–61.
- Blattman, Christopher, Julian Jamison, Tricia Koroknay-Palicz, Katherine Rodrigues, and Margaret Sheridan** (2016) "Measuring the measurement error: A method to qualitatively validate survey data." *Journal of Development Economics* 120: 99–112.
- Boshoff, Willem H., and Johan Fourie** (2010) "The significance of the Cape trade route to economic activity in the Cape Colony: A medium-term business cycle analysis." *European Review of Economic History* 14 (3): 469–503.
- Bruckmeier, Kerstin, Gerrit Müller, and Regina T. Riphahn** (2015) "Survey misreporting of welfare receipt—Respondent, interviewer, and interview characteristics." *Economics Letters* 129: 103–7.
- Buckles, Kasey, Adrian Haws, Joseph Price, and Haley E.B. Wilbert** (2023) "Breakthroughs in historical record linking using genealogy data: The Census Tree Project." National Bureau of Economic Research, No. w31671, DOI: [10.3386/w31671](https://doi.org/10.3386/w31671)
- Card, David, Ciprian Domnisoru, and Lowell Taylor** (2022) "The intergenerational transmission of human capital: Evidence from the golden age of upward mobility." *Journal of Labor Economics* 40 (S1): S39–95.
- Carter, Susan B., and Richard Sutch** (2013) "Why the settlers soared: The dynamics of immigration and economic growth in the 'golden age' for settler societies," in C. Lloyd, J. Metzger, and R. Sutch (eds.) *Settler Economies in World History*. Leiden: Brill: 35–64.
- Chen, Tianqi, and Carlos Guestrin** (2016) "Xgboost: A scalable tree boosting system." *Proceedings of the 2nd ACM SIGKDD international conference on knowledge discovery and data mining*, August 2016. pp. 785–94.
- Chen, Yuyu, Suresh Naidu, Tinghua Yu, and Noam Yuchtman** (2015) "Intergenerational mobility and institutional change in 20th century China." *Explorations in Economic History* 58: 44–73.
- Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez** (2014) "Where is the land of opportunity? The geography of intergenerational mobility in the United States." *The Quarterly Journal of Economics* 129 (4): 1553–623.
- Clark, Gregory, and Neil Cummins** (2014) "Surnames and Social Mobility in England, 1170–2012." *Human Nature* 25 (4): 517–37. <https://doi.org/10.1007/s12110-014-9219-y>.
- Collado, M. Dolores, Ignacio Ortuño-Ortín, and Jan Stuhler** (2023) "Estimating intergenerational and assortative processes in extended family data." *The Review of Economic Studies* 90 (3): 1195–227.
- Collins, William J., Nicholas C. Holtkamp, and Marianne H. Wanamaker** (2022) "Black Americans' landholdings and economic mobility after emancipation: new evidence on the significance of 40 acres." National Bureau of Economic Research, No. w29858. <http://www.nber.org/papers/w29858> (accessed September 27, 2022).
- Collins, William J., and Marianne H. Wanamaker** (2015) "The great migration in black and white: New evidence on the selection and sorting of southern migrants." *The Journal of Economic History* 75 (4): 947–92.

- (2022) “African American intergenerational economic mobility since 1880.” *American Economic Journal: Applied Economics* **14** (3): 84–117.
- Derenoncourt, Ellora** (2022) “Can you move to opportunity? Evidence from the Great Migration.” *American Economic Review* **112** (2): 369–408.
- Dooling, Wayne** (2005) “The making of a colonial elite: Property, family and landed stability in the Cape Colony, c. 1750–1834.” *Journal of Southern African Studies* **31** (1): 147–62.
- Du Plessis, Sophia, and Stan Du Plessis** (2012) “Happy in the service of the Company: The purchasing power of VOC salaries at the Cape in the 18th century.” *Economic History of Developing Regions* **27** (1): 125–49.
- Du Plessis, Sophia, Ada Jansen, and Dieter von Fintel** (2015) “Slave prices and productivity at the Cape of Good Hope from 1700 to 1725: Did everyone win from the trade?” *Cliometrica* **9** (3): 289–330.
- Duin, P. C. van, and Robert Ross** (1987) *The Economy of the Cape Colony in the Eighteenth Century*. Leiden: Centre for the History of European Expansion.
- Engerman, Stanley L., and Kenneth L. Sokoloff** (2012) *Economic Development in the Americas since 1500: Endowments and Institutions*. Cambridge: Cambridge University Press.
- Erikson, Robert, and John H. Goldthorpe** (2002) “Intergenerational inequality: A sociological perspective.” *Journal of Economic Perspectives* **16** (3): 31–44.
- Ferrie, Joseph P.** (2005) “History lessons: The end of American exceptionalism? Mobility in the United States since 1850.” *Journal of Economic Perspectives* **19** (3): 199–215.
- Fields, Gary S., Paul L. Cichello, Samuel Freije, Marta Menéndez, and David Newhouse** (2003) “For Richer or for poorer? Evidence from Indonesia, South Africa, Spain, and Venezuela.” *The Journal of Economic Inequality* **1**: 67–79.
- Fourie, Johan** (2013) “The remarkable wealth of the Dutch Cape Colony: Measurements from eighteenth-century probate inventories.” *The Economic History Review* **66** (2): 419–48.
- (2014) “The quantitative Cape: A review of the new historiography of the Dutch Cape Colony.” *South African Historical Journal* **66** (1): 142–68.
- Fourie, Johan, and Frank Garmon Jr** (2023) “The settlers’ fortunes: Comparing tax censuses in the Cape Colony and early American republic.” *The Economic History Review* **76** (2): 525–50.
- Fourie, Johan, and Erik Green** (2015) “The missing people: Accounting for the productivity of indigenous populations in Cape Colonial History.” *The Journal of African History* **56** (2): 195–215.
- (2018) “Building the Cape of Good Hope Panel.” *The History of the Family* **23** (3): 493–502.
- Fourie, Johan, and Dieter Von Fintel** (2011) “A history with evidence: Income inequality in the Dutch Cape colony.” *Economic History of Developing Regions* **26** (1): 16–48.
- Galenson, David W., and Clayne L. Pope** (1989) “Economic and geographic mobility on the farming frontier: Evidence from Appanoose County, Iowa, 1850–1870.” *The Journal of Economic History* **49** (3): 635–55.
- Ghosh, Arkadev, Sam Il Myoung Hwang, and Munir Squires** (2024) “Links and legibility: Making sense of historical U.S. census automated linking methods.” *Journal of Business & Economic Statistics* **42** (2): 579–90.
- Grusky, David B., and Jesper B. Sørensen** (1998) “Can class analysis be salvaged?” *American Journal of Sociology* **103** (5): 1187–1234.
- Guelke, Leonard** (1976) “Frontier Settlement in Early Dutch South Africa.” *Annals of the Association of American Geographers* **66** (1): 25–42.
- Guelke, Leonard, and Robert Shell** (1983) “An early colonial landed gentry: Land and wealth in the Cape Colony 1682–1731.” *Journal of Historical Geography* **9** (3): 265–86.
- Helgertz, Jonas, Joseph Price, Jacob Wellington, Kelly J. Thompson, Steven Ruggles, and Catherine A. Fitch** (2022) “A new strategy for linking US historical censuses: A case study for the IPUMS multigenerational longitudinal panel.” *Historical Methods: A Journal of Quantitative and Interdisciplinary History* **55** (1): 12–29.
- Kearl, James R., and Clayne L. Pope** (1984) “Mobility and distribution.” *The Review of Economics and Statistics* **66** (2): 192–99.
- Lamar, Howard Roberts** (1981) *The Frontier in History: North America and Southern Africa Compared*. New Haven: Yale University Press.
- Lee, Chul-In, and Gary Solon** (2009) “Trends in intergenerational income mobility.” *The Review of Economics and Statistics* **91** (4): 766–72.

- Long, Jason, and Joseph Ferrie** (2013) "Intergenerational occupational mobility in Great Britain and the United States since 1850." *American Economic Review* **103** (4): 1109–37.
- (2018) "Grandfathers matter(ed): Occupational mobility across three generations in the US and Britain, 1850–1911." *The Economic Journal* **128** (612): F422–45.
- Mann, Ralph** (1984) "Frontier opportunity and the new social history." *Pacific Historical Review* **53** (4): 463–91. <https://doi.org/10.2307/3639416>.
- Massey, Catherine G.** (2017) "Playing with matches: An assessment of accuracy in linked historical data." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* **50** (3): 129–43.
- Mazumder, Bhashkar** (2005) "Fortunate sons: New estimates of intergenerational mobility in the United States using social security earnings data." *Review of Economics and Statistics* **87** (2): 235–55.
- (2016) "Estimating the intergenerational elasticity and rank association in the United States: Overcoming the current limitations of tax data," in L. Cappellari, S. W. Polachek, and K. Tatsiramos (eds.) *Inequality: Causes and Consequences*. Vol. 43. Leeds: Emerald Group Publishing Limited: 83–129.
- McGregor, Craig** (1966) *Profile of Australia*. London: Hodder and Stoughton.
- Mendolia, Silvia, and Peter Siminski** (2016) "New estimates of intergenerational mobility in Australia." *Economic Record* **92** (298): 361–73. <https://doi.org/10.1111/1475-4932.12274>.
- Mentzel, Otto Frederick** (1925) *A Geographical and Topographical Description of the Cape of Good Hope*. Vol. 3. Cape Town: Van Riebeeck Society.
- Morgan, Stephen L.** (2006) "Past themes and future prospects for research on social and economic mobility," in S. L. Morgan, D. B. Grusky, and G. S. Fields (eds.) *Mobility and Inequality: Frontiers of Research from Sociology and Economics*. Stanford: Stanford University Press: 3–20.
- Neumark, S. Daniel** (1957) *Economic Influences on the South African Frontier, 1652-1836*. Miscellaneous Publication (Stanford University. Food Research Institute); 12. Stanford, CA: Stanford University Press.
- Olivetti, Claudia, and M. Daniele Paserman** (2015) "In the name of the son (and the daughter): Intergenerational mobility in the United States, 1850–1940." *American Economic Review* **105** (8): 2695–724.
- Parkerson, Donald H.** (1982) "How mobile were nineteenth century Americans?" *Historical Methods* **15**: 99–110.
- Pérez, Santiago** (2017) "The (South) American dream: Mobility and economic outcomes of first- and second-generation immigrants in nineteenth-century Argentina." *The Journal of Economic History* **77** (4): 971–1006.
- (2019) "Intergenerational occupational mobility across three continents." *The Journal of Economic History* **79** (2): 383–416.
- Price, Joseph, Kasey Buckles, Jacob Van Leeuwen, and Isaac Riley** (2021) "Combining family history and machine learning to link historical records: The Census Tree data set." *Explorations in Economic History* **80**: 101391, DOI: [10.1016/j.eeh.2021.101391](https://doi.org/10.1016/j.eeh.2021.101391).
- Rijppma, Auke, Jeanne Cilliers, and Johan Fourie** (2020) "Record linkage in the Cape of Good Hope panel." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* **53** (2): 112–29.
- Ross, Robert** (1983) "The rise of the Cape gentry." *Journal of Southern African Studies* **9** (2): 193–217.
- Solon, Gary** (1999) "Intergenerational mobility in the labor market," in O. C. Ashenfelter and D. Card (eds.) *Handbook of Labor Economics*. Vol. 3. Amsterdam: Elsevier: 1761–800.
- (2018) "What do we know so far about multigenerational mobility?" *The Economic Journal* **128** (612): F340–52.
- Song, Xi, Catherine G. Massey, Karen A. Rolf, Joseph P. Ferrie, Jonathan L. Rothbaum, and Yu Xie** (2020) "Long-term decline in intergenerational mobility in the United States since the 1850s." *Proceedings of the National Academy of Sciences* **117** (1): 251–58.
- Steckel, Richard H.** (1989) "Household migration and rural settlement in the United States, 1850–1860." *Explorations in Economic History* **26**: 190–218.
- (1990) "Poverty and prosperity: A longitudinal study of wealth accumulation, 1850–1860." *The Review of Economics and Statistics* **72** (2): 275–85.
- Swanepoel, Christie, and Johan Fourie** (2018) "Why local context matters: property rights and debt trading in colonial South Africa." *Studies in Economics and Econometrics* **42** (2): 35–60.
- Van Kerm, Philippe** (2011) "Picturing mobility: Transition probability color plots, United Kingdom Stata User" Group Meetings 2011, Stata Users Group, <https://EconPapers.repec.org/RePEc:boc:usug11:18>, [http://repec.org/usug2011/UK11\\_vankerm.pdf](http://repec.org/usug2011/UK11_vankerm.pdf) (accessed February 28, 2022).

- Von Fintel, Dieter, Sophia Du Plessis, and Ada Jansen** (2013) “The wealth of Cape Colony widows: Inheritance laws and investment responses following male death in the 17th and 18th centuries.” *Economic History of Developing Regions* **28** (1): 87–108.
- Ward, Zachary** (2022) “Internal migration, education, and intergenerational mobility evidence from American history.” *Journal of Human Resources* **57** (6): 1981–2011.
- (2023) “Intergenerational mobility in American history: Accounting for race and measurement error.” *American Economic Review* **113** (12): 3213–48.
- Whitby, Andrew** (2020) *The Sum of the People: How the Census has Shaped Nations, from the Ancient World to the Modern Age*. New York: Basic Books.
- Willebald, Henry, and Javier Juambeltz** (2018) “Land frontier expansion in settler economies, 1830–1950: Was it a Ricardian process?” in V. Pinilla and H. Willebald (eds.) *Agricultural Development in the World Periphery: A Global Economic History Approach*. Cham, CH: Palgrave MacMillan: 439–66.
- Williams, Gavin** (2013) “Who, where, and when were the Cape gentry?” *Economic History of Developing Regions* **28** (2): 83–111.

**Johan Fourie** is professor of economics at Stellenbosch University in South Africa and director of the Laboratory for the Economics of Africa’s Past (LEAP). His book *Our Long Walk to Economic Freedom* was published by Cambridge University Press in 2022. He writes a weekly blog at [www.ourlongwalk.com](http://www.ourlongwalk.com).

**Erik Green** is associate professor in the Department of Economic History at Lund University and a research fellow in the Department of Economics at Stellenbosch University. He is the author of *Creating the Cape Colony: The Political Economy of Settler Colonization* (Bloomsbury Academic 2022).

**Auke Rijpma** is an assistant professor in economic and social history at Utrecht University. He is interested in long-run economic growth, and how this resulted in widely shared gains in wellbeing, and takes a quantitative and computational approach to the historical record to study these questions.

**Dieter von Fintel** is professor of economics at Stellenbosch University in South Africa. He studies historical and contemporary economic development in Africa from a household and firm perspective. His work on South Africa’s Great Divergence was awarded the *Journal of Comparative Economics*’ Montias prize in 2020, while his 2019 study about The Complementarity between Property Rights and Market access in Southern Rhodesia was awarded the Economic History of Developing Regions prize.