Balancing population stocks and flows in the Italian demographic estimation system: a proposal

Diego Zardetto[1], Marco Di Zio[1], Marco Fortini[1]

Abstract

Estimates of population counts ('stocks') should be consistent with counts of demographic events ('flows'). In particular, the Demographic Balancing Equation (DBE) should be satisfied. Istat's modernised production system, underpinned by statistical registers that integrate survey and administrative data, seems ideally positioned to overcome the challenge of producing timely, reliable, and coherent estimates of demographic stocks and flows. However, in Italy, estimates of stocks and flows entering the DBE are currently obtained independently. Population size estimates at subsequent reference times are provided by the integrated system formed by the Permanent Census and the Base Register of Individuals (BRI), whereas birth, death, and migration figures are derived from municipal civil registries. Therefore, owing to sampling and non-sampling errors affecting these "raw" estimates of stocks and flows, the DBE is not trivially fulfilled. Consistency of official estimates can, however, be attained through a suitable macro-integration process that optimally adjusts both stocks and flows to ensure compliance with the DBE. To this end, we propose to use 'balancing' methods that National Statistical Institutes routinely adopt to reconcile large systems of national accounts. We designed and implemented a system that can handle this task at scale, thus allowing Istat to achieve stock and flow consistency for all the subnational (i.e. domain) estimates that must be officially disseminated in accordance with Italian and European statistical regulations. Moreover, simulation results show that our balancing approach determines improved estimates of population counts: besides gaining consistency, they exhibit lower bias and variance as compared to "raw" ones. Finally, we provide considerations and guidance for using balanced population counts as control totals to adjust individual weights of the BRI.

Keywords: Data integration, macro-integration, demographic balancing equation, census estimates, demographic events, data quality, coherence and consistency, statistical registers.

DOI: 10.1481/ISTATRIVISTASTATISTICAUFFICIALE_2.2022.04

1 Diego Zardetto (zardetto@istat.it); Marco Di Zio (dizio@istat.it); Marco Fortini (fortini@istat.it), Italian National Institute of Statistics – Istat.

The views and opinions expressed are those of the authors and do not necessarily reflect the official policy or position of the Italian National Institute of Statistics - Istat.

The authors would like to thank the anonymous reviewers for their comments and suggestions, which enhanced the quality of this article.

1. Background information

The Italian National Institute of Statistics - Istat is currently engaged in completing the transition of its production processes to the model envisioned by the modernisation programme launched some years ago (Istat, 2016). The backbone of the new production system is the infrastructure formed by the 'Integrated System of Statistical Registers' (ISSR), namely a system of connected registers used as a reference for all the statistical activities carried out by Istat (Alleva *et al.*, 2019; Alleva *et al.*, 2021). A pivotal role within the ISSR is played by the 'Base Register of Individuals' (BRI), a comprehensive statistical register that integrates and stores data gathered from disparate sources about people usually or temporarily residing in Italy.

One of the most important achievements of the new statistical production system is the modernisation of the Italian population census. Until the year 2011, traditional population censuses were conducted in Italy every ten years, and their outcomes were employed to correct municipal civil registries once a decade. Starting from 2018, the Italian population census is no longer a complete enumeration survey but rather results from a *twofold* large-scale sample survey that is carried out each year, whose outcomes are integrated with the BRI. Istat has named this new census design 'Permanent Census'. The Permanent Census involves two simultaneous sample surveys: the 'A' survey and the 'L'survey. The L component relies on a list sample: its main objective is to observe variables that are either of insufficient quality or not available at all in the BRI. The A component is instead based on an area sample: it is designed to provide yearly estimates of the under-coverage and over-coverage rates of the BRI, evaluated at national and local levels for different sub-population profiles defined by variables like 'sex', 'age class', 'nationality'.

The new production system enables Istat to deliver official population size estimates more frequently than happened before through traditional censuses. To this end, crude estimates of population counts derived from the BRI are corrected by using individual weights that are functions of the over- and under-coverage probabilities estimated through the A sample linked with the BRI (see Falorsi, 2017, and Righi *et al.*, 2021). This exercise counteracts coverage errors of the BRI, yielding estimates of population counts which, albeit improved, we still consider *"raw"* in this paper.

Official estimates of population counts (*stocks*) should be consistent with civil registry figures about demographic events (*flows*), in such a way that the resulting data system exactly fulfills the Demographic Balancing Equation (DBE).

The DBE states that the population counts at time $t + 1$ must be equal to the population counts at time t plus the sum of the natural increase and the net migration that occurred between t and $t + 1$:

$$P^{(t+1)} = P^{(t)} + N + M \tag{1}$$

where the natural increase, N, is the difference between births and deaths, and the net migration, M, is the difference between immigrants and emigrants:

$$\begin{cases} N = B - D \\ M = I - E \end{cases} \tag{2}$$

In Italy, *raw* estimates of stocks and flows entering the DBE are currently obtained *independently*. Birth, death, and migration figures are derived from municipal civil registries[2], whereas population size estimates at subsequent reference times are provided by the BRI and the Permanent Census, as already noted. Therefore, owing to sampling and non-sampling errors affecting raw estimates of stocks and flows, the DBE is *not* trivially satisfied. Consistency of *official* estimates can, however, be achieved through a suitable process that simultaneously adjusts *both* stocks and flows to ensure compliance with the DBE.

Note that the complexity of this task is inextricably linked to the sample-survey nature of the Permanent Census. Indeed, census estimates of population counts are now affected by sampling uncertainty, which prevents solving the stocks and flows consistency problem with the classical '*flows-first*' approach typically adopted in countries where traditional censuses are conducted. In the *flows-first* approach – called 'component method' (Eurostat, 2003) – the census provides "true" population counts at year t, to which estimates of population flows for the period $[t, t + 1]$ are added, thus obtaining estimates of population stocks at year $t + 1$ that satisfy the DBE by *construction*[3].

2 Indeed, the BRI is not able to correctly register demographic events (births, deaths, internal and cross-border migrations).

3 Note, however, that the flows-first approach has a serious weak point, as any errors or bias in estimating flows will be carried forward from timepoint to timepoint during the intercensal period.

To solve the stock and flow consistency problem in the integrated estimation system formed by the Permanent Census and the BRI, we propose to adjust raw estimates by using methods that are commonly adopted inside National Statistical Institutes (NSI) for *balancing* large systems of national accounts. Indeed, the National Accounts divisions of most NSIs routinely make use of independent initial estimates, which (*i*) are characterised by different degrees of reliability (as is also the case of demographic stocks and flows), and (*ii*) must be adjusted to satisfy a large set of accounting identities (as is the system of DBEs associated to any partition of the overall population into estimation domains). Relevant papers on modelling demographic figures as accounting matrices are Rees (1979), Stone and Corbit (1997), Bryant and Graham (2013), and Bryant and Graham (2015).

We designed and implemented a macro-integration procedure that ensures consistency between official (*i.e. adjusted*) estimates of demographic stocks and flows using balancing methods. The system we developed can handle this task at scale, thus allowing Istat to achieve stock and flow consistency for all the subnational (*i.e.* domain) estimates that must be officially disseminated in accordance with Italian and European statistical regulations. Moreover, simulation results show that our balancing approach also determines improved estimates of population counts: besides gaining consistency, they exhibit lower bias and variance as compared to raw ones.

While our proposal is directly targeted at improving the quality of official demographic statistics at *macro-level*, we see significant scope for leveraging its results at *micro-level* too. Once balanced population counts are obtained through our macro-integration procedure, one could think of using them as control totals to adjust the individual weights tied to records of the BRI. This would determine two beneficial effects. First, estimates of population counts derived from adjusted BRI weights would exactly match officially disseminated estimates, and thereby satisfy the DBE for all the domains addressed by our procedure. Second, BRI weights would receive a second layer of protection against bias, "borrowing strength" from both official population counts disseminated the year before and balanced demographic flows.

The rest of the paper is structured as follows. Section 2 provides the mathematical formulation of the problem, highlighting its computational complexity and the challenges of its extension to domain estimates, along with technical countermeasures put in place by our system. The same Section also introduces the main statistical properties of balanced estimates under ideal conditions. Section 3 illustrates the results of an experimental study based on simulations, where the statistical properties of balanced estimates are investigated under realistic (*i.e.* non-ideal) settings. Section 4 discusses details for the downstream application of our proposal to the BRI. Lastly, Section 5 offers some final remarks.

2. Problem formulation

We formalise the task of finding a system of consistent estimates of demographic stocks and flows as a constrained optimisation problem. This is accomplished along the lines of (Stone *et al.*, 1942) and (Byron, 1978), by suitably reformulating the models and algorithms introduced in those classical papers.

Given *initial* (=raw) estimates of all the aggregates entering the demographic balancing equations (1) defined for all the geographic areas of a given territorial level, we search for *final* estimates that are *balanced*, *i.e.* (*i*) satisfy all the DBEs, and (*ii*) are *as close as possible* to the initial estimates. Therefore, the objective function to be minimised is an appropriate distance metric between the final and initial estimates, while the constraints acting on the final estimates are the area-level DBEs. Moreover, we adopt a *weighted* distance metric such that aggregates whose initial estimates are more *reliable* will tend to be changed less.

Let us suppose we have initial estimates of the population size of k Italian regions[4] U_i at times t and $t + 1$, as well as initial estimates of births, deaths, and natural increase that occurred for each region between time t and $t + 1$[5]:

$$\begin{cases} P^{(t)} = \left(P_1^{(t)}, \dots, P_k^{(t)}\right)' \\ P^{(t+1)} = \left(P_1^{(t+1)}, \dots, P_k^{(t+1)}\right)' \\ B = (B_1, \dots, B_k)' \\ D = (D_1, \dots, D_k)' \\ N = (N_1, \dots, N_k)' \end{cases} \tag{3}$$

Moreover, let us suppose we have initial estimates of the *Migration Flows Matrix F*, whose generic element F_{ij} equals the number of people who *moved* from region i to region j between time t and $t + 1$:

4 The subpopulation notation U_i is adopted because, as we will explain later, “regions” can actually be any partition of the Italian population U, *e.g.* obtained by crossing variables ‘provinces’, ‘sex’, and ‘age classes’.

5 In equation (3), and throughout the whole paper, we denote matrix transposition with a single quote, $'$, to avoid confusion with time superscripts, ^{t}.

$$F = \begin{pmatrix} 0 & F_{1,2} & \cdots & F_{1,k} & F_{1,k+1} \\ F_{2,1} & 0 & \cdots & F_{2,k} & F_{2,k+1} \\ \cdots & \cdots & 0 & \cdots & \cdots \\ F_{k,1} & F_{k,2} & \cdots & 0 & F_{k,k+1} \\ F_{k+1,1} & F_{K+1,2} & \cdots & F_{k+1,k} & 0 \end{pmatrix} \tag{4}$$

Note that the $(k + 1)^{th}$ row and column of F represent migrations from and to any territory *outside* the nation, thus $k + 1$ means '*abroad*'. Note also that matrix F is not, in general, symmetric nor antisymmetric.

Let us indicate with M the *Net Migration Matrix*, whose generic element M_{ij} equals the count of people who *immigrated* in region i from region j *minus* the count of people who *emigrated* from region i to region j, $M_{ij} = F_{ji} - F_{ij}$:

$$M = \begin{pmatrix} 0 & M_{1,2} & \cdots & M_{1,k} & M_{1,k+1} \\ -M_{1,2} & 0 & \cdots & M_{2,k} & M_{2,k+1} \\ \cdots & \cdots & 0 & \cdots & \cdots \\ -M_{1,k} & -M_{2,k} & \cdots & 0 & M_{k,k+1} \\ -M_{1,k+1} & -M_{2,k+1} & \cdots & -M_{k,k+1} & 0 \end{pmatrix} \tag{5}$$

Note that matrix M is *antisymmetric* and actually equal to minus twice the antisymmetric part[6] of F:

$$\begin{cases} M = -M' \\ M = F' - F = -2F \end{cases} \tag{6}$$

Furthermore, let us assume we can attach to each *atomic* initial estimate involved in (3), (4), and (5) a measure of *reliability,* $R \in [0, \infty)$. These reliability measures could be either based on proper statistical measures (*e.g.* proportional to inverse estimated variances) or derived from an assessment made by subject matter experts. For instance, we will indicate the reliability measure of a generic element M_{ij} of the Net Migration Matrix M as $R[M_{ij}]$. Note that $R[\cdot] \to \infty$ will signal *absolute reliability*, and thus *prevent* the corresponding initial atomic estimates from being altered.

Lastly, let us denote *raw estimates* with a *tilde* (*e.g.* $\widetilde{M}_{ij}$) and *balanced estimates* with a *circumflex hat* (*e.g.* $\widehat{M}_{ij}$). Given (3), (4), and (5), we define the objective function, L, for the constrained optimisation problem as follows:

6 Any square matrix, F, can be decomposed into a symmetric part, F^S, and an antisymmetric part, F^A, such that: $F^S = (F^S)'$, $F^A = -(F^A)'$, and $F = F^S + F^A$. Of course: $F^S = (F + F')/2$, $F^A = (F - F')/2$.

$$\begin{aligned}
L\left(\hat{P}^{(t+1)}, \hat{P}^{(t)}, \hat{B}, \hat{D}, \hat{N}, \hat{F}, \hat{M}\right) \\
= \sum_{i=1}^{k}\left(\hat{P}_i^{(t+1)} - \tilde{P}_i^{(t+1)}\right)^2 R\left[\tilde{P}_i^{(t+1)}\right] + \sum_{i=1}^{k}\left(\hat{P}_i^{(t)} - \tilde{P}_i^{(t)}\right)^2 R\left[\tilde{P}_i^{(t)}\right] \\
+ \sum_{i=1}^{k}\left(\hat{B}_i - \tilde{B}_i\right)^2 R\left[\tilde{B}_i\right] + \sum_{i=1}^{k}\left(\hat{D}_i - \tilde{D}_i\right)^2 R\left[\tilde{D}_i\right] \\
+ \sum_{i=1}^{k}\left(\hat{N}_i - \tilde{N}_i\right)^2 R\left[\tilde{N}_i\right] + \sum_{i=1}^{k+1}\sum_{j=1}^{k+1}\left(\hat{F}_{ij} - \tilde{F}_{ij}\right)^2 R\left[\tilde{F}_{ij}\right] \\
+ \sum_{i=1}^{k+1}\sum_{j=1}^{k+1}\left(\hat{M}_{ij} - \tilde{M}_{ij}\right)^2 R\left[\tilde{M}_{ij}\right]
\end{aligned} \tag{7}$$

where $\hat{P}^{(t+1)}$, $\hat{P}^{(t)}$, $\hat{B}$, $\hat{D}$, $\hat{F}$ and $\hat{M}$ are the final (*i.e.* adjusted and balanced) estimates we are looking for. Function L is simply the (squared) weighted Euclidean distance between the vectors of raw and balanced estimates of stocks and flows.

Note that the objective function (7) involves both *gross* and *net* migration flows, and both *gross* and *net* natural flows, as they are all very significant demographic statistics in their own right, which we would like to modify the least during the balancing procedure.

Therefore, the constrained optimisation problem we propose to solve is the following:

$$\begin{cases}
\text{Argmin } L\left(\hat{P}^{(t+1)}, \hat{P}^{(t)}, \hat{B}, \hat{D}, \hat{N}, \hat{F}, \hat{M}\right) & \text{subject to:} \\
\hat{P}_i^{(t+1)} = \hat{P}_i^{(t)} + \hat{N}_i + \sum_{j=1}^{k+1} \hat{M}_{ij} & \text{for } i = 1, \dots, k \\
\hat{N}_i = \hat{B}_i - \hat{D}_i & \text{for } i = 1, \dots, k \\
\hat{M}_{ij} = \hat{F}_{ji} - \hat{F}_{ij} & \text{for } i, j = 1, \dots, k+1
\end{cases} \tag{8}$$

The constraints acting on the problem (8) are, of course, the area-level DBEs, plus structural constraints expressing the relation between births, deaths, and natural increase, and the antisymmetry of the Net Migration Matrix. The solution to problem (8) results in time and space-consistent estimates of population counts, natural flows, and migration flows for all the subpopulations U_i. In addition, owing to the linearity of the DBE, it is evident that the solution to problem (8) ensures consistency of stocks and flows for any other domain that can be obtained by aggregation of the balancing domains U_i, therefore, in particular, for the whole population U.

Problem (8) involves $2(k + 1)^2 + 5k$ unknowns and $(k + 1)^2 + 2k$ linear constraints. If we were to consider as "regions" the subpopulations U_i determined by cross-classifying 'NUTS 3'*'sex'*'5 years age classes', we would need to handle approximately 37,000,000 unknowns. For problems of this size the closed-form solution proposed by Stone *et al.*, 1942, which is essentially derived from the generalised least squares method, is so computationally demanding that cannot be applied in practice. As a viable alternative, an iterative constrained optimisation approach was proposed by Byron, 1978, which exploits the Conjugate Gradient algorithm, see also (van der Ploeg, 1982) and (Nicolardi, 1998). The iterative Conjugate Gradient algorithm is indeed computationally very efficient, see (Greenbaum, 1997), and proved a perfect fit for the stocks and flows reconciliation task (8). To fully automate the solution of this task, we implemented a dedicated software system, based on R (R Core Team, 2022). This software system is publicly available on GitHub[7].

NSIs need to publish population counts by domains that cross territory with covariates like 'sex', 'age class', 'nationality', and so on. We remark that the macro-integration method described here can produce consistent estimates in this context as well. Indeed, when covariates like 'sex', 'age class', and 'nationality' are introduced, we can still write down *generalised DBEs* constraining cell counts of the corresponding N-way classification at subsequent times t and $t + 1$. However, these covariates bring into play:

i. A *more abstract notion of migration flows, e.g.* people can "migrate" from a given 'age class' to the subsequent one or from one 'nationality' to another.

ii. *New structural constraints* (*i.e.* "illicit migrations"), *e.g.* since people cannot get younger, they can only get stuck in their original 'age class', move to the next one, or die[8].

7 https://github.com/DiegoZardetto/Stocks-AND-Flows.

8 Note that structural constraints arising from variable 'age class' can actually be greatly simplified by trading variable 'age class' for variable 'class of cohort'. When using cohorts, the only residual constraint is that the modality of variable 'class of cohort' cannot change between t and $t + 1$.

Fortunately, we can leverage *reliability weights* in equation (7) to prevent "illicit migrations" from being generated within the balanced solution. Since illicit cells have 0 *raw counts*, all we have to do is to let $R[\cdot] \rightarrow \infty$ and the corresponding *balanced counts* will still be 0.

The software system we developed can indeed handle the additional complexity illustrated above at scale, thus allowing Istat to achieve stock and flow consistency for all the subnational estimates that must be officially disseminated in accordance with Italian and European statistical regulations. This has been extensively tested in practice using officially released estimates of population counts and demographic flows referred to the period [2018, 2019]. For instance, our system managed to obtain perfect consistency for all the DBEs related to the *k* = 8,640 subpopulations U_i determined by crossclassifying the following variables:

'NUTS 3' *	'citizenship (Italian/non-Italian)'	* 'sex'	* '5 years age classes'
[108 modalities]	[2 modalities]	[2 modalities]	[20 modalities]

Although in the setting above the optimisation problem (8) involves nearly 150,000,000 unknowns, completion of the balancing task only required about 40 minutes using an ordinary Windows server machine equipped with sufficient RAM (memory usage peaked at 32 GB).

Coming to the statistical properties of the balanced (*i.e.* final) estimates of population stocks and flows, (Theil, 1961) has shown that they are BLUE (best linear unbiased estimates) if:

1. *Errors* affecting raw (*i.e.* initial) estimates are *uncorrelated* and have *zero mean.*
2. *Reliability weights* are equal to *inverse variances* of raw estimates.

When the above assumptions do not hold, *e.g.* because raw estimates are *biased* or reliability weights are *misspecified*, the general properties of balanced estimates are no longer under theoretical control. Yet, of course, they can still be investigated through Monte Carlo simulations, as will be shown in Section 3 (see also Di Zio *et al.*, 2018). Experiments on real Italian data reported herein suggest that, under reasonable assumptions, the proposed approach determines improved estimates of population counts: bias and variance of demographic stocks and flows are both greatly reduced by the balancing process. As will be illustrated in Section 3, key assumptions underpinning this result are the following: (*i*) errors affecting civil registry figures of births and deaths are negligible; (*ii*) high-quality estimates of population counts are available at time t.

3. Simulation study

In this Section, we present a simulation study designed to investigate the behaviour of balanced estimates in a more general setting than the ideal one addressed in Theil (1961) and introduced in Section 2.

We start with official demographic figures (P^t, B, D, N, F, M) of administrative Italian regions (NUTS 2) in 2015 so that $k = 20$. From these data, we compute P^{t+1} using the DBE: this way the set $(P^{t+1}, P^t, B, D, N, F, M)$ *exactly* fulfills the DBE *by construction*. Then, we use such figures as *ground-truth* and perturb them to generate *raw estimates*. Note that in the following, as already stated, a *tilde hat* denotes *raw estimates* (*e.g.* $\tilde{P}_i^{t+1}$), a *circumflex hat* denotes *balanced estimates* (*e.g.* $\hat{P}_i^{t+1}$), whereas *no hat* denotes *ground-truth* values (*e.g.* P^{t+1}). The simulation goes as follows:

a. We assume that *births*, *deaths*, *natural increase,* and *population counts* at time *t* are *known without errors*, *i.e.* $\tilde{B} = B$, $\tilde{D} = D$, $\tilde{N} = N$, and $\tilde{P}^t = P^t$. Therefore, to prevent them from being changed by the balancing algorithm, we set their *reliability weights* to infinite:

$$R[\tilde{B}_i] = R\big[\tilde{D}_i\big] = R\big[\tilde{N}_i\big] = R[\tilde{P}_i^t] \to \infty \tag{9}$$

b. We generate the vector of raw estimates of population counts $\tilde{P}_i^{t+1}$ by adding to P^{t+1} a *Gaussian noise* with a *relative bias* β and a *coefficient of variation α*:

$$\tilde{P}_i^{t+1} = \mathcal{N}(\,\mu = (1+\beta)P_i^{t+1},\ \sigma^2 = (\alpha P_i^{t+1})^2\,) \tag{10}$$

c. We generate the perturbed Migration Flows Matrix $\tilde{F}$ from a *Negative Binomial* distribution centred around F with a *relative bias γ and dispersion parameter δ*:

$$\tilde{F}_{ij} = \mathcal{NB}\big(\,\mu = (1+\gamma)F_{ij},\ v = \mu + \delta\mu^2\,\big) \tag{11}$$

d. We derive the perturbed Net Migration Matrix $\tilde{M}$ from $\tilde{F}$ as generated in (11) following identity (6):

$$\tilde{M} = \tilde{F}' - \tilde{F} \tag{12}$$

e. For the *reliability weights* of $\tilde{P}_i^{t+1}$, $\tilde{F}$ and $\tilde{M}$, we deliberately assume a *naïve and misspecified model*[9], setting their value to the reciprocal of the absolute value of the corresponding raw estimate:

$$R[\tilde{\cdot}] = 1/|\ \tilde{\cdot}\ | \tag{13}$$

f. Lastly, we compute *balanced estimates* $\left(\hat{P}^{t+1}, \hat{F}, \widehat{M}\right)$ by solving the constrained optimisation problem (8).

We repeated all the steps above 5,000 times ($s = 1, \ldots, S$, with S = 5,000) and compared the resulting Monte Carlo distributions of *raw estimates, balanced estimates, and ground-truth figures*. For evaluation, we used standard global accuracy measures:

- **MARB** (Mean Absolute Relative Bias): the average over regions of absolute values of Monte Carlo estimated relative biases (see equations (14)).
- **MRRMSE** (Mean Relative Root Mean Squared Error): the average over regions of absolute values of Monte Carlo estimated relative square roots of MSEs (see equations (15)).

For instance, setting for notational convenience $t \stackrel{\text{def}}{=} 0$ and $t+1 \stackrel{\text{def}}{=} 1$, the accuracy measures for the *balanced population counts* have been computed as follows:

$$RB_i = \frac{1}{S}\sum_{s=1}^{S}\left(\frac{\hat{P}_i^{1(s)} - P_i^1}{P_i^1}\right) \qquad MARB = \frac{1}{k}\sum_{i=1}^{k}|RB_i| \tag{14}$$

$$RRMSE_i = \sqrt{\frac{1}{S}\sum_{s=1}^{S}\left(\frac{\hat{P}_i^{1(s)} - P_i^1}{P_i^1}\right)^2} \qquad MRRMSE = \frac{1}{k}\sum_{i=1}^{k} RRMSE_i \tag{15}$$

We have studied 5 different simulation scenarios: S1, …, S5. The main features of these scenarios are summarised in Table 3.1. Note that different simulation scenarios have been highlighted in Table 3.1 using different colours so to make the presentation of results easier.

9 Model (13) is misspecified in the sense that it defines reliability weights that are clearly *different from the inverse of the variances* of the errors generated by equations (10), (11), and (12). Note that the *naïve* model (13) would instead be appropriate in case raw stocks and absolute values of raw flows follow a Poisson distribution.

Table 3.1 - The investigated simulation scenarios

Scenario	Main Features
S1	No Bias
S2	Only Migration Bias
S3	Both P1 and Migration Biases
S4	Overdispersed Migrations
S5	High Bias - High Variance

Source: Authors' construction

Note also that simulation scenarios S1, …, S5 have to be intended as a *hierarchy*, in the sense that each scenario actually *adds* its main features to those characterising the previous one (*e.g.* S4 switches on overdispersion in perturbed migration flows, but both migration flows and population counts $\hat{P}_i^1$ are already biased owing to S3).

Within each scenario, two combinations of the simulation parameters $(\beta,\ \alpha,\ \gamma,\ \delta)$ defined in (10) and (11) have been investigated. The simulation parameters and the corresponding simulation results expressed in terms of MARB(%) and RRMSE(%) for the *balanced population counts* $\hat{P}_i^1$ are reported in Table 3.2. The rows of Table 3.2 have been consistently highlighted with the same colours that have been used in Table 3.1 to differentiate the simulation scenarios. This makes it straightforward to visually link a given combination of simulation parameters of Table 3.2 to the simulation scenario it belongs to.

Table 3.2 - Main results of the Monte Carlo simulation (5 scenarios, 10 combinations of simulation parameters, 5,000 runs for each combination)

Simulation Parameters						Evaluation Criteria					
P1 Raw		Raw Migration Figures				P1 MARB (%)			P1 MRRMSE (%)		
RBias (%)	CV (%)	Matrix	RBias (%)	Disp (%)	Avg|CV (%)	Bal	Raw	Bal/Raw	Bal	Raw	Bal/Raw
0	10	F	0	0	8	0.0	0.1	-	0.0	10.0	0.2
0	10	M	0	0	15	0.0	0.1	-	0.0	10.0	0.2
0	10	F	-50	0	11	0.1	0.1	-	0.1	10.0	1.1
0	10	M	-50	0	21	0.1	0.1	-	0.1	10.0	1.1
-5	10	M	-50	0	21	0.1	5.0	2.1	0.1	11.2	1.0
5	10	M	-50	0	21	0.1	5.0	2.1	0.1	11.2	1.0
-5	10	F	-50	20	47	0.1	5.0	2.1	0.2	11.2	1.6
-5	10	M	-50	20	53	0.1	5.0	2.1	0.1	11.2	1.0
-10	20	F	-50	20	47	0.1	10.0	1.1	0.2	22.4	0.8
-10	20	M	-50	20	53	0.1	10.0	1.1	0.1	22.4	0.5

Source: Authors' construction

The left panel of Table 3.2 reports simulation parameters used to generate raw values of *population counts* ($\tilde{P}^1$) and *migration figures* (both $\tilde{F}$ and $\widetilde{M}$) for the k = 20 administrative Italian regions: 'RBias' columns indicate β and γ respectively, 'Disp' indicates δ, and 'Avg|CV|' indicates average CVs (in absolute value) of migration figures resulting from a given choice of (γ, δ) in the Monte Carlo. Note that, for migration matrices $\tilde{F}$ and $\widetilde{M}$, we computed 'Avg|CV|' figures from the outcomes of the Monte Carlo simulation and showed them in Table 3.2 because such CVs are not directly controlled by parameters of the simulation (see equations (11) and (12)), at odds with population counts (equation (10)).

The right panel of Table 3.2 reports the MARB(%) and RRMSE(%) for the *balanced* and the *raw estimates* of P^1 (columns 'Bal' and 'Raw' respectively), as well as the ratios of the corresponding accuracy measures (column 'Bal/ Raw', given in percentages once again). The main results of Table 3.2 can be summarised as follows:

- Even though we injected *substantial bias* inside *raw estimates* of population counts ($\tilde{P}^1$) and migration figures ($\tilde{F}$ and $\widetilde{M}$), *balanced estimates* of population counts ($\hat{\tilde{P}}^1$) *are always nearly unbiased*: balancing removed *at least* 98% of the original bias.
- In all simulation scenarios, *balancing dramatically increased the efficiency of* P^1 *estimates*: the MSE of balanced estimates $\hat{\tilde{P}}^1$ is only ~1% of raw estimates' one.

Based on these findings, we are led to the conclusion that the benefits of balancing go beyond the coherence dimension of data quality and extend to the accuracy dimension. In fact, the simulation study showed that our balancing approach also determines improved estimates of population counts: besides gaining consistency, they exhibit lower bias and variance as compared to raw ones, and the accuracy gain seems robust against misspecification of reliability weights. Clearly, the results obtained here pertain to the adopted simulation settings and the investigated simulation scenarios. Real-world applications may exist such that the errors affecting raw estimates of population stocks and flows exhibit distributions that structurally differ from (10) and (11). Furthermore, the errors in the stocks might in principle be somewhat correlated with those in the flows. In the Italian demographic estimation system, for instance, municipal civil registries not only directly provide raw estimates of flows, but also feed the BRI before it is integrated with (and adjusted by) the Permanent Census, thus indirectly contributing to the estimates of stocks as well. In this regard, we plan to conduct further research and explore different simulation settings to improve the robustness of our findings.

The specification of reliability weights adopted in the simulation deserves a few dedicated remarks. Equation (13) implies that larger changes are expected in population counts than in flows, but – in a sense – this is a desirable side-effect, as it tends to more evenly distribute relative changes while restoring the DBEs. Moreover, whenever good information about the variances of error terms affecting raw estimates is unavailable, it is a common choice to set reliability weights to reciprocals of raw estimates (after taking absolute values, if needed). This choice dates back to Stone's seminal works and is often applied in the National Accounts directorates of NSIs for GDP estimation. It is also important to acknowledge that our approach, besides hopefully reducing biases and variances of stocks and flows in simulations, must lead to consistent (*i.e.* compliant with the DBEs) and well-behaved (*i.e.* non-pathological) estimates in concrete applications. The extensive large-scale tests we touched upon at the end of Section 2 (which we conducted using Italian officially released estimates of population counts and demographic flows referred to the period from 2018 to 2019) showed that the discussed specification of reliability weights (13) is indeed beneficial. For instance, balancing methods in themselves cannot protect against the occurrence of sign-changing adjustments. This could, in principle, lead to negative balanced estimates of positive-definite quantities. Arguably, the risk of such a pathological outcome is much higher for raw estimates that are smaller in size, *e.g.* the vast majority of the flows $\tilde{F}_{ij}$ in any highly disaggregated application. From our tests, we gathered compelling empirical evidence that the specification $R\left[\tilde{F}_{ij}\right] = 1/\left|\tilde{F}_{ij}\right|$ of equation (13) plays an important role in preserving the positivity of balanced flows' estimates. In future extensions of this work, we plan to more deeply involve demographers and subject matter experts who can propose different specifications of the reliability weights and test their impact under additional simulation settings.

4. Downstream effects on the Base Register of Individuals

Once balanced population counts are obtained through our macro-integration procedure (8), they are ready to be released as official estimates of population stocks. In addition, we propose to use them as control totals to adjust the individual weights tied to records of the BRI. This would determine two beneficial effects:

- First, estimates of population counts derived from adjusted BRI weights would exactly match officially disseminated estimates, and thereby satisfy the DBE for all the domains addressed by the balancing procedure.
- Second, since balancing has been shown to greatly reduce bias that could possibly affect raw estimates of population counts, the weights of the BRI would receive a second layer of protection against bias, "borrowing strength" from both official population counts disseminated the year before and balanced demographic flows.

In the following, we illustrate how this proposal can be implemented in practice. We also point out some technical details that might become relevant in case, in the foreseeable future, the process that currently generates BRI weights (see Falorsi, 2017 and Righi *et al.*, 2021 for details) is improved to accommodate integrated individual- and household-level weights.

In practice, *raw* estimates of population stocks in equations (7) and (8):

$$\begin{cases} \tilde{P}^{(t)} = \left(\tilde{P}_1^{(t)}, \dots, \tilde{P}_k^{(t)}\right)' \\ \tilde{P}^{(t+1)} = \left(\tilde{P}_1^{(t+1)}, \dots, \tilde{P}_k^{(t+1)}\right)' \end{cases} \tag{16}$$

to be fed as input to the balancing procedure, come from *internally released* BRI versions referred to times t and $t + 1$. As noted in Section 1, each released version of the BRI – referred to a generic time τ – will contain undercoverage and over-coverage corrected weights associated to *individual* records:

$$d_q^{(\tau)} \qquad q = 1, \dots, \mathrm{N}_{\mathrm{BRI}}^{(\tau)} \tag{17}$$

and raw population counts for each balancing cell (= subpopulation) U_i will be obtained by simply adding the weights of BRI's individuals belonging to the cell:

$$\tilde{P}_i^{(\tau)} = \sum\nolimits_{q \in U_i} d_q^{(\tau)} \qquad \text{i = 1, ... , k} \tag{18}$$

Once the balancing procedure has been successfully executed, output *balanced* estimates of population stocks will be available. These balanced estimates can easily be exploited to *adjust* individual weights of the BRI in such a way that estimates of population counts derived from *adjusted BRI weights* fulfill all the DBEs in equation (8):

$$\begin{cases} d_q^{(\tau)} \xrightarrow{\text{BALANCING}} w_q^{(\tau)} & \text{q = 1, ... , } \mathrm{N}_{\mathrm{BRI}}^{(\tau)} \\ & \qquad\qquad \text{such that:} \\ \sum\nolimits_{q \in U_i} w_q^{(\tau)} = \hat{P}_i^{(\tau)} & \text{i = 1, ... , k} \end{cases} \tag{19}$$

To show how this appealing result can be achieved in practice, let us start with a very important distinction between population stocks referred to time t and those referred to time $t + 1$ (namely, the population counts reported in (16) and involved in equations (7) and (8)).

1. At the time the balancing procedure takes place $t^{BAL} > t + 1$, *Istat will have already disseminated to the external audience official population counts referred to time* t. Therefore:

   i. The balancing procedure taking place at time $t^{BAL} > t + 1$ will be performed in such a way that these official population estimates will *not* be altered:

   $$\hat{P}_i^{(t)} \equiv \tilde{P}_i^{(t)} \qquad \text{i = 1, ... , k} \tag{20}$$

   which will simply be obtained by letting $R\left[\ \tilde{P}_i^{(t)}\right] \rightarrow \infty$.

   Note that this will allow Istat to produce balanced estimates on a yearly basis without the need to revise ever again any already disseminated official population counts.

   On the other hand:

   ii. We can assume BRI weights referred to time t to have *already been adjusted* by a balancing procedure run *the year before* and involving stocks and flows referred to times $t - 1$ and t.

2. At the time the balancing procedure takes place $t^{BAL} > t + 1$, Istat will have *not* disseminated official population counts referred to time $t + 1$ *yet*, however *a released version of the BRI referred to time t* + 1 *will be available, along with undercoverage and over-coverage corrected (but still raw) individual weights*. Therefore:

   iii. After successfully executing the balancing procedure, we will use its outputs to *adjust* individual weights of the BRI referred to time $t + 1$ in such a way that estimates of population counts derived from the BRI will henceforth be consistent and fulfill all the DBEs:

$$d_q^{(t+1)} \xrightarrow{\text{BALANCING}} w_q^{(t+1)} \qquad \mathrm{q} = 1, \ldots, \mathrm{N}_{\mathrm{BRI}}^{(t+1)} \tag{21}$$

   iv. Istat will use the *post-balancing adjusted individual weights* $w_q^{(t+1)}$ in (21) to compute *official estimates of population counts for arbitrary estimation domains* $\mathrm{D_e}$:

$$\hat{P}_e^{(t+1)} = \sum_{q \in D_e} w_q^{(t+1)} \qquad \mathrm{e} = 1, \ldots, \mathrm{E} \tag{22}$$

The way to obtain *post-balancing adjusted* individual weights $w_q^{(t+1)}$ appearing in equation (21) is straightforward. It will only take to post-stratify raw individual weights $d_q^{(t+1)}$ using the balanced population estimates $\hat{P}_i^{(t+1)}$ as calibration benchmarks:

$$w_q^{(t+1)} = d_q^{(t+1)} \left[\frac{\hat{P}_i^{(t+1)}}{\tilde{P}_i^{(t+1)}} \right] = d_q^{(t+1)} \left[\frac{\hat{P}_i^{(t+1)}}{\sum_{q \in U_i} d_q^{(t+1)}} \right] \qquad \forall\, \mathrm{q} \in U_i \tag{23}$$

for $i = 1, \ldots, k$ and $q = 1, \ldots, N_{BRI}^{(t+1)}$

Although equation (23) is formally the solution of a calibration problem (which uses the unbounded Euclidean distance, one single auxiliary variable whose values are identically equal to 1, and calibration domains identified by the U_i cells of the balancing partition), its expression is so simple that a trivial PL/SQL Data Base procedure will be enough to accordingly update the weights of the BRI in a real production setting.

Now suppose that undercoverage and over-coverage corrected weights (17) associated to individual records of the BRI were constructed in such a way that *all the individuals belonging to the same household share the same weight*. This property would ensure consistent estimates of individual-level and household-level aggregates computed from the BRI[10]. Should this be the case, it would be desirable that *post-balancing adjusted* individual weights (21) retain that same property. In order to achieve such a goal, the simple formula (23) would *not* in general be suitable. More specifically, formula (23) would produce constant individual weights within households *only if* the balancing cells (= subpopulations) U_i *do not* cut-across households; of course, this would *not* be the case whenever individual-level variables like 'sex' or 'age class' are involved as classification variables in the balancing procedure.

For arbitrary settings – *i.e*. whatever might be the choice of balancing cells (=subpopulations) U_i – the goal of obtaining *post-balancing adjusted* individual weights that are *constant within households* can be easily attained[11] by means of standard calibration software available in Istat, *e.g. ReGenesees* (Zardetto, 2015). Even though the size of the BRI (~6·107 rows) may seem at first sight prohibitively large when compared to typical survey samples for which calibration procedures are routinely carried out in Istat, actually this will not pose any serious computational or technical issues. Indeed, suitable territorial domains can always be identified that would allow us to divide the overall calibration problem into smaller, computationally-affordable subproblems, which can be solved independently, one by one. In general, however, these subproblems will no longer entail a simple post-stratification, but rather require a more complex calibration of BRI weights. Consequently, the mathematical relation between weights $w_q^{(t+1)}$ and $d_q^{(t+1)}$ will be more complex than (23) and, in general, no longer expressible in analytic closed-form.

10 For instance, the estimated number of people living in households of 4 members would be equal to 4 times the estimated number of households of 4 members.

11 Methods to achieve integrated household-individual calibration weights are discussed in, *e.g.* Lemaître and Dufour (1987) and Heldal (1992). Istat's software ReGenesees adopts the approach of Heldal (1992). Put briefly, first the individual-level calibration model-matrix (which can encode simultaneously both household-level and individual-level auxiliary variables) is aggregated at houshold-level, then the calibration task is performed on this aggregated dataset, lastly the obtained household-level calibration weights are re-expanded to the individual-level, attaching to each individual the calibration weights of the household it belongs to.

5. Final remarks

Istat's production system, underpinned by statistical registers that integrate survey and administrative data, is ideally positioned to overcome the challenge of producing timely, reliable, and coherent estimates of demographic stocks and flows. To this end, official estimates of demographic stocks and flows should be consistent and satisfy the demographic balancing equation (DBE). Although this objective is inherently hard in a multi-source estimation system, we showed it can be attained using balancing methods, akin to the Stone-Byron approach often adopted to reconcile estimates in national accounting systems.

We designed and implemented a macro-integration procedure that can handle the demographic balancing task at scale, thus allowing Istat to achieve stock and flow consistency for all the domain estimates that must be officially disseminated in accordance with Italian and European statistical regulations. Consistency promotes credibility in published statistics, thereby enhancing the reputation of the National Statistical Institute. What is more, simulation results showed that our balancing approach also determines improved estimates of population counts: besides gaining consistency, they exhibit lower bias and variance as compared to "raw" estimates (*i.e.* those derived from the integrated system formed by the Permanent Census and the Base Register of Individuals (BRI)). Since the performance of our balancing approach was indeed very good in the illustrated simulation study, it is in our plans to investigate different simulation settings and scenarios in order to make our conclusions even more robust, *e.g*. by considering alternative error models and reliability weights.

In this work, simulation evidence was obtained under two fundamental assumptions (see Section 3, point A): (*i*) errors affecting civil registry figures of births and deaths are negligible; (*ii*) high-quality estimates of population counts at time $t = 0$ are available. Condition (*i*) is surely realistic in Italy. Condition (*ii*) has far-reaching practical consequences. Indeed, as illustrated in our simulation, unbiased estimates $\tilde{P}^t$ induce balanced estimates $\hat{P}^{t+1}$ which are still nearly unbiased. This would allow Istat to produce and publish balanced estimates on a yearly basis, without the need to revise ever again any already disseminated official estimates.

Beyond the simulation context, we did not discuss the uncertainty of balanced estimates. This is essentially for two reasons. First, as of today, Istat neither formally evaluates nor disseminates measures of the uncertainty affecting the inputs to our procedure, namely raw estimates of demographic stocks and flows. We note that this would be a challenging endeavor in itself, considering the variety of sampling and non-sampling errors affecting all the involved data sources (municipal civil registries, statistical registers, and probability sample surveys with complex sampling designs). Second, the macro-integration procedure we proposed does not rely on explicit model assumptions concerning the distributions of the errors affecting raw stocks and flows (although some implicit assumptions admittedly enter the picture via the choice of reliability weights). Some authors (Bryant and Graham, 2013, and Bryant and Graham 2015) adopted a purely model-based approach to demographic accounts (more specifically, a Bayesian approach) that naturally leads to uncertainty measures of output estimates. Further studies are needed to assess the pros and cons of model-based approaches to the Italian demographic estimation system.

Finally, we provided guidance on the scope of, and considerations for, using balanced population counts as control totals to adjust individual weights of the BRI. This downstream feedback would indeed determine two beneficial effects. First, estimates of population counts derived from adjusted BRI weights would exactly match officially disseminated estimates, and thereby satisfy the DBE for all the domains addressed by our procedure. Second, BRI weights would receive a second layer of protection against bias, “borrowing strength” from both official population counts disseminated the year before and balanced demographic flows.

References

Alleva, G., P.D. Falorsi, O. Luzi, and M. Scannapieco. 2019. "Building the Italian Integrated System of Statistical Registers: Methodological and Architectural Solutions". *ESS Workshop on the use of administrative data and social statistics*, Valencia, Spain, 4th and 5th June 2019.

Alleva, G., P.D. Falorsi, F. Petrarca, and P. Righi. 2021. "Measuring the Accuracy of Aggregates Computed from a Statistical Register". *Journal of Official Statistics - JOS*, Volume 37, Issue 2: 481-503.

Bryant, J.R., and P.J. Graham. 2015. "A Bayesian Approach to Population Estimation with Administrative Data". *Journal of Official Statistics - JOS*, Volume 31, Issue 3: 475-487.

Bryant, J.R., and P.J. Graham. 2013. "Bayesian Demographic Accounts: Subnational Population Estimation Using Multiple Data Sources". *Bayesian Analysis*, Volume 8, N. 3: 591-622.

Byron, R.P. 1978. "The Estimation of Large Social Account Matrices". *Journal of the Royal Statistical Society*, Series A (General), Volume 141, N. 3: 359-367.

Di Zio, M., M. Fortini, and D. Zardetto. 2018. "Reconciling Estimates of Demographic Stocks and Flows through Balancing Methods". Presented at the *9th European Conference on Quality in Official Statistics - Q2018,* Krakow, Poland, 26th – 29th June 2018.

Eurostat, and Statistics Netherlands - CBS. 2003. "Demographic statistics: Definitions and methods of collection in 31 European Countries". *Working Papers and Studies - Population and social conditions*, 3/2003/E/n. 25. Luxembourg: Office for Official Publications of the European Communities.

Falorsi, S. 2017. "Census and Social Surveys Integrated System". United Nations Economic Commission for Europe – UNECE, *Conference of European Statisticians - CES, Group of Experts on Population and Housing Censuses,* Nineteenth Meeting, Geneva, Switzerland, 4th – 6th October 2017.

Greenbaum, A. 1997. *Iterative methods for solving linear systems*. Philadelphia, PA, U.S.: Society for Industrial and Applied Mathematics, *Frontiers in Applied Mathematics*.

Heldal, J. 1992. "A Method for Calibration of Weights in Sample Surveys".

Working Papers from Department for Statistics on Individuals and Households. Methods for Collections and Analysis, N. 3/1992. Oslo, Norway: Central Bureau of Statistics.

Istituto Nazionale di Statistica/Italian National Institute of Statistics – Istat. 2016. *Istat's Modernisation Programme*. Roma, Italy: Istat. https://www.istat.it/en/files/2011/04/IstatsModernistionProgramme_EN.pdf.

Lemaître, G., and J. Dufour. 1987. "An integrated method for weighting persons and families". *Survey Methodology*, Volume 13, N. 2: 199-207.

Nicolardi, V. 1998. "Un sistema di bilanciamento per matrici contabili di grandi dimensioni". *Quaderni di ricerca*, N. 4: 5-31. Roma, Italy: Istat.

R Core Team. 2022. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rees, P. 1979. "Regional Population Projection Models and Accounting Methods". *Journal of the Royal Statistical Society*, Series A (General), Volume 142, N. 2: 223–255.

Righi, P., S. Daddi, E. Fiorello, P. Massoli, M.D. Terribili, and P.D. Falorsi. 2021. "Optimal Sampling for the Population Coverage Survey of the New Italian Register-Based Census". *Journal of Official Statistics - JOS*, Volume 37, Issue 3: 655-671.

Stone, R., D.G. Champernowne, and J.E. Meade. 1942. "The Precision of National Income Estimates". *Review of Economic Studies*, Volume 9, N. 2: 111-125.

Stone, R, and J.D. Corbit. 1997. "The Accounts of Society". *The American Economic Review*, Volume 87, N. 6: 17-29.

Theil, H. 1961. *Economic Forecasts and Policy*. Amsterdam, The Netherlands: North-Holland Publishing Company.

Van der Ploeg, F. 1982. "Reliability and the adjustment of sequences of large economic accounting matrices". *Journal of the Royal Statistical Society*, Series A (General), Volume 145, N. 2: 169-194.

Zardetto, D. 2015. "ReGenesees: an advanced R system for calibration, estimation and sampling error assessment in complex sample surveys". *Journal of Official Statistics - JOS*, Volume 31, Issue 2: 177-203.