The household in a coercive society: Evidence from South Africa's Cape Colony*

Johan Fourie†

Abstract

Household size is a standard demographic measurement, but in economies where the household was also the unit of production it conflates two populations with different determinants. Using 243,598 recorded tax units from the Cape Colony opgaafrollen (1692–1829), we separate the free-family core of each unit from its enslaved, Khoe and servant labour and show that recorded size is a weak proxy for family demography: it correlates with the family core at only 0.55, and ranking districts by household size misranks them on family size. Non-family labour accounts for 82 percent of the variance in recorded unit size and the family core for 18 percent; of the roughly fifteen additional persons between the bottom and top deciles of the within-sheet asset ranking, four in five are non-family. The three non-family labour regimes are distinct: the enslaved, Khoe and servant counts each rise two to three-and-a-half times faster with recorded productive capacity than the family core, but along the different production systems that a biplot typology assigns to districts. Family and labour components are jointly organised: their configuration remains concentrated well past conditional nulls that permute components within fine strata of form, period and productive capacity—even when each unit's family block is held fixed and only the labour components are permuted—in every pooled sample and almost every district-decade cell. Household size in this economy measured labour organisation as much as family demography, and demographic inference from size alone is contaminated wherever coerced labour was held.

Keywords: household composition; household production; coerced labour; slavery; Cape Colony; labour organisation

JEL codes: D13; J12; J47; N37; O12; J43

*I thank the transcription teams of the Laboratory for the Economics of Africa's Past (LEAP) for the opgaafrollen transcriptions. All errors are my own. This paper was created with the help of Anthropic's Claude Code (Opus 4.8 and Fable 5) and OpenAI's Codex (GPT-5.6). Cite this paper as: Fourie, Johan. 2026. "The household in a coercive society: Evidence from South Africa's Cape Colony." Working Paper, Department of Economics, Stellenbosch University.

†LEAP, Department of Economics, Stellenbosch University. Email: johanf@sun.ac.za.

1 Introduction

Household size is usually interpreted as a demographic outcome. Larger households are read as evidence of higher fertility, later household division or wider co-residence of kin. This interpretation fails in economies where the household was also the main unit of production and where labour markets allocated workers into households rather than into firms. Family historians have long held that the pre-industrial domestic group was a unit of production whose members included non-kin labour – live-in servants, and in slaveholding societies enslaved people – and not a kin group alone (Hareven, 1991; Ruggles, 2015). In such economies a recorded household contains two populations with different determinants: a family core, governed by marriage and fertility, and a non-family extension of enslaved, indentured and contracted workers, governed by labour-market institutions. A count of household members mixes the two.

Existing methods do not separate them. Household typologies in historical demography classify households by the kin relations among co-residents (Hammel and Laslett, 1974; Laslett and Wall, 1972), so enslaved and contracted workers either disappear from the classification or force every slaveholding household into a residual category; such household-level measures are moreover sensitive to the demographic regime, which limits their comparison across populations (Ruggles, 2009). Economic models of the household take membership as given and study the allocation of time and consumption within a fixed set of members (Becker, 1965; Chiappori, 1992; Greenwood, Guner and Vandenbroucke, 2017; Doepke et al., 2023). Neither approach treats the set of people in the household as the outcome to be explained.

This paper asks how labour-market institutions affected the number and types of people included in a household production unit. A recent survey of the economics of Cape slavery identifies exactly this—the household-level demography and organisation of coerced labour—as a leading open question, observing that the opgaafrollen record household aggregates rather than individual events (Fourie, forthcoming); we ask what those aggregates, once decomposed, reveal about the productive unit. We study the Cape Colony between 1692 and 1829, an economy in which enslaved people, Khoe workers and European servants were held or contracted by settler households, and in which the boundary of the household therefore depended on the institutions that priced and allocated coerced labour. Our data are the opgaafrollen, the annual tax censuses of the Colony. From a validated transcription of 322,502 tax records across nine districts, we construct an analysis sample of 243,598 recorded units in eight districts for which free children and enslaved people are enumerated on the same form. The forms of 191,692 of these units additionally record servants, and those of 81,326 units record Khoe workers (1805–1829); each analysis states the completeness-restricted subsample it uses. Each recorded unit lists the declarant and, where present, a named woman, counts of free sons and daughters, counts of enslaved men, women, boys and girls, and the unit's livestock, crops, vines and vehicles. We measure the family core as the named adults plus free children, and the non-family extension as the sum of the enumerated labour categories.

We report four sets of results. First, recorded household size is a poor guide to family demography. In the validated sample it correlates with the free family core at only 0.55, so size captures

under a third of the variation in family size; the decompositions that follow attribute the dispersion of size itself overwhelmingly to labour. A reading that treats units of ten or more persons as large extended households is mistaken about most of their members: their mean free family core is under six persons and seven in ten of their members are non-family. Ranking the eight districts by mean household size misranks them on family size (Spearman correlation 0.31): the Cape district ranks third by size but seventh by family core, and Graaff-Reinet sixth by size but second. Demographic inference from size is contaminated wherever coerced labour was held, and that contamination is the paper's subject.

Second, an accounting decomposition. Mean recorded unit size is 6.40 persons, of which the family core contributes 3.06 and the non-family extension 3.34. A Shorrocks decomposition that assigns the covariance gives the non-family extension 82 percent of the variance in unit size and the family core 18 percent, and the split is essentially unchanged within district-years. In levels, of the roughly fifteen additional persons between the bottom and top deciles of the within-sheet asset ranking, twelve—four in five—are non-family.

Third, the association between recorded productive capacity and composition, resolved by labour regime. Within district-years we rank units by their productive assets and estimate Poisson regressions of each composition count on this rank. In the family-slavery sample, moving from the bottom to the top of the asset ranking multiplies the expected family core by 2.6 and the expected enslaved count by 15; the difference between the two gradients is 1.78 (standard error 0.05), stable when any district is excluded, confirmed by a cluster bootstrap, and essentially unchanged when the rank is built from livestock and vehicle stocks alone. On the samples whose forms record them, the servant count rises by a factor of 19 and the Khoe count by 8. The three regimes are not interchangeable: a biplot typology of districts locates slavery in the arable and viticultural core and Khoe labour on the pastoral frontier, so the composition of non-family labour tracks the production system. Because assets and labour are jointly determined, we read the gradients as associations, not causal effects.

Fourth, family and labour components were jointly organised beyond common scaling with recorded productive capacity. We treat composition as a joint distribution over binned states and compare it with permutation nulls of increasing strictness: a product-marginal null that permutes each component independently; conditional nulls that permute within fine strata of form family, period and asset rank; and a sharpest variant that holds each unit's family block fixed and permutes only the labour components, so that a departure cannot arise from family-cycle regularities. Observed composition states are far more concentrated than any of these nulls implies: the conditional null is rejected in every pooled sample and all 50 district-decade cells, and the family-block-fixed variant in every pooled sample—including the strict specification—and in 28 of the 29 family-slavery cells. A Poisson log-linear model locates the departure in the pairwise and higher-order associations of the state table. Settler households did not draw family size and labour holdings independently, and not merely because both scaled with productive capacity.

The interpretation is that the boundary of the household in this economy was an economic

outcome. Productive units expanded through the acquisition of coerced and contracted labour rather than through family size, and the recorded distribution of household compositions reflects that organisation. The result implies that household size in slaveholding and labour-coercive settings measures the scale of the production unit, and that inferences about fertility, kinship or living arrangements from size alone conflate family demography with labour institutions.

The paper makes four contributions. Prior research on household production treats the membership of the household as given and studies allocation within it (Becker, 1965; Chiappori, 1992; Greenwood, Guner and Vandenbroucke, 2017; Doepke et al., 2023); we measure variation in membership itself and show that a widely used demographic measure, household size, is contaminated by labour organisation. Research on the boundary between families and firms analyses which transactions are governed by each institution (Coase, 1937; Ben-Porath, 1980; Pollak, 1985); we provide a historical case in which one recorded unit combined both, and quantify how its boundary varied with productive capacity. The economics of coercion has become an active field, spanning serfdom (Buggle and Nafziger, 2021; Matranga and Natkhov, 2026; Korchmina and Malinowski, 2026), the colonial *mita* (Abad and Maurer, 2025), ancient slavery (Guthmann and Scheidel, 2025), the hiring-out of enslaved workers (Piano and Alvarez, 2025), and the interplay of chattel and state coercion (Saleh, 2024; Acemoglu and Wolitzky, 2011). This work measures the enforcement and incidence of compulsion (Naidu and Yuchtman, 2013; Dari-Mattiacci, 2013), the prices and output of slave economies (Domar, 1970; Fogel and Engerman, 1974), and the value of enslaved people as capital (Martins and Green, 2026); it rarely opens the productive unit to ask how coerced labour of different kinds combined with free family labour inside one recorded household. For the Cape, Fourie and Green (2015) and Links, Fourie and Green (2020) show that indigenous, enslaved and settler-family labour were complementary distinct inputs on the farm; we generalise this from a single frontier district into a colony-wide, multi-regime anatomy of the household, distinguishing the enslaved, Khoe and servant margins by their productive gradients and their place in the production system. Research on household measurement classifies households by kin structure or reduces it to a composite index (Hammel and Laslett, 1974; Ruggles, 1994*b*; Kertzer, 1991; Gruber and Szołtysek, 2016); we treat composition as a joint distribution and test its organisation against a conditional null and a Poisson log-linear model, reporting which composition states are over- and under-represented.

Quantitative work on Cape families has so far relied on linked genealogies to study settler fertility, marriage and wealth (Cilliers, 2019; Cilliers et al., 2025). Those sources follow individuals across their lives but record only the free settler population; the opgaafrollen instead enumerate, in a single cross-section, the enslaved and Khoe labour attached to each unit, which is the margin we measure. We do not link units across years, and we return in Section 7 to what such linkage would add.

Three limitations qualify the results. The opgaaf records administrative tax units, not verified residential households, and we therefore write of recorded units throughout. The data are repeated cross-sections; we do not link units across years and we never observe a household changing compo-

sition. Enslaved and Khoe people did not choose their enumeration, and the composition decisions we model are those of the persons controlling the recorded units.

Section 2 describes the institutional setting. Section 3 states the economic framework. Section 4 describes the data and measurement decisions. Section 5 defines the estimators and the tests of joint organisation. Section 6 reports the results, Section 7 the robustness checks and limitations, and Section 8 concludes.

2 Institutional setting

The Dutch East India Company established a settlement at the Cape of Good Hope in 1652. Company servants released from their contracts became free burghers, received or occupied land, and produced wheat, wine and livestock for the Cape Town market and passing fleets. The Colony expanded eastward and northward through the eighteenth century; by the 1820s it comprised more than a dozen districts under British rule. Farm production remained organised through settler households for the whole period we study (Elphick and Giliomee, 1989; Fourie, 2013).

Three labour institutions supplied workers to these households. First, slavery: from 1658 the Company and private owners imported enslaved people from the Indian Ocean world, and enslaved men, women and children were held as property by settler households until emancipation in 1834–1838 (Worden, 1985; Shell, 1994); on the economics of Cape slavery and the land-abundant, labour-scarce setting that made coerced labour attractive, see Nieboer (1910); Green (2014); Fourie (forthcoming); recent work locates the consolidation of such coercion in frontier political economy as much as in land–labour ratios (Matranga and Natkhov, 2026). Second, Khoe labour: indigenous Khoe pastoralists, dispossessed of land and herds as the Colony expanded, worked on settler farms under verbal and later written contracts; the Caledon Code of 1809 required written contracts and pass documents, and Ordinance 50 of 1828 removed the pass requirement (Newton-King, 1999). Khoe workers were absorbed into settler farm production alongside enslaved and family labour (Fourie and Green, 2015). Third, European servants: knechts contracted from the Company or from other settlers, present in small numbers throughout. In northwestern Europe over the same centuries the comparable labour was supplied by life-cycle service, in which young free workers lived and worked in other households before marriage, an arrangement central to the European marriage pattern (Voigtländer and Voth, 2013); at the Cape the equivalent labour was drawn predominantly from coerced and dispossessed populations. The British abolition of the oceanic slave trade in 1807 ended legal slave imports, and slave prices rose thereafter (Dooling, 2007); where the trade was abolished elsewhere, holders are found to have intensified their coercion of the workforce that remained (Saleh, 2024). These institutions determined the cost, the legal status and the availability of each labour category to each household.

The opgaafrollen are the annual returns collected from household heads for taxation. A return lists the head by name, in most years a named woman where present, counts of sons and daughters, counts of enslaved men, women, boys and girls, in some districts and years counts of Khoe men,

women, boys and girls, counts of servants, and the unit's livestock, grain sown and harvested, vines, wine and vehicles. The rolls have been used to measure Cape production and inequality (Fourie and von Fintel, 2010); we use them to measure household composition. Because the returns served tax assessment, the recorded unit is an administrative object: it is the set of people and assets declared by one head. We present evidence in Section 4 on when this object can be read as a household production unit.

3 Economic framework

We state a framework in which the composition of the recorded unit is a choice made under institutional constraints. The framework fixes terms and generates the predictions we test; it is not a structural model to be estimated.

A unit i in district d and year t has a family core F_i, productive assets A_i, unit-specific productivity θ_i and a vector of non-family labour $L_i = (E_i, K_i, S_i)$, where E, K and S denote enslaved, Khoe and servant counts. The institutional environment I_{dt} determines the monetary, legal and enforcement costs of holding each labour category. The person controlling the unit selects a composition state $z_i = (F_i, L_i)$ from a finite set of states to maximise

$$U_i(z) = R(z;\, A_i, \theta_i) - C(z;\, I_{dt}) + \varepsilon_{iz}, \tag{1}$$

where R is the return from operating with composition z, C is the cost of that composition under the institutional environment and ε_{iz} captures unobserved unit-specific factors. The formulation does not assume that the people counted in L_i chose to join the unit; the choice belongs to the unit head, and for enslaved people it was exercised through purchase and inheritance.

The household boundary is endogenous in this framework in a specific sense: L_i is determined jointly with production decisions rather than fixed by family demography. That household membership responds to economic conditions has precedent outside this setting: nineteenth-century intergenerational coresidence in the United States tracked the productive assets of the household rather than old-age dependency (Ruggles, 2011), coresidence insures against labour-market risk (Kaplan, 2012), and the boundary of the family farm contracts and expands with land pressure (Guirkinger and Platteau, 2015). This is a statement about the economic problem, not about econometric endogeneity in a regression, and we test its implications with associations rather than causal designs.

Making the timing explicit sharpens two of the predictions. The family core F_i is set first, by marriage, fertility, survival and household division, and is bounded above by demography. Taking F_i as given, the head then chooses labour $L_i = (E_i, K_i, S_i)$ to maximise the operating surplus $R(A_i, F_i, L_i; \theta_i) - \sum_j p_{jdt} L_j$, where p_{jdt} is the institutional price of regime j in I_{dt}. Assume the return R is supermodular in (L_i, A_i): the marginal product of each labour category rises with assets, and the labour categories do not substitute for one another strongly enough to overturn that complementarity. Under this assumption every component of the optimal labour demand is

nondecreasing in assets, by monotone comparative statics (Milgrom and Shannon, 1994); without it, a rise in assets could raise one labour category while displacing another, so the cross-category restriction is substantive, and Section 6 reports the weakly positive conditional associations among the categories that it requires. Boundedness of the family core yields only an accounting restriction: because F_i is bounded by demography while L_i is not, recorded size beyond the feasible family range must be carried by non-family labour. It does not by itself imply that labour carries more of the variance of size, or a steeper association with recorded scale; those are empirical predictions, and Section 6 tests them. The three prices differ in kind—enslaved people were property, priced by purchase and inheritance, while Khoe and servant labour were contracted—and how coercion selects among these labour forms, and between coerced and free labour, is itself an economic question (Dari-Mattiacci, 2013; Rönnbäck, 2024), so the model does not require the regimes to respond alike, and Section 6 shows they do not.

The framework yields four predictions; the first follows from the assumptions just stated, and the fourth is the empirical counterpart of the accounting restriction. First, conditional on the family core, units with larger A_i select larger L_i. Second, the components of z are jointly selected, so their observed joint distribution differs from the distribution generated by assigning each component independently. Third, changes in I_{dt} alter the prevalence of composition states. Fourth, variation in L_i accounts for more of the variation in total unit size than variation in F_i. Predictions one, two and four are testable in repeated cross-sections and are tested below. Prediction three requires an identification design that separates institutional change from changes in the tax forms; we report the relevant comparisons as descriptive.

4 Data and measurement

4.1 Source and validated release

We use a complete transcription of the opgaafrollen for nine districts: Cape (1660–1825), Stellenbosch (1685–1844), Swellendam (1751–1838), Graaff-Reinet (1788–1828), Tulbagh (1805–1822), Uitenhage (1805–1825), Albany (1821–1825), Worcester (1818–1840) and George (1811–1825). The validated release contains 322,502 tax units from 369 annual district sheets. Every record carries provenance to its source workbook, sheet and row; seventeen automated validation tests, covering identifier uniqueness, nonnegative and integer person counts, and the exclusion of summary rows, all pass. Blank cells are preserved as missing rather than converted to zero.

4.2 Analysis sample

The analysis requires that the components of household composition are recorded on the same form. We audited every sheet for the structural availability of each field. Free sons and daughters and the four enslaved categories are jointly recorded on 263 sheets covering 243,635 units in eight districts between 1692 and 1829; this is the primary sample. Servants are additionally recorded on the forms of 191,692 of these units, and the four Khoe categories on the forms of 81,326 units

in 1805–1829. Structural recording does not guarantee complete composition under the blank-cell rules below: complete validated-zero composition including servants is available for 100,058 units and including Khoe workers for 24,058, and these are the samples of the corresponding state tables, while the regime gradients use the units with recorded counts and usable asset ranks (59,949 and 35,324; Table 3). We exclude 37 units that aggregate more than ten subordinate rows of the source register, because inspection shows them to be administrative blocks rather than declarations by one head; the largest such block chains 582 rows. The primary sample therefore contains 243,598 units.

4.3 Composition measures

The family core is the number of named adults plus the counts of free sons and daughters. A named man appears on 82 percent of declarations; a named woman appears on 52 percent. An audit of name fields by form family shows that in six of the twelve form families a named woman almost always appears as a single name alongside a named man, consistent with a spouse; in the remainder we treat the field only as an indicator that an adult woman was recorded. Because these name fields are administrative, Section 7 re-estimates the headline measurement results with a children-only family core and on the spouse-validated form families alone; the correlations, variance shares and district misrankings are unchanged. The non-family extension is the sum of enslaved, Khoe and servant counts, using only the categories recorded on the relevant form; we never impute a category that a form does not contain. Recorded unit size is the sum of the family core and the non-family extension.

Blank cells require a rule. Transcription conventions differ sharply across districts: on Graaff-Reinet forms enumerators wrote explicit zeros, so a blank cell there is not interpretable as zero, while on Albany, Cape and George forms explicit zeros are essentially absent and blanks are the only way a zero could appear. We therefore classify each form-by-field cell by an audited rule and construct every composition variable under two specifications: a strict specification in which blanks remain missing, and a validated-zero specification in which a blank is set to zero only where the field is recorded on the form and the form family transcribed essentially no explicit zeros. The validated-zero specification yields 214,687 units with complete composition; the strict specification yields 42,103. All results are reported under both.

4.4 Productive capacity

The returns enumerate livestock, grain, vines, wine, brandy and vehicles, but the set of asset fields changes across forms. We therefore avoid a pooled asset index. Within each district-year sheet we sum each of four asset domains (livestock, cultivation, viticulture, vehicles), rank units within the sheet in each domain and average the available domain ranks into a within-sheet percentile rank. The rank is an ordinal measure of recorded productive scale relative to units assessed on the same form in the same year; we do not describe it as wealth, and because the cultivation and viticulture domains partly record realised output rather than predetermined stocks, Section 7

re-estimates every gradient with a stock-only rank built from livestock and vehicles and with each domain excluded in turn.

5 Empirical method

5.1 Decomposition and count models

Recorded unit size satisfies the identity $\text{size}_i = F_i + N_i$, where N_i is the non-family count, the sum of the components of L_i recorded on the relevant form. The variance decomposition

$$\operatorname{Var}(\text{size}) = \operatorname{Var}(F) + \operatorname{Var}(N) + 2\operatorname{Cov}(F, N) \tag{2}$$

attributes the dispersion of unit size to its components. Because the covariance is positive the two variance shares in equation (2) do not sum to one, so we also report the Shorrocks natural decomposition, which assigns each component the share $\operatorname{Cov}(\cdot, \text{size})/\operatorname{Var}(\text{size})$ and sums to one, a within/between district-year split of the variance, and a Kitagawa decomposition of the gradient in persons that divides the top-minus-bottom asset-decile difference in mean size into a family and a non-family part. Decomposing an aggregate household measure into populations with distinct determinants, rather than reading the aggregate directly, has repeatedly overturned demographic conclusions, including about the structure of African-American families and its roots in slavery (Ruggles, 1994*a*).

For the productive-capacity gradient we estimate Poisson pseudo-maximum-likelihood regressions (Santos Silva and Tenreyro, 2006)

$$\mathbb{E}[\, y_i \mid r_i \,] = \exp(\beta\, r_i + \delta_{dt}), \tag{3}$$

where y_i is the family core or the non-family count, r_i is the within-sheet asset rank and δ_{dt} are district-year fixed effects. The coefficient β is the semi-elasticity of the expected count with respect to the full range of the asset distribution. To test whether the two gradients differ we stack the two outcomes and estimate the interaction of r_i with a non-family indicator, with district-year-by-outcome fixed effects. We estimate the same model separately for the enslaved, Khoe and servant counts on the samples where each is recorded. For the non-family share of unit size we estimate a fractional logit (Papke and Wooldridge, 1996), reported as an average partial effect. The within-sheet rank averages the available domain percentile ranks, with ties at the mid-rank and endpoints at zero and one, so $\exp(\beta)$ is the ratio of expected counts between a unit at the bottom and one at the top of its sheet's realised asset support. Standard errors are clustered by district-year sheet throughout; there are 260 such clusters (median 455 units, 3.8 percent below thirty), so cluster-robust inference is well-behaved, and we confirm the gradient difference with a sheet-level cluster bootstrap in addition to the leave-one-district-out check. Of the 214,687 complete-composition units, 67,923 record no asset field at all and so carry no within-sheet rank—rows concentrated in the early Cape and Stellenbosch form families—leaving a gradient sample of 146,764 units, of which

the estimator drops one singleton fixed-effect cell.

5.2 The joint distribution of composition

We represent composition as the joint distribution over composition states, without imposing a household typology. Each component count is assigned to one of five bins (0, 1, 2–3, 4–7, 8+), fixed before estimation, and a composition state is the vector of binned components together with the named-woman indicator. We retain a state when at least five units occupy it; its weight is the number of occupying units. This is a coarsened contingency table of the composition components, not a household typology: no state is privileged, and no unit is observed twice, so the distribution is cross-sectional.

We build four nested state tables: family-only (named woman and children; 224,743 units), family-slavery (adding the enslaved bin; 214,687 units), family-slavery-servant (adding servants; 100,058 units) and full labour (adding Khoe workers; 24,058 units). Tables with more components have mechanically larger state spaces, so we compare each only with its own null distribution.

For each we compute the number of occupied states and the entropy of the state distribution. The entropy is a statistic of the cross-classified composition counts, and its departure from independence is equivalently a total correlation, or multi-information, among the components; we report that interpretable magnitude alongside the tests.

5.3 Null models and validation

We compare each statistic with permutation nulls of increasing strictness. The *product-marginal* null permutes each composition component independently across units, preserving every marginal distribution and removing all within-unit association; it tests mutual independence. Because both margins load on productive capacity, common scaling with assets alone would reject it. The *conditional* nulls therefore permute each component within strata that hold recorded scale fixed: the baseline conditions on form family and asset-rank decile; a tightened variant conditions on form family, decade and asset strata that split the top decile, and since form families are district-specific this conditions on district and period; and the sharpest variant holds each unit's family block—the named-woman indicator and the children bin—fixed and permutes only the labour components, so that a departure cannot arise from family-cycle regularities; in tables with a single labour component this isolates the association between labour and family structure, while in the wider tables it may also reflect association among the labour categories themselves. A departure from these nulls is evidence of joint organisation beyond the conditioning strata; residual variation within strata could in principle still contribute, so we do not claim to exclude every common-cause account. We build each null on 1,000 draws for the pooled tables and 200 per district-decade cell, and rarefy cells to a common size. As a conventional benchmark we also fit Poisson log-linear models to the state table (Agresti, 2013), comparing mutual independence, all two-way associations and higher-order interaction by likelihood-ratio tests, and read cell residuals to identify over- and under-occupied

composition states. Sampling uncertainty is measured by 500 bootstrap draws and the stability of the retained states by split-sample validation (Gotelli and Colwell, 2001).

6 Results

6.1 Recorded size is a poor guide to family demography

We begin with the measurement claim, because it is the paper's reason for decomposing size at all. In the validated-zero sample the recorded unit size correlates with the free family core at only 0.55, so size captures under a third of the variation in family size; that the dispersion of size itself is carried by labour is established by the decompositions of the next subsection, not by this correlation. The error a size reading induces is large where it matters most. Of the 47,910 units that a naive reading would classify as large or extended households, with ten or more members, the mean free family core is 5.78 persons, seven in ten of their members are non-family, and only three percent hold no enslaved person; they are large because they held labour, not because they contained large families. The contamination also distorts comparison across space. Ranking the eight districts by mean recorded household size and by mean family core gives orderings that barely agree (Spearman correlation 0.31): the Cape district ranks third of eight by household size but seventh by family core, and Graaff-Reinet sixth by size but second by family core (Figure 1). A demographer who read household size as family demography would infer high fertility or wide co-residence in the arable and viticultural districts, where the excess is enslaved labour, and would understate it on the pastoral frontier. None of this depends on how the family core is measured: with a children-only core, or restricted to the form families where the named woman is validated as a spouse, the correlation stays between 0.52 and 0.56 and the district rankings still disagree (Section 7). Wherever coerced labour was held, an unadjusted count of household members is not a demographic statistic.

6.2 Composition of recorded unit size

Table 1 reports the decomposition. It is computed on the family-slavery sample, where the non-family extension is the enslaved count alone; the wider labour categories are added below on the samples whose forms record them. Mean recorded unit size in the validated-zero specification is 6.40 persons: 3.06 in the family core and 3.34 in the non-family extension. The variance of unit size is 72.9; the non-family extension accounts for 74.9 percent of it, the family core for 10.2 percent and twice the covariance for 14.9 percent. The distribution of unit size has a long upper tail (the 90th percentile is 16 persons and the 99th percentile 39), and the tail consists mainly of non-family labour: among units in the top decile of size, 76.9 percent of enumerated persons are non-family. The strict specification gives the same allocation (78.0 percent non-family, 6.9 percent family core) on the smaller complete-case sample. Because the components covary positively, these shares leave the covariance unassigned; a Shorrocks decomposition that distributes it gives the non-family extension 82 percent of the variance in unit size and the family core 18 percent. Because

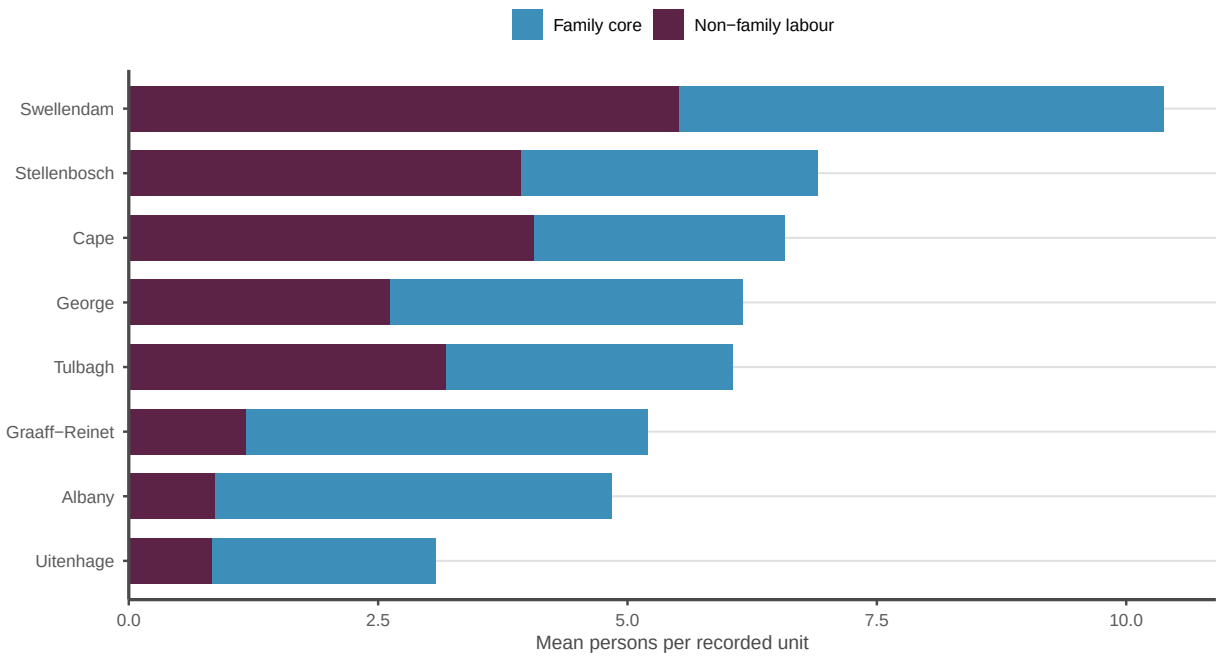

Figure 1: Mean recorded unit size by district, decomposed into the free family core and non-family labour, validated-zero specification. Ranking districts by total size does not rank them by family size.

the extension here counts only the enslaved, 82 percent is a floor: on the servant sample, where the extension adds servants, the Shorrocks non-family share is 84 percent, and on the full-labour sample, where it adds Khoe workers, 97 percent—the last on the labour-intensive 1805–1829 forms. The split is not an artefact of pooling across forms and periods: computed within district-year sheets, the non-family extension still carries 74.7 percent of the within-cell variance against 9.9 percent for the family core, and 95.6 percent of the total variance is within, not between, district-years. Nor is it an artefact of the largest declarations: excluding the flagged multi-row units leaves the shares unchanged, and excluding the top percentile of unit size still leaves the extension with 75 percent (Section 7). Decomposed in levels rather than variances, the gradient of size on productive capacity is almost entirely a labour gradient: of the 15.4 additional persons per unit between the bottom and the top asset decile, 12.2 (four in five) are non-family and 3.2 are family.

Figure 2 plots the two distributions. The family core is concentrated: most units contain between one and five family members, and the density above eight is negligible. Recorded unit size tracks the family core up to about four persons and then departs from it, with substantial mass at every size up to 40. Units of the same size differ in composition: of units with exactly six members, 43.9 percent contain no enslaved person; of units with ten members, 15.1 percent contain none and 46.7 percent are majority non-family. Between sizes six and ten, the mean number of free children rises from 2.8 to 4.1 while the mean number of enslaved people rises from 1.6 to 4.2. Size and family structure are different variables.

Table 1: Decomposition of recorded unit size

	Validated zero	Strict
Units	214,687	42,103
Mean unit size	6.40	10.17
Family core	3.06	4.72
Non-family (enslaved)	3.34	5.45
Variance share (percent): family core	10.2	6.9
Variance share (percent): non-family	74.9	78.0
Variance share (percent): covariance	14.9	15.1
90th percentile of unit size	16	24
99th percentile of unit size	39	52
Non-family share of persons, top size decile	0.77	0.79

Notes: Family-slavery sample, 1692–1829, eight districts; the non-family extension in this sample is the enslaved count (the text reports the decomposition with servants and Khoe workers added on the samples whose forms record them). Validated zero: blanks set to zero only on audited form families; strict: blanks remain missing and units enter only with complete composition. Variance shares follow equation (2). The strict sample is smaller and its units are larger on average because complete rows are concentrated on form families whose enumerators recorded explicit zeros.

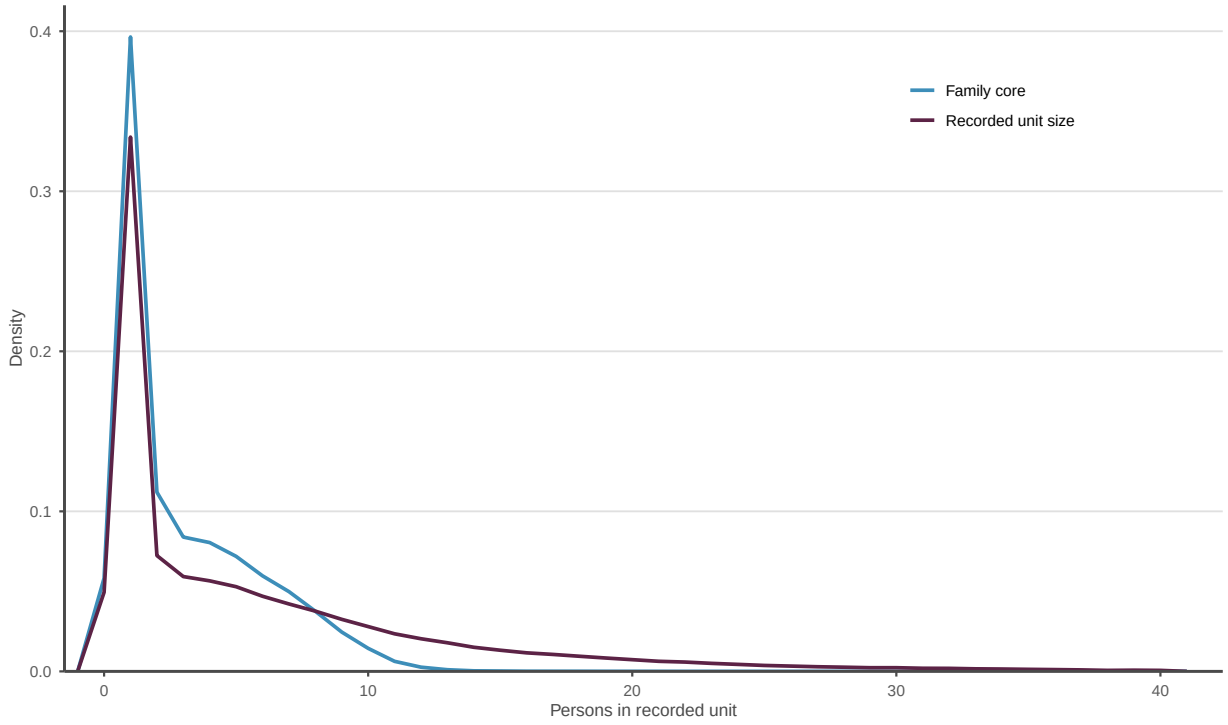

Figure 2: Recorded unit size and family core, validated-zero specification. Densities over persons per unit, truncated at 40 for display.

6.3 Productive capacity and the household boundary

Table 2 reports the count models. Moving from the bottom to the top of the within-sheet asset rank is associated with a family core larger by a factor of $\exp(0.954) = 2.6$ and a non-family count larger by a factor of $\exp(2.737) = 15.4$. Per decile of the asset distribution, the expected family core is 10.0 percent larger and the expected non-family count 31.5 percent larger. The stacked model estimates the difference between the two semi-elasticities at 1.783 (standard error 0.045). The fractional logit for the non-family share gives a coefficient of 2.640 (0.060): higher-capacity units held a larger share of their members outside the family. The strict specification yields the same ordering (0.893 against 2.080). Figure 3 shows the same result without a model, as mean counts by asset-rank vigintile.

Because assets and labour composition are jointly determined, these are conditional associations. Their size is the point: the non-family margin is about three times as responsive to productive capacity as the family margin, which is what prediction one and prediction four of the framework require if households expanded production through labour rather than through family size.

Table 2: Productive-capacity gradients

	Validated zero			Share	Strict	
	Family	Non-family	Stacked	Share	Family	Non-family
Asset rank	0.954	2.737	0.954	2.640	0.893	2.080
	(0.021)	(0.052)	(0.021)	(0.060)	(0.079)	(0.139)
Asset rank × non-family			1.783			
			(0.045)			
Observations	146,763	146,763	293,526	141,122	41,490	41,490

Notes: Columns 1–2, 5–6: Poisson pseudo-maximum likelihood with district-year fixed effects. The stacked column reports one model of both outcomes with district-year-by-outcome fixed effects; its top coefficient (0.954) is the family-core gradient and the interaction row (1.783) is the difference between the non-family and family gradients, precisely estimated and confirmed by a sheet-level cluster bootstrap (95 percent interval 1.70 to 1.88). The share column is a fractional logit for the non-family share of unit size (average partial effect 0.44); it excludes the 5,641 units of recorded size zero, for which the share is undefined. Asset rank is the within-sheet percentile rank of productive assets. Standard errors clustered by district-year sheet in parentheses; there are 260 clusters. All coefficients significant at the 0.1 percent level.

6.4 The three non-family labour regimes

The non-family extension is not one thing. Table 3 estimates the asset-rank gradient separately for each labour regime on the sample where it is recorded; the samples differ by form and period, so the columns describe each regime on its own coverage rather than one common population. All three counts rise far faster with recorded productive capacity than the family core, whose expected size grows by a factor of 2.4 in its own column: the enslaved count grows by a factor of 15, the servant count by 19 and the Khoe count by 8, semi-elasticities between two and three-and-a-half times the family core's. Net of district-year fixed effects and the asset rank, the residual pairwise correlations among the family core and the three labour counts are all weakly positive, between

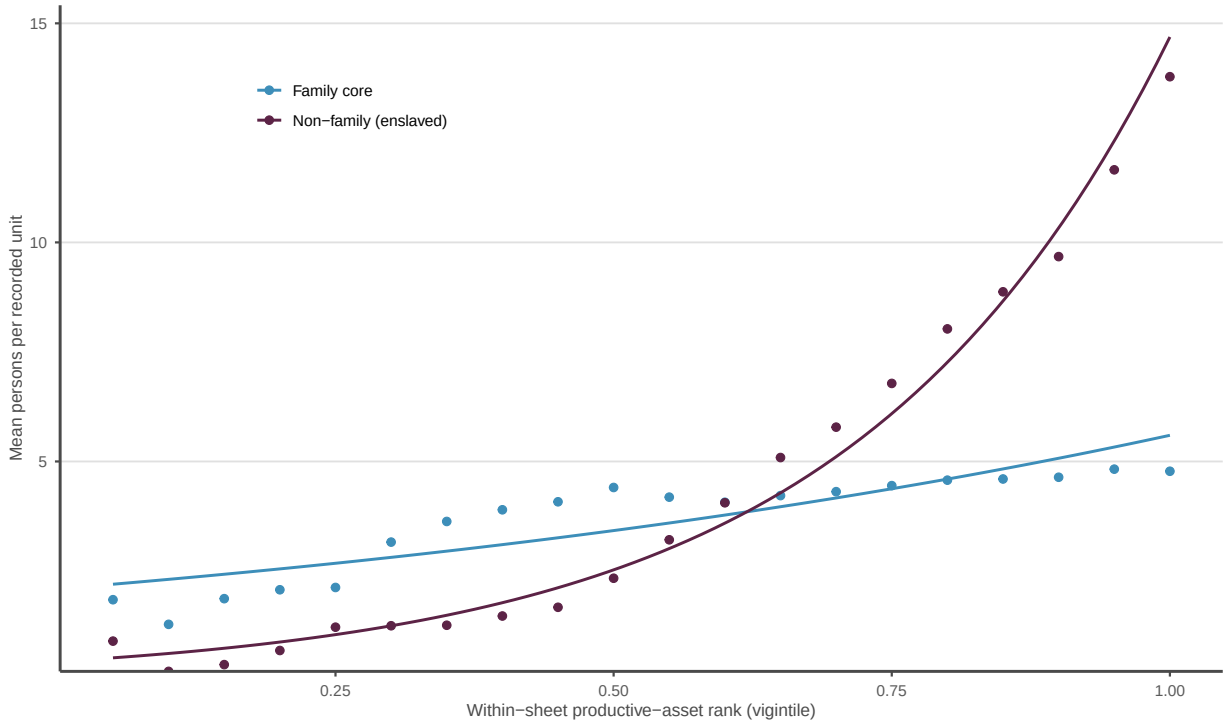

Figure 3: Mean family core and non-family count by vigintile of the within-sheet asset rank, validated-zero specification, with Poisson fits.

0.02 and 0.12, so no negative trade-off between labour categories is apparent at a given recorded scale. Such co-occurrence is evidence against substitution in composition; it is not itself an estimate of production technology. The production-function evidence is prior work: for one frontier district, Links, Fourie and Green (2020) estimate the substitutability of slave, Khoe and settler-family labour and find them complements rather than interchangeable inputs (see also Fourie, forthcoming). The composition of units across the colony is consistent with that finding at the level where household size is recorded.

Where the regimes differ is in the production system they served. We summarise each district's production with the calibrated biplot of Fourie et al. (2026): a correlation-components analysis of the recorded livestock, grain, viticulture and labour holdings that places districts on a pastoral-to-viticultural axis (Figure 4). Slavery concentrates in the arable and viticultural core of Stellenbosch and the Cape district, while Khoe labour concentrates on the pastoral frontier of Graaff-Reinet and Uitenhage. Interacting the asset gradient with a district's position on this axis, the enslaved gradient is if anything steeper towards the frontier (interaction 0.21, standard error 0.13) while the total coerced-labour gradient is flatter there (-0.21, standard error 0.12): frontier productive capacity was extensive, drew less labour per unit of assets, and drew Khoe rather than enslaved workers. The composition of coerced labour, and not only its quantity, tracked the production system—the multi-regime structure that a single "non-family" count hides.

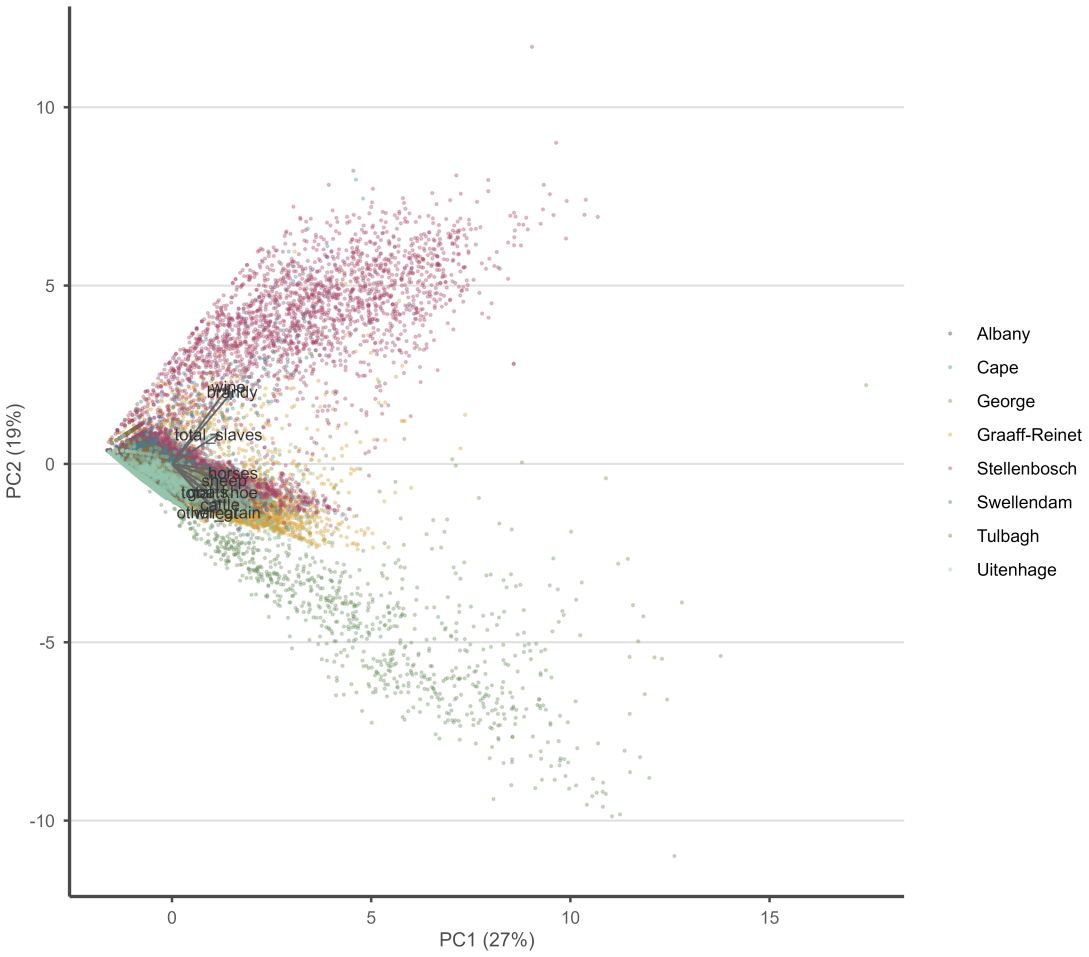

Figure 4: Correlation-components biplot of district production systems, 1820s cross-section, following the calibrated-biplot method of Fourie et al. (2026). Points are recorded units coloured by district; calibrated axes are the production variables. Districts separate along a pastoral (sheep, goats) to viticultural (wine, brandy, slaves) axis.

Table 3: Productive-capacity gradients by labour regime

	Family core	Enslaved	Servant	Khoe
Asset-rank semi-elasticity	0.893	2.709	2.957	2.134
	(0.024)	(0.051)	(0.133)	(0.104)
Factor, bottom to top	2.4	15.0	19.2	8.5
Observations	156,300	147,734	59,949	35,324

Notes: Poisson pseudo-maximum likelihood with district-year fixed effects, validated-zero specification; standard errors clustered by district-year sheet in parentheses. Each column is estimated on the units whose forms record that count and that carry a usable asset rank: the family core wherever it is complete (156,300), the enslaved count on the family-slavery forms (147,734), servants on the servant-recording forms (59,949) and Khoe workers on the Khoe-recording forms of 1805–1829 (35,324). These samples exceed the complete-composition state-table samples of Table 4 because a single count can be complete where the full composition vector is not; the family-core column also exceeds Table 2's 146,763 because Table 2 additionally requires the enslaved count. All coefficients significant at the 0.1 percent level.

6.5 Joint organisation of family and labour

Figure 5 shows the family-slavery composition-state distribution for the largest single-form sample, Stellenbosch in the 1820s (17,596 units), as a heatmap: each cell is a composition state defined by its children and enslaved bins, faceted by the named-woman indicator, shaded by the mean asset rank of the units in it. Forty-eight of the fifty possible states are occupied. Mean asset rank rises along the enslaved dimension at every level of the children dimension, which restates the productive-capacity gradient at the level of composition states.

Table 4 compares the pooled state distributions with both nulls. The two low-dimensional tables occupy their full state space, so the information is in the weights: observed entropy is 1.83 against a product-marginal null mean of 1.99 in the family-only table and 2.93 against 3.19 in the family-slavery table. As z-scores these departures are enormous, but only because the product-marginal null is nearly degenerate; the interpretable magnitude is the total correlation among the components, 0.26 nats, or eight percent of the summed marginal entropy, which leaves 18.8 configurations occupied in effect against 24.4 under independence. The decisive comparisons are with the conditional nulls. Under the baseline conditioning on form family and asset decile, observed entropy departs at the smallest attainable randomisation p (0.001, one draw in a thousand) in every pooled table and in all 50 district-decade cells. Tightening the strata to form family, decade and asset bins that split the top decile changes nothing. The sharpest variant holds each unit's family block fixed and permutes only the labour components, so that family-cycle regularities cannot contribute; observed entropy still departs at $p = 0.001$ in all three pooled labour tables, on the strict family-slavery sample, and in 28 of the 29 family-slavery district-decade cells (Table 4). In the family-slavery table, where the enslaved count is the only labour component, this isolates the association between labour and family structure within district, period, form and fine recorded-scale strata; common causes varying within those strata cannot be excluded. A Poisson log-linear benchmark locates the departure precisely: adding pairwise associations to mutual independence

cuts the deviance of the state table from 111,300 to 3,298, and higher-order interaction removes the remainder (both likelihood-ratio $p < 10^{-15}$). The most over-occupied states are coherent large households—a recorded spouse with many children and many enslaved people—and the most under-occupied are their incongruous opposites. Households combined family and labour in systematically fewer configurations than independent assignment would produce, beyond what the form, period and recorded-scale strata explain; this is prediction two of the framework.

State-level asset ranks confirm where production concentrated: across the three labour tables, the correlation between a state's mean asset rank and its labour bins is 0.82 to 0.88, against 0.21 to 0.40 for its children bin.

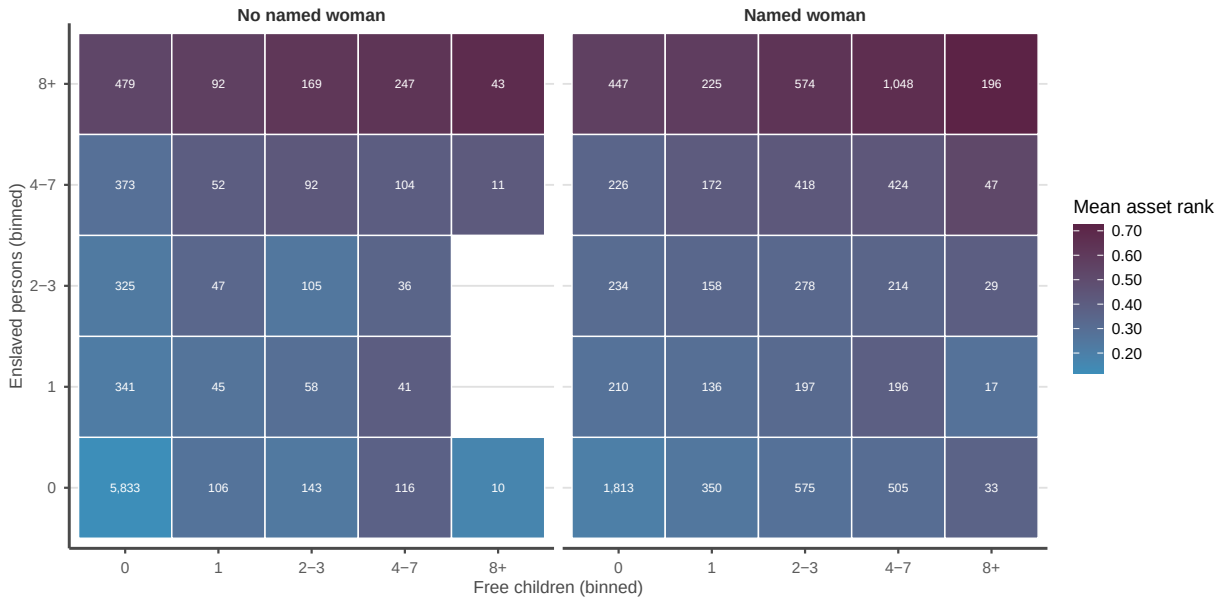

Figure 5: Composition states of the family-slavery sample, Stellenbosch 1820s (17,596 units). Each cell is a composition state (children bin × enslaved bin), faceted by the named-woman indicator; the label is the number of units in the state and the shading is the mean within-sheet asset rank. Only states with at least five units are shown.

6.6 The 1808 abolition: a descriptive audit

Prediction three holds that a change in the institutional environment alters the prevalence of composition states. The clearest such change is the British abolition of the oceanic slave trade, effective at the Cape from 1808, which closed the import margin so that the enslaved population could grow only by natural increase. We cannot identify its effect, because the tax forms change around the same dates, but we can ask descriptively whether the age and sex composition of holdings shifted as imports gave way to reproduction—the demographic footprint that creolisation would leave (Shell, 1994). Within the one Stellenbosch form family that spans 1808 with substantial numbers on each side (3,189 units before, 5,704 after), the child share of the enslaved rises by 5.4 percentage points and the female share by 4.5 across the break, while the free family core, a

Table 4: The joint distribution of composition states against the product-marginal and conditional nulls

	Family only	Family-slavery	+ servants	Full labour
Units	224,743	214,687	100,058	24,058
Components in state	2	3	4	5
Occupied states (threshold 5)	10	50	146	267
product-marginal null mean	10.0	50.0	155.0	264.8
product-marginal p	1.000	1.000	0.005	0.691
conditional null mean	10.0	50.0	155.5	272.3
conditional p	1.000	1.000	0.003	0.176
Entropy	1.831	2.934	3.074	3.698
product-marginal null mean	1.988	3.193	3.400	4.160
product-marginal p	0.001	0.001	0.001	0.001
conditional null mean	1.966	3.138	3.337	4.003
conditional p	0.001	0.001	0.001	0.001
family-block-fixed null mean	—	3.009	3.191	3.879
family-block-fixed p	—	0.001	0.001	0.001
Split-sample state reproduction	1.00	1.00	0.92	0.92

Notes: Pooled state distributions, validated-zero specification, threshold five units. The product-marginal null (1,000 draws) permutes each component independently, testing mutual independence. The conditional null (1,000 draws) permutes each component within strata of form family and asset-rank decile. The family-block-fixed null (1,000 draws) uses tightened strata (form family, decade, and asset bins that split the top decile), holds each unit's named-woman indicator and children bin fixed, and permutes only the labour components; it has no analogue in the family-only table, and on the strict sample it also rejects at $p = 0.001$ (observed entropy 2.978 against a null mean of 2.998). Units without an observed asset rank enter the strata at a common imputed value; assigning them an explicit missing stratum, or excluding them, leaves the family-slavery rejection unchanged at $p = 0.001$. Randomisation p is two-sided under the (1+#)/(draws+1) convention, so 0.001 is the smallest attainable value. Split-sample reproduction is the share of occupied states defined on a random half that reappear in the other half, over 50 splits.

negative control, is flat (Figure 6). The shift is real but not sharp: placebo breaks at 1798 and 1818 give changes of similar size, so what we observe is a gradual movement toward children and women across the period rather than a discontinuity at abolition. The opgaaf records no birthplace, so reading this as creolisation is an inference that runs through the closed import margin rather than direct observation; it is, however, the same post-ban shift that Martins (2026) documents independently with a fixed-effects panel on the opgaaf, as natural increase replaced imports once the trade closed. Our before-and-after, more cautious about dating the shift sharply to 1808, is its household-composition counterpart, though we read the recorded unit as one of labour and production where Martins (2026) emphasises the enslaved as capital. This is consistent with the framework's third prediction and with the historiography, and it is as much as repeated cross-sections, with forms that change alongside the institutions, can establish; a design that separates institutional from recording change is left to subsequent work.

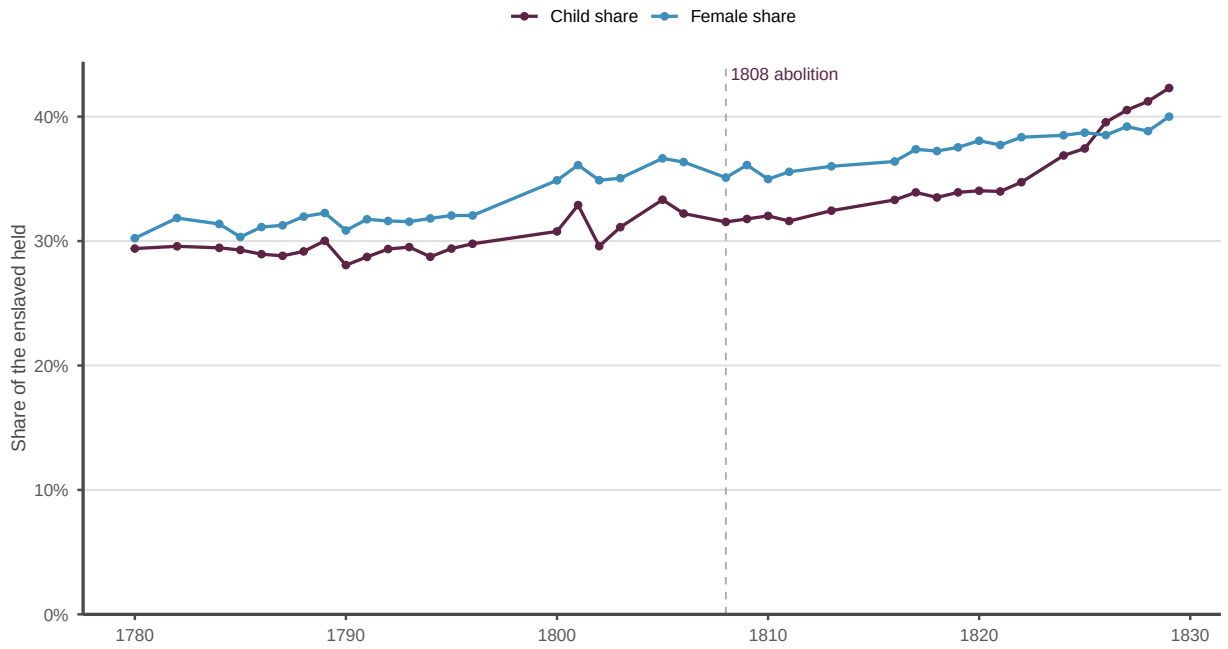

Figure 6: Child and female shares of the enslaved held, Stellenbosch stable form family, by year. Dashed line marks the 1808 abolition of the oceanic slave trade.

6.7 Districts and periods

The decomposition is stable over 130 years: the non-family variance share lies between 67.1 and 78.8 percent in every quarter-century of the sample, and the mean non-family count exceeds the mean count of free children in every period. The 50 district-decade cells reject the conditional null on entropy in every case, on samples rarefied to a common size. We do not attach causal interpretations to differences across districts because form definitions change with district and period.

7 Robustness and limitations

Blank cells. The decomposition and the pooled gradients are reported under a strict specification in which no blank is read as zero, with the variance shares and the gradient ordering unchanged (Tables 1 and 2); the joint-organisation result also holds on the strict family-slavery sample, where the family-block-fixed conditional null still rejects at the smallest attainable p (Table 4, notes).

Asset-rank construction. Because the cultivation and viticulture domains partly record realised output, we re-estimate the gradients with a stock-only rank built from the livestock and vehicle domains and with each domain excluded in turn. The stock-only rank gives a family gradient of 0.958 and a non-family gradient of 2.725 (difference 1.767) on nearly the full sample, indistinguishable from the baseline; across the leave-one-domain-out variants the stacked difference lies between 1.67 and 1.75, always precisely estimated, though these variants also shift the sample to the units on which the reduced rank exists. The association between recorded scale and labour is not an artefact of output entering the rank.

Family-core measurement. Re-estimating the measurement results with a children-only family core, and on the six spouse-validated form families alone (150,520 units), leaves the conclusions intact: the size–family correlation lies between 0.52 and 0.56, the Shorrocks family share between 15 and 19 percent, and the district rankings by size and by family core continue to disagree (Spearman correlations of -0.21 to 0.40 across definitions).

Upper tail. Excluding the 112 units flagged for aggregating six to ten source rows leaves every share unchanged; excluding the top percentile of unit size (above 39 persons) lowers the Shorrocks non-family share from 82 to 75 percent. The dominance of the labour margin is not carried by the extreme tail.

Binning and thresholds. The state tables use bins (0, 1, 2–3, 4–7, 8+) and a threshold of five units, both fixed before estimation. Occupancy and unit coverage at thresholds 1, 5, 10 and 20 are reported in the replication archive; the null comparisons in Table 4 use the threshold of five.

District influence. Excluding any one district changes the stacked gradient difference to between 1.74 and 2.13, always positive and precisely estimated.

Inference. The non-family gradient is monotone across asset deciles, the expected count rising to 21.6 times the bottom decile at the top. The stacked gradient difference of 1.78 has a sheet-level cluster-bootstrap 95 percent interval of 1.70 to 1.88 (299 draws, two-sided $p = 0.007$), so it is robust to few-cluster concerns as well as to district exclusion. There are 260 sheet clusters, of which under four percent hold fewer than thirty units.

Aggregation. Excluded administrative blocks (37 units chaining more than ten source rows) do not affect the reported estimates; including them inflates the upper tail of unit size.

Four limitations are inherent to the source. The units are administrative tax declarations, and although the audited declaration forms list one named head with family and labour attached, co-residence is not verified; we describe recorded units, not verified households. A large declaration may in principle aggregate labour across the dispersed holdings of one owner; the exclusion of the multi-row administrative blocks, and the stability of the decomposition when the flagged units

and the top percentile of size are dropped, bound this concern but cannot eliminate it, so the recorded unit remains a declaration, not a dwelling. The form audit of Section 4 checks that the declared unit has the structure of a single-head declaration with family and labour attached; it cannot establish co-residence or single-site production, and external validation against independent records of residence or against reconstituted genealogies belongs to the linkage programme that this paper's opgaaf-only design deliberately sets aside. The data contain no links across years, so we observe the distribution of compositions and never a change in composition. Linked data would permit the transition analysis we cannot perform, tracing how a unit's composition changes as its assets, labour holdings and family evolve, as studies of coresidence in other settings have done (Ruggles, 2011); linked Cape genealogies have been used to study settler fertility and marriage (Cilliers, 2019; Cilliers et al., 2025), and constructing comparable links from the opgaafrollen is the subject of separate work. Enslaved and Khoe people appear only as counts assigned to the units that held them, with none of their own family structure visible; the composition choices we model were exercised over them, not by them. Productive capacity is measured by within-form asset ranks, so results compare units assessed on the same form and do not measure absolute wealth.

8 Conclusion

In the Cape tax declarations of 1692–1829, variation in recorded household size was dominated by enumerated labour holdings, not by the free family core. Non-family labour accounts for 82 percent of the variance in recorded unit size and four in five of the additional persons associated with higher recorded productive capacity; the enslaved, Khoe and servant counts each rise far more steeply with capacity than the family core, along the distinct production systems of the arable core and the pastoral frontier; and the joint distribution of family and labour was concentrated on far fewer configurations than independent assignment—or common scaling with recorded productive capacity—would produce, even with each unit's family block held fixed. Because size and family demography diverge this sharply, an unadjusted count of household members is not a demographic statistic wherever coerced labour was held: recorded size correlates with the family core at only 0.55 and misranks districts on family demography. For the study of households in slaveholding and labour-coercive economies, counts of household members should be decomposed into family and non-family components before any demographic interpretation, the several labour regimes should be distinguished, and the boundary of the household should be treated as an outcome of labour-market institutions. The correction is portable: wherever production placed workers inside the household—slave societies, serf estates, economies of life-cycle service—the recorded household was a workplace as much as a family, and the demographic content of household size cannot be read until the two are separated. The opgaafrollen show how much the reading changes once they are.

References

Abad, Leticia Arroyo, and Noel Maurer. 2025. "The Long Shadow of History? The Impact of Colonial Labor Institutions on Economic Development in Peru." *Journal of Economic Growth*, 30(4): 521–565.

Acemoglu, Daron, and Alexander Wolitzky. 2011. "The Economics of Labor Coercion." *Econometrica*, 79(2): 555–600.

Agresti, Alan. 2013. *Categorical Data Analysis.* . 3rd ed., Hoboken, NJ:Wiley.

Becker, Gary S. 1965. "A Theory of the Allocation of Time." *The Economic Journal*, 75(299): 493–517.

Ben-Porath, Yoram. 1980. "The F-Connection: Families, Friends, and Firms and the Organization of Exchange." *Population and Development Review*, 6(1): 1–30.

Buggle, Johannes C., and Steven Nafziger. 2021. "The Slow Road from Serfdom: Labor Coercion and Long-Run Development in the Former Russian Empire." *The Review of Economics and Statistics*, 103(1): 1–17.

Chiappori, Pierre-André. 1992. "Collective Labor Supply and Welfare." *Journal of Political Economy*, 100(3): 437–467.

Cilliers, Jeanne. 2019. "The Shaping of a Settler Fertility Transition: Eighteenth- and Nineteenth-Century South African Demographic History Reconsidered." *European Review of Economic History*, 23(4): 421–445. VERIFY end page.

Cilliers, Jeanne, Erik Green, Anne E. McCants, and Auke Rijpma. 2025. "Wealth and Marriage at the Cape: Consanguineous Unions as a Strategy." *The History of the Family*, 30(2): 258–284.

Coase, Ronald H. 1937. "The Nature of the Firm." *Economica*, 4(16): 386–405.

Dari-Mattiacci, Giuseppe. 2013. "Slavery and Information." *The Journal of Economic History*, 73(1): 79–116.

Doepke, Matthias, Anne Hannusch, Fabian Kindermann, and Michèle Tertilt. 2023. "The Economics of Fertility: A New Era." In *Handbook of the Economics of the Family.* Vol. 1, , ed. Shelly Lundberg and Alessandra Voena, 151–254. Amsterdam:North-Holland. VERIFY chapter pages.

Domar, Evsey D. 1970. "The Causes of Slavery or Serfdom: A Hypothesis." *The Journal of Economic History*, 30(1): 18–32.

Dooling, Wayne. 2007. *Slavery, Emancipation and Colonial Rule in South Africa.* Scottsville:University of KwaZulu-Natal Press.

Elphick, Richard, and Hermann Giliomee, ed. 1989. *The Shaping of South African Society, 1652–1840.* . 2nd ed., Middletown, CT:Wesleyan University Press.

Fogel, Robert W., and Stanley L. Engerman. 1974. *Time on the Cross: The Economics of American Negro Slavery.* Boston:Little, Brown.

Fourie, Johan. 2013. "The Remarkable Wealth of the Dutch Cape Colony: Measurements from Eighteenth-Century Probate Inventories." *The Economic History Review*, 66(2): 419–448.

Fourie, Johan. forthcoming. "The New Economics of Cape Slavery." *South African Journal of Economics*. Forthcoming.

Fourie, Johan, and Dieter von Fintel. 2010. "The Dynamics of Inequality in a Newly Settled, Pre-Industrial Society: The Case of the Cape Colony." *Cliometrica*, 4(3): 229–267. VERIFY.

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.

Fourie, Johan, Sugnet Lubbe, Johane Nienkemper-Swanepoel, and Dieter von Fintel. 2026. "Biplots for Historical Household Data: Production Systems in the Cape Colony." Working paper, submitted to *Historical Methods*. VERIFY exact title and status.

Gotelli, Nicholas J., and Robert K. Colwell. 2001. "Quantifying Biodiversity: Procedures and Pitfalls in the Measurement and Comparison of Species Richness." *Ecology Letters*, 4(4): 379–391.

Green, Erik. 2014. "The Economics of Slavery in the Eighteenth-Century Cape Colony: Revising the Nieboer–Domar Hypothesis." *International Review of Social History*, 59(1): 39–70.

Greenwood, Jeremy, Nezih Guner, and Guillaume Vandenbroucke. 2017. "Family Economics Writ Large." *Journal of Economic Literature*, 55(4): 1346–1434.

Gruber, Siegfried, and Mikołaj Szołtysek. 2016. "The Patriarchy Index: A Comparative Study of Power Relations across Historical Europe." *The History of the Family*, 21(2): 133–174.

Guirkinger, Catherine, and Jean-Philippe Platteau. 2015. "Transformation of the Family Farm under Rising Land Pressure: A Theoretical Essay." *Journal of Comparative Economics*, 43(1): 112–137.

Guthmann, Rafael R., and Walter Scheidel. 2025. "The Economics of Greco-Roman Slavery." *Explorations in Economic History*, 97: 101689.

Hammel, Eugene A., and Peter Laslett. 1974. "Comparing Household Structure over Time and between Cultures." *Comparative Studies in Society and History*, 16(1): 73–109.

Hareven, Tamara K. 1991. "The History of the Family and the Complexity of Social Change." *The American Historical Review*, 96(1): 95–124.

Kaplan, Greg. 2012. "Moving Back Home: Insurance against Labor Market Risk." *Journal of Political Economy*, 120(3): 446–512.

Kertzer, David I. 1991. "Household History and Sociological Theory." *Annual Review of Sociology*, 17: 155–179.

Korchmina, Elena, and Mikołaj Malinowski. 2026. "How Extractive Was Russian Serfdom? Income Inequality in Moscow Province in the Early Nineteenth Century." *The Journal of Economic History*, 86(1): 67–108.

Laslett, Peter, and Richard Wall, ed. 1972. *Household and Family in Past Time.* Cambridge:Cambridge University Press.

Links, Calumet, Johan Fourie, and Erik Green. 2020. "The Substitutability of Slaves: Evidence from the Eastern Frontier of the Cape Colony." *Economic History of Developing Regions*, 35(2): 98–122.

Martins, Igor. 2026. "Capital and Coercion: Slavery after the 1807 Import Ban in the Cape Colony." *European Review of Economic History*. Advance article heag012.

Martins, Igor, and Erik Green. 2026. "Theoretical Foundations of the Economics of Slavery: Enslaved People as Capital Investments in the Atlantic World." *Journal of Global History*, 21(1): 22–40.

Matranga, Andrea, and Timur Natkhov. 2026. "All Along the Watchtower: Military Landholders and Serfdom Consolidation in Early Modern Russia." *The Review of Economic Studies*, 93(4): 2784–2818.

Milgrom, Paul, and Chris Shannon. 1994. "Monotone Comparative Statics." *Econometrica*, 62(1): 157–180.

Naidu, Suresh, and Noam Yuchtman. 2013. "Coercive Contract Enforcement: Law and the Labor Market in Nineteenth Century Industrial Britain." *American Economic Review*, 103(1): 107–144.

Newton-King, Susan. 1999. *Masters and Servants on the Cape Eastern Frontier, 1760–1803.* Cambridge:Cambridge University Press.

Nieboer, H. J. 1910. *Slavery as an Industrial System: Ethnological Researches.* The Hague:Martinus Nijhoff.

Papke, Leslie E., and Jeffrey M. Wooldridge. 1996. "Econometric Methods for Fractional Response Variables with an Application to 401(k) Plan Participation Rates." *Journal of Applied Econometrics*, 11(6): 619–632.

Piano, Ennio E., and Sean-Patrick Alvarez. 2025. "Servants of Two Masters: The Economics of 'Slave-Hiring'." *European Economic Review*, 180: 105185.

Pollak, Robert A. 1985. "A Transaction Cost Approach to Families and Households." *Journal of Economic Literature*, 23(2): 581–608.

Rönnbäck, Klas. 2024. "Free or Enslaved Labor: A Theoretical Comparative Model." *Journal of Global Slavery*, 9(3): 303–335.

Ruggles, Steven. 1994*a*. "The Origins of African-American Family Structure." *American Sociological Review*, 59(1): 136–151.

Ruggles, Steven. 1994*b*. "The Transformation of American Family Structure." *The American Historical Review*, 99(1): 103–128.

Ruggles, Steven. 2009. "Reconsidering the Northwest European Family System: Living Arrangements of the Aged in Comparative Historical Perspective." *Population and Development Review*, 35(2): 249–273.

Ruggles, Steven. 2011. "Intergenerational Coresidence and Family Transitions in the United States, 1850–1880." *Journal of Marriage and Family*, 73(1): 136–148.

Ruggles, Steven. 2015. "Patriarchy, Power, and Pay: The Transformation of American Families, 1800–2015." *Demography*, 52(6): 1797–1823.

Saleh, Mohamed. 2024. "Trade, Slavery, and State Coercion of Labor: Egypt during the First Globalization Era." *The Journal of Economic History*, 84(4): 1107–1141.

Santos Silva, J. M. C., and Silvana Tenreyro. 2006. "The Log of Gravity." *The Review of Economics and Statistics*, 88(4): 641–658.

Shell, Robert C.-H. 1994. *Children of Bondage: A Social History of the Slave Society at the Cape of Good Hope, 1652–1838.* Hanover, NH:Wesleyan University Press.

Voigtländer, Nico, and Hans-Joachim Voth. 2013. "How the West "Invented" Fertility Restriction." *American Economic Review*, 103(6): 2227–2264.

Worden, Nigel. 1985. *Slavery in Dutch South Africa.* Cambridge:Cambridge University Press.