BIOMETRIC METHODOLOGY

Continuous time-interaction processes for population size estimation, with an application to drug dealing in Italy

Linda Altieri[1] | **Alessio Farcomeni[2]** | **Danilo Alunni Fegatelli[3]**

[1]Department of Statistical Sciences, University of Bologna, Bologna, Italy

[2]Department of Economics and Finance, University of Rome "Tor Vergata,", Rome, Italy

[3]Department of Public Health and Infectious Diseases, Sapienza - University of Rome, Rome, Italy

Correspondence
Linda Altieri, Department of Statistical Sciences, University of Bologna, via Belle Arti 41, 40126 Bologna, Italy.
Email: linda.altieri@unibo.it

Abstract

We introduce a time-interaction point process where the occurrence of an event can increase (self-excitement) or reduce (self-correction) the probability of future events. Self-excitement and self-correction are allowed to be triggered by the same event, at different timescales; other effects such as those of covariates, unobserved heterogeneity, and temporal dependence are also allowed in the model. We focus on capture-recapture data, as our work is motivated by an original example about the estimation of the total number of drug dealers in Italy. To do so, we derive a conditional likelihood formulation where only subjects with at least one capture are involved in the inference process. The result is a novel and flexible continuous-time population size estimator. A simulation study and the analysis of our motivating example illustrate the validity of our approach in several scenarios.

KEYWORDS
capture–recapture, conditional likelihood, drug-dealing data, Hawkes processes, self-correcting processes, time-interaction processes

1 | INTRODUCTION

Models for capture–recapture data are commonly used to estimate the size of a population that is difficult to sample without bias. Such models find a wide range of applications (Otis *et al.*, 1978; Chao *et al.*, 2001; Silverman, 2020). In this work, we focus on cases in which sampling occurs in continuous time, the and occurrence of sampling can trigger a complex behavioral response that might affect the future sampling risk. Continuous-time models for population size estimation are well established (Chao and Lee, 1993; Yip *et al.*, 1996; Lin and Yip, 1999; Hwang and Chao, 2002; Xi *et al.*, 2007; Farcomeni and Scacciatelli, 2013; Matechou and Caron, 2017; Liu *et al.*, 2018; Schofield *et al.*, 2018) and often involve modeling of the temporal point process (Illian

et al., 2008; Moller *et al.*, 1998) of capture/identification/sighting.

Our motivating application involves assessing the extent of illegal drug trafficking in Italy. In this work, we consider an original dataset about drug dealers identified in Italy between 2005 and 2006. This maps to a continuous-time experiment, with dealers continuously at risk of being observed by Italian police officers from time zero to the end of the available records. Clearly, not all drug dealers have been identified during the observation period. More importantly, we can assume that the occurrence of an interaction with the Italian police brings about a transient and complex behavioral response. This will be numerically verified on the data at hand, and ignoring this effect might lead to biased estimates. For other examples of innovative population size estimation methods to assess the

extent of drug abuse, see Overstall *et al.* (2014), Farcomeni and Scacciatelli (2013), Huggins *et al.* (2016), Hay and Richardson (2016), and references therein. Our final estimate for the number of drug dealers can be of interest to both law and public health authorities. The latter are often mainly interested in the number of illicit drug users for monitoring, planning, and policy; but indeed many official estimates of the number of illicit drug users are based on dealer/consumer ratios (e.g., Bouchard and Tremblay, 2005), which must rely on accurate estimates of the number of dealers.

In the temporal point process framework, behavioral effects correspond to conditional dependence of event rates on times of previous events. This can result in *self-exciting* (*self-correcting*) processes, where the occurrence of an event at a time point affects the probability of occurrence of a later event in an upwards (downwards) fashion. Self-exciting processes are also known as Hawkes processes (Hawkes, 1971). Applications of Hawkes processes span environmental and seismological data (Ogata, 1999), health and genomic studies (Reynaud-Bouret and Schbath, 2010), crime (Mohler *et al.*, 2011), and other areas. Self-correcting processes (Isham and Westcott, 1979; Vere-Jones and Ogata, 1984) have been applied mostly to earthquake data (Rotondi and Varini, 2019; Schoenberg and Bolt, 2000).

Our methodological contribution in this work is twofold. First of all, we contribute to the literature on temporal point processes by defining a combination of self-exciting and self-correcting point processes, allowing each event to increase and/or decrease the likelihood of future ones. A similar proposal can be found in Bonnet *et al.* (2021), but we believe our specification to be more appropriate for the context of population size estimation. Hawkes and self-correcting processes may be derived as special cases. We believe our general model to be of general interest. Second, we show how self-exciting and self-correcting phenomena can be used to specify flexible behavioral responses in continuous-time capture–recapture models. To do so, we optimize the conditional likelihood, as only subjects with at least one capture are observed. The risk of capture of each individual can be thought to depend on a combination of four components: (i) observed heterogeneity (i.e., covariates, producing a class of models known as M_o models), (ii) unobserved heterogeneity (i.e., unmeasured covariates, M_h models), (iii) time effects that are independent of the subject (M_t models), and (iv) behavioral response to capture (M_b models), where the occurrence of being observed can temporarily or permanently change the risk of capture. A combination of the above features is possible, and the most general model is denoted as M_{hotb} (Otis *et al.*, 1978; Farcomeni and Scacciatelli, 2013; Farcomeni, 2016). Therefore, our contribution to the continuous-time capture–recapture

literature is that of a flexible and parsimonious behavioral effect, which has direct interpretation and applicability for both human and animal populations, embedded in a unified framework that can also take into account all other sources of heterogeneity.

The rest of the paper is as follows: in the next section, we introduce our novel time-interaction process, define the observed and the conditional likelihood. This corresponds to model M_b in the population size estimation framework. In Section 3, we specify a general model for the M_{hotb} case. The performance of the new approach is assessed via a simulation study and successfully compared to some existing methods for population size estimation in Section 4. The analysis of our motivating example is reported in Section 5. Concluding remarks are presented in Section 6.

The methods presented in this work have been implemented in R code, which may be found online in the Supporting Information. The data used for the present work must be organized in a list where each element identifies a subject captured at least once; for each subject, a vector of capture times is provided. Optional subject-specific covariates may be included in a separate covariate matrix.

2 | TIME-INTERACTION PROCESSES

A continuous-time point process is a stochastic process modeling a temporal point pattern $\{t_1, \ldots, t_S\}$; that is, a sequence of times of occurrence for an event of interest. The corresponding counting process $N(t)$ records the number of events in the interval $(0, t)$, taking values in **N**. The number of points over a time interval depends on interval length but also on an intensity function that may take a complex form. Under reasonable assumptions, this leads to $N(t) \sim Poisson(\Lambda(t))$, where $\Lambda(t) = \int_0^t \lambda(s)ds$ and the conditional intensity function $\lambda(s)$ satisfies $\lambda(t) = \lim_{h \to 0} \frac{E[N(t+h) - N(t) | \mathcal{H}(t)]}{h}$, with $\mathcal{H}(t)$ being the history of event times up to t. For capture–recapture data in continuous time, $N(t)$ represents the number of captures for one generic subject in $(0, t)$. Conditional intensities may or may not be constant across individuals. In a self-exciting process, the occurrence of an event at time t causes the conditional intensity function to increase in the short term. In capture–recapture contexts, this generates the so-called *trap-happiness* effect. The intensity is modeled as (Hawkes, 1971)

$$\lambda_{SE}(t) = \eta + \sum_{k \,:\, t_k < t} g(t - t_k), \tag{1}$$

for some smooth self-exciting kernel $g(\cdot)$. A standard choice for $g(\cdot)$ is the exponential kernel $g(t) = \sum_{j=1}^{J} \alpha_j e^{-\beta_j t}$, for some $J \geq 1$. A stationarity condition for the exponential kernel is that $\sum_j \alpha_j / \beta_j < 1$,

to avoid explosion of the intensity function. In practice, the exponential kernel with $J = 1$ is often flexible enough to provide a good fit (Oakes, 1975). Model (1) is not suitable for taking *trap-shyness* into account, that is, for considering situations where the conditional intensity function drops after an event. This can be described by self-correcting kernels, where the intensity increases in absence of events and is decreased by each event. The conditional intensity of a self-correcting process may be expressed as (Isham and Westcott, 1979)

$$\lambda_{SC}(t) = \eta \exp\{-\theta(N(t) - \eta t)\}, \qquad (2)$$

where θ is a parameter controlling the degree of trap-shyness and ηt is a target for the number of events at time t.

We here propose a unified *time-interaction process*, where

$$\lambda(t) = \alpha \sum_{k:t_k<t} \exp\{-\beta(t - t_k)\} + \eta \exp\{-\theta(N(t) - \eta t)\}. \qquad (3)$$

In (3), $\alpha, \beta, \eta, \theta$ are positive; and $\alpha < \beta$. It is worth noticing that we are specifying an additive hazard counting process model (Aalen, 1989; Lin and Ying, 1995).

2.1 | Inference

Suppose we have observed K_i events for each subject $i = 1, \ldots, n$, with subject-specific capture times $\{t_{i1}, \ldots, t_{ik}, \ldots, t_{iK_i}\}$, over an interval $(0, T)$. The log-likelihood for the ith subject is

$$l_i(\lambda) = \sum_{k=1}^{K_i} \log(\lambda(t_{ik})) - \int_0^T \lambda_i(t)\, dt, \qquad (4)$$

where $K_i = N_i(T)$, with $t_{i0} = 0$ and $t_{i(K_i+1)} = T$. Under independence, the observed log-likelihood corresponds to the sum of (4) over n subjects

$$l(\lambda) = \sum_{i=1}^n \left\{ \sum_{k=1}^{K_i} \log(\lambda(t_{ik})) - \int_0^T \lambda_i(t)\, dt \right\}. \qquad (5)$$

For our time-interaction process, the subject-specific log-likelihood becomes

$$l_i(\psi) = \sum_{k=1}^{K_i} \log \left[\alpha \sum_{s:t_{is}<t_{ik}} \exp\{-\beta(t_{ik} - t_{is})\} \right.$$

$$\left. + \eta \exp\{-\theta(N_i(t_{ik}) - \eta t_{ik})\} \right]$$

$$+ \frac{\alpha}{\beta} \sum_{k=1}^{K_i} [\exp\{-\beta(T - t_{ik})\} - 1]$$

$$- \frac{1}{\theta} \sum_{k=0}^{K_i} \exp(-\theta k)\{\exp(\theta \eta t_{ik+1}) - \exp(\theta \eta t_{ik})\}, \qquad (6)$$

where $\psi = (\alpha, \beta, \eta, \theta)$ is a short-hand notation for the parameters involved. In order to maximize the conditional log-likelihood, the gradient is needed. This can be derived in closed form, as reported in the Web Appendix. The Hessian can be obtained through a numerical first derivative of the gradient. A Newton–Raphson method proceeds after operating a change of variables so that positive parameters in $\psi = (\alpha, \beta, \eta, \theta)$ are replaced by real-valued parameters in $\psi' = (\alpha', \beta', \eta', \theta')$, where $\alpha = \exp(\alpha')$, $\beta = \exp(\alpha') + \exp(\beta')$, $\eta = \exp(\eta')$ and $\theta = \exp(\theta')$. The value of the Hessian at convergence can be used to obtain standard errors of the parameters as usual. Further details on maximum likelihood estimation are given in the next section.

2.2 | Conditional likelihood and population size estimation

In the capture–recapture framework, we can only observe subjects with $K_i > 0$. We shall resort to the *conditional likelihood* (Sanathanan, 1972), which conditions on the event that there is at least one capture. Since $N_i(t)$ is distributed like a Poisson, the probability of such an event is $\Pr(K_i > 0) = \Pr(K > 0) = 1 - \exp\{-\int_0^T \eta \exp(\theta \eta t) dt\} = 1 - \exp\{-\frac{\exp(\theta \eta T)-1}{\theta}\}$. Consequently, the conditional log-likelihood is given by

$$l^*(\psi) = l(\psi) - n \log \left[1 - \exp \left\{ -\frac{\exp(\theta \eta T) - 1}{\theta} \right\} \right]. \qquad (7)$$

This corresponds to an M_b specification, as commented in the previous section. The closed-form expression for the gradient of (7) is given in the Web Appendix and can be used to set up optimization as described for the observed likelihood (5).

The final goal in a capture–recapture context is to provide an estimate for the total number of subjects N. In our implementation, we compute the Horvitz–Thompson estimator, whose expression is

$$\hat{N} = \frac{n}{\widehat{\Pr(K > 0)}} = \frac{n}{1 - \exp \left\{ -\frac{\exp(\hat{\theta}\hat{\eta}T)-1}{\hat{\theta}} \right\}}. \qquad (8)$$

Details about the standard error of $\hat{N}$ are given in Section 3.2.

3 | INCLUDING TIME EFFECTS, OBSERVED HETEROGENEITY, AND UNOBSERVED HETEROGENEITY

The model outlined previously only takes into account behavioral effects, which might lead to clustering (self-excitement) or spreading (self-correction) of events over time. A general model must consider three further sources of variation in capture rates: time-dependent effects, observed heterogeneity, and unobserved heterogeneity. To this end, we propose the following formulation for the intensity function:

$$
\lambda_i(t) = \alpha_i \sum_{k \,:\, t_{ik} < t} \exp\{-\beta_i(t - t_{ik})\}
$$
$$
+ \lambda_0(t) \exp\left\{ \mathbf{X}_i' \boldsymbol{\gamma} + \mu_i - \theta_i(N_i(t) - \Lambda_0(t)) \right\}, \quad (9)
$$

where $\mathbf{X}_i$ is a $P \times 1$ vector of subject-specific, time constant covariates, and $\boldsymbol{\gamma}$ is a vector of regression coefficients capturing observed heterogeneity (M_o part). Function $\lambda_0(t)$ is a parametric baseline hazard, with $\Lambda_0(t) = \int_0^t \lambda_0(s)\,ds$. In the following, we set $\lambda_0(t) = \eta t^{\eta-1}$, which is the baseline hazard for the Weibull distribution. Different baselines can in principle be specified and compared through likelihood-ratio tests (e.g., Fine, 2002; see also our simulation results). Exogenous time dynamics are captured by the parameter η, which tunes the target number of events for the self-correcting part (M_t). Finally, unobserved heterogeneity (M_h) involves the use of random subject-specific effects. In the present specification, we include a frailty term μ_i, representing unobserved factors that might modify the individual risk of captures and make both behavioral effects subject specific, allowing, for example, part of the population to be trap-shy and part of it to be trap-happy in the short term. Several constraints can be used to obtain any possible submodel (e.g., by assuming $\gamma = 0$ and/or $\eta = 1$). It is also possible to exclude self-exciting ($\alpha_i = 0$) or self-correcting ($\theta_i = 0$) behaviors. The model is completed by the specification of a distribution for subject-specific effects. In this work, we use latent class models (see, e.g., Coull and Agresti, 1999). We assume the existence of a discrete latent variable U_i, with $\Pr(U_i = c) = \pi_c$ for $c = 1, \dots, C$. Parameters $\alpha_i, \beta_i, \theta_i$, and μ_i take values in a discrete set of C elements as indicized by U_i. Therefore, $\alpha_i \in \widetilde{\boldsymbol{\alpha}} = (\widetilde{\alpha}_1, \dots, \widetilde{\alpha}_C), \beta_i \in \widetilde{\boldsymbol{\beta}} = (\widetilde{\beta}_1, \dots, \widetilde{\beta}_C), \theta_i \in \widetilde{\boldsymbol{\theta}} = (\widetilde{\theta}_1, \dots, \widetilde{\theta}_C)$, and $\mu_i \in \widetilde{\boldsymbol{\mu}} = (\widetilde{\mu}_1, \dots, \widetilde{\mu}_C)$. We assume vector $\widetilde{\boldsymbol{\mu}}$ to be in increasing order. This constraint, in addition to $\alpha < \beta$ and $\alpha, \beta, \eta, \theta > 0$ specified in Section 2, ensures the identifiability of the extended model.

3.1 | Estimation of model parameters

The time-interaction M_{hotb} model (9) has $5C + P$ parameters: $4C + P + 1$ elements in $\boldsymbol{\psi} = (\widetilde{\alpha}, \widetilde{\beta}, \eta, \widetilde{\theta}, \widetilde{\mu}, \boldsymbol{\gamma})$, plus $C - 1$ latent class probabilities π_c. The log-likelihood for subject i conditional on a generic class c is

$$
l_i(\boldsymbol{\psi}|c) = \sum_{k=1}^{K_i} \log\Bigg[\widetilde{\alpha}_c \sum_{s\,:\,t_{is} < t_{ik}} \exp\{-\widetilde{\beta}_c(t_{ik} - t_{is})\}
$$
$$
+ \eta t_{ik}^{\eta-1} \exp\left\{ \mathbf{X}_i'\boldsymbol{\gamma} + \widetilde{\mu}_c - \widetilde{\theta}_c\big(N_i(t_{ik}) - t_{ik}^{\eta}\big)\right\}\Bigg]
$$
$$
+ \frac{\widetilde{\alpha}_c}{\widetilde{\beta}_c} \sum_{k=1}^{K_i} [\exp\{-\widetilde{\beta}_c(T - t_{ik})\} - 1]
$$
$$
- e^{\mathbf{X}_i'\gamma + \widetilde{\mu}_c} \sum_{k=0}^{K_i} \eta e^{-\widetilde{\theta}_c k} \int_{t_k}^{t_{k+1}} t^{\eta-1} e^{\widetilde{\theta}_c t^{\eta}}\,dt. \quad (10)
$$

The observed likelihood is therefore $\prod_i \sum_c \pi_c \exp(l_i(\boldsymbol{\psi}|c))$. The latter is slightly cumbersome to compute and might lead to numerical issues during optimization. For this reason, we prefer setting up an expectation-maximization algorithm, which is based on the complete log-likelihood $\sum_i l_i(\boldsymbol{\psi}) = \sum_i \sum_{c=1}^C \{l_i(\boldsymbol{\psi}|c)I(U_i = c) + I(U_i = c)\log(\pi_c)\}$, where $I(\cdot)$ is the indicator function. The latter is the likelihood one would observe in case the discrete latent variables U_i were measured.

The complete *conditional* log-likelihood corresponds to $\sum_{i=1}^n \sum_{c=1}^C l_i^*(\boldsymbol{\psi}|c) = \sum_{i=1}^n \sum_{c=1}^C l_i(\boldsymbol{\psi}|c) - \log(P(K_i > 0|c)) + I(U_i = c)\log(\pi_c)$, where, for the time-interaction M_{hotb} model,

$$
P(K_i > 0) = \sum_c \pi_c \left(1 - \exp\left(-e^{\mathbf{X}_i'\gamma + \widetilde{\mu}_c} \eta \int_0^T t^{\eta-1} e^{\widetilde{\theta}_c t^{\eta}}\,dt \right)\right)
$$
$$
= \sum_c \pi_c P(K_i > 0|c). \quad (11)
$$

The E step consists of computing the posterior expected complete log-likelihood values, given observed time patterns and the current parameter values. Since the complete likelihoods are linear in $I(U_i = c)$, the E step involves simply replacing the indicators with

$$
\widehat{w}_{ic} = \frac{\pi_c e^{l_i(\boldsymbol{\psi}|c)}}{\sum_{c'=1}^C \pi_{c'} e^{l_i(\boldsymbol{\psi}|c')}}, \quad (12)
$$

and an analogous expression is obtained for the conditional log-likelihood. At the M step, a closed-form

expression is obtained for updating probability masses as

$$\widehat{\pi}_c = \frac{\sum_{i=1}^n \widehat{w}_{ic}}{\sum_{c'=1}^C \sum_{i=1}^n \widehat{w}_{ic'}}. \tag{13}$$

For all other parameters, we use a numerical optimization algorithm at each iteration.

A deterministic starting solution is obtained by performing cluster analysis of the time patterns to assign initial latent classes, and then estimating our model with $C = 1$ on each subgroup. Additional starting solutions are obtained through random perturbations of the deterministic solution.

3.2 | Estimation of the population size

In order to estimate the population size with our general M_{hotb} model (or submodels), we use a Horvitz–Thompson type estimator. This would involve a plug-in estimate of (11), but in our experience the resulting estimates are sometimes unstable (see also Farcomeni and Scacciatelli, 2013, on this point). For this reason, we use the maximum-A-posteriori strategy, which we will now describe.

First, we use posterior probabilities (12) at convergence to assign each subject to its most likely latent class, denoted as c_i^*. Then, we compute

$$\widehat{N} = \sum_{i=1}^n \frac{1}{1 - \exp\left(-e^{X_i'\widehat{\gamma} + \widehat{\overline{\mu}}_{c_i^*}} \widehat{\eta} \int_0^T t^{\widehat{\eta}-1} e^{\widehat{\overline{\theta}}_{c_i^*} t^{\widehat{\eta}}} dt\right)}. \tag{14}$$

In our experience (both on simulated and real data), such a procedure leads to stable population size estimates for sampling fractions as small as 5%. There are additional methods that can be used, if necessary, to improve the stability of population size estimates, such as trimming or truncating large weights. For a complete survey, refer to Chen *et al.* (2017).

We now focus on the standard error of (14). As noted in Van Der Heihden *et al.* (2003) and Böhning (2008),

$$\text{Var}[\widehat{N}] = \text{E}[\text{Var}[\widehat{N}|n]] + \text{Var}[\text{E}[\widehat{N}|n]], \tag{15}$$

acknowledging that extra variability is brought about by the random sample size. It is straightforward to check (Farcomeni and Scacciatelli, 2013) that in our case one can express $\widehat{N} = g(\widehat{\psi})$ as a function of model parameters and use the Delta method to estimate $\text{E}[\text{Var}[\widehat{N}|n]]$ as $\nabla g(\widehat{\psi})' J(\widehat{\psi}) \nabla g(\widehat{\psi})$, where $\nabla g(\widehat{\psi})$ denotes the gradient of $\widehat{N}$ as a function of $\widehat{\psi}$, and $J(\widehat{\psi})$ is the observed Fisher information. The latter is a direct by-product of the

model estimation procedure, while the former has been in part derived in closed form and is reported in the Web Appendix. All other cases can be easily derived from the expressions reported as the Supporting Information in the Web Appendix. Finally, the second addend of (15) can be approximated as

$$\text{Var}[\text{E}[\widehat{N}|n]] \approx \sum_{i=1}^n \frac{1 - P(K_i > 0|c_i^*)}{P(K_i > 0|c_i^*)^2}, \tag{16}$$

(see also Section 2.1 of Böhning, 2008; Van Der Heihden *et al.*, 2003; Farcomeni and Scacciatelli, 2013). In our experiments, Wald confidence intervals (CI) on the log-scale had a satisfactory coverage, as reported in the next section. Namely, we first obtain Wald CI for $\log(N)$, where, using the Delta method, $\text{Var}[\log(\widehat{N})]$ is approximately estimated by $\text{Var}[\widehat{N}]/\widehat{N}^2$. Finally, we take the exponential of confidence limits.

4 | SIMULATION STUDY

In this section, we report a simulation study which assesses the comparative performance of our time-interaction (*TI*) approach over different scenarios. We compare our estimator, in terms of root mean square error (RMSE) for the population size, to a selection of literature proposals: the M_t and M_{th} models available in the R package ctime (Schofield *et al.*, 2018), Chao's lower bound estimator (*Ch*) (Chao, 1987), the Poisson model with unobserved heterogeneity (*Poi*) (Rivest and Baillargeon, 2007), the Generalized Chao estimator (*GC*), which includes covariates and a behavioral effect (Böhning *et al.*, 2013; Farcomeni, 2018), and an additive version of the continuous-time model by Lin and Yip (1999) (*LY*). The latter corresponds to an M_{htb} model with Weibull baseline hazard and a permanent behavioral effect and is a variation of the model by Hwang and Chao (2002). For our approach, we evaluate the most general time interaction process (*TI*$_{hotb}$) together with a model that does not include behavioral effects (*TI*$_{hot}$). A self-exciting-only and a self-correcting-only version of the *TI* process have also been tested and are not reported here, as they have a similar performance to the other methods and are outperformed by the complete model in all cases.

We fix $C = 2$ latent classes and $T = 1$; $\mu = (-7.5, 0)$, $\alpha_1 = 0.2$, $\beta_1 = 0.5$, $\beta_2 = \alpha_2 + 4$, $\theta_1 = 1$, and $\gamma_1 = -3$. We then generate different scenarios through combinations of $\alpha_2 \in \{1, 2\}$, $\theta_2 \in \{0.5, 2\}$, $\gamma_2 \in \{0, 1\}$, and $\eta \in \{1.2, 1.3, 1.4, 1.5\}$. Parameters are calibrated to lead to different sampling ratios and degrees of heterogeneity. The sampling ratios range from 5% to nearly 90% of the population size, which also supports the stability of the estimates in the application in Section 5. In addition, we

generated data under three scenarios which are special cases of a general M_{hotb} process: no self-exciting effect ($\alpha_1 = \alpha_2 = 0$), no self-correcting effect ($\theta_1 = \theta_2 = 0$), and no time effect ($\eta = 1$), to show the performance of our approach even when data are generated under a simpler model. We let $N \in \{500, 1000, 5000\}$, leading to a hundred and five simulation settings. Additionally, we test the first four parameter combinations for $N = 10,000$ for a complete evaluation of the estimator CI. For each parameter combination, we generate 1000 datasets, over which we average results to estimate the RMSE for each population size estimator.

Tables 1 and 2 report the simulation settings, the sampling ratios, and the resulting RMSEs, with the best performing estimator highlighted in bold for each scenario. It can be seen that our *TI* proposal achieves the smallest RMSE in almost all scenarios, with the remaining being a close call between TI_{hotb} and model M_{th}. The TI_{hotb} seems to find a competitive alternative only when the contribution of the time effect is at its strongest, and the population size is small. The *TI* submodel without behavioral effects TI_{hot} has a worse performance, which is similar to the M_t and *LY* models. For $N = 5000$, the *LY* model performs better but never becomes a close competitor of our proposal. In general, as N increases from 500 to 1000 and 5000, the gap with other estimators increases. In Table 2, we also check sensitivity to the specification of $\lambda_0(t)$ by generating data with a baseline step function with five jumps equally spaced between 0 and T, of sizes 0.067, 0.133, 0.200, 0.267, and 0.333, respectively. Our model is still fitted using a Weibull baseline. It can be seen that, at least with this setting and over a variety of simulation scenarios, there is little sensitivity to the misspecification of the baseline rate; our approach still shows the best performance.

The good theoretical properties of our approach are also testified by the fact that the RMSE decreases with the sample size at the expected rate; the same was successfully checked for the model parameters. Additionally, for the first scenarios we have evaluated the coverage of 95% CI (CIs) after computing the standard error of $\hat{N}_{TI_{hotb}}$ as described in the previous section. We obtain an empirical coverage of 0.999 for $N = 500$, 0.989 for $N = 1000$, 0.976 for $N = 5000$, 0.963 for $N = 10,000$, and 0.979 for $N = 500$ and the misspecified baseline.

5 | ESTIMATION OF THE NUMBER OF DRUG DEALERS IN ITALY

We consider a dataset of drug dealers sanctioned under the Italian law 309/90 (Bossi–Fini law). We have data about all arrests by any of the Italian police forces, and we fix time zero as the publishing of the law in "Gazzetta Uffi-

ciale," which is the moment in which new laws enter into force in Italy. Our main target for analysis is the estimation of the number of drug dealers in Italy at the time. A secondary aim is assessing the features of the observation process, which can give information about what modulates the risks of identification of drug dealers by police forces. Our analysis is based on few rather strong assumptions. First of all, our target population is made of all dealers that were not in jail at time zero. We assume that no new dealer entered or exited the drug market during the observation period, which clearly cannot be precisely true. Additionally, we assume all drug dealers continued their activity after being identified for the observation period (even after they served a short time in jail, in case this happened). We restrict the dataset to 2 years to make these assumptions more credible. Our choice for the time span is comparable to that of previous studies (e.g., Chiang *et al.*, 2007; Bouchard and Tremblay, 2005). It shall be further added that our estimates are fairly stable with respect to this choice, due to the fact that many dealers were identified shortly after the onset of the new law. Clearly, the standard error of our estimates is inversely associated with the length of the observation period.

The original dataset is provided as a matrix, where each row describes a single capture; for the considered capture, some subject-specific information is reported (sex, age) together with the capture date. We reorganize the dataset in the format of an R list, where each element contains capture times for each subject. Sex and age are organized in a separate covariate matrix with the number of rows is equal to the number of elements of the list (one for each subject). The number of observed subjects identified at least once is $n = 4271$; 3831 dealers were males, while only 440 were females. Figure 1 summarizes some preliminary information about some covariate age and capture times, in the form of histograms. The median age was 32 (mean 33), with minimum 14, first quartile 26, third 39 and maximum 82. The great majority of subjects, that is 4170, are captured only once, 99 dealers are captured twice, and two are stopped three times.

The recapture pattern may be a consequence of a combination of factors: the probability of meeting the same dealer within 2 years is low, especially in big cities. Moreover, sanctions and legal consequences become more severe after the first capture; therefore, subjects are expected to be trap-shy. The counting distribution shows therefore one inflation (Farcomeni and Scacciatelli, 2013; Godwin and Bohning, 2017; Bohning and Van Der Heijden, 2019; Bohning and Friedl, 2021). This is not a particularly problematic issue with our continuous-time approach, which takes into account the time interval between time zero and the event. The total number N of drug dealers in Italy for the considered time interval is estimated

TABLE 1 Simulation study, part 1: RMSE of estimators of models M_t ($\hat{N}_{M_t}$) and M_{th} ($\hat{N}_{M_{th}}$) (in ctime library), Chao estimator ($\hat{N}_{Ch}$), Poisson estimator ($\hat{N}_{Poi}$), generalized Chao estimator ($\hat{N}_{GC}$), additive hazard Lin and Yip's estimator ($\hat{N}_{LY}$) compared to our proposal with the complete model ($\hat{N}_{TI_{hotb}}$), and a submodel without behavioral effects ($\hat{N}_{TI_{hot}}$)

Setting				Coverage	RMSE							
α_2	η	θ_2	γ_2	n/N	$\hat{N}_{M_t}$	$\hat{N}_{M_{th}}$	$\hat{N}_{Ch}$	$\hat{N}_{Poi}$	$\hat{N}_{GC}$	$\hat{N}_{LY}$	$\hat{N}_{TI_{hotb}}$	$\hat{N}_{TI_{hot}}$
N = 500												
1	1.2	0.5	0	0.05	550.35	364.84	365.99	518.23	4820.97	547.06	**291.87**	495.32
1	1.3	0.5	0	0.16	300.67	274.31	284.35	266.44	272.72	295.47	**213.15**	255.92
1	1.4	0.5	0	0.36	190.29	142.00	162.25	163.19	153.23	193.47	**105.09**	142.98
1	1.5	0.5	0	0.69	66.38	40.72	47.89	57.44	47.91	64.09	**32.58**	51.7
1	1.2	2	0	0.05	492.06	621.95	351.33	516.50	5742.29	508.94	**281.06**	559.76
1	1.3	2	0	0.15	304.67	278.99	288.11	265.28	284.20	310.45	**231.58**	255.78
1	1.4	2	0	0.36	192.92	144.56	164.69	165.16	157.50	191.38	**93.42**	148.22
1	1.5	2	0	0.70	64.88	39.42	47.03	56.04	46.21	63.24	**31.54**	50.44
1	1.2	0.5	1	0.15	338.31	314.48	316.08	298.48	316.66	337.33	**277.37**	284.47
1	1.3	0.5	1	0.30	248.04	204.39	217.04	220.83	196.41	249.89	**110.71**	125.63
1	1.4	0.5	1	0.56	129.16	75.02	92.28	110.74	84.04	129.60	**44.66**	47.05
1	1.5	0.5	1	0.85	29.49	16.62	19.98	25.47	20.42	29.49	**13.30**	22.92
1	1.2	2	1	0.15	339.86	315.82	319.69	305.01	301.79	346.04	**281.69**	284.24
1	1.3	2	1	0.30	251.52	205.95	217.96	222.44	197.13	248.33	**68.67**	86.89
1	1.4	2	1	0.56	129.37	74.45	90.86	111.27	82.62	128.54	**16.59**	47.27
1	1.5	2	1	0.85	29.18	16.77	19.74	25.43	19.95	28.78	**13.42**	22.89
2	1.2	0.5	0	0.05	367.36	368.43	361.08	366.64	13917.45	377.58	**248.30**	283.88
2	1.3	0.5	0	0.15	304.36	277.73	284.98	274.20	275.39	292.96	**219.36**	256.48
2	1.4	0.5	0	0.36	188.94	141.33	161.31	162.11	153.57	193.51	**12.32**	33.72
2	1.5	0.5	0	0.69	65.76	38.70	46.41	56.29	46.17	66.09	**30.96**	48.77
2	1.2	2	0	0.05	1220.48	368.24	367.34	438.76	13288.37	1190.87	**307.71**	394.88
2	1.3	2	0	0.15	306.23	280.41	288.02	263.65	272.70	309.16	**224.56**	225.62
2	1.4	2	0	0.36	192.29	142.87	164.24	163.78	157.52	197.35	**32.66**	120.82
2	1.5	2	0	0.69	64.80	39.65	47.69	56.35	48.58	65.02	**31.72**	43.72
2	1.2	0.5	1	0.15	335.06	311.36	315.98	298.18	296.62	327.00	**143.75**	215.18
2	1.3	0.5	1	0.30	253.11	206.09	221.13	224.59	204.34	257.62	**93.19**	106.95
2	1.4	0.5	1	0.56	126.98	72.22	89.72	108.38	81.86	127.12	**38.40**	43.95
2	1.5	0.5	1	0.85	28.46	15.74	18.17	24.43	17.99	27.71	**12.59**	21.99
2	1.2	2	1	0.15	338.21	314.60	318.53	300.67	300.17	327.66	**245.46**	283.14
2	1.3	2	1	0.30	252.51	205.05	219.20	225.72	197.29	259.70	**79.07**	203.46
2	1.4	2	1	0.56	130.60	78.77	93.54	112.61	83.97	131.32	**37.16**	40.74
2	1.5	2	1	0.85	29.95	18.39	20.71	26.30	21.59	30.80	**14.71**	23.67
0	1.4	0.5	0	0.38	138.59	86.18	115.70	109.90	114.83	132.79	**68.94**	104.13
1	1.6	0	0	0.16	357.13	311.77	296.02	319.27	284.80	364.48	**228.32**	280.59
2	1	0.5	0	0.50	156.80	117.52	135.43	143.46	134.51	155.53	121.89	**72.26**
N = 1000												
Setting				Coverage	RMSE							
α_2	η	θ_2	γ_2	n/N	$\hat{N}_{M_t}$	$\hat{N}_{M_{th}}$	$\hat{N}_{Ch}$	$\hat{N}_{Poi}$	$\hat{N}_{GC}$	$\hat{N}_{LY}$	$\hat{N}_{TI_{hotb}}$	$\hat{N}_{TI_{hot}}$
1	1.2	0.5	0	0.05	741.58	715.64	714.16	639.91	687.64	724.91	**511.93**	644.08
1	1.3	0.5	0	0.17	560.31	481.93	515.39	472.97	508.42	568.15	**331.41**	457.58
1	1.4	0.5	0	0.40	315.20	213.49	270.10	271.61	267.38	317.50	**132.23**	203.05
1	1.5	0.5	0	0.73	98.55	72.37	80.71	92.27	80.34	98.16	**57.90**	83.04
1	1.2	2	0	0.05	754.05	729.77	730.93	684.14	747.80	753.82	**303.52**	669.24

(Continues)

TABLE 1 (Continued)

Setting				Coverage	RMSE							
α_2	η	θ_2	γ_2	n/N	$\widehat{N}_{M_t}$	$\widehat{N}_{M_{th}}$	$\widehat{N}_{Ch}$	$\widehat{N}_{Poi}$	$\widehat{N}_{GC}$	$\widehat{N}_{LY}$	$\widehat{N}_{TI_{hotb}}$	$\widehat{N}_{TI_{hot}}$
1	1.3	2	0	0.17	557.94	478.48	513.51	474.35	505.86	563.75	**310.77**	455.27
1	1.4	2	0	0.40	320.57	222.37	275.24	280.33	271.99	319.65	**69.95**	165.47
1	1.5	2	0	0.73	95.56	67.86	75.01	89.03	74.99	95.93	**54.29**	80.13
1	1.2	0.5	1	0.16	630.68	559.88	581.21	549.12	567.95	634.18	**153.67**	309.56
1	1.3	0.5	1	0.33	422.86	306.30	365.58	370.18	347.00	428.15	**79.30**	152.32
1	1.4	0.5	1	0.60	201.15	133.77	158.52	180.98	146.07	200.74	**83.14**	92.83
1	1.5	0.5	1	0.87	47.69	39.37	39.49	48.16	35.56	50.12	**28.45**	42.92
1	1.2	2	1	0.16	630.70	562.16	586.50	554.42	570.60	619.98	**308.57**	414.04
1	1.3	2	1	0.33	423.70	318.69	374.08	376.40	352.40	427.79	**130.58**	135.67
1	1.4	2	1	0.60	201.03	138.03	159.89	181.53	145.39	201.75	**96.22**	96.75
1	1.5	2	1	0.87	48.49	40.43	39.12	49.08	34.27	48.31	**27.42**	43.64
2	1.2	0.5	0	0.05	755.83	729.29	720.35	638.13	715.87	752.35	**606.32**	648.32
2	1.3	0.5	0	0.17	563.02	487.62	519.14	481.37	507.82	579.42	**292.00**	457.04
2	1.4	0.5	0	0.40	320.20	212.96	270.31	275.86	268.09	318.63	**115.58**	187.74
2	1.5	0.5	0	0.73	97.67	69.88	77.15	91.72	78.75	96.09	**55.90**	82.55
2	1.2	2	0	0.05	758.08	731.53	731.05	643.02	1205.09	757.85	**560.02**	658.38
2	1.3	2	0	0.17	557.34	477.07	510.20	460.32	500.02	583.20	**361.35**	450.02
2	1.4	2	0	0.40	318.01	215.82	272.03	272.25	268.55	317.78	**172.66**	245.02
2	1.5	2	0	0.73	97.78	70.39	79.25	91.04	80.54	98.67	**56.31**	81.94
2	1.2	0.5	1	0.16	627.98	562.89	581.60	555.38	567.58	633.52	**444.30**	523.44
2	1.3	0.5	1	0.33	415.01	306.45	359.26	363.50	337.42	414.75	**245.16**	327.15
2	1.4	0.5	1	0.60	197.37	135.84	154.05	178.11	142.99	201.01	**108.67**	160.3
2	1.5	0.5	1	0.87	47.97	40.22	40.28	48.48	36.71	47.02	**29.37**	43.17
2	1.2	2	1	0.16	629.16	562.18	587.63	560.57	573.46	636.66	**448.46**	528.87
2	1.3	2	1	0.33	420.14	312.63	369.10	369.76	349.02	404.43	**250.10**	332.78
2	1.4	2	1	0.60	202.11	139.81	162.51	183.36	148.68	202.41	**111.85**	165.02
2	1.5	2	1	0.87	49.48	41.65	41.71	49.83	37.78	48.68	**30.22**	44.53
0	1.4	0.5	0	0.38	278.10	197.22	246.73	244.26	242.35	276.08	**157.78**	222.06
1	1.6	0	0	0.16	761.65	571.67	630.76	739.75	623.37	757.91	**457.34**	665.78
2	1	0.5	0	0.48	322.82	245.57	282.79	301.20	279.57	316.02	271.08	**196.46**

Note: For each parameter setting, we report the sampling ratio (n/N/coverage) and the root mean square error (RMSE) for each estimator averaged over 1000 replicates. In bold, the smallest RMSE for each scenario.
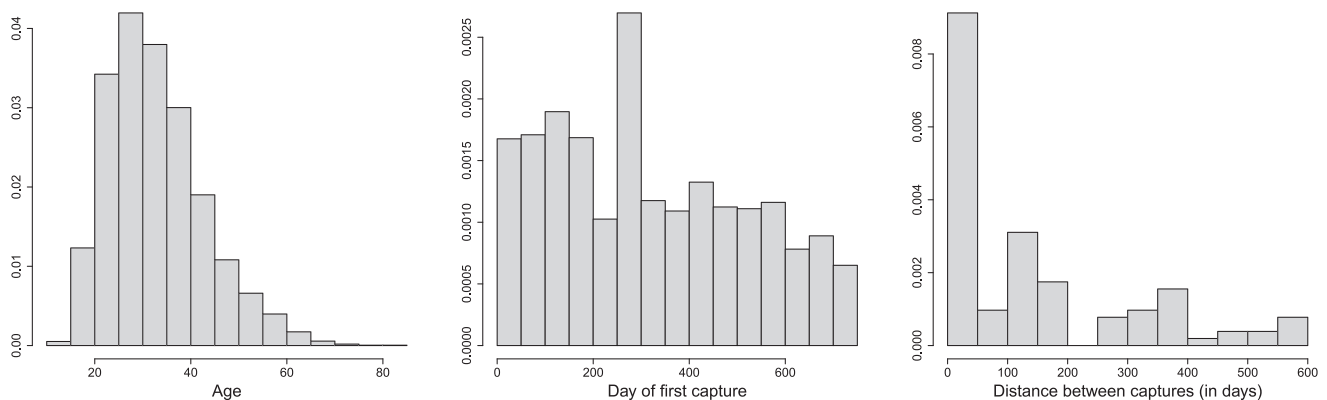

FIGURE 1 Drug dealers data: left, histogram of *age*; middle, day of the first capture; right, days between captures

TABLE 2 Simulation study, part 2: RMSE of estimators of models M_t ($\hat{N}_{M_t}$) and M_{th} ($\hat{N}_{M_{th}}$) (in ctime library), Chao estimator ($\hat{N}_{Ch}$), Poisson estimator ($\hat{N}_{Poi}$), Generalized Chao estimator ($\hat{N}_{GC}$), additive hazard Lin and Yip's estimator ($\hat{N}_{LY}$) compared to our proposal with the complete model ($\hat{N}_{TI_{hotb}}$) and a submodel without behavioral effects ($\hat{N}_{TI_{hot}}$)

Setting				Coverage	RMSE							
α_2	η	θ_2	γ_2	n/N	$\hat{N}_{M_t}$	$\hat{N}_{M_{th}}$	$\hat{N}_{Ch}$	$\hat{N}_{Poi}$	$\hat{N}_{GC}$	$\hat{N}_{LY}$	$\hat{N}_{TI_{hotb}}$	$\hat{N}_{TI_{hot}}$
$N = 5000$												
1	1.2	0.5	0	0.05	3864.41	3589.18	3667.37	3470.44	3073.08	3839.26	**3062.67**	3351.45
1	1.3	0.5	0	0.16	2942.32	2425.83	2786.90	2739.17	1462.42	2909.68	**1169.94**	1811.00
1	1.4	0.5	0	0.37	1776.07	1373.58	1605.85	1643.19	1444.11	1759.23	**294.26**	529.22
1	1.5	0.5	0	0.71	589.67	504.69	501.57	569.33	2693.14	573.38	**401.26**	530.70
1	1.2	2	0	0.05	3952.23	3687.87	3809.13	3713.66	3516.85	4037.68	**2813.48**	3342.29
1	1.3	2	0	0.16	2915.40	2316.41	2714.55	2616.65	1204.62	3017.15	**963.70**	2537.19
1	1.4	2	0	0.38	1691.87	1222.17	1515.75	1540.39	2626.65	1664.05	**337.47**	441.36
1	1.5	2	0	0.70	598.51	501.62	479.29	582.15	3077.60	609.80	**383.43**	523.93
1	1.2	0.5	1	0.15	3218.94	2576.93	2970.03	2951.13	2020.01	3296.69	**1616.01**	2611.98
1	1.3	0.5	1	0.31	2184.78	1503.56	1943.16	1989.22	497.30	2154.20	**430.27**	593.90
1	1.4	0.5	1	0.57	1125.58	826.53	933.66	1056.00	965.71	1151.79	**354.59**	365.18
1	1.5	0.5	1	0.85	298.89	269.98	222.84	301.83	777.62	300.02	**178.27**	271.65
1	1.2	2	1	0.15	3210.5	2710.43	2998.97	2887.14	2239.51	3314.29	**1791.61**	2598.43
1	1.3	2	1	0.31	2204.43	1560.39	1987.93	2044.32	678.27	2173.26	**542.62**	1839.89
1	1.4	2	1	0.58	1103.81	800.27	904.77	1027.83	1024.42	1068.10	**640.22**	925.05
1	1.5	2	1	0.85	298.51	265.23	224.41	297.62	782.64	296.02	**179.53**	268.66
2	1.2	0.5	0	0.05	3908.30	3650.38	3752.98	3595.59	3169.60	3944.76	**2535.68**	3377.68
2	1.3	0.5	0	0.16	2917.32	2364.81	2732.00	2628.16	1235.24	2937.47	**988.19**	2458.8
2	1.4	0.5	0	0.38	1704.82	1170.81	1499.44	1538.78	1752.17	1735.83	**936.65**	1534.34
2	1.5	0.5	0	0.71	582.27	466.79	463.29	557.28	2837.61	565.60	**370.63**	524.04
2	1.2	2	0	0.05	3948.44	3726.15	3833.17	3627.49	3248.95	3954.16	**2599.16**	3449.85
2	1.3	2	0	0.16	2891.45	2467.07	2746.01	2710.41	1539.96	3030.67	**1231.97**	2471.41
2	1.4	2	0	0.38	1709.95	1335.28	1566.91	1578.73	1683.84	1732.72	**1068.22**	1515.46
2	1.5	2	0	0.70	606.22	510.19	492.19	589.98	2702.18	601.55	**393.75**	545.60
2	1.2	0.5	1	0.15	3253.96	2625.55	2997.78	2947.97	2178.98	3241.91	**1743.18**	2698.00
2	1.3	0.5	1	0.31	2255.84	1669.49	2051.29	2091.57	538.91	2202.99	**431.13**	1882.41
2	1.4	0.5	1	0.57	1152.75	875.42	956.84	1089.82	1037.73	1194.69	**700.34**	980.84
2	1.5	0.5	1	0.85	309.60	285.97	244.26	313.52	646.62	309.17	**195.41**	282.17
2	1.2	2	1	0.14	3229.74	2641.70	3008.07	2950.52	2116.56	3314	**1693.25**	2707.26
2	1.3	2	1	0.31	2255.15	1532.63	2001.28	2059.99	474.16	2217.16	**379.33**	1853.99
2	1.4	2	1	0.57	1136.53	873.74	957.91	1065.17	984.55	1129.47	**698.99**	958.65
2	1.5	2	1	0.85	289.77	257.49	220.45	288.69	671.34	292.35	**176.36**	260.79
0	1.4	0.5	0	0.38	1273.58	802.91	1110.76	1120.43	10625.57	1264.21	**642.33**	1146.22
1	1.6	0	0	0.16	3882.53	906.03	3073.35	3834.27	1512.4	3899.04	**724.82**	3450.84
2	1	0.5	0	0.49	1631.22	1149.12	1384.37	1559.65	491.72	1623.02	1403.69	**393.38**
$N = 10,000$												

Setting				Coverage	RMSE							
α_2	η	θ_2	γ_2	n/N	$\hat{N}_{M_t}$	$\hat{N}_{M_{th}}$	$\hat{N}_{Ch}$	$\hat{N}_{Poi}$	$\hat{N}_{GC}$	$\hat{N}_{LY}$	$\hat{N}_{TI_{hotb}}$	$\hat{N}_{TI_{hot}}$
1	1.2	0.5	0	0.05	7883.09	7164.23	7557.58	7344.50	6649.97	7828.35	**5319.98**	6610.05
1	1.3	0.5	0	0.16	5957.84	4369.74	5504.60	5483.58	2435.80	5819.48	**1948.64**	4517.28
1	1.4	0.5	0	0.38	3450.95	2217.47	2931.26	3149.13	4616.29	3539.94	**537.89**	830.10
1	1.5	0.5	0	0.71	1168.99	909.77	906.19	1119.19	5276.41	1160.68	**724.95**	1052.09

$N = 500$ - alternative $\lambda_0(t)$

TABLE 2 (Continued)

Setting				Coverage	RMSE							
α_2	η	θ_2	γ_2	n/N	$\hat{N}_{M_t}$	$\hat{N}_{M_{th}}$	$\hat{N}_{Ch}$	$\hat{N}_{Poi}$	$\hat{N}_{GC}$	$\hat{N}_{LY}$	$\hat{N}_{TI_{hotb}}$	$\hat{N}_{TI_{hot}}$
1	1.2	0.5	1	0.10	431.22	412.39	409.71	421.35	401.20	442.77	**320.96**	377.66
1	1.3	0.5	1	0.10	429.93	409.39	403.15	420.41	392.97	442.96	**314.38**	375.42
1	1.4	0.5	1	0.11	426.70	407.35	407.62	416.95	401.88	435.93	**321.54**	373.36
1	1.5	0.5	1	0.12	423.81	401.50	400.92	415.29	393.52	419.67	**314.82**	370.21
1	1.2	2	1	0.10	429.26	408.24	407.41	419.68	400.37	435.47	**320.30**	374.13
1	1.3	2	1	0.11	426.70	407.90	404.37	416.39	399.06	422.31	**319.25**	370.62
1	1.4	2	1	0.11	427.97	410.44	408.37	418.28	401.75	426.59	**321.40**	375.18
1	1.5	2	1	0.11	424.56	404.18	406.33	416.02	399.03	426.01	**319.22**	373.11
2	1.2	0.5	1	0.10	431.44	412.37	412.58	424.52	399.45	442.94	**319.56**	378.14
2	1.3	0.5	1	0.11	425.53	404.99	400.13	415.02	386.41	417.58	**309.13**	372.65
2	1.4	0.5	1	0.11	424.65	402.22	399.36	412.03	390.23	413.97	**312.18**	366.68
2	1.5	0.5	1	0.12	423.40	403.80	406.23	414.21	402.88	423.38	**322.30**	371.76
2	1.2	2	1	0.11	425.83	407.64	403.67	413.79	396.31	420.58	**317.05**	372.41
2	1.3	2	1	0.11	425.42	406.34	404.08	416.94	386.21	428.83	**308.97**	369.86
2	1.4	2	1	0.11	425.95	404.07	402.86	416.56	395.04	434.36	**316.03**	368.71
2	1.5	2	1	0.11	426.05	405.94	406.27	416.97	398.00	421.32	**318.40**	371.47

Note: For each parameter setting, we report the sampling ratio (n/N/coverage) and the root mean square error (RMSE) for each estimator averaged over 1000 replicates. In bold, the smallest RMSE for each scenario.

with our approach; we compare our estimator with Chao, generalized Chao, M_t, and M_{th} models. In Table 3, we report the estimated population size $\hat{N}$, the corresponding 95% CI and Akaike information criterion (AIC). We only show the results for a selection of the models we have estimated, omitting several other ones which had a clearly worse performance in terms of AIC. For Chao's estimator, we report AIC computed through its likelihood-based representation (Böhning *et al.*, 2013; Farcomeni, 2018; Dotto and Farcomeni, 2018). In Table 3 letters h, o, t, and b stand for unobserved heterogeneity, covariates, time effects, and behavioral effects, respectively. Numbers refer to the number of latent classes C for models with unobserved heterogeneity (see Equation 10). The notation TI_{htb2}^{corr} refers to a constrained model, which includes unobserved heterogeneity based on $C = 2$ latent classes, time, and self-correcting behavioral effects; but with fixed $\alpha = (0,0)$, $\beta = (1,1)$, and $\theta_1 = 0$.

Covariates do not seem to be important (it shall be remarked that they are not also considering them separately), which is a mild evidence that there is little age and gender profiling. The gender effect might also be lacking due to the very limited number of female drug dealers, making it difficult to detect the effect. For the data at hand TI models give fairly stable estimates of population size regardless of the formulation. The best model in terms of AIC is TI_{htb2}^{corr}, which has no self-exciting effect and shows a mild improvement over the unconstrained TI_{htb2} model. The estimated population size $\hat{N} = 91,368$

drug dealers in Italy for years 2005 and 2006, with a 95% CI ranging from 81,032 to 101,704. Parameter estimates for the chosen model can be found in Table 4. One of the latent classes (about $\hat{\pi}_1 = 39.5\%$ of the dealers) shows no self-correcting effect, and hence no behavioral effects overall. This provides evidence that one subgroup of dealers does not modify its behavior after the first capture. The other latent class, regarding $\hat{\pi}_2 = 60,5\%$ of the dealers, shows a self-correcting effect, according to which dealers have a decreased risk of being stopped after the first capture. This trap-shyness effect is reasonable in the context of criminal offenses, where many dealers become more cautious after being caught by the police. In Table 4, one can see that the behavioral parameter θ_2 is small but relevant, being linked to the scale of the data. It can be concluded that the correcting effect slowly decreases over time. The fitted intensity is shown in Figure 2. For the first latent class (left panel), with no behavioral effects, the intensity only depends on η and μ_1 and smoothly decreases over time irrespective of the captures. In the second latent class, we also have θ_2 and the number of captures enters the computations. The right panel of the figure shows three fitted intensities, partitioning subjects by the number of captures; for each group, the mean capture times are taken to show the jumps in the intensity function representing the trap-shyness effect.

Other models used for comparison, with the exception of M_{th} which might be possibly overestimating the population size, give similar estimates. On the other hand, Chao's and M_t estimators have larger CI and can be thus
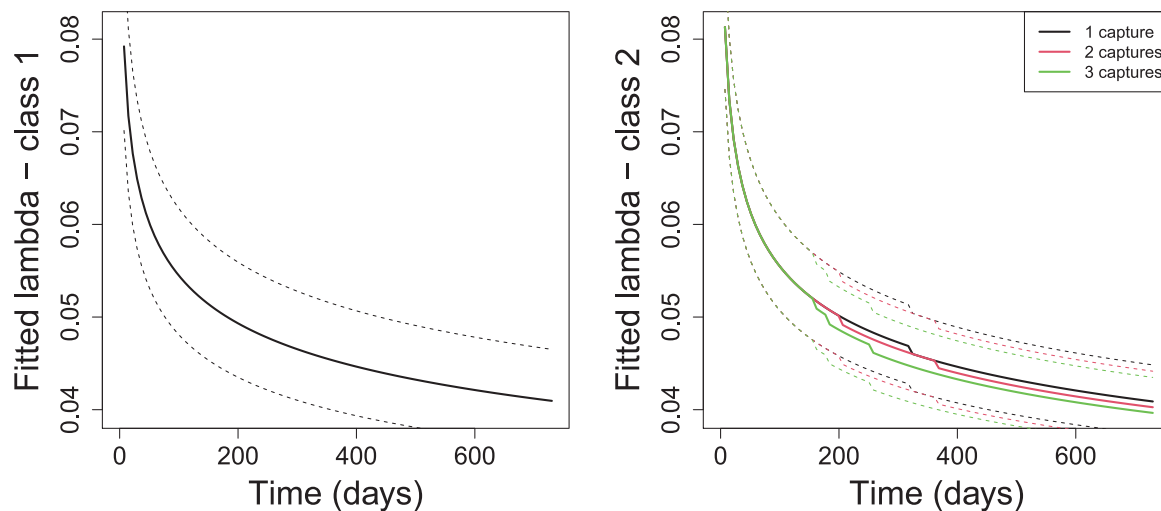
FIGURE 2 Drug dealers data: The fitted intensity function for latent class 1 (left) and latent class 2 (right). Dashed lines show the 95% CI. This figure appears in color in the electronic version of this article, and any mention of color refers to that version

TABLE 3 Drug dealers data: Population size $\hat{N}$, 95% CI, and AIC values for a selection of approaches

	$\hat{N}$	95% CI	AIC
Time interaction processes			
TI_{htb2}^{corr}	91,368	(81,032; 101,704)	720.03
TI_{hotb2}^{corr} (sex, age)	92,113	(81,999; 102,227)	721.26
TI_{ht2}	91,368	(75,357; 107,379)	728.03
TI_{hot2} (sex, age)	92,241	(71,199; 113,283)	729.26
TI_{tb}	91,389	(74,969; 107,809)	732.03
TI_{otb} (sex, age)	92,186	(74,832; 109,540)	733.26
TI_{htb2}	91,386	(75,515; 107,257)	740.03
TI_{hotb2} (sex, age)	92,056	(82,025; 102,087)	744.03
TI_{htb3}	91,650	(69,636; 113,664)	748.03
TI_{hotb3} (sex, age)	92,183	(81,930; 102,436)	749.26
TI_{hob2} (sex, age)	93,737	(50,861; 136,613)	809.5
TI_{hb2}	91,410	(51,284; 131,536)	849.72
Literature models			
	$\hat{N}$	95% CI	AIC
LY	91,369	(73,940; 108,798)	798.03
Chao	92,094	(73,982; 110,206)	839.62
Poisson	103,491	(24,040; 182,943)	839.62
M_t	92,573	(72,910; 112,236)	856.67
M_{th}	105,674	(74,083; 137,265)	861.26
GC (sex,age)	93,986	(91,189; 96,783)	945.08

Note: For time interaction processes, the number indicates the number of latent classes in models with heterogeneity. The absence of a number indicates a single latent class and the absence of unobserved heterogeneity components. The ^{corr} constrained model only contains the self-correcting behavioral effect. For the generalized Chao (GC) estimator, we indicate the covariates used in parenthesis.

Abbreviations: AIC, Akaike information criterion; CI, confidence interval; GC, generalized Chao; LY, Lin and Yip model.

TABLE 4 Drug dealers data: parameter estimates, CI, *p*-values for model TI_{htb2}^{corr}

Estimated value		95% CI	*p*-value
θ_2	1.533e−05	(1.533e−05;1.533e−05)	<0.001
μ_1	−3.040	(−3.166; −2.915)	<0.001
μ_2	−3.039	(−3.130; −2.949)	<0.001
η	0.856	(0.854; 0.858)	<0.001
π_1	0.395	(0.392; 0.398)	
π_2	0.605	(0.602; 0.608)	

Abbreviation: CI, confidence intervals.

deemed to be slightly less precise. The GC estimator gives a very short CI, which might be suspected not to cover the true population size, especially in the light of the evidence against the importance of covariates. The fact that covariates might increase the Mean Square Error (MSE) of population size estimates was also noted in Farcomeni (2018).

We finally assign, using posterior probabilities, each subject to her/his most likely latent class. This allows us to clarify the operating characteristics of our model. A proportion of 39.5% (first latent class) of subjects do not show behavioral response. These can be seen to have been observed only once: all subjects with more than one observation time are assigned to the second latent class. Additionally, observation times for the second latent class are generally smaller than those of the first one. This, under a standard Weibull baseline hazard, is consistent with a self-correcting behavior. The very few subjects with three captures lead to the slow loss of memory as estimated. Hence, subjects are either captured late and only once (first latent class) or earlier, maybe more than once (second latent

class). The data can be expected not to be consistent with other configurations: for instance, in order to claim trap-happiness for the second latent class, one would have to obtain a very low common hazard early on, which under the Weibull model might not be consistent with a high later hazard.

6 | CONCLUSIONS

We have proposed a general framework for estimating the size of closed populations with continuous-time capture–recapture experiments. A number of proposals are present in the recent literature regarding behavioral models for capture–recapture data in discrete time (e.g., Yang and Chao, 2005; Farcomeni, 2011; Alunni Fegatelli and Tardella, 2016). The case of continuous-time remains as yet relatively unexplored. The proposed time-interaction processes are arguably useful for modeling complex behavioral effects in continuous-time population size estimation. Time-interaction processes embed both self-exciting and self-correcting processes and are of independent interest in our opinion. Our model for population size estimation is based on a conditional likelihood formulation for time-interaction processes, in which units with no events are never observed. It allows for covariates, unobserved sources of heterogeneity, and time effects; and it is particularly useful in situations where a complex behavioral effect may be expected, as in our motivating application. Our model compares to Hwang and Chao (2002), which can also take into account time effects, behavioral response, and individual heterogeneity. There are two main differences: unlike Hwang and Chao (2002), and more in the spirit of Lin and Yip (1999), we specify a parametric baseline hazard. On the other hand, in Hwang and Chao (2002) a classical constant and permanent effect are specified for the behavioral response, which is a special case of our time-interacting framework. For the data at hand regarding the number of drug dealers in Italy over the years 2005 and 2006, our best population size estimate is equal to 91,368 dealers. Over a population of 58.14 million (Italy official data 2006), this equals 1.5 dealers for each 1000 Italians. We believe this estimate to be credible when we consider the definition of a drug dealer according to the Italian law at the time of the study. In addition, our conclusions compare well with Bouchard and Tremblay (2005), whose estimates range from 1 to 2.7 for each 1000 Quebec inhabitants in 1998.

We find that about 60% of the dealers underwent a self-correcting behavioral response according to which they were slightly less likely to be stopped again by police, shortly after the event, all else fixed. It could be argued that the very limited number of recaptured individuals in

the real data example could make it difficult to estimate behavioral effects. It shall be kept in mind, though, that the model takes into account the time to first capture, the time between captures (if any), and the time between the last capture and the end of follow-up. Since subjects are continuously at risk of being captured, a subject that is observed only once but shortly after time zero is different from a subject that is observed only once but shortly before the end of follow-up. The latter is unlikely to have the possibility to show a behavioral response and will mostly contribute to the baseline hazard. The former can be deemed to be trap-shy, trap-happy, or not to have a behavioral response, depending on what is consistent with the common baseline hazard. There is no risk of confounding with the baseline hazard, since it has a parametric form and it is common to all subjects (up to equal covariate profile).

A limitation of the current work, as common in this literature, is that we must assume that the population is closed. This has led us to restrict the observation period and discard valuable information about dealers stopped under the 309/90 law after 2006. Moreover, due to obvious privacy reasons we only have limited information on the captured subjects and none about the officers. In addition, further work might be devoted to the derivation of profile CI, joint population size, and total estimation (Farcomeni, 2022), or how to exploit information about the location of identifications (e.g., Borchers and Fewster, 2016; Stevenson *et al.*, 2021).

DATA AVAILABILITY STATEMENT
The data that support the findings of this paper are not publicly available due to privacy restrictions.

ORCID
Linda Altieri https://orcid.org/0000-0001-5667-3157
Alessio Farcomeni https://orcid.org/0000-0002-7104-5826

REFERENCES

Aalen, O.O. (1989) A linear regression model for the analysis of life times. *Statistics in Medicine*, 8, 907–925.

Alunni Fegatelli, D. and Tardella, L. (2016) Flexible behavioral capture-recapture modeling. *Biometrics*, 72, 125–135.

Böhning, D. (2008) A simple variance formula for population size estimators by conditioning. *Statical Methodology*, 5, 410–423.

Bohning, D. and Friedl, H. (2021) Population size estimation based upon zero-truncated, one-inflated and sparse count data. *Statical Methods and Applications*, 30, 1197–1217.

Bohning, D. and Van Der Heijden, P.G.M. (2019) The identity of the zero-truncated, one-inflated likelihood and the zero-one-truncated likelihood for general count densities with an application to drink-driving in Britain. *Annals of Applied Statistics*, 13, 1198–1211.

Böhning, D., Vidal-Diez, A., Lerdsuwansri, R., Viwatwongkasem, C., and Arnold, M. (2013) A generalization of Chao's estimator for covariate information. *Biometrics*, 69, 1033–1042.

Bonnet, A., Herrera, M.M., and Sangnier, M. (2021) Maximum likelihood estimation for Hawkes processes with self-excitation or inhibition. Preprint. *arXiv* 2103.05299.

Borchers, D. and Fewster, R.M. (2016) Spatial capture-recapture models. *Statistical Science*, 31, 219–232.

Bouchard, M. and Tremblay, P. (2005) Risks of arrests across drug markets: A capture-recapture analysis of hidden dealer and user populations. *Journal of Drug Issues*, 35, 733–754.

Chao, A. (1987) Estimating the population size for capture-recapture data with unequal catchability. *Biometrics*, 43, 783–791.

Chao, A. and Lee, S.-M. (1993) Estimating population size for continuous time capture-recapture models via sample coverage. *Biometrical Journal*, 35, 29–45.

Chao, A., Tsay, P.K., Lin, S.-H., Shau, W.-Y. and Chao, D.-Y. (2001) The applications of capture-recapture models to epidemiological data. *Statistics in Medicine*, 20, 3123–3157.

Chen, Q., Elliott, M.R., Haziza, D., Yang, Y., Ghosh, M., Little, R.J.A., Sedransk, J. and Thompson, M. (2017) Approaches to improving survey-weighted estimates. *Statistical Science*, 32, 227–248.

Chiang, S.-C., Chen, C.-Y., Chang, Y.-Y., Sun, H.-J. and Chen, W. (2007) Prevalence of heroin and methamphetamine male users in the northern Taiwan, 1999–2002: Capture-recapture estimates. *BMC Public Health*, 7, 292.

Coull, B.A. and Agresti, A. (1999) The use of mixed logit models to reflect heterogeneity in capture-recapture studies. *Biometrics*, 55, 294–301.

Dotto, F. and Farcomeni, A. (2018) A generalized Chao estimator with measurement error and external information. *Environmental and Ecological Statistics*, 25, 53–69.

Farcomeni, A. (2011) Recapture models under equality constraints for the conditional capture probabilities. *Biometrika*, 98, 237–242.

Farcomeni, A. (2016) A general class of recapture models based on the conditional capture probabilities. *Biometrics*, 72, 116–124.

Farcomeni, A. (2018) Fully general Chao and Zelterman estimators with application to a whale shark population. *Journal of the Royal Statistical Society (Series C)*, 67, 217–229.

Farcomeni, A. (2022) How many refugees and migrants died trying to reach Europe? Joint population size and total estimation. *Annals of Applied Statistics*, Available online.

Farcomeni, A. and Scacciatelli, D. (2013) Heterogeneity and behavioral response in continuous time capture-recapture, with application to street cannabis use in Italy. *The Annals of Applied Statistics*, 7, 2293–2314.

Fine, J.P. (2002) Comparing nonnested Cox models. *Biometrika*, 89, 635–647.

Godwin, R. and Bohning, D. (2017) Estimation of the population size by using the one-inflated positive Poisson model. *Journal of the Royal Statistical Society (Series C)*, 66, 425–448.

Hawkes, A.G. (1971) Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58, 83–90.

Hay, G. and Richardson, C. (2016) Estimating the prevalence of drug use using mark-recapture methods. *Statistical Science*, 31, 191–204.

Huggins, R.M., Yip, P.S.F., and Stoklosa, J. (2016) Nonparametric estimation of the number of drug users in Hong Kong using repeated multiple lists. *Australian and New Zealand Journal of Statistics*, 58, 1–13.

Hwang, W. and Chao, A. (2002) Continuous-time capture-recapture models with covariates. *Statistica Sinica*, 12, 1115–1131.

Illian, J., Penttinen, A., Stoyan, H., and Stoyan, D. (2008) *Statistical Analysis and Modelling of Spatial Point Patterns*. Chichester: Wiley.

Isham, V. and Westcott, M. (1979) A self-correcting point process. *Stochastic Processes and Their Applications*, 8, 335–347.

Lin, D. and Ying, Z. (1995) Semiparametric analysis of general additive-multiplicative hazard models for counting processes. *Annals of Statistics*, 5, 1712–1734.

Lin, D. and Yip, P. (1999) Parametric regression models for continuous time removal and recapture studies. *Journal of the Royal Statistical Society (Series B)*, 61, 401–411.

Liu, Y., Liu, Y., Li, P. and Qin, J. (2018) Full likelihood inference for abundance from continuous time capture-recapture data. *Journal of the Royal Statistical Society (Series B)*, 80, 995–1014.

Matechou, E. and Caron, F. (2017) Modelling individual migration patterns using a Bayesian nonparametric approach for capture–recapture data. *Annals of Applied Statistics*, 11, 21–40.

Mohler, G.O., Short, M.B., Brantingham, P.J., Schoenberg, F.P. and Tita, G.E. (2011) Self-exciting point process modeling of crime. *Journal of the America Statistical Association*, 106, 100–108.

Moller, J., Syversveen, A.R. and Waagepetersen, R.P. (1998) Log Gaussian Cox processes. *Scandinavian Journal of Statistics*, 25, 451–482.

Oakes, D. (1975) The Markovian self-exciting process. *Journal of Applied Probability*, 12, 69–77.

Ogata, Y. (1999) Seismicity analysis through point-process modeling: a review. *Pure and Applied Geophysics*, 155, 471–507.

Otis, D.L., Burnham, K.P., White, G.C. and Anderson, D.R. (1978) *Statistical Inference from Capture Data on Closed Animal Populations*. Wildlife Monographs, Vol. 62. Louisville, KY: The Wildlife Society.

Overstall, A.M., King, R., Bird, S.M., Huchinson, S.J. and Hay, G. (2014) Incomplete contingency tables with censored cells with application to estimating the number of people who inject drugs in Scotland. *Statistics in Medicine*, 33, 1564–1579.

Reynaud-Bouret, P. and Schbath, S. (2010) Adaptive estimation for Hawkes processes; application to genome analysis. *Annals of Statistics*, 38, 2781–2822.

Rivest, L.P. and Baillargeon, S. (2007) Applications and extensions of Chao's moment estimator for the size of a closed population. *Biometrics*, 63, 999–1006. https://doi.org/10.1111/j.1541-0420.2007.00779.x

Rotondi, R. and Varini, E. (2019) Failure models driven by a self-correcting point process in earthquake occurrence modeling. *Stochastic Environmental Research and Risk Assessment*, 33, 709–724.

Sanathanan, L. (1972) Estimating the size of a multinomial population. *The Annals of Mathematical Statistics*, 43, 142–152.

Schoenberg, F. and Bolt, B. (2000) Short-term exciting, long-term correcting models for earthquake catalogs. *Bulletin of the Seismological Society of America*, 90, 849–858.

Schofield, M.R., Barker, R.J. and Gelling, N. (2018) Continuous-time capture-recapture in closed populations. *Biometrics*, 74, 626–635.

Silverman, B.W. (2020) Multiple systems analysis for the quantification of modern slavery: classical and Bayesian approaches (with discussion). *Journal of the Royal Statical Society (Series A)*, 183, 691–736.

Stevenson, B.C., Fewster, R.M. and Sharma, K. (2021) Spatial correlation structures for detections of individuals in spatial capture–recapture models. *Biometrics*. Early view. https://doi.org/10.1111/biom.13502

Van Der Heihden, P.G.M., Bustami, R., Cruyff, M.J.L.F., Engbersen, G. and van Houwelingen, H.C. (2003) Point and interval estimation of the population size using the truncated Poisson regression model. *Statistical Modeling*, 3, 305–322.

Vere-Jones, D. and Ogata, Y. (1984) On the moments of a self-correcting process. *Journal of Applied Probability*, 21, 335–342.

Xi, L., Yip, P. and Watson, R. (2007) A unified likelihood-based approach for estimating population size in continuous-time capture-recapture experiments with frailty. *Biometrics*, 63, 228–236.

Yang, H.-C. and Chao, A. (2005) Modeling animals' behavioral response by Markov chain models for capture-recapture experiments. *Biometrics*, 61, 1010–1017.

Yip, P., Huggins, R. and Lin, D. (1996) Inference for capture-recapture experiments in continuous time with variable capture rates. *Biometrika*, 83, 477–483.

SUPPORTING INFORMATION

Web Appendices referenced in Sections 3.2 are available with this paper at the Biometrics website on Wiley Online Library. The R code for implementing the methods presented in this work can also be found as Supporting Information with this paper at the Biometrics website on Wiley Online Library. The code is also freely available at https://github.com/afarcome/captureTIP.