An analysis of recidivism by gender based on Brazilian prison microdata*

Francis Petterini

PPGEco/UFSC [f.petterini@ufsc.br]

Jeniffer Gonçalves

PPGEco/UFSC [jeniffer.g@posgrad.ufsc.br]

Silvio Sallum

TCE/SC [silvio.sallum@tcesc.tc.br]

May 17, 2022

Abstract

In many countries, recidivism of offenders is analyzed using prison microdata to formulate public policies, but in Brazil this is still rare; and even in the international literature there are few studies by gender. This study helps fill both of these gaps by tabulating thousands of prison records in Santa Catarina in an unprecedented way. We find, for example, that women enter prison at an older age than men, are more likely to participate in work and education programs and psychological treatments, and receive more family visits. Recidivism rates are lower and the length of freedom (after release from prison) is longer for women. However, as the number of prior convictions increases, everyone gets caught in a cycle of incarceration and release. The conclusion is that a better policy to reduce recidivism would be to reduce the number of young offenders, taking into account gender.

Keywords**:** recidivism, gender economics, prison microdata.

JEL**:** C41, J12, K36.

*Research sponsored by CNPq 304653/2018-7.

1 Introduction

The statistical information program for the Brazilian prison system (Infopen) shows that the number of female and male prisoners increased by 900% and 209%, respectively, between 2000 and 2020, while the Brazilian population as a whole grew by 21%. IPEA [2015] and Depen [2019] suggest that this substantial increase in arrests is related to the high recidivism rate of offenders in the country. Moreover, these two papers stimulate numerous discussions about the problems that society's negligent handling of these issues can cause – e.g., more violence – and show that the literature dealing with prison microdata is still sparse in Brazil.

There are hundreds of studies in the international literature that model recidivism using prison microdata from different countries to promote the formulation of more effective public policies to combat this misfortune. To briefly introduce the reader to this line of research, we note that such studies began with Partanen [1969], who conducted a follow-up study of former prisoners in Finland in the 1950s. Since then, recidivism and the effects of work training, education, psychological support, prison infrastructure, family ties, religiosity, gang influence, type of crime committed, sentence length, electronic monitoring, and other topics have been the subject of repeated studies – see reviews in Chung et al. [1991], Gendreau et al. [1996], Hanson and Bussiere [1998], Uggen [2000], Collins [2010], Crutchfield [2011], Conklin [2013], Dooley et al. [2014], Aarten et al. [2015], Mitchell et al. [2016], Walters [2016], Katsiyannis et al. [2018] and Yukhnenko et al. [2019].

The statistical method that has prevailed over time in these papers is the survival analysis of post-release freedom. The motivation for this modeling is rooted in the fact that virtually all prison databases show that recidivism increases with time after release, and that this depends on some individual and local characteristics – e.g, individuals who had problems with the police before the incident tend to return to prison more quickly, just as recidivism rates are higher in areas with high unemployment. As a result, not only has the analysis of a binary indicator of recidivism become popular but also the modeling of various probability density functions for the duration of freedom after incarceration, which depends on criminal history and other covariates.

In this literature, in addition to the rare analyzes for Brazil, there are also few studies comparing female and male prison populations, although there are potentially ambiguous features of women's criminal recidivism [Jurik, 1983, Bloom et al., 2004, Kruttschnitt, 2010, Walmsley, 2014]. For example, on the one hand, motherhood would reduce the likelihood of (re)arrest because the mother would make an effort not to separate from her children; on the other hand, it would increase the likelihood of recidivism because she would have to care for the children while having low employability, which would force a return to crime. Or criminal groups would use women for activities that, while less likely to result in arrest, would be no less demeaning – e.g., as a means of transporting items inserted into the body to give to male inmates during conjugal visits.

In this context, the study presented in this article tabulated, in an unprecedented manner, the prison records of all individuals released between 2013 and 2017 in the Brazilian state of Santa Catarina between 2013 and 2017, and later verified which of these individuals were re-entered into the Infopen system. A total of 2,017 women and 22,981 men were observed who had been legally convicted of some type of crime were observed. By analyzing this information in terms of indicators of recidivism, length of freedom after incarceration, and covariates, the article aims to contribute to the literature in two ways: to promote the description of a Brazilian case based on prison microdata, and to examine gender differences in recidivism and length of freedom after incarceration. Given these goals, some facts stand out:

- When a person is arrested, women are on average older than men – 30 versus 26 years old; 35% of them are mothers, while 19% of men are fathers; 48% have a spouse, compared to 39% of men; and 19% of women and 33% of men have (early) criminal records;
- Women receive more family visits while incarcerated and are more likely to participate in study, work and psychological programs offered in prison;
- Upon release, 24% of women and 40% of men reoffend within five years of release – i.e., a new conviction or preventive detention occurs;
- Once the person reoffends, the duration of freedom (after release) is twice as long for women as for men with similar covariates; and,
- Regardless of gender, the presence of a criminal record appears to be the most important characteristic that increases the likelihood of recidivism and decreases the expected duration of freedom.

In addition to this introduction, the paper consists of five other sections. Section 2 briefly reviews the literature on prison microdata analysis, focusing on the case of women. Section 3 discusses econometrics. Section 4 describes the data. Section 5 examines the results of the post-prison freedom estimates. Finally, Section 6 presents the conclusions and suggestions for future research.

2 Literature

Smart [1977, 1979] (in criminology) and Bartel [1979] (in economics) were probably the first authors to take up the idea that the social sciences (in the broadest sense) and politics had long considered women incapable of committing crimes because it was assumed that their main task was to take care of the family, which would be something completely different from transgressive behavior. From this perspective, female misbehavior was seen as a rare social deviation that should be treated as a physical pathology or even a psychological problem.

This imperceptibility contributed to the introduction of public policies against criminal behavior and recidivism that did not take gender into account. To this day, the architecture of prisons and the design of prisoner transport vehicles bear witness to this inattention, as most of these buildings and vehicles (around the world) are designed for men, with no room for pregnant or nursing women, for example. From another perspective, male prisoners typically have their spouses outside of prison to care for their children. In contrast, mothers in prison often have to witness their children being placed in a non-parental environment, usually under their maternal grandparents' care or in orphanages [Heidensohn, 2012, Tripathi, 2014, Luallen et al., 2018].

Many criminal justice systems do not recognize these disparities, which affects the likelihood that these women will leave prison in worse conditions than when they entered and then recidivate. The need to focus public policy on gender issues is increasingly evident today, especially as the female prison population is growing significantly in many countries, and typically faster than the male prison population [Kruttschnitt and Gartner, 2003, Bloom et al., 2004, Kruttschnitt, 2010, Walmsley, 2014, Gama-Araujo et al., 2020].

Although there have been hundreds of studies of men's recidivism and duration of freedom after incarceration and their covariates since Partanen [1969], there are only slightly more than two dozen analyzes for a female sample. With this in mind, Table 1 shows all the studies found in journals that have analyzed microdata on female prisoners (occasionally compared to male information) and sample characteristics. Most of these studies use data from U.S. and Canadian prisons.

The number of observations varies from a few dozen to thousands, and recidivism rates differ widely across samples. The last point lies in the analysis scenario because prisons with more prior offenders have higher recidivism rates, places and times with more unemployment also have more recidivism, etc.

Paper	Country	Decade	Observations	Recidivism
Jurik [1983]	USA	1970	125 [w]	19% $^{(i)}$
Bonta et al. [1995]	Canada	1980	136 [w]	46% $^{(iii)}$
Jones and Sims [1997]	Canada	1990	1.485 [w] 11.002 [m]	38% [w] 51% [m] $^{(iii)}$
Holtfreter et al. [2004]	USA	1990	134 [w]	37% $^{(iii)}$
Benda [2005]	USA	1990	300 [w] 300 [m]	20% [w] 30% [m] $^{(ii)}$
Reisig et al. [2006]	USA	1990	402 [w]	49% $^{(iii)}$
Rettinger and Andrews [2010]	Canada	1990	531 [w]	52% $^{(iii)}$
Huebner et al. [2010]	USA	2000	506 [w]	47% $^{(iii)}$
Makarios et al. [2010]	USA	2000	1.012 [w] 1.040 [m]	59% [w] 67% [m] $^{(iii)}$
Cobbina et al. [2012]	USA	2000	169 [w] 401 [m]	55% [w] 65% [m] $^{(iii)}$
Eisenbarth et al. [2012]	Germany	2000	80 [w]	31% $^{(iii)}$
Mears et al. [2012]	USA	2000	468 [w] 3.534 [m]	32% [w] 46% [m] $^{(iii)}$
McCoy and Miller [2013]	USA	2000	164 [w] 164 [m]	35% [w] 41% [m] $^{(iii)}$
Steiner and Wooldridge [2014]	USA	2000	570 [w] 5.059 [m]	39% [w] 50% [m] $^{(iii)}$
Scott et al. [2014]	USA	2000	624 [w] 5.011 [m]	58% [w] 69% [m] $^{(iii)}$
Barrick et al. [2014]	USA	2000	255 [w]	44% $^{(iii)}$
Greiner et al. [2015]	Canada	2000	470 [w]	45% $^{(iii)}$
Hedderman and Jolliffe [2015]	England	2000	3.041 [w]	55% $^{(iii)}$
Stalans and Lurigio [2015]	USA	2000	257 [w]	59% $^{(iii)}$
Morash et al. [2016]	USA	2010	226 [w]	42% $^{(iii)}$
Olson et al. [2016]	USA	2010	3.014 [w] 23.520 [m]	57% [w] 73% [m] $^{(iii)}$
Morash et al. [2019]	USA	2010	402 [w]	27% $^{(iii)}$

[w] and [m] indicate women and men, respectively; (i), (ii), (iii) until 1, 2 and 3 years.

Table 1: Papers analyzing female prison microdata and sample characteristics.

Four points stand out in the listed papers and their literature reviews. First, women tend to re-offend less often than men. Generally, this is attributed to motherhood, as they fear a new (or greater) distance from their children. Second, regardless of sample or gender, individuals with lower education and adverse family circumstances (e.g., growing up in poor communities) are more likely to relapse. Third, regardless of sample or gender, criminal history explains much of the recidivism. Fourth, the most commonly used statistical method is duration analysis (of freedom after incarceration), and papers rarely discuss endogeneity issues.

Regarding the last point, Mears et al. [2012] discuss the endogeneity of variables related to the visit. Specifically, they note the existence of a general intuition that visiting family members during incarceration (and the number of visits) is negatively associated with recidivism, but this is generally not observed in the data. The explanation is that there is an omitted variable related to the quality and purpose of the visits that would mask the expected effect. For example, some visits might increase prisoners' psychological resilience and thus reduce recidivism, while other visits might have more diffuse goals, such as bringing drugs into the prison or providing information to leaders of criminal gangs.

At the Brazilian level, Gonçalves Jr and Shikida [2013] investigated cases of 262 former male prisoners in Paraná. IPEA [2015] provides an overview of criminal recidivism, compiling 1,498 cases tried in criminal courts – males accounted for 91.9% of the observations, and a recidivism rate of 24.4% was found, which was higher for young people with low education and lower for females. Sapori et al. [2017] examined a database of former detainees of the Civil Police of Minas Gerais, with 44 female and 756 male cases—the five-year recidivism rate for females was 22.7%, while it was 53.0% for males. Apart from these three studies (and two others from the 1980s cited by these authors), there is no other analysis of microdata on recidivism and length of freedom after incarceration in the Brazilian literature.

3 Modeling

Following the literature, let us model the time from a person's release from prison to some recidivism. First, we define m_0 and m_1 as moments of release and recidivism, respectively, and the variable of interest is $t = m_1 - m_0 > 0$, considering:

$$\ln t = X\beta + u \tag{1}$$

Where: X and β are row and column vectors containing explanatory variables (and a constant) and parameters, respectively; and u is an error term.

The equation (1) is convenient for two reasons: logarithmic (and binary) covariates allow β to be interpreted as elasticity (and semi-elasticity) with respect to t; and estimates of β can be obtained by simply fitting Tobit models. The first point needs no further explanation, and the second is that there are cases of censoring in t, so the logic of Tobit fits well here. Given a sample of n individuals indexed by i, we thus define: $r_i = \mathbb{1}(m_{1i} \leq m^*)$ for recidivism observed in m^*, where $\mathbb{1}$ is the indicator function; $y_i = \min\{t_i, t_i^*\}$; F and f as cumulative and probability density functions on $t_i \mid X_i$ with auxiliary parameter $\sigma > 0$; and normalization $z_i = (\ln y_i - X_i\beta)/\sigma$. Assuming that $u_i \mid X_i$ is independent and identically distributed [i.d.d.], the maximum likelihood estimate of β (and σ) is thus:

$$[\hat{\beta}', \hat{\sigma}]' = \underset{[\beta', \sigma]'}{\operatorname{argmax}} \left\{ \sum_{i=1}^{n} \Big[r_i \times \ln f(z_i) + (1 - r_i) \times \ln(1 - F(z_i)) \Big] \right\} \tag{2}$$

There are three well-documented ways to operationalize (2), depending on the distribution of u_i [Wooldridge, 2010, pp. 997-1000]:

I. if $u_i \mid X_i$ is normal with mean zero and standard deviation σ [so $t_i \mid X_i$ is log-normal], we get $F(z_i) = \Phi(z_i)$ and $f(z_i) = \varphi(z_i)/(\sigma y_i)$ [where Φ and φ are the cumulative and density functions of the standard normal].

II. if $u_i \mid X_i$ is logistic with mean zero and standard deviation $\sigma(\pi/\sqrt{3})$ [so $t_i \mid X_i$ is logistic], we have $F(z_i) = [1 + \exp(-z_i)]^{-1}$ and $f(z_i) = [F(z_i)(1 - F(z_i))]/(\sigma y_i)$.

III. if $u_i \mid X_i$ is Gumbel [type I extreme value] with mean $\sigma\gamma$ [where γ is the Euler-Mascheroni constant] and standard deviation $\sigma(\pi/\sqrt{6})$ [in this case $t_i \mid X_i$ is a Weibull value with a certain type of reparametrization [Greene, 2012, pp. 907-908]], we have $F(z_i) = 1 - \exp(-\exp(z_i))$ and $f(z_i) = [\exp(z_i - \exp(z_i))]/(\sigma y_i)$.

The correspondence of I, II, and III to the data can be assessed using the Akaike Information Criterion (AIC) or similar. Gumbel's alternative may have additional appeal because t_i can thus be interpreted as the realization of an extreme value random variable T in the following sense: if $c \geq 0$ represents association with a criminal grouping, such that $c = 0$ means no involvement and a higher value represents a stronger association; if T is a function of c, such that $T'(c) < 0$ because a higher c increases the probability of prison; and if there is a debt payable to the gang represented by a function $d(c)$; then, $c^* = \operatorname{argmax}\big\{T(c), \text{s.r.: } d(c) = 0\big\}$, and the duration of freedom observed by the researcher would thus be $t = T(c^* = d^{-1}(0))$, reflecting an individual optimization decision, and would then follow one of the extreme value distributions – e.g., Weibull [Beirlant et al., 2006].

As for endogeneity, following the example above, it is assumed that an important variable explaining t is the unobserved c that enters into u. Thus, we see that studies analyzing the duration of freedom after incarceration often use a binary control, $d = 1$, indicating the presence of criminal record, since this usually explains much of t. In this case, c and d would be positively correlated because many gang members have a criminal history. However, having a criminal past does not mean being a gang member, because there are people who are repeatedly arrested for crimes and released without being involved in a criminal organization, otherwise d would be a *proxy* of c. Consequently, the hypothesis that $u \mid X$ is i.d.d. is not valid if d is carelessly embedded in X because the errors $u \mid d$ are interdependent by c.

To illustrate the consequences of this, we take the context of an uncensored linear regression in t and rewrite (1) as $\ln t = X\beta + u = [d, X_1][\delta, \beta_1']' + (\theta c + v)$, where: X_1 is a vector of exogenous controls – i.e., there is no correlation between any element of X_1 and u; δ, β_1, and θ are parameters; and v is an error term independent of d, X_1, and c. Omitting c, in OLS plim $\hat{\delta} = \delta + \theta \times \text{Cov}(d, c)/\text{Var}(d)$ [Wooldridge, 2010, pp. 65-67].In the case studied here, $\delta < 0$ is to be expected because those who have a background have, on average, already shown a criminal tendency, which reduces the chance of remaining free; $\theta < 0$ because everyone who is a member of a gang is a criminal; and $\text{Cov}(d, c) > 0$, for the reasons given above. In absolute terms, then, the OLS estimator would inflate the effect of d if c were omitted.

In connection with a Tobit, Wooldridge [2010] [pp. 681-685] details the case that d would be continuous, and develops a system of simultaneous equations between the dependent and endogenous variables based on Smith and Blundell [1986]. On the other hand, Maddala [1986] [pp. 120-122] presents a proposal for the binary case, whose parsimonious implementation is achieved by fitting a probit with continuous endogenous variable in the spirit of Greene [2012] [pp. 746-750]:

$$\begin{cases} \ln y = [d, X_1][\delta, \beta_1']' + u \\ d = \mathbb{1}(d^* = X_2\beta_2 + u_2 > 0) \end{cases} \quad ; \quad [u/\sigma, u_2]' \sim N\Big([0, 0]', \left[\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right]\Big) \tag{3}$$

Where: X_2 and β_2 are row and column vectors containing exogenous explanatory variables (and a constant) and parameters, respectively; δ is the parameter of the endogenous variable; u_2 is an error term; $\rho \in (-1, 1)$ is the correlation coefficient between u and u_2 assuming a bivariate normal distribution; and the rest follows as before.

The specification (3) differs from that discussed in Greene [2012] in two points: here endogeneity comes into the equation of the continuous dependent variable, and there is censoring in the continuous dependent variable. This specification makes modeling easier, since the problem is reduced to fitting the mass functions of the reference structure. In particular, when ($r = 1$ and $d = 1$) and ($r = 1$ and $d = 0$) occurs $\varphi(z)/(\sigma y) \times \Phi(z_2)$ and $\varphi(z)/(\sigma y) \times \Phi(-z_2)$, respectively, where $z_2 = (X_2\beta_2 + \rho z)/\sqrt{1 - \rho^2}$, because $u_2 \mid u/\sigma \sim N(\rho z, 1 - \rho^2)$. When $r = 0$ and $d = 1$, $\Pr(t > t^*, d^* > 0) = \Pr(u/\sigma > z, u_2 > -X_2\beta_2) = \Phi_2(-z, X_2\beta_2; \rho)$, where Φ_2 is the standard bivariate cumulative normal function. Finally, $r = 0$ and $d = 0$ implies $\Pr(t > t^*, d^* \leq 0) = \Phi_2(-z, -X_2\beta_2; -\rho)$. So here the maximum-likelihood will be:

$$\begin{aligned} [\hat{\beta}', \hat{\sigma}, \hat{\beta}_2', \hat{\rho}]' = \underset{[\beta', \sigma, \beta_2', \rho]'}{\text{argmax}} \Bigg\{ \sum_{i=1}^{n} \Big[& \mathbb{1}(r_i = 1 \text{ e } d_i = 1) \times \ln[\varphi(z_i)/(\sigma y_i) \times \Phi(z_{2i})] \\ & + \mathbb{1}(r_i = 1 \text{ e } d_i = 0) \times \ln[\varphi(z_i)/(\sigma y_i) \times \Phi(-z_{2i})] \\ & + \mathbb{1}(r_i = 0 \text{ e } d_i = 1) \times \ln \Phi_2(-z_i, X_{2i}\beta_2; \rho) \\ & + \mathbb{1}(r_i = 0 \text{ e } d_i = 0) \times \ln \Phi_2(-z_i, -X_{2i}\beta_2; -\rho) \Big] \Bigg\} \end{aligned} \tag{4}$$

Interestingly, it is possible to estimate the entire system without instruments. This becomes clear when the variable of interest is transformed from $\ln y$ to $r^* = t^* - t$, where $r = \mathbb{1}(r^* = X\beta + u > 0)$. This procedure makes (3) into a recursive biprobit, and it is well documented that such model can be estimated without instruments, provided there is sufficient variability in the data – details, for example, in Wilde [2000], Greene [2012] [pp. 785-787] or Filippini et al. [2018]. Consequently, the same is true here, just remember that the crucial point for the complete identification of a system is the impossibility of its equations being written as linear combinations of each other. To see this, we take $\lambda \in (0, 1)$ and write:

$$\lambda \ln y + (1-\lambda)d^* \Leftrightarrow \ln y = -\frac{\lambda}{1-\lambda}d^* + \delta d + X_1\beta_1 + X_2\Big(\frac{\lambda}{1-\lambda}\beta_2\Big) + \Big(\frac{\lambda}{1-\lambda}u_2 + u\Big) \tag{5}$$

This statement shows that the specification of $\ln y$ is structurally different from the first equation of the system (3) due to the term $-(\lambda/(1-\lambda))d^*$, even if X_1 contains all the elements of X_2, and the same is true for the other equation. Thus, the classical identification problem does not exist here; however, it would exist if $d^* = d$ and $X_1 = X_2$ – i.e., if the endogenous variable were treated as continuous and there was no instrumentalization.

It remains to adapt the cases where u follows Logistic and Gumbel distributions, and for this we can apply a Gaussian copula [Nelsen, 2007, Trivedi and Zimmer, 2007]. Specifically, in (3) the joint cumulative distribution of $[u/\sigma, u_2]'$ must be fitted to $\Phi_2(\Phi^{-1}(F(z)), u_2; \rho)$, where Φ^{-1} is the inverse of the standard normal distribution. Thus, u is assumed to follow one of the distributions defined at points I, II, and III in the marginal region, and u_2 follows a standard normal distribution. Taking advantage of the fact that the copula is defined from the margins, the problem (4) is rewritten as follows:

$$\begin{aligned}[\hat\beta', \hat\sigma, \hat\beta_2', \hat\rho]' = \underset{[\beta', \sigma, \beta_2', \rho]'}{\operatorname{argmax}} \Bigg\{ \sum_{i=1}^{n} \Big[& \mathbb{1}(r_i = 1 \text{ e } d_i = 1) \times \ln[f(z_i) \times \Phi\Big(\frac{X_2\beta_2 + \rho\Phi^{-1}(F(z_i))}{\sqrt{1-\rho^2}}\Big)] \\ & + \mathbb{1}(r_i = 1 \text{ e } d_i = 0) \times \ln[f(z_i) \times \Phi\Big(-\frac{X_2\beta_2 + \rho\Phi^{-1}(F(z_i))}{\sqrt{1-\rho^2}}\Big)] \\ & + \mathbb{1}(r_i = 0 \text{ e } d_i = 1) \times \ln \Phi_2(-\Phi^{-1}(F(z_i)), X_{2i}\beta_2; \rho) \\ & + \mathbb{1}(r_i = 0 \text{ e } d_i = 0) \times \ln \Phi_2(-\Phi^{-1}(F(z_i)), -X_{2i}\beta_2; -\rho)\Big] \Bigg\}\end{aligned} \tag{6}$$

If F and f then take specifications I, II, and III, we model u as Normal, Logistic and Gumbel, respectively, and d as probit. In these terms, $\rho = 0$ implies that (6) collapses into a problem adiviously separated by a part identical to (2) and a Probit for the second line of the system (3), ignoring possible endogeneity – because $\Phi_2(-\Phi^{-1}(F(\cdot)), \cdot; 0) = \Phi(-\Phi^{-1}(F(\cdot))) \times \Phi(\cdot) = (1 - F(\cdot)) \times \Phi(\cdot)$; and if it is necessary to calculate the intercept of the first equation in the case of Gumbel, one must subtract $\sigma\gamma$ in the numerator of z to reach the mean zero in (3), which is irrelevant in any other case.

Finally, the proposal of Sedgley et al. [2010] and Mears et al. [2012] is applied in a complementary way, in the following sense: if you consider the possibility that being a woman/man might be associated with some sort of bias in incarceration, punishment, or the like, then a male/female propensity score matching (PSM) based on observational data might define a subsample in which this bias is mitigated; repeating the modeling in this subsample would therefore only scale the gender effect. It would be an exercise in robustness to use an additional method to test whether there is a gender difference in the pattern of recidivism in the sample.

4 Data

The data comes from Infopen, a system created in 2004 to consolidate and standardize information about prisons in Brazil and coordinated by the National Penitentiary Department (Depen). It works as follows: in each prison, jail, etc., daily activities are recorded in prisoners' entries and exits, interviews with social services, medical care, participation in work and educational activities, visits, court decisions, escape attempts, etc. These reports are produced by a local government agency that often uses its own systems. Finally, the information is standardized and entered into Infopen, usually by a Department of Public Safety or a Department of Corrections.

Based on these records, Figure 1 shows the evolution of the prison population in Brazil, by gender, over the period 2000-2020. On the left axis is the scale for women, on the right for men. The number of women starts with 4 thousand in 2000 and reaches 40 thousand in 2020 – a growth of 900%. For men, it starts with 233 thousand in 2000 and reaches 720 thousand in 2020, increasing 209%.

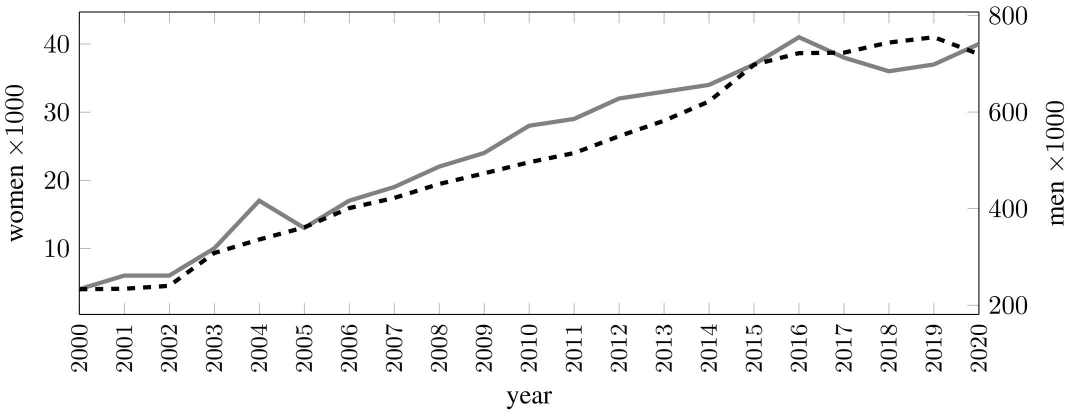

Figure 1: Prison population in Brazil, by gender, period 2000-2020.

The trends are almost identical, and the explanation can be found in Depen [2019]. About 2/3 of the cases involve young people caught in the act, first-time offenders who serve criminal groups as pawns in theft and drug trafficking. They are usually poor and are supported by the public defender's office, which in turn is overwhelmed with many cases. Once arrested, these young people fall into a kind of trap, because it is very common to join a criminal group (or strengthen the relationship if it already exists) in order to maintain physical integrity in prison. The groups demand money and require participation in criminal activities until the debt is paid. This, in turn, increases the likelihood of re-arrest and the overall recidivism rate, while a mass of similar individuals continue to be arrested and the process continues.

In Santa Catarina, the Secretary of Penitentiary Administration (SAP) is in charge of prisons, whose information is recorded in its own system called Ipen, which subsidizes Infopen. In November 2018, we gained access to the list of all prisoners released from the state between January 2013 and December 2016, totaling more than 70 thousand observations. After deduplicating and removing cases of preventive measures, temporary detention, disciplinary incidents, and other proceedings without a final judgment, 24,998 cases remained for analysis.

We then reviewed which individuals had a new conviction or temporary detention ordered – this is the definition of recidivism used here– and disaggregated them by gender and Criminal Record (CR) in the following sense: we checked whether the person had been arrested at least once in the analyzed period before the process leading to release – Table 2 summarizes this. Of the 2,017 women observed, 393 (19.5%) had CR. Of the 22,981 men observed, 7,620 (33.2%) had

CR. Among women with and without CR, 5.9% and 99.2%, respectively, reoffended within five years; for men, these rates were 10.6% and 99.9%, respectively.

Duration of freedom (t in years)	Women						Men					
	Without CR			With CR			Without CR			With CR		
	n	%	%†	n	%	%†	n	%	%†	n	%	%†
$t \leq 1$	18	1.1	1.1	230	58.4	58.4	380	2.5	2.5	5,110	67.1	67.1
$1 < t \leq 2$	38	2.3	3.4	84	21.4	79.8	435	2.8	5.3	1,570	20.6	87.7
$2 < t \leq 3$	13	.8	4.2	44	11.2	91.0	339	2.2	7.5	635	8.3	96.0
$3 < t \leq 4$	19	1.2	5.4	29	7.4	98.4	272	1.8	9.3	273	3.6	99.6
$4 < t \leq 5$	8	.5	5.9	3	.8	99.2	206	1.3	10.6	21	.3	99.9
$t > 5$	3	.2	6.1			99.2	61	.4	11.0	1	.0	99.9
Cases $m_1 > m^*$	1,525	93.9		3	.8		13,668	89.0		10	.1	
Total	1,624	100		393	100		15,361	100		7,620	100	

† Cumulative.

Table 2: Observations per range of duration of freedom, gender and Criminal Record (CR).

A higher recidivism rate among those who have previously offended had already been expected. Still, the figure of 99% for both women and men leads to the fatal conclusion that a person who recidivizes is almost certainly in constant movement in and out of prison. This conclusion, coupled with the fact that the number of offenses continues to rise, makes it clear that the upward trend shown in Figure 1 will continue permanently unless the state intervenes vigorously. Table 2 also shows instances of censoring ($m_1 > m^*$): no recidivism was observed in 75.8% of female and 59.6% of male observations between the date of release from prison and November 2018. Thus, this suggests that females are less likely to recidivate, consistent with what has been documented in the literature.

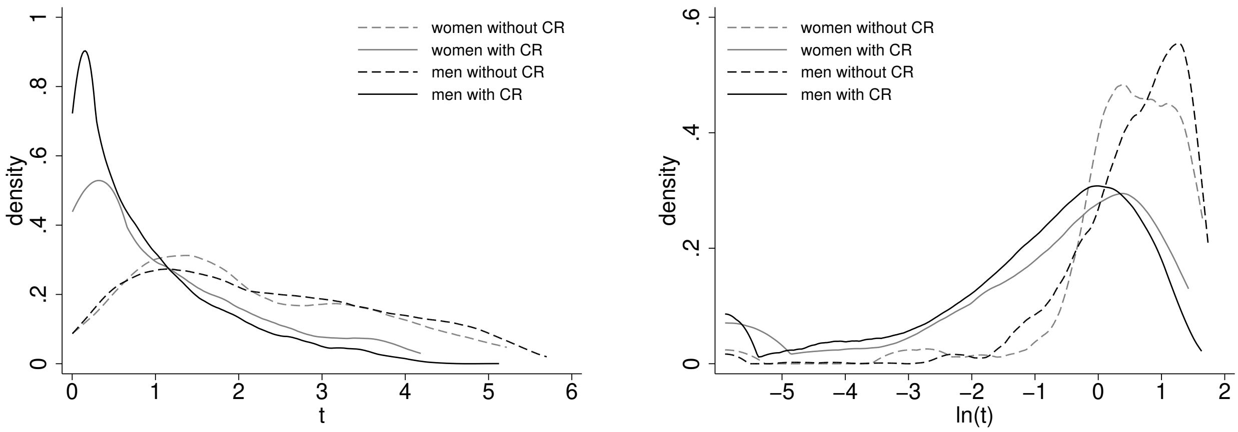

Figure 2: Smoothed histograms of t and $\ln t$ among cases $m_1 \leq m^*$, by gender and CR.

Looking only at recidivism, Figure 2 (left) shows smoothed histograms (i.e., *kernel* densities) of t, in years, by gender and CR. For individuals with CR, densities are higher at low values t, reflecting the 99% rate discussed above. Densities do not show significant visual differences between women and men in this sense, but they do in terms of CR. Figure 2 (right) depicts smoothed histograms of $\ln t$, as specified in the equation (1), in which the problem of censoring is more apparent: there is an asymmetry to the left because the results further to the right at time m^* could not be observed in more than half of the cases.

Table 3 shows the means of observed covariates by gender, as well as the differences and their statistical significance. These indicators were taken mainly from prison social services follow-up forms, and only the last two variables are not dummies. First, the proportion of those with CR is reported. Then, it is noted that about 60% of the individuals are white, which is not surprising since, according to the 2010 census, 4/5 of the population of Santa Catarina is white. In 48.0% of female cases, the social service indicated that there was a spouse (or similar affective relationship), and 35.3% of them had at least one child. In contrast, 39.5% of male cases had a spouse, and the proportion of those with children was 18.9%. So, on average, there are more cases of women with partners and children than men.

Covariate	Women (a)	Men (b)	Difference (b-a)
CR = 1	.195	.332	.137***
White = 1	.612	.608	-.004
Spouse = 1	.480	.395	-.086***
Child = 1	.353	.189	-.163***
Elementary (school) = 1	.051	.047	-.003
Intermediate (school) = 1	.459	.481	.021
Middle (school) = 1	.137	.146	.009
High (school) = 1	.138	.133	-.005
Undergraduate (school) = 1	.182	.156	-.026*
Employment = 1	.333	.364	.031*
Father unknown = 1	.177	.171	-.007
Father arrested = 1	.060	.054	-.006
Sibling arrested = 1	.177	.143	-.034***
Study (at prison) = 1	.116	.047	-.069***
Job (at prison) = 1	.357	.224	-.134***
Psycho (at prison) = 1	.191	.116	-.075***
Visits = 1	.570	.502	-.068***
Try escape = 1	.034	.078	.044***
Age at entry (years)	30.1	26.4	-3.7***
Prison time (years)	1.72	1.86	.14**

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3: Mean of observed covariates, by gender.

The next variables listed in Table 3 concern the level of education. The highest formal level attended was the elementary and intermediate school ("primeiro" and "segundo" cycles of "ensino fundamental" in Brazil) and the middle school (the person completed the "ensino fundamental"), respectively. The sequence indicates that the person attended secondary school and higher education, respectively, and it can be seen that the modal situation is an intermediate education, regardless of gender.

The variable indicating employment concerns self-report to social services that they had a regular job before their arrest, and about 1/3 answered affirmatively. The following three variables indicate whether the father's name is not on the person's birth records and whether it was observed that the father or a sibling is or was incarcerated. For nearly 17% of these individuals, the father's name is not in the civil records. The National Association of Civil Registry Officers indicates that between 5% and 7% of the Brazilian population has this characteristic, which, in agreement with reports in the international literature, suggests that incarcerated individuals tend to have less structured families. Similar percentages are registered for the father, when known, is or was incarcerated. For siblings, this percentage is about 15%, with slightly higher rates for women.

The following variables concern daily life in prison. The first indicates whether they participated in programs to improve education and vocational training and whether they needed psychological support. It can be seen that women are more committed to participating in these programs. The next question refers to whether a family member has signed up for visits. It shows that about half of the families do not and that women receive more visits than men. Next, we examine whether an escape attempt has occurred and found that women are less involved in these cases. Finally, the average age at incarceration is 30 for women and 26 for men, and the average length of imprisonment is close to two years, regardless of gender.

Dummy (D)	dependent variable: r				dependent variable: t			
	cte	w	D	$w \times D$	cte	w	D	$w \times D$
CR = 1	.11***	-.04***	.88***	.04*	2.30***	-.16	-1.44**	.03
White = 1	.45***	-.13***	.02	-.04	1.07***	.21**	.08	-.07
Spouse = 1	.41***	-.17***	-.01	.03	1.12***	.21*	-.06	-.13
Child = 1	.39***	-.15***	.02	-.02	1.08***	.15*	.09	.01
Elementary = 1	.40***	-.16***	-.02	.04	1.12***	.15*	.04	.14
Intermediate = 1	.37***	-.15***	.06	-.01	1.15***	.15*	-.05	.02
Middle = 1	.40***	-.16***	.01	.02	1.11***	.16*	.07	-.03
High = 1	.40***	-.15***	-.01	-.05	1.14***	.16*	-.07	-.07
Undergraduate = 1	.41***	-.16***	-.06*	-.02	1.11***	. 16**	.08	-.01
Employment = 1	.36***	-.12***	.11*	-.09*	1.11***	.18*	.02	-.06
Father unknown = 1	.41***	-.17***	.02*	.07*	1.12***	.16*	.03	-.02
Father arrested = 1	.41***	-.16***	.07*	.03*	1.12***	.15**	-.02	.13
Sibling arrested = 1	.39***	-.15***	.10***	.04*	1.13***	.18**	-.03	-.09
Study = 1	.40***	-.16***	.06***	-.04	1.14***	.20***	-.31***	-.13
Job = 1	.39***	-.18***	.03***	.05	1.18***	. 28***	-.26***	-.16
Psycho = 1	.39***	-.16***	.07***	-.02	1.15***	.20***	-.24***	-.08
Visits = 1	.42***	-.15***	-.04	-.01	1.23***	.27***	-.24*	-.19
Try escape = 1	.38***	-.15***	.20***	-.03	1.15***	.14*	-.22***	.15
Young = 1	.32***	-.08***	.15**	-.15***	1.23***	.13*	-.18***	.01
Short penalty = 1	.34***	-.14***	.12***	-.02	1.24***	.18*	-.12*	-.04

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 4: Estimated parameters for linear regressions r or $t = \beta_0 + \beta_1 w + \beta_2 D + \beta_3 w \times D + u$, where: the dependent variable is an indicator $r = 1$ in this case of recidivism, or the duration of post-prison freedom (t, if $r = 1$); $w = 1$ is a dummy indicating female; D is a dummy for intercept and slope; and, u is an error. In all cases, the number of observations is 24,998.

To see how gender and observed controls may affect recidivism and duration of freedom after incarceration, Table 4 provides estimated parameters for linear regressions r or $t = \beta_0 + \beta_1 w + \beta_2 D + \beta_3 w \times D + u$, where: the dependent variable is an indicator $r = 1$ of recidivism or duration of freedom after incarceration (t, case $r = 1$); $w = 1$ is a dummy for woman; D is an auxiliary dummy to analyze the shift in intercept and slope; each β is parameter; and u is error – in all cases, the number of observations is 24.998. In the first row, $D = 1$ represents the presence of CR, and the numbers are as follows: for a man without CR ($w = 0$ and $D = 0$), the expectation is .11 (or 11%) probability of recidivism; for a woman with without CR ($w = 1$ and $D = 0$), the expectation is .11-.04 = .07 (or 7%) probability of recidivism;

for a man with CR ($w = 0$ and $D = 1$), the expectation is .11+.88=.99 (or 99%) probability of recidivism; and for a woman with CR ($w = 1$ and $D = 1$), the expectation is .11-.04+.88+.04 = .99 (or 99%). In the next block, the shift in intercept and slope for $w = 1$ were not statistically different from zero; hence the expectation of years of freedom after incarceration for men/women with and without CR is 2.30 and 2.30-1.44 = .86, respectively.

In the following rows of the table, the dummy D represents the fact that the individual is white, has a spouse, is a father/mother, has a different level of education, had a regular job before being arrested, has an unknown or incarcerated father, or has an incarcerated sibling. In this order, a pattern emerges: men have a recidivism rate of nearly 40% and a duration of freedom of about 1.1 years, and women have a recidivism rate of nearly 25% and a duration of freedom of about 1.3 years. None of these characteristics has a significant effect on this pattern.

Finally, in the last rows of Table 4, the dummy D represents facts that occurred in prison, namely whether the person voluntarily participated in study, work, and psychological assistance programs, whether they received visits, and whether there was an escape attempt, and whether they were young, and had a short sentence - in the last two cases $= 1$ if they are below the average reported in table 3. In this block, it is also observed that women have a lower recidivism rate and a longer duration of freedom after incarceration. In addition, however, it is also observed that: men, but not women, who participate in programs offered in prison have a higher recidivism rate and a shorter duration of freedom after release.

An explanation for this last point can be found in Sedgley et al. [2010] and Mears et al. [2012], in the following sense: in the USA, the variables “Study”, “Job” and “Psycho” help in obtaining parole, and this is also true for Brazil. So the hypothesis is that a large number of inmates participate in these programs to get out of prison faster, and this behavior could be driven by criminal gangs. So if there is a high proportion of gangsters among these people, this would explain the pattern of higher recidivism and shorter duration of freedom.

5 Estimated results

Results for duration analysis were estimated using the following controls (X and X_1): intercept and slope dummies with respect to women, w, to identify gender differences; all variables of Table 4; and X_2 with indicators “White”, “Father unknown”, “Job” and “Young”. The modeling (3) allows this reduced specification of X_2, and it is the only one with logical consistency that can be used because the other variables were computed during the period when the person was incarcerated (the last time), and CR is by definition something from the person’s past.

In addition, as mentioned in the methodology section, the results were estimated using the whole sample and a subsample adjusted using PSM. This last procedure was operationalized as follows: a probit $w = 1$ using all controls, and then an adjustment was made between women and men using the nearest neighbor criterion. The probit results are reported in the first column of Table A1 in the appendix, and reflect what was already found in the descriptive analysis: an inmate with a spouse and children is more likely to be a woman; and if they participated in vocational/educational programs, received psychosocial services, and received visits, they are more likely to be a woman. The other controls had no statistically significant effect on these probabilities. Before matching, the sample included 24,998 observations; after matching, the sample included 3,324 observations – 1,307 women and 2,017 men. The results of the probit test for the paired sample are in the second column of the table. They show that there is no statistically significant change in the probability of observing a particular gender by another control.

The discussion of the endogeneity of CR (i.e., the variable d) revolves around the following configuration: when $\delta = \rho = 0$, the problem in question does not exist at the expense of the effect of d on the duration of freedom after incarceration (i.e., the variable t); when $\delta \neq 0$ and $\rho = 0$, the problem exists; and when $\delta \neq 0$ and $\rho \neq 0$, endogeneity is addressed. Then, specifications I, II, and III were estimated taking this into account to calculate the log-maximum likelihood ($\ln L$) and the Akaike criterion (AIC) to evaluate these cases. Table 5 shows the results of the problem (6) operationalized in the whole sample (unpaired) and in the sample paired by propensity score between women and men.

In Model I, where t would have a lognormal distribution, $\ln L =$ yields -40,537.7 and -30,908.0 under the constraints $\delta = \rho = 0$ and $\rho = 0$, respectively. Therefore, the inclusion of d in the specification leads to a gain of 23.8 percentage points with respect to pseudo-$R^2 = 1 - \ln L' / \ln L''$, where L' and L'' are the least and most constrained likelihoods, respectively. This shows that CR is an important covariate for explaining t. The gain between the constrained specification $\delta = \rho = 0$ and the unconstrained specification is 25.4 percentage points in pseudo-R^2, and the improvement is confirmed by the smallest AIC. Moreover, the AICs are consistently smaller under the assumption that t has a lognormal distribution; and this does not depend on whether one uses the whole sample or only the matched subsample. Therefore, from now on, we will only consider model I.

Model	Reference	Without PSM ($n = 24,998$) $\delta = 0$, $\rho = 0$	Without PSM $\delta \neq 0$, $\rho = 0$	Without PSM $\delta \neq 0$, $\rho \neq 0$	With PSM ($n = 3,324$) $\delta = 0$, $\rho = 0$	With PSM $\delta \neq 0$, $\rho = 0$	With PSM $\delta \neq 0$, $\rho \neq 0$
I	AIC	81,173.4	61,916.0	60,546.8	9,740.0	7,455.8	7,338.9
	$\ln L$	-40,537.7	-30,908.0	-30,222.4	-4,821.	-3,677.9	-3,618.4
II	AIC	81,389.4	61,999.1	60,990.3	9,980.0	9,855.8	9,389.0
	$\ln L$	-40,889.7	-31,009.1	-30,988.3	-4,899.9	-3,978.0	-3,788.8
III	AIC	82,002.1	63,009.0	62,981.1	9,999.8	9,905.3	9,890.0
	$\ln L$	-41,896.5	-32,204.3	-32,197.1	-4,984.1	-4,151.9	-4,007.0

Table 5: AIC and $\ln L$ of the three specifications of the duration of post-prison freedom by type of constraint.

Table 6 contains the estimates of the first equation of the system (3), and the estimates of the second equation can be found in Table A2 in the appendix. Regarding the effect of CR ($d = 1$) on the duration of freedom, estimates of δ (in absolute values) decrease from 4.699 to .086 when controlling for endogeneity with $\rho \neq 0$ in the full sample; in the paired subsample, they decrease from 4.968 to .081. As for semi-elasticity, the presence of CR would reduce the expectation of the duration of freedom (conditional on idiosyncrasies) by more than 400%; and controlling for endogeneity, this semi-elasticity would be close to 8%. Since in the first case the semi-elasticities are more than 100%, the following interpretation emerges: the presence of CR implies a rightward curve for the prison. This is consistent with the 99% reversal rate mentioned earlier, which does not depend on explanatory variables and is biased by unobserved splitting or some other confounding element. Moreover, the statistically significant semi-elasticities near 8% compared to CR are consistent with what is usually found in the literature. This is explained by the stigma of being an ex-convict – individuals have lower employability and eventually return to crime, whether they have a relationship with a criminal group or simply because they prove to be persistent criminals.

	Without PSM			With PSM		
Covariate	$\delta = 0$ $\rho = 0$	$\delta \neq 0$ $\rho = 0$	$\delta \neq 0$ $\rho \neq 0$	$\delta = 0$ $\rho = 0$	$\delta \neq 0$ $\rho = 0$	$\delta \neq 0$ $\rho \neq 0$
CR ($d = 1$)		-4,699***	-.086*		-4,968***	-.081*
Woman ($w = 1$)	.227*	.461*	.984**	.714*	.988*	1,042***
White $= 1$	.006	.023	.047	-.077	-.038	.006
White $\times w$	.210	.231	.294	.303	.310	.358
Spouse $= 1$	-.086	.046	.048	.442*	.307*	.280
Spouse $\times w$	.195	.003	-.004	-.333	-.261	-.235
Child $= 1$	.036	-.043	.014	.245	.099	.170
Child $\times w$	-.056	.032	-.060	-.266	-.111	-.210
Elementary $= 1$	.300	.026	.023	.242	-.334	-.281
Elementary $\times w$	.015	.078	-.006	-.091	-.080	-.005
Intermediate $= 1$	-.369***	.010	-.038	.373	.475	.362
Intermediate $\times w$	.061	-.176	-.156	-.695	-.647	-.557
Higher $= 1$	-.277*	.009	-.060	-.067	.433	.287
Higher $\times w$	.510	-.078	-.063	.312	-.507	-.404
Undergraduate $= 1$	-.148	.022	.001	-.097	.326	.142
Undergraduate $\times w$	.567	.114	.045	.528	-.194	-.090
Employment $= 1$	-.148	.022	.001	-.097	.326	.142
Employment $\times w$	.567	.114	.045	.528	-.194	-.090
Father unknown $= 1$	.087	.018	.112	.383	-.128	.282
Father unknown $\times w$	-.498*	-.017	-.513	-.808*	.141	-.689
Father arrested $= 1$	-.182	.098	.082	-.066	-.370	-.296
Father arrested $\times w$	.012	-.054	.065	-.117	.410	.430
Sibling arrested $= 1$	-.504***	-.092*	-.062	-.074	-.243	-.226
Sibling arrested $\times w$	.234	-.093	-.212	-.205	.051	-.047
Study $= 1$	-.330**	-.423***	-.348***	-1,153***	-. 924**	-.830
Study $\times w$	.044	-.333	-.186	.872*	.140	.257
Job $= 1$	.059	-.176***	-.195***	.310	-.206	-.237
Job $\times w$	-.992***	-.248	-.212	-1,271***	-.218	-.176
Psycho $= 1$	-.204**	-.107*	-.170***	.050	-.126	-.124
Psycho $\times w$	-.135	-.050	.084	-.396	-.026	.037
Visits $= 1$	.092	-.498***	-.459***	.530	-.381	-.400
Visits $\times w$	.356	-.051	-.140	-.057	-.211	-.216
Try escape $= 1$	-.952***	-.112	-.060	-.749	.164	.063
Try escape $\times w$	.141	-.160	-.093	-.092	-.439	-.243
Young $= 1$	-.810***	-.107**	-.095	-.826***	-.031	-.075
Young $\times w$	.716***	.051	.866	.732	-.021	.691
Short penalty $= 1$	-.554***	-.052	-.028	-.299	-.201	-.254
Short penalty $\times w$	-.034	.025	.124	-.310	.191	.340
Constant	3,489***	4,077***	3,315***	2,250***	3,950***	2,962***
$\ln \hat{\sigma}$	1,096***	.631	1,410***	1,142***	.655***	1,193***
$\tanh^{-1}(\hat{\rho})$			-1,869***			-1,794***
n		24,998			3,324	

* $p < .05$, ** $p < .01$, *** $p < .001$.

Table 6: Estimated results for the first equation of the system (3) from specification I.

Regarding the gender effect, the parameters estimated for $w = 1$ in all models show that the expectation of the duration of freedom is larger for women than for men – regardless of whether one works with the whole sample or only with the subsample. It also happens that in the models that better represent the data (after the AIC), with $\rho \neq 0$, semi-elasticities of .984 and 1.042 are present; therefore, the expected duration of their freedom after incarceration would be about twice as high for women as for men with similar covariates.

Next, we look more closely at the effects of "Study", "Job", "Psycho" and "Visits" because their effects were statistically significant for men (but not for women) and because the state intervenes directly here while the person is in prison – the first three variables are an offer of opportunity, and visits are part of the work of social services to maintain family ties in the face of incarceration. The point is that the semi-elasticities concerning these four variables were negative, and it seems counterintuitive.

To understand this better, consider the context of an uncensored linear regression on t, as before, and take x_K as one of these four controls. Thus, in OLS, plim $\hat{\beta}_K = \beta_K + \theta \times \text{Cov}(x_K, c)/\text{Var}(x_K)$, where β_K and θ are the parameters of the linear projection of x_K and c, respectively, onto $\ln t$ – and the other assumptions formulated similarly above hold. Here, $\beta_K > 0$ is expected because "Study", "Job", "Psycho" and "Visits" would increase the chance of remaining free by increasing people's employability and/or resilience; $\theta < 0$, as before; and $\text{Cov}(x_K, c) > 0$, assuming that the fractions are instructed how to reduce penalties and their visits work in favor of the fraction. In this case, even if $\text{Cov}(x_K, c)$ is relatively small, a sufficiently large θ would show a negative $\hat{\beta}_K$ even though β_K is positive.

Thus, when a more accurate estimate of β_K is critical – e.g., in evaluating programs – Mears et al. [2012] suggests repeating more propensity score matching procedures to mitigate these biases. Note in the strategy used here that the system (3) can be extended to a multivariate structure, but the number of vectors of explanatory variables and correlation parameters would increase linearly and exponentially, respectively. Therefore, even with the addition of a few more endogenous variables, the system estimates would lose operability. Indeed, Nelsen [2007] and Trivedi and Zimmer [2007] (among other authors) show that three or more-dimensional copulas (as well as any other joint probability distribution function) become empirically impractical from a numerical costing perspective due to the so-called "curse of dimensionality". Thus, the cost of modeling more than one endogenous variable is not negligible.

Finally, it should be noted that contrary to the widely accepted hypothesis in the literature, the results presented in Table 6 show zero effect of motherhood (and fatherhood) on the duration of freedom after incarceration – an issue for future research. Moreover, the correlation coefficient between the error terms (both containing c) takes values above -.9 when calculating the hyperbolic tangent; an absolute number close to one, which is likely due to the strong negative effect of unobserved covariates, and it also is an issue for further research.

6 Conclusion

In the last two decades, the number of female and male prisoners in Brazil has increased significantly, due in large part to the high recidivism rate of offenders in the country. This study analyzed the files of thousands of former prisoners in Santa Catarina to contribute to the debate on how to remedy this situation. The main findings were that women enter prison at an older age than men, they participate more in social programs during their incarceration, recidivism rates are lower for women, and they remain free longer after prison. However, once criminal records accumulate, everyone gets caught in a cycle of prison and incarceration, regardless of gender. At first glance, one might think that better policies to reduce recidivism would include a more appropriate form of crime prevention for young people that takes into account the gender differences documented here.

With this exercise, we show that there is an important problem that requires attention and that the microdata exists to analyze it. Every prison system in Brazil feeds Infopen through some kind of public security secretariat, penitentiary administration, or similar. And this information can (and should) be analyzed by researchers for the benefit of society as a whole, but perhaps the administrators of these systems do not know how science can help them. In addition, other states' databases may offer much more information than those examined here, because criminal justice social services operate differently in each system, and perhaps in some places there is much more detail than was examined here.

As for suggestions for future research, the role of the endogeneity of the regressor "having a criminal record" was particularly highlighted here because it may be related to the unobserved variable "belonging to a criminal organization´´. However, this is not the only confounding element in regressions that have recidivism and/or length of post-prison freedom as dependent variables. The literature also discusses other issues that were not explored in detail here. For example, participation in job skills and education programs with the need for parole to get out of prison sooner [Sedgley et al., 2010]; the relationship between participation and obscure interests, such as the introduction of drugs into prison [Mears et al., 2012]; relationships between youth and gangs [Dooley et al., 2014]; and, motherhood with family problems [Kruttschnitt, 2010]. Thus, the discussion of endogeneity problems in this literature seems to be an open question.

Finally, it should be noted that the unemployment rate in Santa Catarina was lower than the national average during the period studied (2013-2018). Since the results on recidivism and the duration of freedom after incarceration must be related to labor market issues, studies with databases from other states – in other periods and socioeconomic contexts – would greatly complement the discussion here.

References

P. G. Aarten, A. Denkers, M. J. Borgers, and P. H. van der Laan. Reconviction rates after suspended sentences: Comparison of the effects of different types of suspended sentences on reconviction in the netherlands. *International journal of offender therapy and comparative criminology*, 59(2):143–158, 2015.

K. Barrick, P. K. Lattimore, and C. A. Visher. Reentering women: The impact of social ties on long-term recidivism. *The Prison Journal*, 94(3):279–304, 2014.

A. P. Bartel. Women and crime: An economic analysis. *Economic Inquiry*, 17(1):29–51, 1979.

J. Beirlant, Y. Goegebeur, J. Segers, and J. L. Teugels. *Statistics of extremes: theory and applications*. John Wiley & Sons, 2006.

B. B. Benda. Gender differences in life-course theory of recidivism: A survival analysis. *International Journal of Offender Therapy and Comparative Criminology*, 49(3):325–342, 2005.

B. Bloom, B. Owen, and S. Covington. Women offenders and the gendered effects of public policy. *Review of policy research*, 21(1):31–48, 2004.

J. Bonta, B. Pang, and S. Wallace-Capretta. Predictors of recidivism among incarcerated female offenders. *The Prison Journal*, 75(3):277–294, 1995.

C.-F. Chung, P. Schmidt, and A. D. Witte. Survival analysis: A survey. *Journal of Quantitative Criminology*, 7(1):59–98, 1991.

J. E. Cobbina, B. M. Huebner, and M. T. Berg. Men, women, and postrelease offending: An examination of the nature of the link between relational ties and recidivism. *Crime & Delinquency*, 58(3):331–361, 2012.

R. E. Collins. The effect of gender on violent and nonviolent recidivism: A meta-analysis. *Journal of Criminal Justice*, 38(4):675–684, 2010.

J. D. Conklin. *Criminology*. Pearson Publishing, 2013.

R. D. Crutchfield. *Labor Force Participation, Labor Markets, and Crime*. DIANE Publishing, 2011.

Depen. *Plano Nacional de Política Criminal e Penitenciária*. Departamento Penitenciário Nacional, Brasília-DF, 2019., 2019.

B. D. Dooley, A. Seals, and D. Skarbek. The effect of prison gang membership on recidivism. *Journal of Criminal Justice*, 42(3):267–275, 2014.

H. Eisenbarth, M. Osterheider, N. Nedopil, and C. Stadtland. Recidivism in female offenders: Pcl-r lifestyle factor and vrag show predictive validity in a german sample. *Behavioral Sciences & the Law*, 30(5):575–584, 2012.

M. Filippini, W. H. Greene, N. Kumar, and A. L. Martinez-Cruz. A note on the different interpretation of the correlation parameters in the bivariate probit and the recursive bivariate probit. *Economics Letters*, 167:104–107, 2018.

I. S. Gama-Araujo, J. G. Bezerra, L. Kerr, C. Kendall, R. H. M. Macena, R. S. Mota, M. Ferreira, and R. d. J. Pires. Physical violence inside female prisons in brazil: prevalence and related factors. *Ciencia & saude coletiva*, 25:623–632, 2020.

P. Gendreau, T. Little, and C. Goggin. A meta-analysis of the predictors of adult offender recidivism: What works! *Criminology*, 34(4):575–608, 1996.

C. A. Gonçalves Jr and P. F. A. Shikida. Determinantes da reincidência penal no estado do paraná: uma análise empírica da economia do crime/determinants of criminal recidivism in the state of paraná: an empirical analysis of the economics of crime. *Economic Analysis of Law Review*, 4(2):315, 2013.

W. H. Greene. Econometric analysis 7th edition. *International edition, New Jersey: Prentice Hall*, 2012.

L. E. Greiner, M. A. Law, and S. L. Brown. Using dynamic factors to predict recidivism among women: A four-wave prospective study. *Criminal Justice and Behavior*, 42(5):457–480, 2015.

R. K. Hanson and M. T. Bussiere. Predicting relapse: a meta-analysis of sexual offender recidivism studies. *Journal of consulting and clinical psychology*, 66(2):348, 1998.

C. Hedderman and D. Jolliffe. The impact of prison for women on the edge: paying the price for wrong decisions. *Victims & Offenders*, 10(2):152–178, 2015.

F. Heidensohn. The future of feminist criminology. *Crime, Media, Culture*, 8(2):123–134, 2012.

K. Holtfreter, M. D. Reisig, and M. Morash. Poverty, state capital, and recidivism among women offenders. *Criminology & Public Policy*, 3(2):185–208, 2004.

B. M. Huebner, C. DeJong, and J. Cobbina. Women coming home: Long-term patterns of recidivism. *Justice Quarterly*, 27(2):225–254, 2010.

IPEA. Reincidência criminal no brasil: relatório de pesquisa. *Ipea. Rio de Janeiro*, 2015.

M. Jones and B. Sims. Recidivism of offenders released from prison in north carolina: A gender comparison. *The Prison Journal*, 77(3):335–348, 1997.

N. C. Jurik. The economics of female recidivism: A study of tarp women ex-offenders. *Criminology*, 21(4):603–622, 1983.

A. Katsiyannis, D. K. Whitford, D. Zhang, and N. A. Gage. Adult recidivism in united states: A meta-analysis 1994–2015. *Journal of Child and Family Studies*, 27(3):686–696, 2018.

C. Kruttschnitt. The paradox of women's imprisonment. *Daedalus*, 139(3):32–42, 2010.

C. Kruttschnitt and R. Gartner. Women's imprisonment. *Crime and Justice*, 30:1–81, 2003.

J. Luallen, J. Edgerton, and D. Rabideau. A quasi-experimental evaluation of the impact of public assistance on prisoner recidivism. *Journal of Quantitative Criminology*, 34(3):741–773, 2018.

G. S. Maddala. *Limited-dependent and qualitative variables in econometrics*. Number 3. Cambridge university press, 1986.

M. Makarios, B. Steiner, and L. F. Travis III. Examining the predictors of recidivism among men and women released from prison in ohio. *Criminal Justice and Behavior*, 37(12):1377–1391, 2010.

L. A. McCoy and H. A. Miller. Comparing gender across risk and recidivism in nonviolent offenders. *Women & Criminal Justice*, 23(2):143–162, 2013.

D. P. Mears, J. C. Cochran, S. E. Siennick, and W. D. Bales. Prison visitation and recidivism. *Justice Quarterly*, 29(6): 888–918, 2012.

M. M. Mitchell, K. Spooner, D. Jia, and Y. Zhang. The effect of prison visitation on reentry success: A meta-analysis. *Journal of Criminal Justice*, 47:74–83, 2016.

M. Morash, D. A. Kashy, S. W. Smith, and J. E. Cobbina. The connection of probation/parole officer actions to women offenders' recidivism. *Criminal Justice and Behavior*, 43(4):506–524, 2016.

M. Morash, D. A. Kashy, S. W. Smith, and J. E. Cobbina. Technical violations, treatment and punishment responses, and recidivism of women on probation and parole. *Criminal Justice Policy Review*, 30(5):788–810, 2019.

R. B. Nelsen. *An introduction to copulas*. Springer Science & Business Media, 2007.

D. E. Olson, L. J. Stalans, and G. Escobar. Comparing male and female prison releasees across risk factors and postprison recidivism. *Women & Criminal Justice*, 26(2):122–144, 2016.

J. Partanen. On waiting time distributions. *Acta Sociologica*, 12(3):132–143, 1969.

M. D. Reisig, K. Holtfreter, and M. Morash. Assessing recidivism risk across female pathways to crime. *Justice Quarterly*, 23(3):384–405, 2006.

L. J. Rettinger and D. A. Andrews. General risk and need, gender specificity, and the recidivism of female offenders. *Criminal Justice and Behavior*, 37(1):29–46, 2010.

L. F. Sapori, R. F. Santos, and L. W. D. Maas. Fatores sociais determinantes da reincidência criminal no brasil: o caso de minas gerais. *Revista Brasileira de Ciências Sociais*, 32, 2017.

C. K. Scott, C. E. Grella, M. L. Dennis, and R. R. Funk. Predictors of recidivism over 3 years among substance-using women released from jail. *Criminal Justice and Behavior*, 41(11):1257–1289, 2014.

N. H. Sedgley, C. E. Scott, N. A. Williams, and F. W. Derrick. Prison's dilemma: Do education and jobs programmes affect recidivism? *Economica*, 77(307):497–517, 2010.

C. Smart. Criminological theory: its ideology and implications concerning women. *British Journal of Sociology*, pages 89–100, 1977.

C. Smart. The new female criminal: reality or myth? *The British Journal of Criminology*, 19(1):50–59, 1979.

R. J. Smith and R. W. Blundell. An exogeneity test for a simultaneous equation tobit model with an application to labor supply. *Econometrica: journal of the Econometric Society*, pages 679–685, 1986.

L. J. Stalans and A. J. Lurigio. Parenting and intimate relationship effects on women offenders' recidivism and noncompliance with probation. *Women & Criminal Justice*, 25(3):152–168, 2015.

B. Steiner and J. Wooldridge. Sex differences in the predictors of prisoner misconduct. *Criminal Justice and Behavior*, 41(4):433–452, 2014.

B. Tripathi. Feminist criminology: Some reflections. *Vidhigya: The Journal of Legal Awareness*, 9(1):1–6, 2014.

P. K. Trivedi and D. M. Zimmer. *Copula modeling: an introduction for practitioners*. Now Publishers Inc, 2007.

C. Uggen. Work as a turning point in the life course of criminals: A duration model of age, employment, and recidivism. *American sociological review*, pages 529–546, 2000.

R. Walmsley. World female imprisonment list. *London: Kings College International Centre for Prison Studies*, 2014.

G. D. Walters. Predicting recidivism with the criminal sentiments scale: A meta-analysis of a putative measure of criminal thought content. *Criminal Justice and Behavior*, 43(9):1159–1172, 2016.

J. Wilde. Identification of multiple equation probit models with endogenous dummy regressors. *Economics letters*, 69 (3):309–312, 2000.

J. M. Wooldridge. *Econometric analysis of cross section and panel data*. MIT press, 2010.

D. Yukhnenko, S. Sridhar, and S. Fazel. A systematic review of criminal recidivism rates worldwide: 3-year update. *Wellcome Open Research*, 4, 2019.

Covariate	Unmatched	Matched
White = 1	-.001	.005
Spouse = 1	.086***	.050
Child = 1	.375***	-.007
Elementary = 1	-.041	.014
Intermediate = 1	-.066	.004
High = 1	.001	.001
Undergraduate = 1	.078	.002
Employment = 1	-.007	-.005
Father unknown = 1	-.001	.001
Father arrested = 1	-.012	.068
Sibling arrested = 1	.012	.013
Study = 1	.369***	.049
Job = 1	.284***	.006
Psycho = 1	.210***	-.009
Visits = 1	.008**	-.004
Try escape = 1	-.050	-.017
Young = 1	.048	.009
Short penalty = 1	-.079	-.094
Constant	-1,513***	.454***
n	24,998	3,324

* $p < .05$, ** $p < .01$, *** $p < .001$.

Table A1: Estimated results for a probit $w = 1$ (dummy for women), in order to support PSM.

Covariate	Without PSM $\rho = 0$	Without PSM $\rho \neq 0$	With PSM $\rho = 0$	With PSM $\rho \neq 0$
$w = 1$	-.195**	-.202**	-.338***	-.323**
White $= 1$	-.014	-.017	-.301	-.003
White $\times w$	-.192	-.020	-.013	-.016
Employment $= 1$	.230**	.212**	.244***	.231***
Employment $\times w$	-.175	-.120	-.123	-.132
Father unknown $= 1$	-.015	-.021	-.038	-.032
Father unknown $\times w$	-.001	-.002	-.003	-.003
Young $= 1$	.376***	.335***	.331***	.293***
Young $\times w$	-.366**	-.321**	-.308***	-.313***
Constant	-.702**	-.940**	-.558***	-.563***

* $p < .05$, ** $p < .01$, *** $p < .001$.

Table A2: Estimated results for the second equation of the system (3), complementing Table 6.