Integration of MNO data with survey data to produce commuting statistics

Di Consiglio, L. Pichiorri T., Piovani A. Tuoto T. , Zhang L.-C.

Abstract

Historically, decennial population censuses have served as the foundation for official statistics on commuter flows between municipalities. However, to obtain more timely data, Istat, the Italian National Institute for Statistics, among other statistical agencies, has transitioned to conducting annual census surveys with smaller sample sizes. This shift introduces several challenges in deriving accurate commuting statistics.

The availability of Mobile Network Operator (MNO) data enables the identification of recurring flows between home and work or study locations. These flows could potentially become the primary source for generating official statistics by properly adjusting the MNO coverage in a quasi randomisation (QR) approach.

In this paper, we explore the use of MNO data to estimate the number of commuters between the municipalities of an Italian region, with a focus on the pseudo-randomisation approach.

1 Introduction

Historically, decennial population censuses have served as the foundation for official statistics on commuter flows between municipalities. However, to obtain more timely data, ISTAT has transitioned to conducting annual census surveys with smaller sample sizes. This shift introduces several challenges in deriving accurate commuting statistics.

The availability of Mobile Network Operator (MNO) data enables the identification of recurring flows between home and work or study locations. These flows could potentially become the primary source for generating official statistics by properly adjusting the MNO coverage in a quasi-randomisation (QR) approach.

On the other hand, in a superpopulation framework, aggregated MNO data can serve as covariates, complementing target variables obtained from census sample surveys. Spatial interaction models are widely used in transportation planning and urban studies to describe the movement of people or goods across different places. These models describe the relationship between the origin and destination of flows within a geographic area, accounting for factors such as distance, attractiveness, and connectivity of the different locations, also including spatial auto-regressive components. To reduce bias that may arise due to misspecification of the previous regression models, small area models can be applied (see Rao and Molina, 2015).

In addition, a transfer learning approach can also be useful to combine sample surveys with MNO data, where MNO counts can serve as proxies, rather than covariates, of the target variables obtained from the survey.

Here we apply these approaches to estimate the number of commuters between the municipalities of an Italian region with a focus on the pseudo-randomisation approach.

2 MNO data

The MNO data exploited in this work refer to the so-called Call Detail Records (CDRs), collected by a local operator in the Tuscany region for six weeks, from January 2017 to February 2017. The CDR data usually include calls and text messages. The data available to this study are composed as follows: the caller ID, that is a numeric code associated to each SIM (Subscriber Identity Module) by an algorithm that guarantees anonymity; the antenna where the call originated; the date and time at which the call originated; the duration of the call; the antenna where the call ended. For text messages, data report the date and time of the text message and the antenna from which the text message was sent. The CDRs are processed so that anonymity is ensured. During the 6 weeks of observation, the number of CDRs is just fewer than two hundred million, divided into: more than one hundred million calls and seventy million SMS.

The large number of observations limits the eventual impact of missing data due to random device inactivity or inability to transmit data due to network glitches. Clearly, device inactivity may also be non-random, reflecting specific user behaviors, for example, switching off the device during rest periods. The impact of such inactivity on our analysis is mitigated by conducting observations over a sufficiently long period. Indeed, the anonymised SIM-level data are observed almost continuously over time for 6 weeks, allowing us to estimate some meaningful locations, such as “home” and “work/study”. As a result, the longer the observation period, the more accurate the labeling of meaningful locations will be. In this study, a very simple approach has been followed:

- “home” is the municipality where a device is more frequently located during the nighttime, defined from 8 pm to 7 am;
- “work/study” is the municipality where a device is repeatedly observed during the daytime, defined from 7 am to 8 pm.

As a matter of fact, antenna coverage is not constrained by municipal administrative boundaries. When an antenna serves two or more municipalities, the MNO counts are distributed among them in proportion to the area of each municipality covered. For example, if an antenna covers municipalities A, B, and C by 15%, 35%, and 50%, respectively, the counts collected by this antenna are allocated according to these percentages. This situation occurs quite frequently in rural areas, where a single antenna may cover large portions of land. However, this simple allocation approach may introduce errors in device location, which in turn affect the accuracy of MNO counts at the municipal level. More sophisticated algorithms can be employed to infer device locations, drawing on information about the network infrastructure and various operational contingencies. For the purposes of this study, these location errors are treated as part of the general measurement error affecting the MNO counts, and their methodological implications will be discussed in the next section.

By aggregating device data for which home and work/study have been derived, it is possible to produce home and work/study origin-destination (OD) flows at the municipal level, where only movements within the region are taken into account.

Further aggregating the flows for mobility outside the municipality of the home location, it is possible to obtain a measure of the number of commuters outside the municipality for work or study.

One drawback of this data is that it is not possible to detect the reason for mobility, although it can indicate the frequencies of mobility. In this paper, we focus on the estimation of commuters by municipality. Estimation of the flows is left for future work.

3 Methods

3.1 Superpopulation models

Let Y denote an $m \times m$ square matrix of OD flows from each of the m origin zones to each of the m destination zones as shown in Equation (1). The elements on the main diagonal of the matrix represent intra-zonal flows.

$$Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1m} \\ y_{21} & y_{22} & \cdots & y_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mm} \end{bmatrix} \tag{1}$$

Here we consider estimating the number of commuters originated from each municipality $y_i = \sum_{j \neq i} y_{ij}$. In our setting, only a two-stage sample is considered: i.e. a sample m_{os} of the m municipalities is first selected and for each sampled municipality a subsample of individuals is selected to record the mobility. In this setting a direct estimate of y_i can be calculated only for the m_{os} municipalities, while for the other $m - m_{os}$ municipalities counts the direct estimates cannot be obtained. When auxiliary variables X are available, such as administrative data, or MNO data as in the illustrated scenario, the relationship between these and the target variable Y can improve its estimation.

In particular to reduce the potential bias of the estimates relying only on the relationship between the covariates and the target variable y, a small area model can be formulated as follows (see Rao and Molina, 2015 for an extensive illustration of the small area methods).

First, a sampling model is assumed:

$$\hat{Y}_{\text{Dir},i} = Y_i + e_i, \quad i = 1, \ldots, m_{os} \tag{2}$$

where $\hat{Y}_{\text{Dir},i}$ is the direct estimator of Y and e_i the sampling errors distributed with mean 0 and variance σ_ϵ^2,

Secondly, a relationship between the target variable and the known auxiliary variables (linking model) is assumed:

$$Y_i = \eta + \beta X_i + u_i. \tag{3}$$

where u_i are independent from the e_i and normally with mean zero and variance σ_u^2.

The empirical best linear unbiased estimator (EBLUP) under the model (2 and 3) is the so-called Fay-Herriot estimator (Fay and Herriot, 1979):

$$\hat{Y}_{FH} = \gamma \hat{Y}_{\text{Dir}} + (1 - \gamma) X\beta$$

with

$$\gamma = \hat{\sigma}_u^2 / (\sigma_e^2 + \hat{\sigma}_u^2)$$

For out-of-sample units, it reduces to the synthetic estimator $X\hat{\beta}$.

Note that in this superpopulation model, the MNO data play a secondary role and are used as covariates in equation 3. Errors in the MNO data, such as those arising from individuals carrying multiple phones or SIM cards, commuters traveling without a phone, or inaccurate device location, have a smaller impact on the final estimates than they would if MNO data were used as the primary data source for directly deriving the statistics of interest. This holds under the assumption that such errors affect the entire area uniformly. In the context of superpopulation modeling, these errors primarily increase the variability of the estimates without introducing significant bias.

In addition, the SP modeling benefits from features that are able to explain the different relationship between MNO data and Y in different area of the region. This is the case, for instance, of the *degree of urbanisation* variable. As it will be discussed later, this variable seems relevant to explain the selection mechanism of the MNO data users from the target population of commuters, and it also improves the model performances.

3.1.1 An alternative estimator for empty sample domains

As seen above, the EBLUP reduces to $X_i^\top \hat{\beta}$ whenever the area is not observed. An alternative to the synthetic estimator is proposed in this section, by transfer learning given source $\{Y_i^*\}$, a proxy to Y_i, which may be available from recent census or other concurrent data sources.

Let us consider the domains with or without sample sizes in two steps. First let $\mathcal{D}_1 = \{i : n_i > 0\}$ contain the m_{os} out of m domains with sample observations. Apply EBLUP to $\mathcal{D}_1$ as usual to obtain $(\hat{\beta}, \hat{\sigma}_u^2)$ as the parameter estimates and $\hat{Y}_i^H = x_i^\top \hat{\beta} x_i + \hat{u}_i$ as the resulting EBLUP of Y_i. Let $\mathcal{D}_0 = \{k : n_i = 0\}$, the subset of $m - m_{os}$ domains not included in the sample. As an alternative to synthetic estimator for $\mathcal{D}_0$, which results from minimising $\sum_{i \in \mathcal{D}_0} u_i^2$ trivially, consider estimating $\{u_i : i \in \mathcal{D}_0\}$ by minimising

$$L(\mathcal{D}_0) = \sum_{i \in \mathcal{D}_0} u_i^2 + \alpha \sum_{i \in \mathcal{D}_0} (x_i^\top \hat{\beta} + u_i - Y_i^*)^2 \tag{4}$$

where α is a tuning constant, $\alpha \geq 0$, and given $\hat{\beta}$ obtained from $\mathcal{D}_1$ above. As α increases, u_i is pulled towards $Y_i^* - x_i^\top \hat{\beta}$. It is possible to choose α as follows, based on $\mathcal{D}_1$ and the EBLUP $\{\hat{u}_i : i \in \mathcal{D}_1\}$ obtained above. Given α, let $\tilde{u}_i(\alpha)$ be obtained by minimising $L(\mathcal{D}_1)$ for the domains with $n_i > 0$, just like one would have done for the domains with $n_i = 0$. Choose the value of α, such that the corresponding $\{\tilde{u}_i(\alpha) : i \in \mathcal{D}_1\}$ are closest to the EBLUP $\{\hat{u}_i : i \in \mathcal{D}_1\}$ according to some chosen metric. This two-step approach preserves the EBLUP for the domains with $n_i > 0$ and transfer learning is only applied to the rest domains with $n_i = 0$ as an alternative to the synthetic estimation.

3.2 Quasi Randomisation approach

In the Quasi Randomization approach (QR), the MNO data are used as primary source of information, hence negligible measurement errors are expected to affect the MNO counts related to the commuters. Despite the MNO data are not selected according to some known probabilities, as in random sample surveys, a model can explain the observation mechanism of the MNO data, as if they had been obtained by designed randomisation.

For our application, let us assume that among all detected flows by MNO, the flows due to commuting for work and study m_{ij} can be identified and we can aggregate the flow to obtain the total number of commuters from municipality i, $i = 1 \cdots m$: $m_i = \sum_j m_{ij}$. If this is not the case, e.g. when the home location or the work/study location are assigned on the basis of short observations or by algorithms that do not take into consideration the NSO requirements and the official definitions, the MNO-derived counts m_{ij} risk to be affected by relevant measurement errors, and they are not suitable to be used as primary source of information in official statistics

Quasi-randomisation (QR) estimation adjusts for selection error of the devices to the target population.

Here we consider two different QR adjustments: in the first one, the pseudo probabilities for selection are estimated on the direct estimates and the MNO counts at some aggregated level. We refer to this approach as *internal* adjustment, and it is variable-specific.

In the second one, we rely on information derived from *external* sources to derive the conversion from MNO data to the target population, such as the deduplication factor coming from an ad-hoc survey, to account for some users carring multiple devices, or the proportion of subscribers of the MNO in a given population sub-group z, which is provided by the MNO itself.

3.2.1 Internal QR adjustment: direct estimator disaggregation

Given an accurate direct estimate of the total number of commuters at a given level l, e.g., aggregate of municipalities such as region or province, an estimate of the pseudo probability of being included in the MNO sample is

$$m_{l|z}/\hat{Y}_{l|z} \tag{5}$$

where $\hat{Y}_{l|z}$ are the direct estimates of commuters originating from area (province or region) l in a group z and $m_{l|z}$ the corresponding MNO count, where z can be for example age classes groups or other proper sub classes where uniform inclusion probability of being in the MNO sample can be assumed. The target estimate follows as

$$\hat{Y}^{INT}_{i|z} = \hat{Y}_{l|z}\frac{m_{i|z}}{m_{l|z}}. \tag{6}$$

and

$$\hat{Y}^{INT}_{i} = \sum_{z}\hat{Y}^{INT}_{i|z} \tag{7}$$

Unfortunately, in this study, the MNO data lack socio-demographic information on the users, such as age or gender. We therefore let z depend only on geography or features of the municipalities. In our application, province and degree of urbanisation were considered.

3.2.2 External QR adjustment: the role of the survey

Alternatively, coverage adjusting factor(s) can be estimated through a survey where the usage of mobile phone data is being asked, subject to the condition that the MNOs are able to provide their aggregates broken down by the same groups considered by the adjusting factors.

The previous adjustment (6) breaks down the direct estimates of commuters $\hat{Y}_{l|z}$ by the MNO ratios $m_{i|z}/m_{l|z}$, here instead we use some generic adjusting factors to derive MNO-based target estimates:

$$\hat{Y}^{EXT}_{i|z} = \frac{m_{i|z}}{\tau_z \nu_z \theta_z} \tag{8}$$

where τ_z is the MNO propensity usage of the target population in area z and it can be obtained from a survey, ν_z is the proportion of subscribers of the specific MNO in area z eventually provided by the MNO itself, and θ_z is an adjustment (deduplication factor) to account for multiple use of mobile phones or SIM cards in area z, obtained from an ad hoc survey.

Note that the adjustment factor ν_z should refer to the MNOs' users and not to the global population. In our application scenario, the proportion of subscribers given by the MNO referred to the general population and, for this reason, we applied:

$$\hat{Y}^{EXT}_{i|z} = \frac{m_{i|z}}{\nu^*_z \theta_z} \tag{9}$$

assuming the value ν^*_z equals the proportion for the ad-hoc target population of commuters.

In Italy, the survey on Aspects of Daily Life is an annual sample survey that collects data on the penetration and use of mobile devices. The survey also collects basic demographics for all the individuals in the household, as well as data related to commuting for studying or working, working conditions, and

Table 1: Summary statistics of CV (%) of direct estimates

Year	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2021	2.180	5.092	6.828	7.972	9.532	20.650
2018	2.222	5.132	6.791	8.739	10.023	47.113

income. The survey sample has not been planned to provide reliable estimates for the variables mentioned above, more detailed than at the national level; hence, for this study, we compute the adjustment for coverage and duplication at the national level. As already said, the available MNO data do not contain any information related to the device owner, which would be ideal for computing correction factors for homogeneous groups.

4 Results

The results shown and discussed in this section can serve to illustrate the potential of the aforementioned methods, but not the quality of the estimates, due to the nature of the MNO data (CDR instead of signalling), and the temporal and geographical mismatch with the statistics of interest.

To estimate the number of commuters of the municipalities of an Italian region we have applied small area methods and the quasi randomisation approach on the following real data.

- Direct flow estimates from the Italian Population permanent census survey in 2021 and 2018. Direct estimates are used both in the super-population (SP) modeling and in the adjustment factors of the QR approach in (6);
- Call Detail Records generated in 6 weeks (January 2017 - February 2017) from an Italian mobile operator, used as auxiliary variables in the SP models and as primary source of information in the QR approach;
- The previous census counts 2011 have been used as auxiliary in regression models as well as proxy Y^* for transfer learning.
- The survey on Aspects of Daily Life has been used as alternative source to obtain the adjustment factors in the QR approach (9).

Table 1 reports some summary statistics on the Coefficients of Variation (CVs) of direct estimates in 2021 and 2018. The municipality direct estimates of outbound commuters are reliable both in 2021 and 2018, however about 100 (or 150) municipalities are not in the sample in 2021 (or 2018) for which alternatives to direct estimates are needed.

Figures 1 and 2 show the relationships of the direct flow estimates and the covariates by census, MNO or administrative data. Note that whereas in 2021 the administrative data refer to the same reference year as the survey, for 2018 there is a three year mismatch. The opposite condition applies for MNO where the data reference time is closer to the 2018 than 2021.

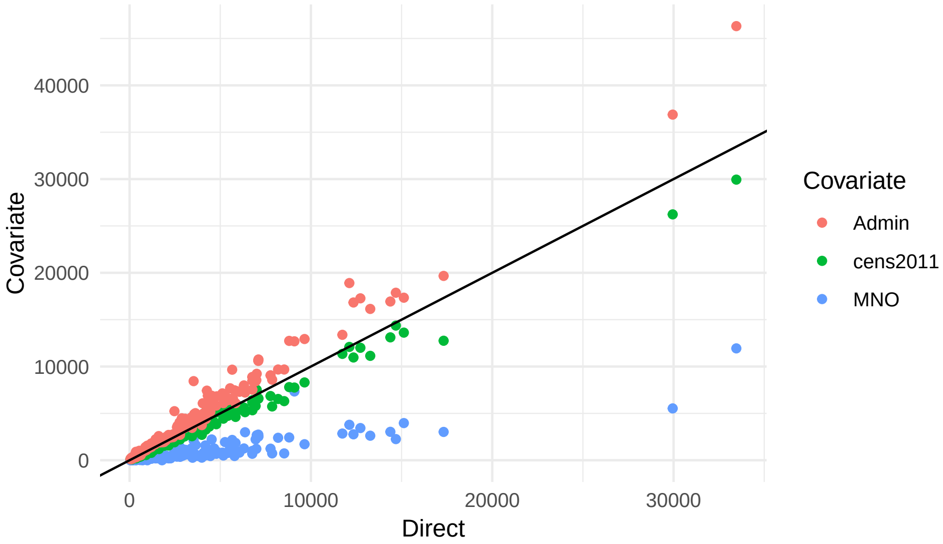

Figure 1: Direct flow estimates vs covariates, year 2021

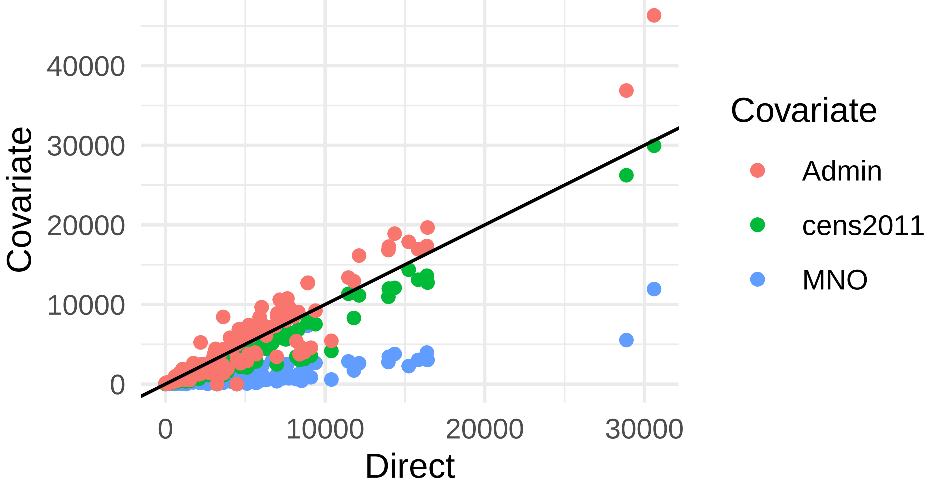

Figure 2: Direct flow estimates vs covariates, year 2018

Although the three auxiliary variables, MNO, previous census, and administrative data, exhibit linear correlation, indicating potential multicollinearity, the empirical results reported later show that their joint inclusion in the model reduces the mean square error (MSE) of the FH estimators.

4.1 Results for SP modelling

Plot 3 shows the direct estimates of the flows vs the Fay-Herriot (FH) model estimates which use as covariates the MNO and the Admin data, both for 2021 and 2018. The plot areas are limited to Direct estimates less than 5000 to highlight the pattern of the relationship.

Table 2 reports some summary statistics of the shrinkage values γ of the estimated FH models, which is the weight given to the direct vs the synthetic estimator in the EBLUP estimator, using different sets of covariates, both for

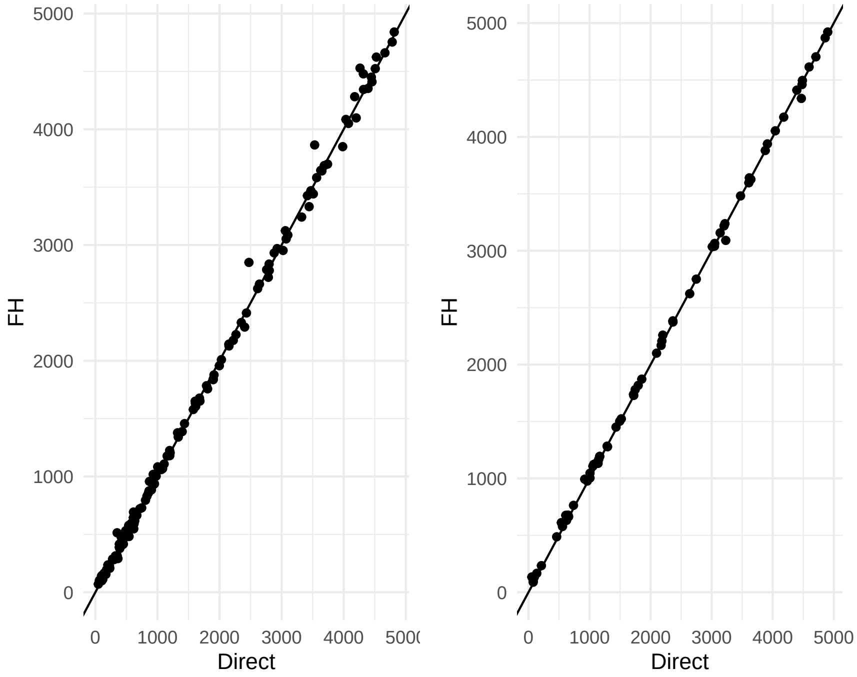

Figure 3: Direct and FH model estimates with MNO and Admin covariates, year 2021 on the left, year 2018 on the right

2021 and 2018. All the FH models have been computed on a logarithmic scale, with a bias adjusted back-transformation.

Table 2: Summary statistics of γ values of the FH models

Covariates	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2021						
MNO + Admin	0.35	0.71	0.83	0.78	0.90	0.98
MNO + census11 + Admin	0.25	0.62	0.76	0.71	0.85	0.97
2018						
MNO + Admin	0.69	0.98	0.99	0.98	0.99	0.99
MNO + census11 + Admin	0.63	0.97	0.99	0.97	0.99	0.99

Table 3 reports some summary statistics of the relative Square Root Mean Squared Error (MSE) of the the EBLUP by the FH models, with best choices of covariates. The results on the MSEs confirm the results already shown in Table 1, i.e. the direct estimates perform well in terms of variance and the EBLUP does not provide much improvement for 2021, a bit larger for 2018. Still, the direct estimates are unavailable for a large number of municipalities, for which alternative estimators are needed and the synthetic component of the EBLUP can be applied on these municipalities. It is worth noting that in 2018, the use of MNO data with the 2011 Census is more effective in enhancing the direct estimates, given its closer proximity to the reference period with respect to 2021.

Regarding transfer learning for the out-of-sample municipalities, the tuning parameter α is 0.46 for 2021 for the model without Census 2011 among

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2021						
MNO+Admin+GU	1.83	4.81	6.38	6.86	8.01	17.13
Census11+MNO+Admin + GU	1.98	4.60	6.17	6.46	7.62	13.64
2018						
Census11 + MNO	2.32	4.89	6.18	6.59	7.98	14.89

Table 3: relative sqrt MSE % - 2021- 2018

the auxiliary covariates. Figures 4 and 5 depict the direct estimates, the FH model estimates and the transfer learning estimates for the in-sampled and out-of-sample municipalities. The transfer learning estimates seem a viable alternative to the synthetic estimates for the out-of-sample municipalities in two respects: the estimated out-of-sample outbound flows are smaller on average than the in-sample outbound flows, and there are no implausible outlying estimates compared to the EBLUP (i.e., synthetic) estimates.

4.2 Results for QR estimation

Figure 6 compares the four different QR estimates with the direct estimates, for the in-sample municipalities in year 2021, depending on

- the 10 provinces (NUTS3) of the Tuscany region as l aggregation level, labelled with *prov*, vs. the whole region as a single level l, labelled with *reg*;
- adjustment factor by MNO data and direct estimates, labelled with *int*, vs. adjustment factors by ad-hoc survey data, labelled with *ext*.

It is worth noting that estimates based on adjustment at regional level *reg* are very similar using internal or external sources for the adjustment. This suggests that the proportion of subscribers in the area provided by the operator, together with the deduplication rate that was applied, is indeed estimating something very similar to $m_l/\hat{Y}_l$. We also considered the naive adjustment with the coverage provided by the MNO, and in this case the regional number of commuters were too large with respect to the estimated counts from the survey.

From conversation with the MNO we know indeed that they estimate the proportion of subscribers in the area at the regional level with n_l/N_l where n_l is the subscribers resident in the area l and the N_l is the total resident in the region l according to the figures from the national statistical office. On the other side, the estimates based on adjustments at province level *prov* show larger differences, with the *ext* providing higher values for the highest commuting counts. The relationship with the direct estimates is quite good for all the estimates, including the *ext* that indeed do not use this information in the QR adjustment, despite the relationship is stronger between direct estimates and *int* QR estimates, as expected and shown in Figure 7.

The two QR estimators produce very similar results.

Finally we consider a further QR adjustment where the pseudo-inclusion probabilities are estimated conditionally on varying degrees of urbanisation

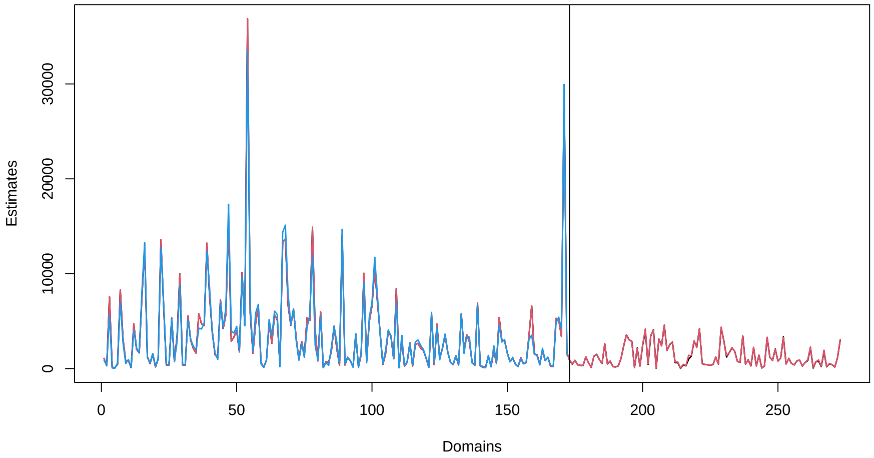

Figure 4: Direct Estimates, FH with covariates MNO+Admin, TL with proxy census2011, survey year 2021

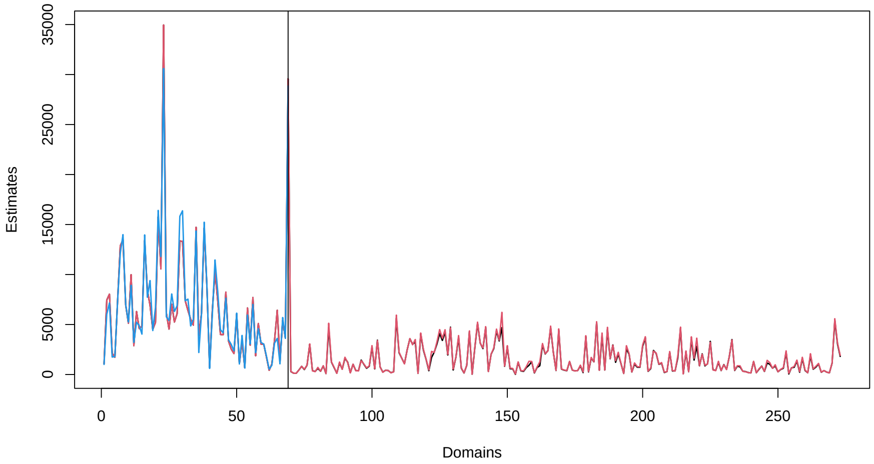

Figure 5: Direct Estimates, FH covariate MNO, TL with proxy census2011, survey year 2018

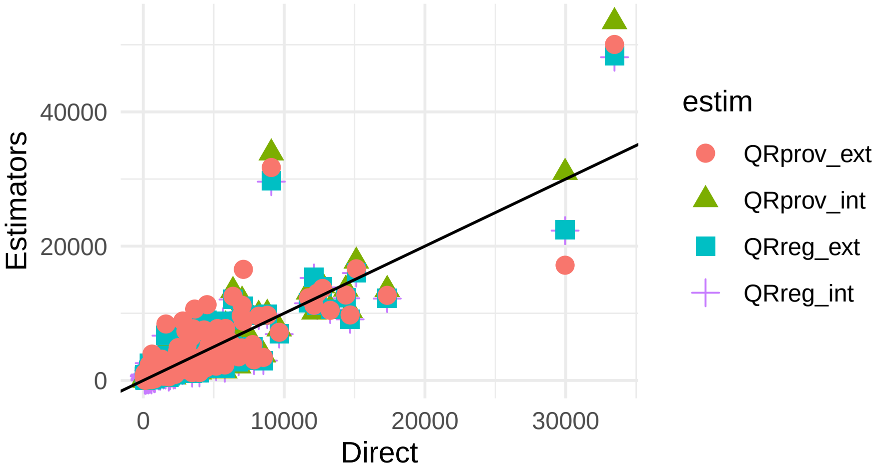

Figure 6: Comparison of QR estimates and direct estimates, 2021 survey.

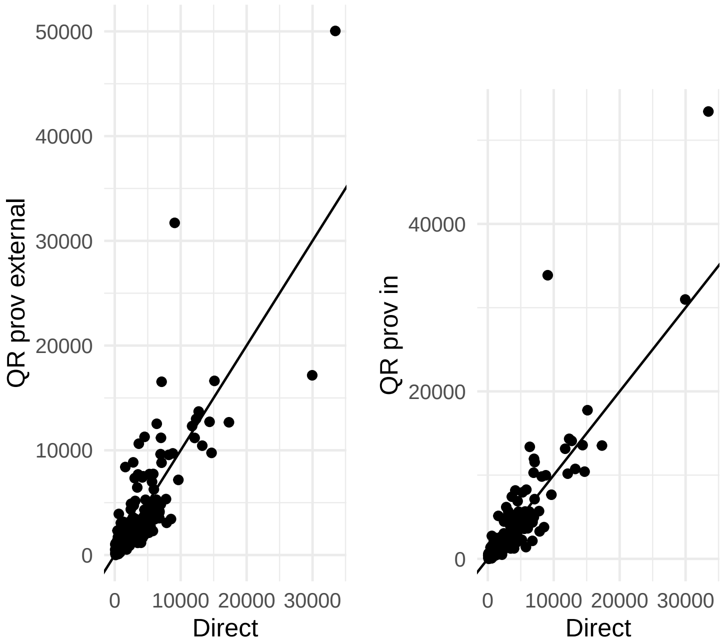

Figure 7: Comparison of QR estimates and direct estimates, 2021 survey, focus on external and internal QR adjustments at province level.

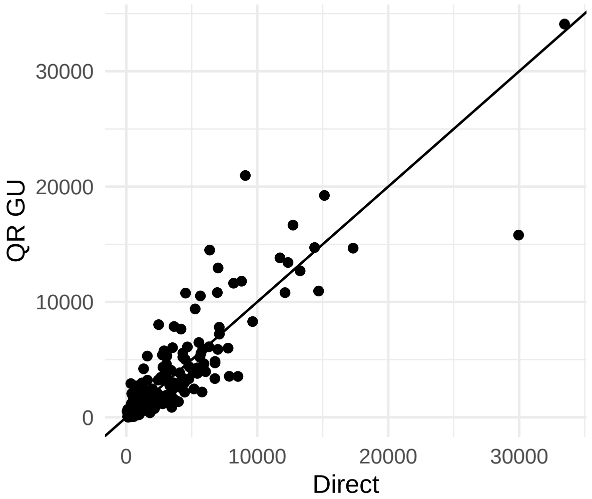

Figure 8: Comparison of QR-DU estimates and direct estimates, 2021 survey.

(DU) of the municipalities. While this approach does not yet incorporate population group-level stratification, accounting for municipal-level characteristics offers a viable alternative for capturing heterogeneity in mobile network operator (MNO) data coverage. Additional covariates—such as municipal altitude—may be incorporated to better represent rural localities. It is important to note that this extension is applicable solely within the framework of internal QR adjustments. MNOs could potentially provide coverage rates stratified by the same set of characteristics or this might be estimated by the ad hoc survey; however, such data are not currently available.

As figure 8 shows this adjustment seems preferable with respect to the previous ones examined. In particular for the outlying municipalities of Florence and Pisa appearing in figure 7.

For the out-of-sample municipalities, we can compare the QR estimates with the administrative data and the previous census data. Figure 9 shows the comparison between QR estimates based on province level adjustment with internal information (MNO data and census survey), Fay-Herriot estimates and adjusted Fay-Herriot estimates with transfer learning, with census 2011 counts and administrative data, respectively. The Figure clearly shows how the FH estimates in the out-of-sample units are quite close to the TL adjustment. In addition, the figure illustrates how QR estimates can yield biased results in the presence of measurement errors. This might be the case for the blue squared point, which stands out in the upper area of the graph. It corresponds to the municipality of Barberino del Mugello, a mountainous area crossed by a highway, both factors that, we suspect, influence home location and commuting patterns.

5 Concluding remarks and future works

This study has applied the methods to produce commuter statistics by combining MNO data with other sources, namely survey, census and administrative data. Despite the available MNO data are far from what we hope to be able to

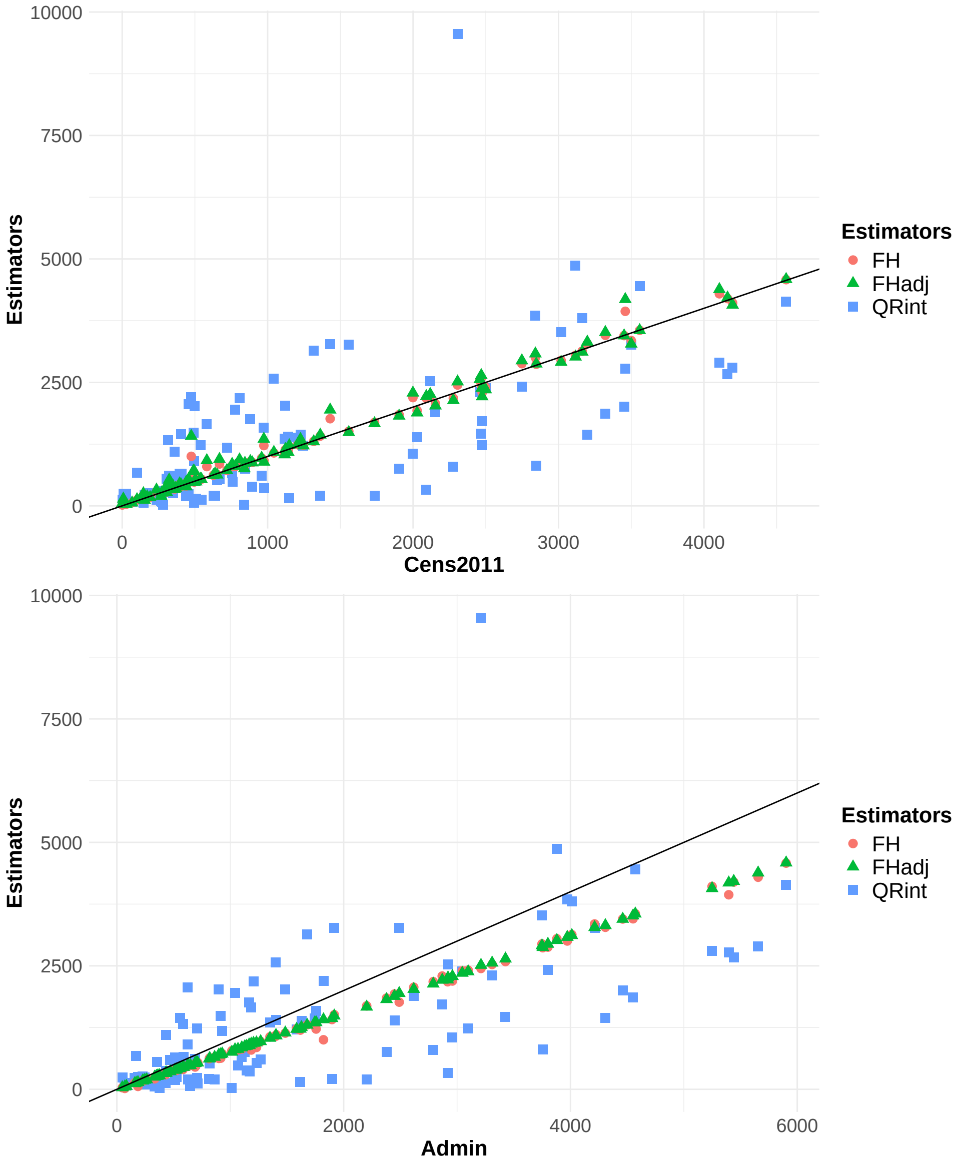

Figure 9: Comparison of QR and FH estimates with census 2011 (top) and administrative data (bottom), year 2021 survey.

exploit in the next future, it has demonstrated the potentials of MNO data for commuting statistics.

For the SP modelling approach, one would expect the MNO flow counts to be the most useful covariates, due to their high population coverage and limited measurement errors for inter-municipalities mobility, i.e. compared to past census or concurrent administrative data on home and work/study locations. However, we have only CDR data from one MNO over 6 weeks, whose coverage and measurement errors are naturally far greater than what would be admissible for real applications. Although the MNO data are outdated for both the years 2018 and 2021, for which we have other data sources, there is clear evidence that the MNO data lead to better results for 2018 than 2021. This may be taken as an indication of the potential value of MNO data, provided they can be processed and available according to the official statistical authorities' requirement.

In addition, in this experiment we could rely on the 2011 census data that still relate well to the recent surveys. But one cannot expect this to last forever, whereas one can expect to access more updated MNO data in future.

Regarding the quasi-randomisation approach, likely the adjustment that has been applied is not truly sensible, since the available MNO data do not contain any demographic characteristics on the mobile users, which vary more closely with the target population coverage by users and heterogeneous mobile device usage.

Several topics can be mentioned for future investigation, aided with more appropriate MNO data. First, since plausible estimates may be possible by different assumptions, without any of them clearly outperforming the others, we would like to combine the different estimates into a robust estimator, as outlined in the deliverable 3.2 of the ESSnet MNO MINDS.

Finally, this preliminary work lays the foundation for the estimation of a complete origin-destination (OD) matrix among municipalities, which is an highly attractive output of the current official statistics production. In this context, leveraging mobile network operator (MNO) flow data constitutes a valuable means of enriching the informational content of mobility estimates, offering a complementary source to traditional survey- or register-based approaches.

Acknowledgments

This work was co-funded by the European Commission Project "MNO-MINDS" - 101132744 - 2022-IT-TSS-METH-TOO https://cros.ec.europa.eu/mno-minds

References

ESSNet MON-MINDS (2025) Deliverable 3.2 Methodologies and open source tools for integrating MNO and non-MNO data sources, Work Package 3 https://cros.ec.europa.eu/book-page/mno-minds-wp3

Fay, R.E. and Herriot, R.A. (1979) Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data. *Journal of the American*

Statistical Association, 85, 398-409.

Rao, J.N. and Molina, I. (2015) Small Area Estimation. John Wiley & Sons, Inc., Hoboken.