CEIS TOR VERGATA Research Papers

Università di Roma Tor Vergata

ISSN 2610-931X

CEIS Tor Vergata

RESEARCH PAPER SERIES

Vol. 24, Issue 5, No. 631 – November 2026

Estimating Local Income Distributions from Historical Tax Records

Federico Belotti, Mathias Silva, Giovanni Vecchi

Estimating Local Income Distributions from Historical Tax Records

Federico Belotti
Mathias Silva
Giovanni Vecchi*

November 11, 2026

Abstract

In 1868, the Italian government authorized nearly 8,000 municipalities to levy a household tax, the *focatico*. Over the following century, its registers preserved unusually rich evidence on household incomes, occupations, residential addresses, and more. We focus on the income data, often truncated and grouped into brackets, with the highest bracket left open-ended. We assess alternative parametric maximum-likelihood estimators and show how these records, combined with population censuses, can be used to reconstruct the entire household income distribution. An application to Modena in 1875–76 illustrates the approach at the sub-municipal level, exploring the income divide between town and countryside.

Keywords: Income distributions, Inequality, Poverty, Truncated data, Interval censoring, Historical tax records.

JEL classification: C13; C24; D31; I32; N13; N93.

*All authors are affiliated with the Department of Economics and Finance, Tor Vergata University of Rome. Corresponding author: Giovanni Vecchi (giovanni.vecchi@uniroma2.it). We thank Brian A'Hearn, Giulia Mancini and Alberto Rinaldi for helpful comments and suggestions. This research was supported by the European Union - NextGenerationEU through the GRINS project (PE00000018, CUP E83C22004690001) and the project "Economic Growth, Inequality and Poverty in the Long Run: Italy, 1861-2021" (2022BBMA3T). The views expressed are those of the authors and do not necessarily reflect those of the European Union.

1 Introduction

Detailing the distributional history of a country or a locality from the perspective of incomes requires the same basic ingredient: a reconstruction of the entire personal, or size, distributions of income, with households or individuals as the units of analysis (Cowell, 2011). This is the first best. It allows us to investigate living standards across the population and across time, to examine poverty alongside affluence, and to assess how the gains and losses associated with economic change were distributed. Obtaining this information is difficult even today. Modern household surveys contend with non-response, misreporting, and incomplete coverage of the richest households, with consequences for the distributions they describe (Korinek et al., 2007; Burkhauser et al., 2018). For the economic historian, these difficulties are compounded by sources collected for other purposes, uneven survival, and definitions that change across time and place.

At the national level, a path towards this objective has been opened recently for the Italian case. As discussed in A'Hearn et al. (2026b), the Historical Household Budget (HHB) project brings together household income and expenditure records from a wide range of historical sources, seeking to reconstruct a database that resembles modern household surveys as closely as the evidence permits (A'Hearn et al., 2016; Chianese and Vecchi, 2017). Important gaps remain, but the approach provides a basis for studying Italy's distributional history since Unification. Moving below the national level, the evidence becomes much thinner. A collection sufficiently rich to support national estimates need not contain enough observations to reconstruct the income distribution of a particular region, province, or municipality.

The difficulty is partly familiar from contemporary statistics. Italy's EU-SILC survey supports regional estimates, but reconstructing entire distributions for individual municipalities requires information beyond what national household surveys ordinarily provide.[1] For the historical scope, the distance between the questions we would like to ask and the evidence available to answer them is greater still. Research has substantially improved our knowledge of regional and provincial differences in production, wages, and living standards (Ciccarelli and Fenoaltea, 2013; Federico et al., 2019; Chiaiese, 2024; Fiore Melacrinis and Felice, 2026). These

[1] EU-SILC (European Union Statistics on Income and Living Conditions) collects harmonized household and individual information on income, poverty, and living conditions; Italy has participated since 2004, with a survey designed to support regional estimates (Istat, 2026b). For methods of extending survey-based poverty and deprivation indicators to subnational areas, see Betti et al. (2012).

reconstructions tell us much about differences *between* places; they tell us considerably less about the distribution of household incomes within them. Even at the regional level, the latter remains poorly documented over the long run. At the *provincial* level, the evidence is more fragmentary, while a systematic reconstruction of *municipal* income distributions in post-Unification Italy remains largely to be undertaken. The scale of this last ambition should not be underestimated. Italy already counted 7,720 municipalities in 1861, encompassing communities that differed profoundly in population, economic structure, and administrative resources. Moving from provinces to municipalities means multiplying the distributions to be reconstructed while often reducing the evidence available for each.

This paper contributes to that task by drawing on a source still largely unused in distributional history: the registers of the *tassa di famiglia*, or *focatico* (A'Hearn et al., 2026a). In 1868, Italian municipalities were authorized to levy a tax on households' *agiatezza*, or economic circumstances, commonly assessed through household income. Over the following century, the administration of this tax produced annual registers identifying taxpayers by name, often recording their occupations and addresses, and assigning them an income or an income class. Municipalities enjoyed considerable discretion in defining and administering the tax (Casu, 2026; Salsano, 2026). Some recorded precise income assessments; others divided households into a handful of broad income classes or several dozen narrower ones. Households with incomes below a local exemption threshold were generally absent, and the highest income class was often open-ended. The registers preserve exceptionally rich evidence on household incomes, but using it to reconstruct income distributions requires statistically informed handling.

This work makes two related contributions. First, it provides a framework for establishing what the *focatico* registers measure. The records were created for one purpose; we use them for another. The starting point is therefore historical: what counted as income, whose resources entered the assessment, which households were exempt, and how were taxpayers assigned to classes? Answering these questions requires patient reconstruction from municipal regulations, the minutes of council and executive-board meetings, administrative correspondence, and other archival documents, many of them handwritten. These materials reveal the decisions, disputes, and practices through which households' circumstances became entries in a tax register. Taking those entries seriously as evidence means investigating how they came into being, a point forcefully made by Fenoaltea (2019). In doing so,

the economic historian performs a task familiar to the statistician: understanding the process that generated the observations before deciding how to model them. Here, this connection takes a concrete form: the rules governing income assessment, exemption, and classification determine how the historical evidence enters the statistical analyses. Historical source criticism helps devise the suitable form of empirical methods.

Second, we develop and assess an econometric method for recovering the entire household income distribution from these incomplete records. The approach combines the income thresholds and household counts reported in the tax registers with census-based estimates of the total number of resident households. Under explicit assumptions about coverage, the latter provide information on the size of the population missing below the exemption threshold. We use a parametric model, which represents the income distribution through a small number of parameters governing its shape and scale, and devise two maximum-likelihood estimators incorporating this external information in different manners. The constrained maximum-likelihood estimator requires the fitted distribution to reproduce the estimated share of exempt households exactly; the complete-data estimator treats those households as an additional income class to be fitted alongside the observed classes. Both retain the full distribution as the object of estimation, making it possible to examine poverty, inequality, and income levels within a common framework.

The paper is methodological, and its central argument requires a technical treatment. The historical questions motivating it, however, are straightforward. How much can incomplete tax records tell us about the distribution of incomes in a local population? What additional information is needed? And which conclusions depend most heavily on the assumptions used to fill the gaps? Our aim is to make the passage from archival evidence to statistical reconstruction explicit enough for economic historians to assess and use.

An application to Modena in 1875–76 illustrates this passage. The registers distinguish between urban and rural population sub-groups, allowing us to take the reconstruction below the municipal level. This unusually fine geographical detail opens to quantitative investigation a divide that municipal averages conceal: the distance in living standards between the city and its surrounding countryside. Growth incidence curves (Ravallion and Chen, 2003) trace income changes across the distribution. Their decomposition, inspired by Ferreira et al. (2019), distinguishes changes in mean income from changes in the shape of the distribution and

their interaction. The application illustrates the comparisons these reconstructions make possible, and provides an assessment of their sensitivity to the demographic information used to account for exempt households. Two consecutive years provide a limited window on Modena's history, but a useful demonstration of what a longer reconstruction could make possible.

The remainder of the paper is organized as follows. Section 2 examines the historical sources and their statistical implications. Section 3 develops the estimation framework. Section 4 presents the Modena application and sensitivity analyses. Section 5 discusses the scope and limitations of the approach.

2 Difficult data, elusive income distributions

A household enters a *focatico* register because the municipality has decided that it can afford to pay a tax. Its entry may record a name, an occupation, an address, and an assessed income or income class. The following year's register may contain the same household, at the same address or elsewhere, with an assessment that has remained unchanged or moved up or down. Where long runs of registers survive, these entries offer something unusual in economic history: the possibility of observing household circumstances repeatedly, within the same local population, over many years. Read together, they promise a much richer account of living standards than the tax authorities needed to collect their revenue.

Part of this promise lies in the possibility of recovering the connections between entries. In Modena, progressive and roll numbers help link households across successive registers. The resulting histories could reveal how often families moved into and out of economic difficulty, how persistent their relative positions were, and how these experiences differed across occupations or generations.[2] Following these families is not straightforward. Names and numbers must be read correctly from handwritten records, and a household that disappears may have moved, divided, or ceased to exist. Transcription and record linkage therefore require their own historical and statistical work (Abramitzky et al., 2020). Yet the prospect of following ordinary families year after year remains one of the source's most compelling features.

Their addresses open another line of inquiry. Placed on historical maps, the entries could show where households of different means lived, how closely they

[2]For an analysis of vulnerability to poverty in Italian economic history using pseudo-panel data, see Rossi et al. (2017).

shared the same streets and neighbourhoods, and how the geography of income changed as towns expanded. They could bring to Italian settings questions about residential segregation and urban development explored elsewhere by Logan and Parman (2017) and Heblich et al. (2020). The map would nevertheless have a revealing blank: households exempt from the tax generally left no address in its registers. The very source that makes it possible to locate taxpayers can leave the poorest residents out of sight.

In this work we do not pursue these possibilities but rather focus on income information only, asking how far the registers allow us to reconstruct the distribution of incomes among *all* resident households in a particular year and municipality. The distinction between taxpayers and residents is central, but it is not the first difficulty the historian encounters. Before deciding whose income is missing, we must establish what the recorded incomes mean.

The law made *agiatezza* the object of taxation, leaving municipalities considerable freedom to translate economic circumstances into an assessment Casu (2026); Salsano (2026). An amount expressed in Italian lire may therefore look more familiar than it really is. It may refer to income before taxes, income after taxes, or taxable income after deductions that varied from one municipality to another. Nor does the heading "household income" establish whether the resources of all household members were included, whether earnings in kind entered the calculation, or how income from property was treated. These questions lead from the register to municipal regulations, council proceedings, and administrative correspondence. Only by reading them together can we determine what lies behind the apparent precision of the fiscal entry.

The consequences can be substantial. In Augusta, north of Syracuse in Sicily, registers from the late 1940s and early 1950s report taxable incomes that averaged only about one-third of gross incomes. A historian who took the taxable amount as a measure of household resources would obtain a very different picture of local living standards. Here the distinction can be established, and the problem becomes one of bringing the recorded figures to a common income definition. The framework proposed by the Canberra Group provides a useful standard for this work (United Nations Economic Commission for Europe, 2011). Ideally, we seek total gross household income, including the resources of all members and the various components through which those resources were received. How closely we can approach that definition depends on the evidence available in each case.

Vitorchiano, in the province of Viterbo, presents a difficult example. The municipality collected the family tax from 1872 until after the Second World War, but its assessments appear to have rested on a broader understanding of *agiatezza*. Property and other signs of economic standing mattered alongside income. The evidence includes state taxes paid, the presence of domestic servants, and even ownership of dogs, themselves subject to taxation. Such entries may tell us a great deal about the circumstances of a family. They do not, however, give us an income figure that can be made comparable with another municipality's by adjusting for taxes or deductions. For the purposes of this paper, we set such cases aside until further research establishes whether household income can be recovered from the assessment.

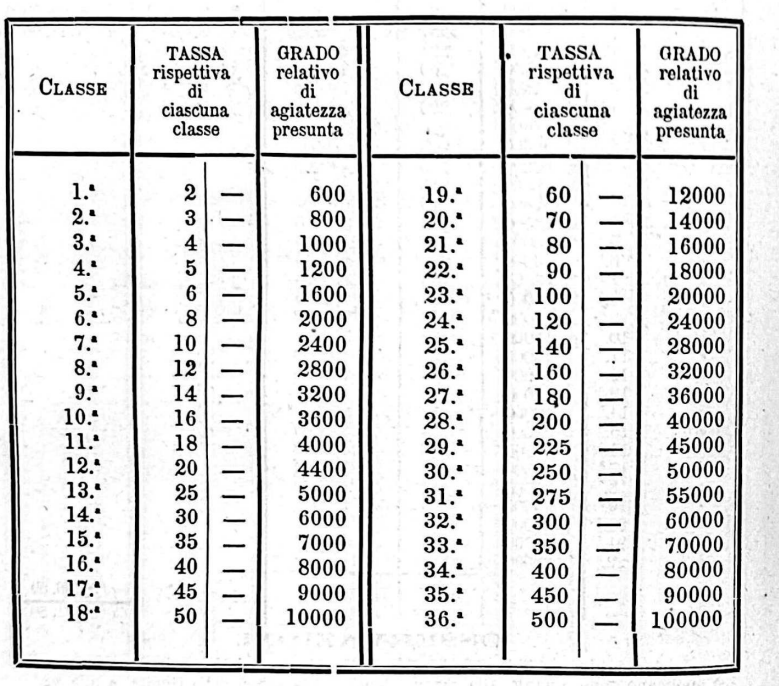
Even when the meaning of an entry is clear, its accuracy remains a matter for investigation. Municipal administrations differed in resources and practices, and assessments were made within local political and social relationships. As Maas and Morgan (2012, p. 15) observe, social statistics are "always the outcome of strategic interactions that are politically loaded." The influence of administrative priorities on historical evidence is also familiar from the history of household surveys (Serra, 2014). Independent records sometimes allow us to examine an assessment more closely. The incomes of public employees or bank officials, for example, may be checked against personnel files, provided that the comparison concerns the same resources.[3] These checks cannot certify an entire register, but they can help establish what its figures represent and how much confidence to place in them.

Suppose that this work has been done: the income definition is understood, the necessary adjustments have been made, and the assessments are judged usable. We still do not have the income distribution of the municipality. The reason is readily apparent in Modena's tax schedule for 1875, reproduced in Figure 1. A family receiving less than 600 lire a year paid no tax and generally did not appear in the register. Above that threshold, households were assigned to one of 36 income classes. The intervals were relatively narrow in the central part of the distribution and became wider towards the top. The final class began at 100,000 lire, with no upper limit. We can count the families in that class, but cannot tell how far their

[3]For Luigi Bodio, the statistician and public administrator discussed by Soresina (2001), the Rome *focatico* records household income of 9,840 lire in 1891, while personnel records held at the Archivio Centrale dello Stato report 9,000 lire. Establishing whether the difference reflects additional household income or a discrepancy in assessment requires comparing the coverage of the two sources. For Armando Boetti, an employee of the Banca Popolare di Modena, personnel records preserved in the bank's historical archive can likewise be compared with the income class assigned in the *focatico*.

incomes extended beyond its lower boundary.

Figure 1: Income-class table for the family tax in Modena, 1875



Classe	TASSA rispettiva di ciascuna classe	GRADO relativo di agiatezza presunta	Classe	TASSA rispettiva di ciascuna classe	GRADO relativo di agiatezza presunta
1.ª	2 —	600	19.ª	60 —	12000
2.ª	3 —	800	20.ª	70 —	14000
3.ª	4 —	1000	21.ª	80 —	16000
4.ª	5 —	1200	22.ª	90 —	18000
5.ª	6 —	1600	23.ª	100 —	20000
6.ª	8 —	2000	24.ª	120 —	24000
7.ª	10 —	2400	25.ª	140 —	28000
8.ª	12 —	2800	26.ª	160 —	32000
9.ª	14 —	3200	27.ª	180 —	36000
10.ª	16 —	3600	28.ª	200 —	40000
11.ª	18 —	4000	29.ª	225 —	45000
12.ª	20 —	4400	30.ª	250 —	50000
13.ª	25 —	5000	31.ª	275 —	55000
14.ª	30 —	6000	32.ª	300 —	60000
15.ª	35 —	7000	33.ª	350 —	70000
16.ª	40 —	8000	34.ª	400 —	80000
17.ª	45 —	9000	35.ª	450 —	90000
18.ª	50 —	10000	36.ª	500 —	100000

Source: ASCMO, Atti a stampa del Consiglio Comunale, minutes of 21 January 1875, Appendix IX, p. 131.

The municipality could administer its tax without resolving these uncertainties. The economic historian cannot ignore them when measuring living standards. Excluding families below the exemption threshold changes the population being described; assigning the remaining families to classes conceals differences within each class. In statistical language, the first arrangement produces *left truncation*, the second *interval censoring*. The open-ended class further constitutes *right censoring*. These terms describe different ways in which the fiscal record falls short of a complete distribution. The poor are missing from it; the taxpayers are present, but their precise incomes are often unknown.[4]

To learn how many households are missing, we must look beyond the tax records. Population censuses provide, or help us reconstruct, the total number of households residing in the municipality. Comparing that total with the number in the register gives an estimate of the omitted population. In Modena in 1875, the comparison suggests that approximately 31.3 percent of households were exempt. This interpretation is conditional on the two sources covering the same population and on

[4]Some municipalities, including Rome in 1891, recorded individual income assessments rather than income classes. These figures may nevertheless exhibit rounding, introducing a further loss of precision that we do not address separately in the estimation framework.

the difference being attributable to low-income exemptions, rather than evasion, other exemptions, or incomplete records. Even under these conditions, the census tells us how many households lie below the threshold, not how their incomes are distributed.

The importance of these omissions is easier to appreciate by illustrating a hypothetical distribution that we can observe and then remove information from it. This is the purpose of the exercise in Figure 2. Using the Bank of Italy's Survey on Household Income and Wealth (SHIW) for 2022, we construct a benchmark distribution and ask what we would learn from it if incomes had been recorded under rules resembling those of the *focatico*. The modern survey supplies a reference against which the loss of information can be measured. It need not be an error-free account of Italian incomes for the comparison to be informative.

Panel A shows the benchmark: a Singh–Maddala distribution fitted to equivalized disposable household incomes using survey weights, together with the corresponding measures of inequality and poverty.[5] In Panel B, the poorest 31.3 percent of households have been removed, reproducing the estimated extent of exemption in Modena. Calculating indicators from those who remain now describes the taxable population. In this example, poverty disappears altogether: the poverty line falls below the exemption threshold, so every poor household has been excluded from observation.

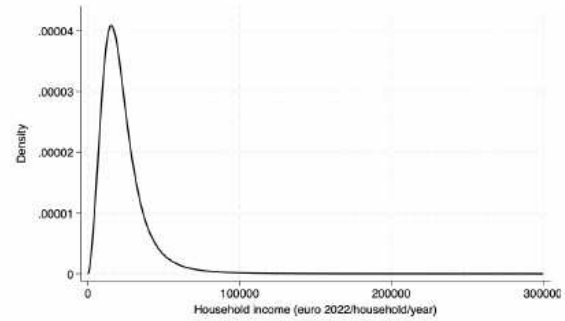
Panel C also conceals differences above an upper threshold, recording all incomes beyond it at the threshold itself. The resulting concentration of observations is an artefact of the recording rule. Panel D then groups the observed incomes into 36 classes, bringing the available information closer to that in the Modena register. The indicators alongside the panels show how the account of poverty and inequality changes as information is lost. The magnitudes belong to this particular exercise; they are not estimates of the bias in every *focatico* register. They nevertheless demonstrate why a direct calculation from the fiscal entries can give a misleading account of the population from which those entries were drawn.

The problem is therefore not simply that the registers contain too little information. They contain a great deal, but its meaning and coverage must first be estab-

[5]The benchmark indicators are parametric estimates and therefore differ from the non-parametric estimates published by Bank of Italy (2024). The poverty line, approximately 8,800 euros per equivalent adult per year, is calibrated to reproduce the aggregate incidence reported by ISTAT (2023). This is an illustrative calibration. Istat's official measure uses consumption expenditure and thresholds that vary with household characteristics, rather than a single income threshold.

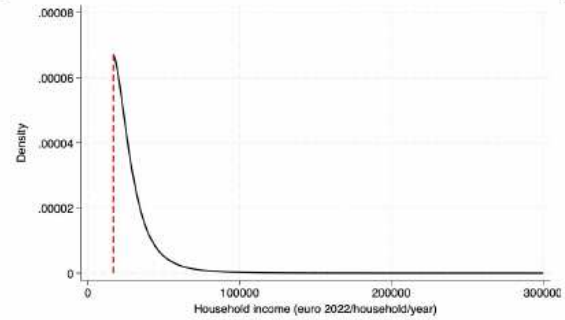
Figure 2: Estimating inequality and poverty: modern data compared with *focatico* data

A. Complete data



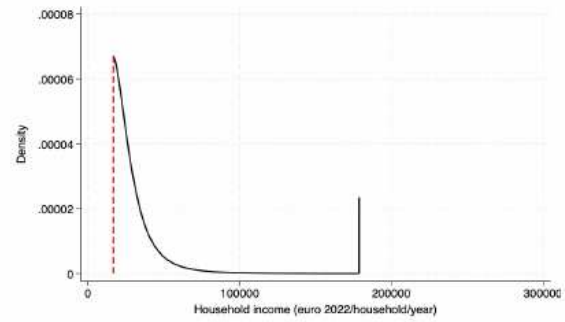
Gini	32.5
MLD	18.5
Theil	18.4
Top 1%	4.8
H	10.4
PG	3.0

B. Left-truncated data, due to exemption threshold (bottom 40%)



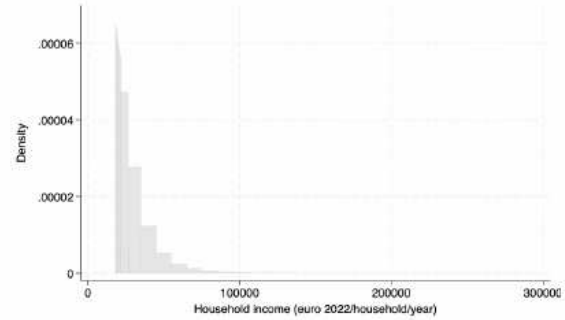
Gini	22.9
MLD	8.5
Theil	10.1
Top 1%	4.1
H	0.0
PG	0.0

C. Left-truncated and right-censored data



Gini	22.8
MLD	8.4
Theil	9.7
Top 1%	3.9
H	0.0
PG	0.0

D. Left truncation and interval censoring (36 classes)



Gini	22.5
MLD	8.2
Theil	9.6
Top 1%	4.0
H	0.0
PG	0.0

Note: "MLD" denotes the mean log deviation; "Theil" is the Theil index; "Top 1%" denotes the share of income held by the richest 1 percent; "H" denotes the poverty headcount rate, with the poverty line set at approximately 8,800 euros per equivalent adult per year; "PG" denotes the poverty gap. Source: Authors' calculations based on SHIW 2022 data.

lished, and its form must then be respected in estimation. No statistical method can turn an unsuitable measure of *agiatezza* into household income merely by fitting a curve to it. Where the assessments can be used, however, the thresholds and household counts preserve information about the underlying distribution, while censuses help establish how much of the population is missing. The following section brings these pieces together. Its task is to reconstruct the full income distribution under explicit assumptions, without allowing the boundaries of the tax-paying population to become the boundaries of the history we write.

3 Statistical framework

Once we have established what the income figures recorded in the *focatico* registers represent, the reconstruction begins with a relatively small amount of information: the income thresholds used by the municipality, the number of households assigned to each class, and an estimate of the total number of resident households. The registers tell us how taxpayers were distributed across the fiscal schedule; the census helps establish how many households were left outside it. Neither source reveals the incomes within each class or below the exemption threshold. Our task is to bring these pieces of evidence together without treating the distribution of taxpayers as if it described the whole population.

3.1 From fiscal records to income distributions

Economic historians have long worked with income information preserved in groups rather than as individual observations. Fiscal tabulations underlie much of the early research on income distributions, from Pareto's investigations to Kuznets's reconstructions (Persky, 1992; Kuznets, 1953). Social tables offer another familiar route, describing populations through the numbers and average incomes of occupational or social groups (Milanovic et al., 2011; Modalsli, 2015; Allen, 2019). These sources differ in what they preserve, but share a difficulty: variation within the reported groups remains partly or wholly unobserved.

The statistical literature provides several ways of working with such evidence. Some methods establish bounds on distributional measures; others estimate a distribution by imposing additional structure (Gastwirth, 1972; Kakwani and Podder, 1976; Cowell and Flachaire, 2015). Grouping, censoring, and truncation also belong to a broader literature on incomplete observations (Turnbull, 1976). More recent

work has extended the possibilities for reconstructing income distributions from tabulations, while showing that the appropriate method depends on the information those tabulations contain (Hajargasht et al., 2012; Jorda et al., 2021; Eckernkemper and Gribisch, 2021; Blanchet et al., 2022).

A *focatico* scheme thresholds identify intervals in monetary units, and its registers count how many taxable households fall within each interval. The principal difficulty is that the intervals do not cover the whole resident population: households with incomes below the exemption threshold are generally missing, while the uppermost class has no recorded ceiling. We build on a conventional multinomial maximum-likelihood approach to grouped income data (McDonald, 1984), incorporating census-based information on the missing population in two alternative ways. Related work on incomplete historical data provides useful precedents, including the reconstruction of height distributions from truncated samples and the estimation of inequality from censored or truncated incomes (Wachter and Trussell, 1982; A'Hearn, 2004; Hong et al., 2018). Our concern is with an implementation suited to the information preserved by the tax registers and with the historical assumptions needed to use it.

3.2 Modelling income distributions

Let N denote the number of households residing in a municipality in a given year, and let y_i denote the income of household i. The register contains n_g households in income class g, for $g = 1, \ldots, G$, with boundaries z_g and z_{g+1}. The lowest boundary, z_1, is the exemption threshold; the highest class is open-ended, so that $z_{G+1} = \infty$. The total number of registered households is therefore

$$n = \sum_{g=1}^{G} n_g.$$

For consistency, we represent the taxable classes as $(z_g, z_{g+1}]$. Under a continuous income model, the choice of whether to include an endpoint does not affect the class probabilities.

A population census, supplemented where necessary by demographic reconstruction, provides an estimate $\widehat{N}$ of the resident household total. The difference between this estimate and the number of registered households gives an estimated count of

exempt households, $\widehat{n}_0 = \widehat{N} - n$, and an estimated exempt share:

$$\widehat{p}_0 = \frac{\widehat{n}_0}{\widehat{N}} = 1 - \frac{\sum_{g=1}^{G} n_g}{\widehat{N}}. \tag{1}$$

The interpretation of this difference is historical before it is statistical. It requires the census and register to refer to comparable populations and assumes that the missing households were omitted because their incomes fell below z_1. Evasion, exemptions unrelated to low income, and gaps in the surviving records can invalidate that interpretation. Throughout this section, $\widehat{N}$ and $\widehat{p}_0$ denote external reconstructions, rather than quantities directly observed in the register.

Even an exact count of exempt households would leave their incomes unknown. It would establish how much of the population lies below the threshold, but not how income is distributed within that group. The same difficulty arises within the taxable classes and above the lower boundary of the highest class. To reconstruct a continuous distribution, we therefore need an assumption linking its observed and unobserved parts.

We represent household incomes by a parametric cumulative distribution function $F(y|\theta)$, where θ is a small set of parameters. For any income level y, this function gives the proportion of households with incomes at or below that level. The probability of belonging to taxable class g is consequently

$$p_g(\theta) = F(z_{g+1}|\theta) - F(z_g|\theta), \qquad g = 1, \ldots, G. \tag{2}$$

This is the link between the historical record and the statistical model. The municipal schedule supplies the boundaries; the register supplies the counts against which the probabilities are fitted. Thresholds and frequencies together inform the scale and shape of the reconstructed distribution.

We use the Singh–Maddala distribution (Singh and Maddala, 1976), a three-parameter member of the generalized beta family widely employed in income-distribution analysis (Kleiber and Kotz, 2003; Chotikapanich et al., 2018). Its cumulative distribution function is

$$F(y|\theta) = 1 - \left[1 + \left(\frac{y}{\beta}\right)^{\alpha}\right]^{-q}, \qquad y > 0, \quad \theta = (\alpha, \beta, q), \quad \alpha, \beta, q > 0. \tag{3}$$

The parameter β governs scale, while α and q govern shape. The specification offers a compromise between flexibility and parsimony, particularly relevant when a mu-

nicipality reports only a limited number of income classes. Its analytical properties also allow income levels, inequality, poverty, and income shares to be calculated within a common framework.

This convenience should not obscure the role of the assumption. The fitted curve supplies the variation that the register does not record. A close match to taxable class frequencies is evidence that the model describes those frequencies well; it does not independently establish the distribution of incomes among exempt households. The reconstruction of that part of the population remains conditional on the functional form and on the external population estimate.

Once the distribution has been estimated, the historian need not choose a single summary measure in advance. The same curve yields mean income, the Lorenz curve, top income shares, and inequality indices with different sensitivities to changes across the distribution (Jenkins, 2009, 2024). For the Singh–Maddala distribution, the Gini coefficient is

$$G(\alpha, q) = 1 - \frac{\Gamma(q)\Gamma\left(2q - \frac{1}{\alpha}\right)}{\Gamma(2q)\Gamma\left(q - \frac{1}{\alpha}\right)}, \tag{4}$$

provided the mean exists. Its independence from β reflects the familiar distinction between income levels and relative inequality.

Poverty brings the level of income back into the comparison. Given a poverty line z, the headcount ratio is simply $F(z|\theta)$: the proportion of households below that line. The estimated distribution also permits calculation of poverty depth and severity through the measures proposed by Foster et al. (1984). Appendix A reports the expressions for moments, Lorenz curves, generalized entropy indices, and income quantiles. Together, these quantities allow the reconstruction to support a range of historical comparisons rather than produce only another estimate of the Gini coefficient.

3.3 Using information on exempt households

A natural starting point for estimating the parameters of the Singh–Maddala distribution from income-class counts and known thresholds is the multinomial maximum-likelihood approach of McDonald (1984). The intuition is straightforward. The parametric $p_g(\theta)$ provide the theoretical probability that a household belongs to each income class while the *focatico* registers, in turn, provide the corresponding counts n_g, for $g = 1, \ldots, G$. The likelihood maximization approach then seeks the

parameter values that make the observed distribution of households across income classes most likely.

The application of this approach to the *focatico* registers is complicated by left truncation, since households exempt from taxation are not observed.[6] An apparently straightforward adjustment is to divide the theoretical probability of each observed income class by $1 - \hat{p}_0$, the share of non-exempt households estimated from external sources. We refer to the resulting specification as the *naive* estimator because this adjustment does not actually incorporate the external information into the estimation of the model parameters: since $\hat{p}_0$ is treated as fixed, the denominator contributes only a constant to the objective function. The resulting optimization problem is given in Equation (5):

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{MLE}^{Naive} &= \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} -\sum_{g=1}^{G} n_g \ln\left(\frac{F(z_{g+1}|\theta) - F(z_g|\theta)}{1-\hat{p}_0}\right) \\ &= \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} -\sum_{g=1}^{G} n_g \ln\left(F(z_{g+1}|\theta) - F(z_g|\theta)\right) \\ &\qquad + \sum_{g=1}^{G} n_g \ln\left(1-\hat{p}_0\right)\end{aligned} \tag{5}$$

The naive estimator in Equation (5) does not require the fitted distribution to reproduce the externally estimated share of exempt households. A second limitation is that the adjusted class probabilities are not guaranteed to sum to one, leaving the likelihood function unnormalized. To overcome these limitations, we consider two alternatives. The first imposes the census-based estimate of the exempt share as an exact constraint on the fitted distribution. For any given values of the shape parameters α and q, this constraint determines the scale parameter β, leaving only two parameters to be estimated. This yields the *constrained maximum-likelihood estimator* in Equation (6):

$$\hat{\boldsymbol{\theta}}_{MLE}^{CML} = \underset{(\alpha,q)\in\tilde{\boldsymbol{\Theta}}}{arg\ min}\ -\sum_{g=1}^{G} n_g \times \ln(F(z_{g+1}|\theta) - F(z_g|\theta)) \tag{6}$$

where the shape parameters α and q are estimated within a constrained pa-

[6] Hong et al. (2018) consider related problems but treat censoring and truncation as alternative schemes. Moreover, their parametric estimator requires individual observations rather than class frequencies, making it unsuitable for direct application to the *focatico* registers.

rameter space $\tilde{\boldsymbol{\Theta}}$ enforcing the constraint of the fitted CDF matching exactly the external data on the left-truncation mass at the lower threshold z_1 as follows:

$$\tilde{\boldsymbol{\Theta}} = \{(\alpha, q) \ : \ F(z_1|(\alpha, \beta(\alpha, q), q) = \hat{p}_0 \ , \ (\alpha, \beta(\alpha, q), q) \in \boldsymbol{\Theta}\} \ , \ \tilde{\boldsymbol{\Theta}} \subset \boldsymbol{\Theta}.$$

Under the constrained maximum-likelihood estimator in Equation (6), the *focatico* registers are used to estimate the two shape parameters, α and q. The scale parameter, β, is instead determined by imposing an additional requirement: the fitted distribution must reproduce the share of exempt households estimated by comparing the population census with the *focatico* registers. In effect, this census-based estimate fixes one point of the cumulative income distribution. For any given values of α and q, the constraint then determines β uniquely, reducing the estimation problem from three unknowns to two. For any given values of α and q, this condition determines β uniquely.[7] The estimation problem is thus reduced from three parameters to two.

$$\hat{\beta}^{CML}_{MLE} = \beta(\hat{\alpha}^{CML}_{MLE}, \hat{q}^{CML}_{MLE}) = \frac{z_1}{((1-\hat{p}_0)^{-\frac{1}{\hat{q}^{CML}_{MLE}}} - 1)^{\frac{1}{\hat{\alpha}^{CML}_{MLE}}}}$$

A second strategy treats the households missing from the *focatico* registers as an additional income class with counts $n_0 = \hat{n}_0$. The counts for this class are then combined with those observed in the *focatico* registers, allowing all three parameters of the Singh–Maddala distribution to be estimated jointly. This yields the *complete-data estimator* in Equation (7):

$$\hat{\boldsymbol{\theta}}^{CD}_{MLE} = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{arg\ min} \ -\sum_{g=0}^{G} n_g \times \ln(F(z_{g+1}|\theta) - F(z_g|\theta)) \tag{7}$$

Unlike the constrained estimator, this approach does not require the fitted distribution to reproduce the estimated share of exempt households exactly. Instead, it uses the estimated number of exempt households as additional information when fitting the full distribution.

The distinction between the two procedures concerns the status assigned to the external information. CML requires exact agreement with the reconstructed exempt share. CD fits the additional class jointly with the taxable classes, allowing the fitted exempt share to depart from its external estimate. This flexibility may

[7] The expression for $\hat{\beta}^{CML}_{MLE}$ is obtained by substituting the Singh–Maddala CDF in Equation (3) into the constraint $F(z_1|\alpha, \beta, q) = \hat{p}_0$ and solving for β.

be attractive when census totals and tax registers are difficult to reconcile, as they often are in historical applications. It does not, however, amount to a model of the uncertainty surrounding that reconciliation: the reconstructed count is still treated as fixed in the estimation. We therefore compare both procedures and examine how the results change when the demographic reconstruction is altered.

Experience with the historical *focatico* registers suggests a few practical precautions for implementing the CML and CD estimators in Equations (6) and (7). Both can be estimated using standard numerical optimization methods, but rescaling income-class thresholds and restricting the parameter space can improve numerical stability.[8]

These restrictions have implications that extend beyond numerical optimization. First, the joint constraints $\alpha > 1$ and $q > 2/\alpha$ limit the Gini coefficient of the fitted Singh–Maddala distribution to values strictly smaller than 2/3 (Equation (4)). This upper bound is a consequence of the restrictions imposed on the model, rather than a property of the historical data. Its potential importance is illustrated by the high levels of inequality documented in historical wealth distributions, particularly when households with no taxable property are included (Alfani and Ammannati, 2017; Alfani et al., 2026). Although wealth and income are different variables, these findings suggest caution in imposing restrictions that may rule out highly unequal distributions before the estimation begins.

Second, the restriction $\alpha > 1$ has important implications for the lower tail of the income distribution. It excludes Singh–Maddala distributions whose density has its mode at zero, even though the *focatico* registers provide no direct information on the distribution of incomes among exempt households. When a substantial share of households is exempt, different combinations of parameters may fit the observed income classes similarly while implying quite different distributions for the missing population. Estimates may therefore be sensitive to the restriction $\alpha > 1$, particularly when the estimated parameter approaches this boundary. Assessing this sensitivity is important before interpreting the resulting distributional estimates.

We report point estimates without standard errors. Left truncation and income grouping reduce information but do not, by themselves, preclude model-based

[8]In our applications, the shape parameters α and q were restricted to $(0, 10]$. Since incomes in the *focatico* registers are expressed in lire, income-class thresholds can be numerically much larger than the shape parameters. We therefore divided the thresholds by a fixed constant so that all positive, finite thresholds fell within $(0, 10]$, allowing the same bounds to be imposed on the rescaled scale parameter β. Two additional restrictions, $q > 2/\alpha$ and $\alpha > 1$, ensure a finite variance and a strictly positive mode, respectively (Kleiber and Kotz, 2003). Unlike rescaling, these restrictions also constrain the range of admissible income distributions.

inference. For regular interior solutions, likelihood-based or estimating-equation methods could provide a covariance matrix for the free parameters, with uncertainty propagated to distributional indicators through the delta method. For CML, this calculation must account for the restriction linking the scale parameter to the two shape parameters and the reconstructed exempt share. A parametric bootstrap could instead generate multinomial class counts and repeat the estimation, preserving the income thresholds and reproducing the construction of the exempt class. Its finite sample performance would need to be assessed, particularly when estimates approach the boundaries imposed by the parameter restrictions, where neither conventional normal approximations nor an unmodified bootstrap can be assumed reliable.

Holding the demographic inputs fixed would make such inference conditional on the reconstructed household totals and exempt shares. Propagating their uncertainty would require additional evidence or explicit assumptions about their reliability and their relationship to the register counts. Neither approach would, by itself, quantify reconstruction errors arising from distributional misspecification or departures from the assumed recording process. We therefore limit the empirical analysis to point estimates and sensitivity to alternative demographic assumptions, without assessing the statistical significance of differences across areas or years.

3.4 Implementation and goodness of fit

A fitted distribution should first be judged against what the sources actually preserve. We compare observed and fitted household proportions across income classes using total variation (TV), a distance ranging from zero for identical distributions to one for distributions with no overlap. Two comparisons serve different purposes.

The first compares the fitted probabilities with the completed table, including the reconstructed exempt class:

$$\mathrm{TV}_C(\theta) = \frac{1}{2}\left[\left|\frac{\widehat{n}_0}{\widehat{N}} - p_0(\theta)\right| + \sum_{g=1}^{G}\left|\frac{n_g}{\widehat{N}} - p_g(\theta)\right|\right]. \tag{8}$$

The second considers only the households recorded in the *focatico*, comparing observed and fitted proportions conditional on being taxable:

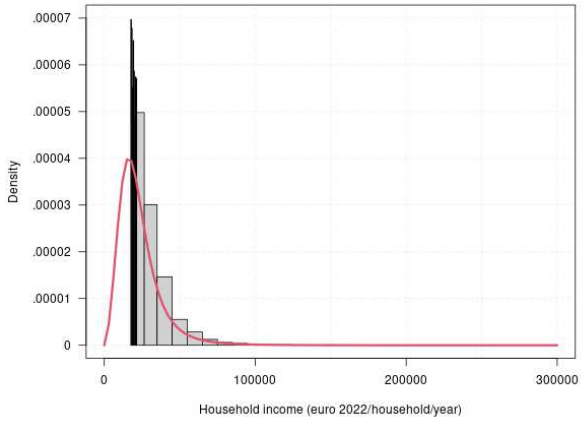
$$\mathrm{TV}_F(\theta) = \frac{1}{2}\sum_{g=1}^{G}\left|\frac{n_g}{n} - \frac{p_g(\theta)}{1 - F(z_1|\theta)}\right|. \tag{9}$$

The subscripts refer to the completed table and the *focatico*, respectively.[9]

For comparing the fit of CML and CD to the fiscal evidence, TV_F is the more direct criterion. TV_C additionally assesses agreement with the externally reconstructed exempt class, which CML reproduces by construction. Neither statistic tests the distribution of incomes within classes. In particular, a good fit to the completed table cannot verify the shape of the unobserved lower tail.

A simulation illustrates what the estimators can recover when their assumptions hold. We use parameters fitted to SHIW 2022 incomes to generate a sample of $N = 9{,}616$ household incomes from a Singh–Maddala distribution. We then remove the lowest incomes and group the remaining observations into 36 classes, including an open-ended highest class. The resulting table resembles the information available in a *focatico* register, while the full simulated sample provides a benchmark against which to assess the reconstruction. The proportion excluded is known in this exercise, rather than estimated from a separate historical source.

Figure 3: Reconstructing income distributions from simulated *focatico* data



	Benchmark	CD eq. (7)	CML eq. (6)
Gini	0.317	0.317	0.317
MLD	0.175	0.178	0.178
Theil	0.174	0.171	0.171
Top 1%	0.046	0.043	0.043
H	0.087	0.095	0.095
PG	0.024	0.028	0.028
TV_F	-	0.019	0.019
TV_C	-	0.013	0.013

Note: The red curve represents the Singh–Maddala distribution fitted to SHIW 2022 microdata and used to generate a sample of $N = 9,616$ household incomes. The complete simulated sample serves as the "true" benchmark. To reproduce the information available in the *focatico* registers, the lowest 40% of incomes are excluded and the remaining observations grouped into 36 income classes, including an open-ended highest class. The CML and CD estimates are obtained from these grouped data and the known share of excluded households. H and PG denote the poverty headcount ratio and poverty gap, respectively. Source: Authors' calculations based on SHIW 2022 data.

Figure 3 compares the benchmark indicators with those recovered by CML and

[9] An equivalent measure has long been used in the analysis of residential segregation (Duncan and Duncan, 1955). In the empirical calculations, sparsely populated upper classes are pooled, starting from the open-ended class, until each pooled group contains at least five observed households. The corresponding fitted probabilities are aggregated over the same intervals. This pooling can conceal discrepancies within the combined classes and thus affect the resulting TV values, while it avoids artifically inflating the value of the statistic with comparisons against empty cells.

CD. Both estimators reproduce the Gini coefficient of 0.317 to three decimal places, and their reported estimates coincide at that precision. The other inequality measures and the top income share are also close to the benchmark. Poverty measures show somewhat larger differences: the headcount rises from 0.087 in the complete sample to 0.095 in the reconstruction, while the poverty gap rises from 0.024 to 0.028. The exercise shows that class frequencies and information on the excluded population can preserve enough information for a useful reconstruction, even when individual incomes are no longer available.

This is a demonstration under favourable conditions. The simulated incomes follow the same distribution used for estimation, and the excluded share is known. A single exercise of this kind does not establish how the estimators perform under different income distributions, inaccurate population totals, or other forms of omission. Those are precisely the concerns that return when we move from simulation to historical records. The value of the approach is that it makes the passage from incomplete fiscal evidence to a full distribution explicit. The resulting curve supports comparisons of income levels, poverty, inequality, and changes across the distribution, all grounded in the same reconstruction. How much confidence to place in those comparisons depends on the source, the fit, and the assumptions governing what remains unobserved. The Modena application brings these elements together and examines what happens when one consequential piece of the historical reconstruction is changed.

4 Modena, 1875–76

The Modena registers offer an opportunity to examine differences in household incomes at a geographical scale rarely accessible to historical research. The surviving series covers 1873–1923 and distinguishes between *città* and *campagna*: the city and its surrounding countryside within the same municipality. We use the first two consecutive years for which the registers survive in full, 1875 and 1876, to reconstruct distributions for these two areas and for Modena as a whole. The exercise brings into view differences that municipal averages conceal, while allowing us to examine how the reconstruction depends on information beyond the tax records themselves.

The historical literature describes a setting marked by widespread deprivation.[10]

[10] Azzi (1970) provides a rich account of economic and social conditions in the Modena area. The *tassa di famiglia* receives only a passing mention on p. 55, without further discussion. Nor does it appear in his account of the unpublished sources consulted for this meticulous study of

Drawing on local newspapers and contemporary accounts, Muzzioli (1993, p. 119) portrays a province in which "niches of comfortable wealth coexisted with large pockets of unspeakable misery. One third of the population was afflicted by unemployment and insufficient access to food." The evidence subsequently collected by the Jacini Agrarian Inquiry offers an equally stark image: in 1881, it reported that "animals enjoy better health than men" (Jacini, 1884). These accounts concern the province and a broader period than the two years examined here, but help establish the circumstances in which the registers were compiled. The province's limited industrial base in the early 1870s was concentrated in activities such as tobacco processing, hat and cap manufacture, and silk reeling (Guaraldi, 2009, p. 18). In the countryside, the agricultural modernization described by Rinaldi (1998) would gather momentum only in the following decade. The registers allow us to examine this landscape of deprivation more closely, measuring the distance between city and countryside and the differences within each.

Table 1 presents the recorded households alongside the income intervals used in the reconstruction. The schedule comprised 36 classes, beginning at 600 lire per year and becoming progressively wider towards the top. The final class began at 100,000 lire and had no upper limit. Most taxpayers were concentrated in the lower classes, particularly in the countryside; almost all households in the highest classes were recorded in the city. These are distributions of taxpayers, however, rather than of all resident households. The families below the exemption threshold, whose absence matters most for the measurement of poverty, do not appear in the table.

The number of recorded households fell from 8,288 in 1875 to 7,824 in 1876. This change cannot by itself establish a decline in the resident population: movements across the exemption threshold or changes in assessment practices could also alter the number of taxpayers. Since these registers belong to the early years of the tax, administrative adjustment is a particular concern. The comparisons that follow therefore describe changes in reconstructed distributions; attributing them entirely to changes in economic circumstances would require further documentary evidence.

To move from taxpayers to the resident population, we need an estimate of the number of households missing from each register. The demographic reconstruction gives municipal totals of 12,065 households in 1875 and 12,066 in 1876. Census information does not directly establish household counts for the two territorial divisions used by the tax administration. Our benchmark therefore assumes the municipal

local history.

Table 1: Recorded households by income class: Modena, 1875–76

Class g	Income interval (annual lire)	1875 City	1875 Countryside	1875 Total	1876 City	1876 Countryside	1876 Total
1	600–800	1883	1132	3015	1648	1014	2662
2	800–1,000	683	448	1131	646	346	992
3	1,000–1,200	542	301	843	579	259	838
4	1,200–1,600	283	271	554	218	199	417
5	1,600–2,000	375	246	621	348	242	590
6	2,000–2,400	395	156	551	358	216	574
7	2,400–2,800	198	86	284	217	120	337
8	2,800–3,200	220	41	261	179	79	258
9	3,200–3,600	120	22	142	180	39	219
10	3,600–4,000	92	19	111	76	26	102
11	4,000–4,400	115	6	121	106	7	113
12	4,400–5,000	84	6	90	97	6	103
13	5,000–6,000	79	8	87	107	8	115
14	6,000–7,000	90	7	97	108	7	115
15	7,000–8,000	37	1	38	36	4	40
16	8,000–9,000	75	4	79	57	5	62
17	9,000–10,000	10	2	12	15	2	17
18	10,000–12,000	68	0	68	56	3	59
19	12,000–14,000	39	2	41	50	2	52
20	14,000–16,000	20	0	20	32	0	32
21	16,000–18,000	28	1	29	27	1	28
22	18,000–20,000	5	0	5	12	0	12
23	20,000–24,000	25	0	25	18	0	18
24	24,000–28,000	11	0	11	18	0	18
25	28,000–32,000	11	0	11	8	0	8
26	32,000–36,000	5	0	5	4	0	4
27	36,000–40,000	0	1	1	4	1	5
28	40,000–45,000	7	0	7	7	0	7
29	45,000–50,000	3	0	3	2	0	2
30	50,000–55,000	7	0	7	5	0	5
31	55,000–60,000	0	0	0	2	0	2
32	60,000–70,000	3	0	3	3	0	3
33	70,000–80,000	0	0	0	0	0	0
34	80,000–90,000	5	0	5	6	0	6
35	90,000–100,000	2	0	2	2	0	2
36	100,000 and over	8	0	8	7	0	7
Recorded households, n		5528	2760	8288	5238	2586	7824

Note: City and countryside correspond to the registers' *città* and *campagna* divisions. Income intervals follow the interpretation of the schedule adopted in the estimation: each listed income value is treated as a lower boundary, and the next value as the upper boundary. Intervals are lower-inclusive and upper-exclusive; the final class is open-ended. Households below 600 lire are absent from the recorded distribution. Source: Authors' calculations from the Modena *focatico* registers. Class boundaries: ASCMO, *Atti a stampa del Consiglio Comunale*, minutes of 21 January 1875, Appendix IX, p. 131.

average household size, approximately 4.7 persons, in both areas. This yields 6,520 city households and 5,545 countryside households in 1875, and 6,510 and 5,556, respectively, in 1876.

Comparing these totals with the recorded taxpayers implies that approximately half of countryside households were exempt, against 15–20% in the city. The reconstruction consequently relies much more heavily on the unobserved lower tail for the countryside. These proportions are estimates derived from demographic information, rather than counts supplied by the tax administration. Section 4.2 examines how the results change when average household size is allowed to differ between the two areas.

We estimate a separate Singh–Maddala distribution for each area and year, and another for the municipality as a whole, using CML (Eq. (6)) as the benchmark estimator. The municipal-level distribution is fitted to the pooled class counts rather than constructed as a mixture of the two fitted local distributions; its indicators therefore need not reproduce the corresponding aggregation of the local estimates. Results obtained with CD are reported in Table 3 in Appendix B. Mean incomes in Table 2 are expressed in current lire; poverty comparisons and the figures use constant 1875 prices, obtained using Istat's historical consumer price index (Istat, 2026a).[11]

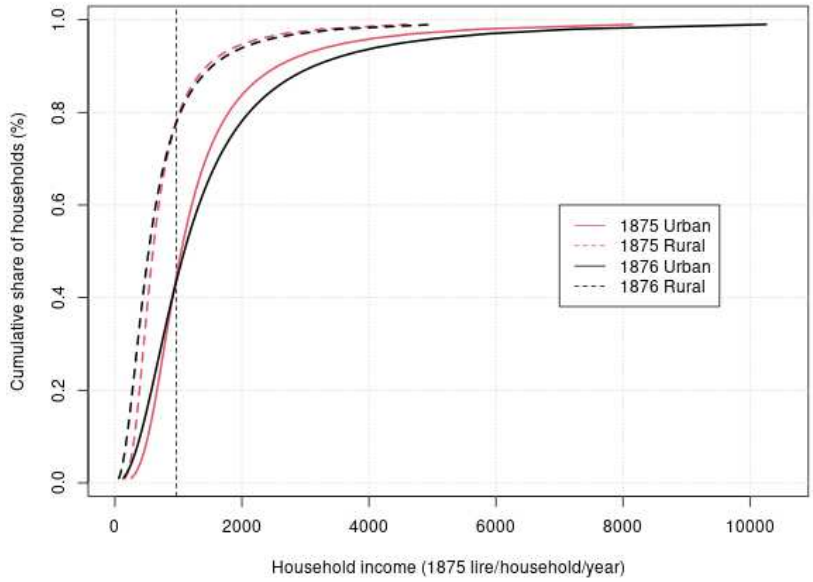
Figure 4 offers a first comparison of the reconstructed distributions. At any income threshold, the height of a cumulative distribution curve gives the proportion of households below it. The vertical distance between the city and countryside curves thus measures the difference in poverty incidence at that threshold. Across the income range displayed, the countryside curves lie above those for the city: low household incomes were substantially more prevalent outside the city.

For an illustrative poverty comparison, we use a threshold derived from Amendola et al. (2017), approximately 970 lire per household per year at 1875 prices. This is a benchmark rather than a locally established subsistence threshold. It lies above the 600-lire exemption threshold, so the estimated poor include both households absent from the registers and taxpayers in the lowest income classes. The incidence of poverty depends on how many households were omitted; its depth and severity also depend on the incomes assigned to them by the fitted distribution.

Table 2 quantifies the differences. In 1876, estimated mean household income

[11]The index rises from 0.912 in 1875 to 0.965 in 1876 (1913 = 1), implying inflation of approximately 5.8%. Incomes recorded in 1876 are therefore multiplied by $0.912/0.965 \simeq 0.94508$ to express them in 1875 lire. The same national index is applied to the city and countryside.

Figure 4: Household income distributions in the city and countryside: Modena, 1875–76



Note: Cumulative distribution functions estimated using CML (Eq. (6)). Incomes are expressed in constant 1875 lire. The vertical line marks the illustrative poverty threshold of approximately 970 lire per household per year, derived from Amendola et al. (2017). Source: Authors' calculations.

was approximately twice as high in the city as in the countryside, while poverty incidence was lower by about 34 percentage points. Yet the two Gini coefficients were almost identical, at 0.482 and 0.485. Similar relative inequality could therefore coexist with markedly different levels of deprivation. The distinction also matters for changes over time. In the countryside, the poverty headcount barely changed between 1875 and 1876, while the poverty gap and its squared measure increased. The proportion below the threshold remained almost unchanged, but the estimated shortfall from that threshold became larger and more unevenly distributed.

These comparisons give quantitative content to the distance between the city and its surrounding countryside. They also show why a single inequality index cannot adequately describe that distance. Their interpretation nevertheless requires caution. The fit statistics in Table 2 reveal discrepancies between observed and fitted class frequencies, while even a close fit to taxpayers would not validate the reconstruction below the exemption threshold. As discussed in Section 3.3, we report no standard errors, so the statistical significance of differences across areas or years has not been assessed. Separately, the CD estimates in Appendix B produce different temporal patterns for some indicators, particularly in the city. The

Table 2: Income, poverty, and inequality: Modena, 1875–76

	Modena		City		Countryside	
	1875	1876	1875	1876	1875	1876
Mean household income (current lire)	1,217	1,346	1,484	1,769	846	863
Poverty						
Headcount	0.592	0.597	0.445	0.435	0.781	0.778
Poverty gap	0.239	0.285	0.136	0.171	0.350	0.412
Squared poverty gap	0.127	0.174	0.059	0.091	0.190	0.262
Inequality						
Gini	0.452	0.517	0.414	0.482	0.407	0.485
Mean logarithmic deviation	0.354	0.493	0.290	0.413	0.279	0.419
Theil	0.460	0.584	0.401	0.514	0.391	0.520
Atkinson (0.5)	0.178	0.229	0.152	0.200	0.148	0.203
Atkinson (2)	0.476	0.640	0.390	0.550	0.375	0.557
Top 1% income share	0.117	0.132	0.110	0.123	0.109	0.124
Top 10% income share	0.367	0.407	0.347	0.385	0.344	0.386
Population coverage						
Estimated households, $\widehat{N}$	12065	12066	6520	6510	5545	5556
Recorded households, n	8288	7824	5528	5238	2760	2586
Estimated exempt share, $\widehat{p}_0$	0.313	0.352	0.152	0.195	0.502	0.535
Goodness of fit						
TV_F	0.202	0.197	0.280	0.242	0.103	0.182
TV_C	0.139	0.128	0.237	0.195	0.051	0.085

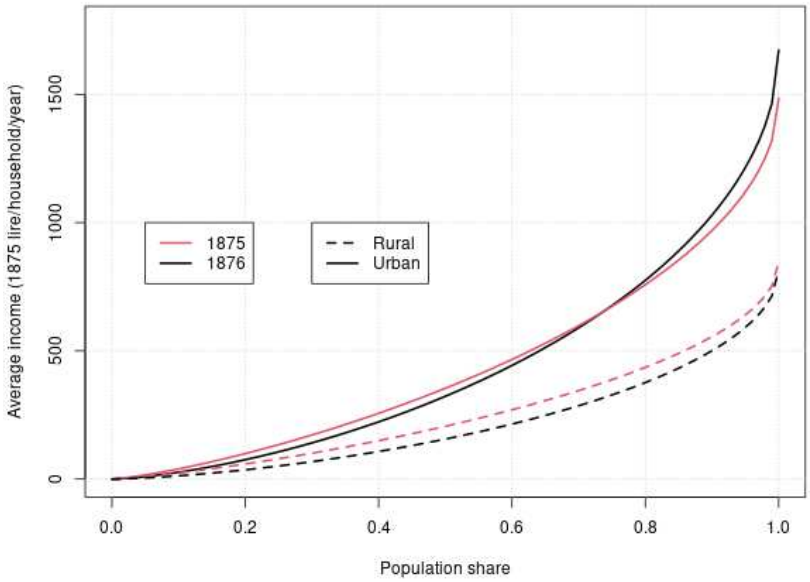
Note: Estimates use CML (Eq. (6)), with separate fits for each area and year. Mean incomes are in current lire. Poverty measures follow Foster et al. (1984), using a threshold of approximately 970 lire per household per year at 1875 prices (Amendola et al., 2017). The parameters in parentheses for the Atkinson indices indicate inequality aversion. The estimated exempt share is $\widehat{p}_0 = 1 - n/\widehat{N}$. The goodness-of-fit measures TV_C and TV_F are defined in Eqs. (8) and (9), respectively. Corresponding CD estimates appear in Table 3 in Appendix B. Source: Authors' calculations.

reported changes thus depend on the estimation method as well as on the distributional assumptions and demographic inputs.

To compare income levels and inequality together, Figure 5 presents generalized Lorenz curves (Shorrocks, 1983). These multiply the cumulative income shares shown by an ordinary Lorenz curve by mean income. A curve lying above another consequently identifies a distribution preferred by all increasing, inequality-averse social welfare functions. This permits comparisons that account for both the resources available and their distribution, without selecting a particular inequality index. Ordinary Lorenz curves are reported in Figure 8 in Appendix B.

The estimated city curves lie above their countryside counterparts in both years,

Figure 5: Generalized Lorenz curves: Modena, 1875–76



Note: Curves are computed from distributions estimated using CML (Eq. (6)), with incomes expressed in constant 1875 lire. Source: Authors' calculations.

reinforcing the contrast in income levels already apparent in the poverty comparisons. Comparisons over time are less uniform. In the city, the curves cross at approximately the 70th percentile: the 1875 curve is higher below that point, while the 1876 curve is higher above it. The increase in mean income therefore does not produce an unambiguous generalized Lorenz ranking. In the countryside, the 1876 curve lies below the 1875 curve throughout the plotted range. The slight increase in nominal mean income reported in Table 2 becomes a decline when expressed at constant prices. These are comparisons of fitted curves, rather than statistical tests of dominance, but they direct attention to the parts of the distribution in which the estimated changes occurred.

4.1 Where do incomes rise and fall?

Growth incidence curves (GICs) make those changes directly visible. Introduced by Ravallion and Chen (2003), they measure income growth at corresponding percentiles of two distributions. Figure 6 presents them for Modena as a whole and separately for the city and countryside. All three curves slope upwards: higher percentiles experienced more favourable changes than lower ones. For the municipality as a whole, estimated real income fell by roughly 25% around the 20th percentile

and rose by nearly 10% around the 90th. Such large movements within a single year warrant caution, particularly at the lower percentiles, where incomes are inferred below the exemption threshold. They nevertheless illustrate how changes concealed by averages can be located within the reconstructed distribution.

The figure also decomposes each GIC into changes in mean income and changes in the income of each percentile relative to that mean, together with their interaction. This helps distinguish an increase in income from an improvement in relative position: a percentile can gain in absolute terms while falling behind the average, or lose income while gaining relative to it.

Let $Q_t(u)$ denote income at percentile u in year t, with $u \in (0,1)$, and let μ_t denote mean income, both at constant 1875 prices. Writing $Q_t(u) = \mu_t r_t(u)$ separates the mean from relative income, $r_t(u) = Q_t(u)/\mu_t$. If $\gamma = \mu_{1876}/\mu_{1875} - 1$ is mean income growth and $\rho(u) = r_{1876}(u)/r_{1875}(u) - 1$ is the change in relative income, then

$$\mathrm{GIC}(u) = \frac{Q_{1876}(u)}{Q_{1875}(u)} - 1 = \gamma + \rho(u) + \gamma\rho(u). \tag{10}$$

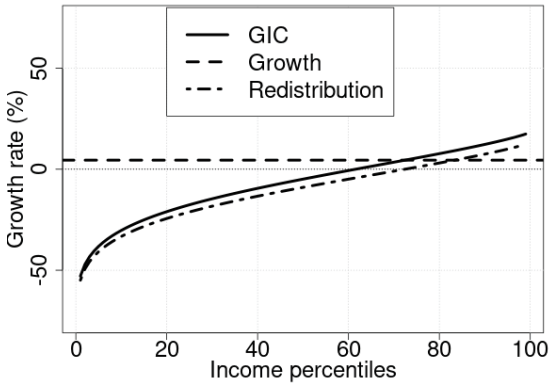
The decomposition is exact. Its connection with changes in the shape of the distribution follows from the fact that relative income at a percentile is the slope of the Lorenz curve.[12]

The horizontal line in each panel represents mean income growth, γ. The component labelled "Redistribution", $\rho(u)$, shows whether income at a percentile rose or fell relative to the mean. Redistribution here describes a change in the shape of the distribution; it does not identify a transfer between households or the effect of a particular policy. The interaction term, $\gamma\rho(u)$, accounts for the fact that relative incomes and the mean change together. It is not plotted separately, so the two displayed components need not add up to the GIC.

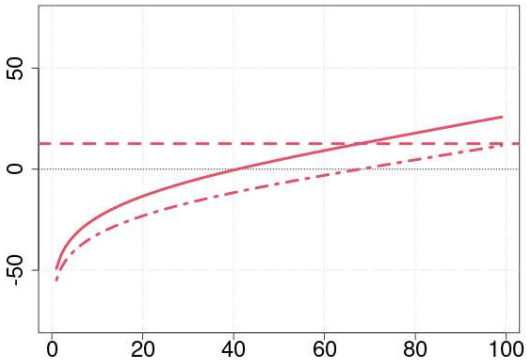
The city offers a clear example of why the distinction matters. Some middle percentiles experienced positive income growth while losing ground relative to the mean: their incomes rose, but less rapidly than the city average. The countryside shows the opposite possibility. With mean income declining, some upper-middle percentiles improved relative to the mean even though their incomes fell. The decomposition thus distinguishes changes in purchasing power from changes in relative income that a conventional GIC combines.

[12] Writing the Lorenz curve at any percentile u for the period in time t as $LC_t(u) = \mu_t^{-1} \int_0^u Q_t(v)\,dv$ gives $LC_t'(u) = Q_t(u)/\mu_t = r_t(u)$. For the related relationship between growth incidence curves and changes in the Lorenz curve, see Ferreira et al. (2019). Appendix A gives the corresponding expressions for the Singh–Maddala distribution.

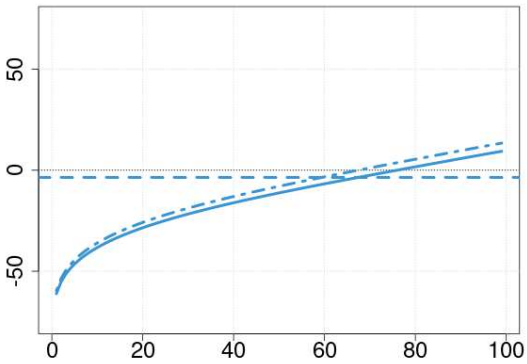
Figure 6: Growth incidence curves and their decomposition: Modena, 1875–76



(a) Modena



(b) City



(c) Countryside

Note: Curves are computed from distributions estimated using CML (Eq. (6)), with incomes expressed in constant 1875 lire. "Growth" denotes γ and "Redistribution" denotes $\rho(u)$ in Eq. (10). The interaction term $\gamma\rho(u)$ is not plotted separately. Source: Authors' calculations.

These curves compare positions in the distribution, not the fortunes of particular households. The household occupying a given percentile in 1876 need not be the one occupying it in 1875. Non-anonymous GICs instead follow income changes according to households' initial positions (Grimm, 2007; Bourguignon, 2011). The nominative registers offer a potential route towards such comparisons, as discussed in Section 2. Record linkage would not, however, resolve all the difficulties: incomes remain grouped, and exempt households are absent. Our estimates reconstruct the separate distributions for each year, not the joint distribution of household incomes across years. Studying mobility and the persistence of poverty would require both longitudinal linkage and further assumptions about the income trajectories that the registers leave unobserved.

4.2 Sensitivity to the demographic reconstruction

The comparison between city and countryside depends on how many households are assigned to each area, including those absent from the tax records. The benchmark assumes an average of 4.7 persons per household in both. This is convenient, but it need not describe a setting in which household formation and living arrangements differed between the city and its agricultural surroundings.

We therefore consider an alternative informed by evidence on household size in nearby municipalities. Among the local cases available to us, Calderara di Reno provides an agricultural comparator with an average household size of 6.5 persons.[13] Applying this value to the countryside population of Modena reduces its estimated number of households. Keeping the municipal household total unchanged and assigning the remainder to the city implies an average city household size of approximately 3.8 persons.

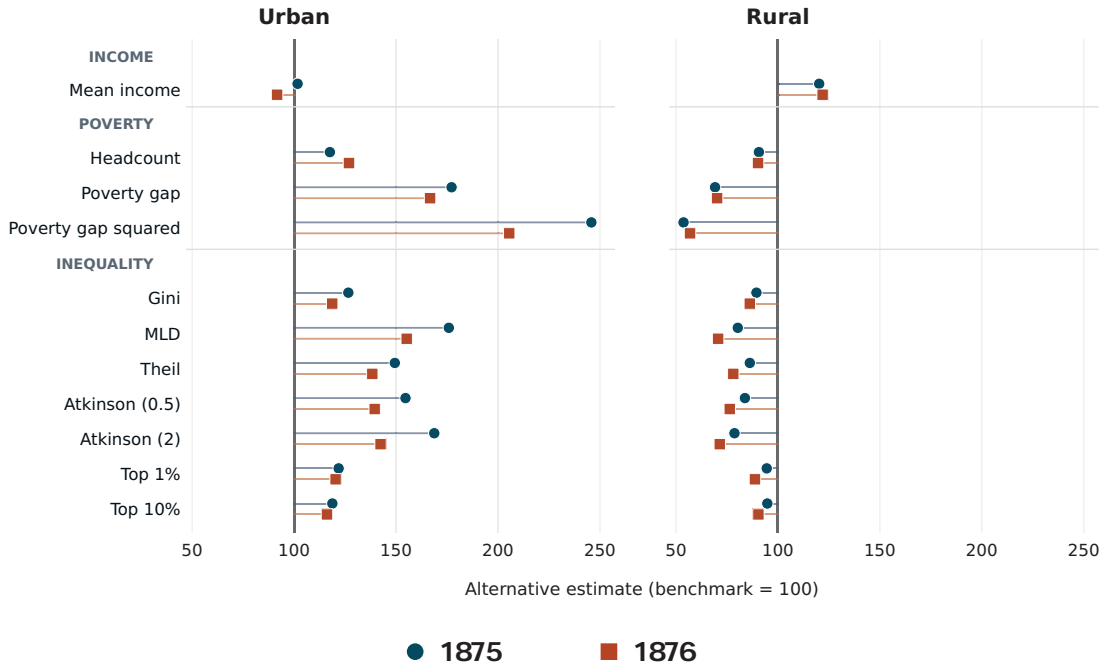
The consequences begin with the allocation of exempt households. In 1875, the estimated exempt share rises from 15.2% to 31.4% in the city and falls from 50.2% to 31.1% in the countryside. In 1876, the corresponding changes are from 19.5% to 35.0% and from 53.5% to 35.5%. Nothing in the recorded class frequencies has changed. What changes is the demographic information used to establish how much of each distribution lies outside the registers. Under CML, this information constrains the fitted distribution as a whole, so its effects need not be confined to

[13]Calderara di Reno was selected as a nearby, predominantly agricultural municipality broadly comparable to the Modena countryside, for which population census data allowed a first approximation to average household size. The resulting value provides an indicative benchmark for the sensitivity exercise.

the unobserved lower tail.

Figure 7 compares the resulting CML estimates with the benchmark. Each benchmark value is normalized to 100, separately by indicator, area, and year. The municipal estimates remain unchanged because neither the pooled tax records nor the municipal household totals change.

Figure 7: Sensitivity to alternative household-size assumptions: Modena, 1875–76



Note: For each indicator, area, and year, the benchmark estimate in Table 2 is normalized to 100. Markers show estimates obtained assuming an average countryside household size of 6.5 persons, with the city household count calculated as the residual from the unchanged municipal total. All estimates use CML (Eq. (6)). Source: Authors' calculations.

The estimated income gap narrows. In the city, mean household income changes little in 1875, from 1,484 to 1,509 lire, but falls from 1,769 to 1,621 lire in 1876. The nominal increase between the two years is consequently reduced from approximately 19% to 7%. Estimated countryside mean incomes rise under the alternative assumption in both years.

The change in poverty estimates is equally consequential. In 1876, the city headcount rises from 43.5% to 55.2%, while the countryside headcount falls from 77.8% to 70.2%. The difference between the two areas narrows from about 34 to 15 percentage points. These comparisons hold the household poverty threshold fixed to isolate the effect of changing the estimated population counts. They do not adjust the threshold for differences in household size or composition, which would be a

separate requirement for comparing living standards across differently constituted households.

For inequality, the alternative reconstruction changes the relative position of the two areas. The almost identical Gini coefficients estimated for 1876 under the benchmark, 0.482 in the city and 0.485 in the countryside, become 0.572 and 0.418. A comparison suggesting similar relative inequality thus becomes one of substantially greater inequality in the city. The demographic assumption affects not only the magnitude of an estimated gap, but also the conclusion drawn from it.

Some findings survive both reconstructions: mean household income remains higher in the city, and poverty incidence remains higher in the countryside. Their magnitude, and the comparison of relative inequality, are much less stable. The exercise does not determine which demographic allocation is correct. It identifies the historical information needed to make the comparison more secure. Reconstructing household numbers and territorial boundaries is therefore part of estimating the income distribution itself. The unusually fine geographical detail of the registers makes the comparison possible; establishing who is missing from each register determines how confidently it can be interpreted.

5 Conclusions

The starting point of this paper was a source whose potential for distributional history remains largely unexplored. The *focatico* registers recorded household economic circumstances year after year, across municipalities with widely differing populations and economic structures. Their survival offers the prospect of reconstructing local income distributions over periods for which even national evidence is scarce. Their value extends beyond adding observations to the history of inequality. By bringing income levels, poverty, and affluence into the same reconstruction, they can change both the geographical scale of distributional history and the questions it is able to address.

Realizing this potential requires understanding how the records were produced. Exemptions, income classes, and open-ended upper brackets determine what can be learned from the entries and what must be inferred from other evidence. This paper has shown how those features can enter an estimation framework that combines fiscal records with demographic information to reconstruct the household income distribution. The connection between archival work and statistical mod-

elling is therefore concrete: the historian's reconstruction of the assessment rules helps determine the form of the estimator. Although developed for the *focatico*, the framework is not tied to this particular tax. Other historical or contemporary fiscal and administrative sources may present the same combination of grouped incomes, exclusion below a threshold, and an open upper class. Where the recording rules can be established and external information helps quantify the omitted population, the approach offers a basis for reconstructing distributions from such records.

The Modena application illustrates what becomes possible when the reconstruction moves below the municipal level. In both demographic scenarios considered, estimated mean household income remains higher in the city and poverty more widespread in the surrounding countryside. The size of these differences, however, changes substantially, as does the comparison of relative inequality. Under the benchmark reconstruction, nearly identical Gini coefficients in 1876 coexist with sharply different mean incomes and poverty rates. Growth incidence curves and their decomposition reveal another distinction: income at a given percentile can rise while falling behind the mean, or decline while improving relative to it. These comparisons show what is gained by reconstructing the distribution as a whole. A history of average incomes or of a single inequality index would leave important parts of the experience unexamined.

The application also clarifies the limits of the reconstruction. Some arise from the statistical problem itself: class counts do not reveal incomes within brackets, and a close fit to taxpayers cannot establish the distribution of incomes among exempt households. The assumed distribution and the restrictions placed on its parameters supply information that the source does not contain. Other limitations concern the historical evidence available in a particular application. At Modena, assumptions about household size change the estimated allocation of exempt families between city and countryside; the short interval and the early administration of the tax also complicate the interpretation of changes between 1875 and 1876. Elsewhere, small populations, broader income classes, or changes in assessment practices may impose different constraints. The method is general, but the strength of the historical conclusions depends on the evidence supporting each reconstruction.

A further step is to quantify uncertainty in the reconstructed distributions and their summary indicators. This requires distinguishing estimation variability under the maintained model from uncertainty in the demographic inputs and sensitivity to assumptions about the unobserved incomes. Where the demographic evidence

supports an explicit model of its uncertainty, that uncertainty could be propagated jointly with the variability of the register counts, preserving the relationship between reconstructed household totals and exempt shares. Future work could explore the extension of our proposed approach under a Bayesian formulation, combining the grouped-income likelihood with an explicit model for the demographic evidence and prior distributions for unknown quantities. Either approach requires assumptions about the reliability of the external information. Neither would, by itself, resolve evasion, exemptions unrelated to income, or mismatches between fiscal and census populations. Sensitivity analysis would therefore remain necessary alongside any reported standard errors or credible intervals.

The larger prospect is a history of local income distributions that can follow economic change across places, years, and income levels. Consider the gradual abolition of the *macinato* tax between 1879 and 1884 (Marongiu, 2009). Annual reconstructions could help establish how local income distributions evolved around the reform, whether changes were concentrated among poorer households, and whether cities and their surrounding countryside followed different paths. They would not, by themselves, identify the causal effects of abolition, but would provide distributional evidence against which historical explanations could be assessed. The two Modena years offer only a glimpse of what longer series might reveal. The promise of the *focatico* lies in making local income distributions a recurring object of historical observation, allowing us to investigate how economic change reached different parts of the population.

References

Abramitzky, R., R. Mill, and S. Pérez (2020). Linking individuals across historical sources: A fully automated approach. *Historical Methods 53*(2), 94–111.

A'Hearn, B. (2004). A restricted maximum likelihood estimator for truncated height samples. *Economics & Human Biology 2*(1), 5–19.

A'Hearn, B., N. Amendola, and G. Vecchi (2016). On historical household budgets. *Rivista di storia economica 32*(2), 137–176.

A'Hearn, B., G. Mancini, and G. Vecchi (Eds.) (2026a). *La tassa di famiglia e la vulnerabilità economica degli italiani: una ricostruzione storica*. Roma-Bari: Editori Laterza. Forthcoming.

A'Hearn, B., G. Mancini, and G. Vecchi (2026b). Rethinking italy's distributional history, 1861–2022. Manuscript prepared for the special issue of Rivista di storia economica.

Alfani, G. and F. Ammannati (2017). Long-term trends in economic inequality: The case of the Florentine state, c. 1300–1800. *The Economic History Review 70*(4), 1072–1102.

Alfani, G., M. Di Tullio, and M. Fochesato (2026). Wealth inequality and epidemics in the republic of venice (1400–1800). *The Economic History Review 79*(2), 811–849.

Allen, R. C. (2019). Class structure and inequality during the industrial revolution: Lessons from england's social tables, 1688–1867. *The Economic History Review 72*(1), 88–125.

Amendola, N., F. Salsano, and G. Vecchi (2017). *Poverty*. In *Measuring Wellbeing: A History of Italian Living Standards*. Oxford University Press.

Azzi, G. (1970). *Modena 1859–1898: condizioni economiche, sociali, politiche. Il movimento clericale dopo l'unificazione, le origini del socialismo modenese*. Modena: STEM.

Bank of Italy (2024). *Indagine sui bilanci delle famiglie italiane*. Fascicolo IBF collana *Statistiche - Indagine sui bilanci delle famiglie italiane 2022*. Banca d'Italia.

Betti, G., F. Gagliardi, A. Lemmi, and V. Verma (2012). Sub-national indicators of poverty and deprivation in Europe: Methodology and applications. *Cambridge Journal of Regions, Economy and Society 5*(1), 149–162.

Blanchet, T., J. Fournier, and T. Piketty (2022). Generalized pareto curves: Theory and applications. *Review of Income and Wealth 68*(1), 263–288.

Bourguignon, F. (2011). Non-anonymous growth incidence curves, income mobility and social welfare dominance. *The Journal of Economic Inequality 9*(4), 605–627.

Burkhauser, R. V., N. Hérault, S. P. Jenkins, and R. Wilkins (2018). Top incomes and inequality in the UK: Reconciling estimates from household survey and tax return data. *Oxford Economic Papers 70*(2), 301–326.

Casu, S. (2026). A new source for the history of italian living standards: the "tassa di famiglia", 1868-1973. Manuscript in preparation for Rivista di Storia Economica / Italian Review of Economic History.

Chiaiese, D. (2024). Provincial estimates of the Italian value-added in the Liberal Age, 1871–1911. *Rivista di storia economica* (1), 3–42.

Chianese, S. and G. Vecchi (2017). Household budgets. In *Measuring Wellbeing: A History of Italian Living Standards*, pp. 492–523. New York: Oxford University Press.

Chotikapanich, D., W. E. Griffiths, G. Hajargasht, W. Karunarathne, and D. Rao (2018). Using the gb2 income distribution. *Econometrics 6*(2), 21.

Ciccarelli, C. and S. Fenoaltea (2013). Through the magnifying glass: Provincial aspects of industrial growth in post-Unification Italy. *The Economic History Review 66*(1), 57–85.

Cowell, F. A. (2011). *Measuring inequality*, Volume 3. Oxford University Press Oxford.

Cowell, F. A. and E. Flachaire (2015). Statistical methods for distributional analysis. In A. B. Atkinson and F. Bourguignon (Eds.), *Handbook of Income Distribution*, Volume 2, pp. 359–465. Elsevier.

Duncan, O. D. and B. Duncan (1955). A methodological analysis of segregation indexes. *American sociological review 20*(2), 210–217.

Eckernkemper, T. and B. Gribisch (2021). Classical and Bayesian inference for income distributions using grouped data. *Oxford Bulletin of Economics and Statistics 83*(1), 32–65.

Federico, G., A. Nuvolari, and M. Vasta (2019). The origins of the Italian regional divide: Evidence from real wages, 1861–1913. *The Journal of Economic History 79*(1), 63–98.

Fenoaltea, S. (2019). Spleen: The failures of the cliometric school. *Annals of the Fondazione Luigi Einaudi 53*(2), 5–24.

Ferreira, F. H., S. Firpo, and A. F. Galvao (2019). Actual and counterfactual growth incidence and delta lorenz curves: estimation and inference. *Journal of Applied Econometrics 34*(3), 385–402.

Fiore Melacrinis, F. M. S. and E. Felice (2026). Diverging consequences of protectionism: Tariffs and agricultural GDP across Italian provinces, 1871–1911. *Cliometrica 20*(3), 807–845.

Foster, J., J. Greer, and E. Thorbecke (1984). A class of decomposable poverty measures. *Econometrica: journal of the econometric society*, 761–766.

Gastwirth, J. L. (1972). The estimation of the lorenz curve and gini index. *The Review of Economics and Statistics 54*(3), 306–316.

Grimm, M. (2007). Removing the anonymity axiom in assessing pro-poor growth. *The Journal of Economic Inequality 5*(2), 179–197.

Guaraldi, E. (2009). *Breve storia della Provincia di Modena 1859–2009*. Roma: Ediesse.

Hajargasht, G., W. E. Griffiths, J. Brice, D. P. Rao, and D. Chotikapanich (2012). Inference for income distributions using grouped data. *Journal of Business & Economic Statistics 30*(4), 563–575.

Heblich, S., S. J. Redding, and D. M. Sturm (2020). The making of the modern metropolis: Evidence from london. *The Quarterly Journal of Economics 135*(4), 2059–2133.

Hong, L., G. Alfani, C. Gigliarano, and M. Bonetti (2018). giniinc: A stata package for measuring inequality from incomplete income and survival data. *The Stata Journal 18*(3), 692–715.

ISTAT (2023). *Le statistiche de l'Istat sulla povertà. Anno 2022*. Statistiche report, Italian National Insitute of Statistics (Istat).

Istat (2026a). Il valore della moneta in italia dal 1861 al 2025. Nota informativa e tavole statistiche.

Istat (2026b). Indagine sul reddito e le condizioni di vita (EU-SILC). Informazioni sulla rilevazione. Accessed 25 October 2026.

Jacini, S. (1884). *Atti della Giunta per la inchiesta agraria e sulle condizioni della classe agricola*, Volume 1. Forzani.

Jenkins, S. P. (2009). Distributionally-sensitive inequality indices and the GB2 income distribution. *Review of Income and Wealth 55*(2), 392–398.

Jenkins, S. P. (2024). Getting the measure of inequality. *Oxford Open Economics 3*(Supplement_1), i156–i166.

Jorda, V., J. M. Sarabia, and M. Jäntti (2021). Inequality measurement with grouped data: parametric and non-parametric methods. *Journal of the Royal Statistical Society Series A: Statistics in Society 184*(3), 964–984.

Kakwani, N. C. and N. Podder (1976). Efficient estimation of the lorenz curve and associated inequality measures from grouped observations. *Econometrica 44*(1), 137–48.

Kleiber, C. and S. Kotz (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*. Hoboken, NJ: John Wiley & Sons.

Korinek, A., J. A. Mistiaen, and M. Ravallion (2007). An econometric method of correcting for unit nonresponse bias in surveys. *Journal of Econometrics 136*(1), 213–235.

Kuznets, S. (1953). *Shares of Upper Income Groups in Income and Savings*. New York: National Bureau of Economic Research. Assisted by Elizabeth Jenks.

Logan, T. D. and J. M. Parman (2017). The national rise in residential segregation. *The Journal of Economic History 77*(1), 127–170.

Maas, H. and M. S. Morgan (2012). Observation and observing in economics. *History of Political Economy 44*(Supplement 1), 1–24.

Marongiu, G. (2009). La tassa sul macinato: un nome vecchio per un'imposta nuova. *Rivista di Storia dell'Agricoltura 49*(2), 25–98.

McDonald, J. B. (1984). Some generalized functions for the size distribution of income. *Econometrica 52*(3), 647–663.

Milanovic, B., P. H. Lindert, and J. G. Williamson (2011). Pre-industrial inequality. *The Economic Journal 121*(551), 255–272.

Modalsli, J. (2015). Inequality in the very long run: Inferring inequality from data on social groups. *The Journal of Economic Inequality 13*(2), 225–247.

Muzzioli, G. (1993). *Modena.* Storia delle città italiane. Roma-Bari: Laterza.

Persky, J. (1992). Retrospectives: Pareto's Law. *Journal of Economic Perspectives 6*(2), 181–192.

Ravallion, M. and S. Chen (2003). Measuring pro-poor growth. *Economics letters 78*(1), 93–99.

Rinaldi, A. (1998). Le rotazioni agrarie nel modenese dalla fine del settecento all'età giolittiana. *Rivista di Storia dell'Agricoltura 38*(2), 115–151.

Rossi, M., L. Latino, and G. Vecchi (2017). Vulnerability. In *Measuring Wellbeing: A History of Italian Living Standards*, pp. 415–453. New York: Oxford University Press.

Salsano, F. (2026). The household tax in the italian fiscal system. Manuscript in preparation for Rivista di Storia Economica / Italian Review of Economic History.

Serra, G. (2014). An uneven statistical topography: The political economy of household budget surveys in late colonial ghana, 1951–1957. *Canadian Journal of Development Studies 35*(1), 9–27.

Shorrocks, A. F. (1983). Ranking income distributions. *Economica 50*(197), 3–17.

Singh, S. K. and G. S. Maddala (1976). A function for size distribution of incomes. *Econometrica 44*(5), 963–70.

Soresina, M. (2001). *Conoscere per amministrare: Luigi Bodio. Statistica, economia e pubblica amministrazione.* Milano: FrancoAngeli.

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society: Series B (Methodological) 38*(3), 290–295.

United Nations Economic Commission for Europe (2011). *Canberra Group Handbook on Household Income Statistics* (2nd ed.). United Nations.

Wachter, K. W. and J. Trussell (1982). Estimating historical heights. *Journal of the American Statistical Association 77*(378), 279–293.

A Mathematical appendix

This appendix reports a set of technical results that complement the analysis in the main text. The moments of the Singh-Maddala distribution can be computed as Kleiber and Kotz 2003, Chapter 6

$$E[y^k|\alpha,\beta,q] = \frac{\beta^k \Gamma(1+\frac{k}{\alpha})\Gamma(q-\frac{k}{\alpha})}{\Gamma(q)} \ , \ -\alpha < k < \alpha q \ ,$$

with $\Gamma(x)$ denoting the gamma function[14] and exist for the k-th order whenever $-\alpha < k < \alpha q$ holds. The mode of the distribution can be respectively computed as (Kleiber and Kotz, 2003, Chapter 6)

$$y_{mode}(\alpha,\beta,q) = \begin{cases} \beta\left(\frac{\alpha-1}{\alpha q+1}\right)^{\frac{1}{\alpha}} \ , \text{ if } \alpha > 1 \\ 0 \ , \text{ if } 0 < \alpha \leq 1 \end{cases}$$

The Lorenz curve of this distribution representing the share of total incomes $LC(u|\alpha,q) \in [0,1]$ held by the bottom proportion $u \in [0,1]$ of the population ordered by income, admits a closed-form expression making it practical to compute income shares (e.g., see Kleiber and Kotz 2003, Chapter 6):

$$LC(u|\alpha,q) = \frac{\Gamma(1+q)}{\Gamma\left(1+\frac{1}{\alpha}\right)\Gamma\left(q-\frac{1}{\alpha}\right)} \int_0^{\nu} t^{\frac{1}{\alpha}}(1-t)^{q-\left(\frac{\alpha+1}{\alpha}\right)} dt \ , \ \nu \equiv \frac{[(1-u)^{-\frac{1}{q}}-1]}{1+[(1-u)^{-\frac{1}{q}}-1]} \ , \ u \in (0,1). \tag{11}$$

In practice, computing $LC(u|\alpha,q)$ under this formulation is made simple by noting that it corresponds to the CDF of a beta distribution as $F^{Beta}\left(\nu|1+\frac{1}{\alpha},q-\frac{1}{\alpha}\right)$ a function available in most statistical software packages[15]. The corresponding generalized Lorenz curve is then obtained as $GLC(u|\alpha,\beta,q) = LC(u|\alpha,q) \times E[y]$.

Widely used indices belonging to the Generalized Entropy family can be computed in parametric form under the assumption of a Singh-Maddala distribution as

[14] The gamma function is formally defined as $\Gamma(x) = \int_0^{\infty} t^{x-1}e^{-t}dt \ , \ x > 0$.

[15] The CDF of a beta distribution $F^{Beta}(x|a,b)$ corresponds to the regularized incomplete beta function defined as

$$F^{Beta}(x|a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^x t^{a-1}(1-t)^{b-1}dt$$

follows Jenkins (2009) and Chotikapanich et al. (2018):

$$GEI_a(\alpha,\beta,q) = \begin{cases} -\frac{1}{2} + \left(\frac{\Gamma(1-\frac{1}{\alpha})\Gamma(q+\frac{1}{\alpha})\Gamma(1+\frac{1}{\alpha})\Gamma(q-\frac{1}{\alpha})}{2\Gamma(q)^2}\right) \ , \text{ if } a=-1 \\ \gamma(1+\frac{1}{\alpha}) + \gamma(q-\frac{1}{\alpha}) - \gamma(q) - \frac{\psi(1)}{\alpha} + \frac{\psi(q)}{\alpha} \ , \text{ if } a=0 \\ \frac{\psi(1+\frac{1}{\alpha})}{\alpha} - \frac{\psi(q-\frac{1}{\alpha})}{\alpha} - \gamma(1+\frac{1}{\alpha}) - \gamma(q-\frac{1}{\alpha}) + \gamma(q) \ , \text{ if } a=1 \\ -\frac{1}{2} + \frac{\Gamma(q)\Gamma(1+\frac{2}{\alpha})\Gamma(q-\frac{2}{\alpha})}{2\Gamma(1+\frac{1}{\alpha})^2\Gamma(q-\frac{1}{\alpha})^2} \ , \text{ if } a=2 \end{cases} \tag{12}$$

where $\gamma(x) = \log(\Gamma(x))$ denotes the log-gamma function, and $\psi(x) = \Gamma'(x)/\Gamma(x)$ denotes the digamma function.

Another particularly appealing virtue of the Singh-Maddala distribution is that it has an analytically tractable quantile function, defined as (Kleiber and Kotz, 2003, Chapter 6)

$$F_y^{-1;SM}(u|\alpha,\beta,q) = \beta[(1-u)^{-\frac{1}{q}} - 1]^{\frac{1}{\alpha}} \ , \ u \in (0,1). \tag{13}$$

In Equation (13), given a percentile u of the distribution, $F_y^{-1;SM}(u|\alpha,\beta,q)$ returns the corresponding income quantile. This result is particularly useful for constructing growth incidence curves (GICs). To construct a GIC, two distributions referring to different years are required, denoted here, for simplicity, by $t-1$ and t. The first distribution is defined by the parameter vector $(\alpha^{(t-1)}, \beta^{(t-1)}, q^{(t-1)})$, and the second by $(\alpha^{(t)}, \beta^{(t)}, q^{(t)})$. The GIC for the period $(t-1, t)$ can then be computed as the growth rate of each percentile $u \in (0,1)$ of the income distribution:

$$GIC_y^{SM}(u|(\alpha^{(t-1)},\alpha^{(t)});(\beta^{(t-1)},\beta^{(t)});(q^{(t-1)},q^{(t)})) = \frac{F_y^{-1;SM}(u|\alpha^{(t)},\beta^{(t)},q^{(t)})}{F_y^{-1;SM}(u|\alpha^{(t-1)},\beta^{(t-1)},q^{(t-1)})} - 1$$
$$= \left(\frac{\beta^{(t)}}{\beta^{(t-1)}}\right) \times \left(\frac{[(1-u)^{-\frac{1}{q^{(t)}}} - 1]^{\frac{1}{\alpha^{(t)}}}}{[(1-u)^{-\frac{1}{q^{(t-1)}}} - 1]^{\frac{1}{\alpha^{(t-1)}}}}\right) - 1 \ , \ u \in (0,1).$$

This formulation of the GIC is particularly useful for analyzing changes in the income distribution over long periods of time, an especially relevant feature given the time span covered by the *focatico* registers. In addition to allowing growth rates to be computed locally along the income distribution, it also makes it possible to decompose these rates into an income growth effect and a distributional effect. The income growth effect captures changes in the scale of the distribution, which affect all incomes in the same proportion and may be especially large over long time intervals. This corresponds simply to the growth rate of mean incomes

$\gamma = (E[y|\alpha^{(t)}, \beta^{(t)}, q^{(t)}]/E[y|\alpha^{(t-1)}, \beta^{(t-1)}, q^{(t-1)}]) - 1$. The redistribution component corresponds to the changes in the shape of the income distribution leaving aside changes in income levels. These can be computed from the GIC as [16] :

$$\rho(u|(\alpha^{(t-1)}, \alpha^{(t)}); (\beta^{(t-1)}, \beta^{(t)}); (q^{(t-1)}, q^{(t)})) = \frac{GIC(u|(\alpha^{(t-1)}, \alpha^{(t)}); (\beta^{(t-1)}, \beta^{(t)}); (q^{(t-1)}, q^{(t)})) - \gamma}{1 + \gamma}$$

for $u \in (0, 1)$ and can be interpreted as the growth rate of the income share of the u-th percentile in the distribution (i.e., changes in the slope of the Lorenz curve at this percentile).

Finally, it is important to note that these expressions devise a three-component decomposition of the GIC following:

$$\begin{aligned} GIC(u|(\alpha^{(t-1)}, \alpha^{(t)}); (\beta^{(t-1)}, \beta^{(t)}); (q^{(t-1)}, q^{(t)})) = \gamma \\ + \rho(u|(\alpha^{(t-1)}, \alpha^{(t)}); (\beta^{(t-1)}, \beta^{(t)}); (q^{(t-1)}, q^{(t)})) \\ + \gamma \times \rho(u|(\alpha^{(t-1)}, \alpha^{(t)}); (\beta^{(t-1)}, \beta^{(t)}); (q^{(t-1)}, q^{(t)})) \end{aligned}$$

such that the component ρ captures redistributive changes net of both the income growth changes and net of any changes in income levels strictly due to the joint interaction between growth and redistribution.

[16] For the closely related relationship between growth incidence curves and changes in the Lorenz curve, see Ferreira et al. (2019).

B Additional results

This appendix presents supplementary estimates and figures for Modena and its urban and rural populations.

Table 3: Poverty and inequality indicators for Modena, 1875 and 1876.

Indicator	**1875** Modena	**1876** Modena	**1875** Urban	**1876** Urban	**1875** Rural	**1876** Rural
Mean income (current lire/household)	1,209	1,161	1,602	1,412	843	861
Absolute poverty						
Headcount	0.602	0.653	0.449	0.535	0.781	0.778
Poverty Gap	0.257	0.298	0.173	0.202	0.354	0.416
Poverty Gap Squared	0.143	0.174	0.091	0.102	0.194	0.268
Inequality						
Gini	0.470	0.478	0.473	0.446	0.412	0.492
MLD	0.387	0.404	0.394	0.342	0.287	0.433
Theil	0.491	0.506	0.497	0.450	0.398	0.532
Atkinson (0.5)	0.191	0.197	0.193	0.173	0.151	0.208
Atkinson (2)	0.518	0.539	0.526	0.461	0.385	0.573
Top 1%	0.120	0.122	0.121	0.115	0.109	0.125
Top 10%	0.377	0.382	0.379	0.363	0.346	0.390
Population coverage						
N	12065	12066	6520	6510	5545	5556
$\sum_{g=1}^{G} n_g$	8288	7824	5528	5238	2760	2586
$F_y(z_1)$	0.313	0.352	0.152	0.195	0.502	0.535
Goodness-of-fit						
TV_F	.197	.209	.246	.254	.105	.183
TV_C	.147	.144	.233	.217	.055	.087

Source: Authors' calculations.

Note: The table presents select estimated features of the income distributions obtined under the Singh-Maddala distribution using $\hat{\boldsymbol{\theta}}^{CD}_{MLE}$ parameter estimates for the 1875-1876 *focatico* registers of Modena. The 1875 absolute poverty line derived from the estimates in Amendola et al. (2017) is used for estimating poverty and is valued at approximately 970 lire (household/year) in 1875 lire. The headcount, poverty gap, and poverty gap squared indicators correspond to the Foster et al. (1984) class of poverty measures. Population size from census sources and population coverage of the *focatico* sources are also presented. The TV_F and TV_C goodness-of-fit statistics are computed as detailed in Subsection 3.3.

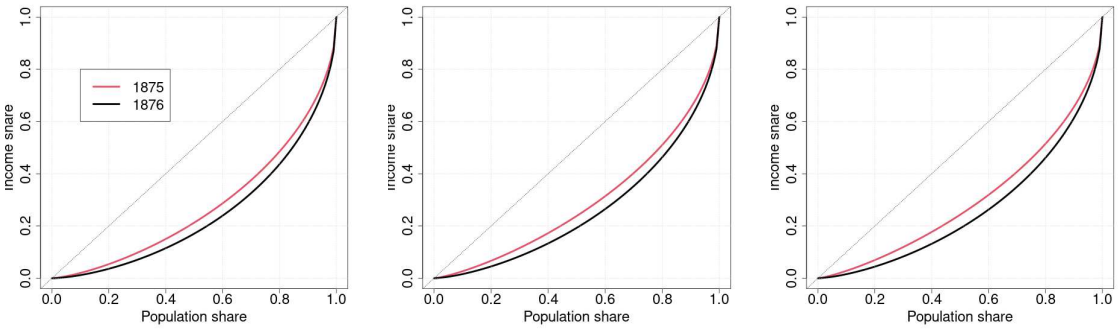
Table 4: Poverty and inequality indicators for Modena, 1875 and 1876. Alternative population splits across Urban and Rural.

Indicator	**1875** Modena	**1876** Modena	**1875** Urban	**1876** Urban	**1875** Rural	**1876** Rural
Mean income (current lire/household)	1,209	1,161	1,493	1,127	1,019	1,043
Absolute poverty						
Headcount	0.602	0.653	0.551	0.664	0.705	0.706
Poverty Gap	0.257	0.298	0.277	0.281	0.243	0.295
Poverty Gap Squared	0.143	0.174	0.177	0.151	0.104	0.155
Inequality						
Gini	0.470	0.478	0.556	0.443	0.367	0.422
MLD	0.387	0.404	0.596	0.338	0.227	0.302
Theil	0.491	0.506	0.673	0.446	0.340	0.413
Atkinson (0.5)	0.191	0.197	0.265	0.172	0.126	0.157
Atkinson (2)	0.518	0.539	0.743	0.455	0.300	0.407
Top 1%	0.120	0.122	0.143	0.115	0.103	0.111
Top 10%	0.377	0.382	0.436	0.362	0.327	0.351
Population coverage						
N	12065	12066	8060	8054	4005	4012
$\sum_{g=1}^{G} n_g$	8288	7824	5528	5238	2760	2586
$F_y(z_1)$	0.313	0.352	0.314	0.350	0.311	0.355
Goodness-of-fit						
TV_F	.197	.209	.216	.233	.102	.193
TV_C	.147	.144	.163	.155	.071	.130

Source: Authors' calculations.

Note: The table presents select estimated features of the income distributions obtined under the Singh-Maddala distribution using $\hat{\boldsymbol{\theta}}^{CD}_{MLE}$ parameter estimates for the 1875-1876 *focatico* registers of Modena. The total population count of Modena obtained from census sources is split into urban and rural household populations by dividing the total rural population count into households of average size of 6.5 components assuming demographic similarities to the nearby municipality of Calderara di Reno, while the urban household population is obtained as the difference between the resulting number of households for the rural population and the total number of households of Modena. This population split amounts to an average household size for urban households of approximately 3.8 components. The 1875 absolute poverty line derived from the estimates in Amendola et al. (2017) is used for estimating poverty and is valued at approximately 970 lire (household/year) in 1875 lire. The headcount, poverty gap, and poverty gap squared indicators correspond to the Foster et al. (1984) class of poverty measures. Population size from census sources and population coverage of the *focatico* sources are also presented. The TV_F and TV_C goodness-of-fit statistics are computed as detailed in Subsection 3.3.

Figure 8: Lorenz curves for Modena households in the second half of the nineteenth century.



(a) Modena. (b) Modena - Urban. (c) Modena - Rural.

Source: Authors' calculations.
Note: The figure presents estimated Lorenz curves under the Singh-Maddala distribution using $\hat{\boldsymbol{\theta}}_{MLE}^{CML}$ parameter estimates for the 1875-1876 *focatico* registers of the total population of Modena, and for urban and rural Modena population sub-groups separately.

RECENT PUBLICATIONS BY *CEIS Tor Vergata*

Leaving Precarity, Not Academia: Gender Gaps among Early-Career Researchers in Italy
Marianna Brunetti, Annalisa Fabretti and Mariangela Zoli
CEIS *Research Paper*, 630 November 2026

How Many Vulnerability Lines Do We Need? Measuring Vulnerability as Expected Poverty
Nicola Amendola, Federico Belotti, Domenico Moramarco and Giovanni Vecchi
CEIS *Research Paper*, 629 November 2026

Investment Mandates and the Value of Distribution Timing in Currency Portfolios Choice
Mattia Alfero, Stefano Grassiosimo Petracchi and Francesco Violante
CEIS *Research Paper*, 628 October 2026

Exchange-Rate Regimes and the Behaviour of Exporters
Cosimo Petracchi, Luca Riva and Marco S. Petterson
CEIS *Research Paper*, 627 August 2026

Pricing options on illiquid assets using liquid market benchmarks: an application to energy markets
Federico Aluigi, Lucia Caramellino, Paolo Pigato and Edoardo Scrima
CEIS *Research Paper*, 626 August 2026

Image Considerations in Second-Hand Adoption: A Vignette Experiment
Luca Congiu, Luisa Lorè and Mariangela Zoli
CEIS *Research Paper*, 625 July 2026

Communication Failure: The Hidden Face of The Tragedy of the Commons
Philippe Colo and Guillaume Pommey
CEIS *Research Paper*, 624 July 2026

Female Mayors and Public Procurement
Erminia Florio and Giancarlo Spagnolo
CEIS *Research Paper*, 623 July 2026

Optimal Structure of Penalties with Judgment-Proof Injurers
Guillaume Pommey
CEIS *Research Paper*, 622 June 2026

Flexibility Versus Security in Agency Contracts with Moral Hazard
Lorenzo Bozzoli and Guillaume Pommey
CEIS *Research Paper*, 621 June 2026

DISTRIBUTION

Our publications are available online at www.ceistorvergata.it

DISCLAIMER

The opinions expressed in these publications are the authors' alone and therefore do not necessarily reflect the opinions of the supporters, staff, or boards of CEIS Tor Vergata.

COPYRIGHT

Copyright © 2026 by authors. All rights reserved. No part of this publication may be reproduced in any manner whatsoever without written permission except in the case of brief passages quoted in critical articles and reviews.

MEDIA INQUIRIES AND INFORMATION

For media inquiries, please contact Barbara Piazzi at +39 06 72595584 or by e-mail at barbara.piazzi@ceis.uniroma2.it. Our web site, www.ceistorvergata.it, contains more information about Center's events, publications, and staff.

DEVELOPMENT AND SUPPORT

For information about contributing to CEIS Tor Vergata, please contact at +39 06 72595587 or by e-mail at segr.ceis@economia.uniroma2.it