



Munich Personal RePEc Archive

**Supervised Principal Components and  
Factor Instrumental Variables. An  
Application to Violent Crime Trends in  
the US, 1982-2005.**

Travaglini, Guido

Università di Roma La Sapienza, Italy

April 2010

Online at <https://mpra.ub.uni-muenchen.de/22077/>

MPRA Paper No. 22077, posted 14 Apr 2010 23:18 UTC

# **Supervised Principal Components and Factor Instrumental Variables. An Application to Violent Crime Trends in the US, 1982-2005.**

Guido Travaglini, Università di Roma La Sapienza, Italy.

Provisional version: April 2010.

## **Abstract**

Supervised Principal Component Analysis (SPCA) and Factor Instrumental Variables (FIV) are competing methods addressed at estimating models affected by regressor collinearity and at detecting a reduced-size instrument set from a large database, possibly dominated by nonexogeneity and weakness. While the first method stresses the role of regressors by taking account of their data-induced tie with the endogenous variable, the second places absolute relevance on the data-induced structure of the covariance matrix and selects the true common factors as instruments by means of formal statistical procedures. Theoretical analysis and Montecarlo simulations demonstrate that FIV is more efficient than SPCA and standard Generalized Method of Moments (GMM) even when the instruments are few and possibly weak. The preferred FIV estimation is then applied to a large dataset to test the more recent theories on the determinants of total violent crime and homicide trends in the United States for the period 1982-2005. Demographic variables, and especially abortion, law enforcement and unchecked gun availability are found to be the most significant determinants.

JEL classification: C01, C22, K14.

Keywords: Principal Components, Instrumental Variables, Generalized Method of Moments, Crime, Law and Order.

## 1. Introduction

In Ordinary Least Squares (OLS) estimation of a single equation, collinearity, errors in variables, simultaneity and omitted variables typically produce inconsistent coefficient estimates. While the first two produce parameter attenuation and downward bias their  $t$  statistics, the other two determine the reverse effects. In all cases, distorted significance tests and confidence intervals ensue.

Principal Component Analysis (PCA) combats collinearity by orthogonalizing and reducing through Singular Value Decomposition (SVD) the dimension of a possibly large dataset, and utilizes the ensuing principal components for estimation. Sometime after Jolliffe's suggestion (Jolliffe, 1982) that "people should have buried forever the idea of selection based solely on size of variance", SPCA (Bair et al., 2006) has been devised to account for the correlation existing between principal components and the endogenous variable.

The basic idea of SPCA is to reduce in a first place the number of regressors, chosen among the widest possible available set, including variables that may be either justified or unjustified on theoretical grounds. The regressors that exhibit from data analysis low correlation with the endogenous variable are thrown away, and then standard PCA is applied and finally the appropriate regression is performed.

Specifically, SPCA involves the following steps: *i*) performing univariate OLS of each regressor with respect to the endogenous variable; *ii*) finding a reduced-size regressor set consisting of the variables whose coefficient in (*i*) exceeds a preselect threshold value (Bair et al., 2006); *iii*) computing standard PCA of the set to obtain the size-reduced orthogonalized principal components; *iv*) performing multivariate OLS of the principal components with respect to the endogenous variable.

If endogeneity is detected, however, multivariate OLS no longer applies and the practitioner must adopt standard Instrumental Variable Estimation (IVE) or alternatively Factor Instrumental Variables (FIV) (Bai and Ng, 2006; Kapeitanos and Marcellino, 2007; Bai and Ng, 2008). The available instruments in IVE – quite frequently very large in number – are reduced in FIV to the fewer homogenous variables (the true factors) which contain most of the model's information and selected 'endogenously' via infor-

mation criteria testing that places penalties on large datasets (Bai and Ng, 2002, 2007) or via eigenvalue relative magnitude testing (Onatski, 2009; Ahn and Horenstein, 2009). FIV is claimed to be more efficient than standard GMM and also robust to weak instruments. SPCA, instead, sticks to traditional selection methods of the instruments which are nothing more than the variables dropped in the first three steps.

While SPCA has been demonstrated by its authors to produce consistent estimates of the coefficients as the number of regressors and of observations tend to infinity, FIV analysts prove the contrary, and debate over this issue is heating up especially after the revival or introduction of competing techniques, e.g. ridge regression and Partial Least Squares (Groen and Kapeitanos, 2009). Needless to say, SPCA and FIV significantly differ in their approach and most likely also in their results. Both methods carry their own advantages although at a cost: SPCA takes account of the tie between regressors and the endogenous variable but does not place relevance on the data-induced structure of the covariance matrix and, furthermore, is not immune to the (many) weak instruments curse should its arbitrary selection criterion (e.g. size of the eigenvalues or dominant shares) be too loose. FIV does exactly the opposite while risking, however, to record too few common factors relative to the number of regressors and produce an identification problem. This may be practically superseded, however, by augmenting the instrument set with lagged common factors (Bai and Ng, 2006) and/or a small set of key observed instruments (Kapeitanos and Marcellino, 2007).

Sect. 2 introduces the econometric framework and summarizes the major features of SPCA, FIV and GMM. Sect. 3, after showing and describing the figure plots of the major crimes, law-and-order, and preselect demographic variables, provides a brief analysis of the recent theories on the determinants of crime. Sect. 4 briefly introduces the dataset utilized and the time-aggregation transformation technique applied to the available annual data. Sect.5 produces the empirical results and Sect.6 concludes.

## **2. The Econometric Framework**

This section sequentially examines the properties of SPCA, FIV and GMM computation methods. By making use of Montecarlo simulation with the same data, the three methods are eventually compared to test for efficiency and consistency.

## 2.1. Supervised Principal Component Analysis (SPCA)

For  $n$  observations and  $p$  variables, let the time-series matrix of the centered and scaled regressors be

$$X_p : (n \times p) = (x_1, \dots, x_p) \quad (1)$$

where

$$x_i : (n \times 1) = (x_i^D - \text{mean}(x_i^D)) / \text{std}(x_i^D); i \in p \quad (2)$$

where the suffix  $D$  indicates the original time series. Eq. (2) has the following properties:  $x_i \sim N.I.D.(0,1)$ ,  $X_p'X_p$  is p.s.d., and  $\text{trace}(X_p'X_p) = p^1$ . Let also the endogenous-variable time series be

$$y : (n \times 1) \quad (3)$$

with the same distributional property as above.

The  $p$ .th OLS equation invoked by step ( $i$ ) of SPCA is

$$y = x_i \alpha_i + 1 \gamma + v_i; i \in p \quad (4)$$

where  $\alpha_i$  and  $\gamma$  are scalars,  $1$  is a  $n$ -vector of ones, and  $v_i : (n \times 1) \sim I.I.D.(0, \sigma_i)$ . For practical purposes, however, step ( $i$ ) can also be performed by means of multiple regression, without loss of generality (Koch and Naito, 2008). From all of the  $p$  equations of eq.(4) the regressor set (eq.1) may be split as follows

$$X_p = (X_f : X_h); f + h = p \quad (5)$$

by virtue of a given  $t_c$ , a  $t$  statistic critical value chosen to be 1.96 for the two-tailed 5% level with  $n < \infty$ . Thus

---

<sup>1</sup> Centering and scaling by the standard error, a typical data transformation of PCA, automatically produces normally distributed variables with  $X_p'X_p$  a correlation matrix.

$$\begin{cases} X_f = X_p & | & t(\alpha_i) \leq t_c \\ X_h = X_p & | & t(\alpha_i) > t_c \end{cases} \quad (6)$$

where  $t(\alpha_i)$  is the absolute  $t$  statistic associated to the  $i$ .th coefficient of eq. (4). SPCA in step (ii) suggests retaining  $X_h$  and dropping  $X_f$  in eq.(6), although the criterion used by its authors (Bair et al., 2006) is somewhat different<sup>2</sup>. Due to the likelihood of collinearity within  $X_h$ , the SVD of  $X_h$  is written as

$$US_hV = X_h \quad (7)$$

whose principal components are

$$P_r : (n \times r) = X_r V_r; \quad r < h \quad (8)$$

where  $r$  is the number of the largest ordered eigenvalues or shares usually taken, respectively, to lay above unity and 5%. The matrix  $P_r$  represents the set of mutually orthogonal principal components, such that

$$P_r' P_r = \text{diag}(S_r) < \text{diag}(S_h) \quad (9)$$

Eq. (8) can be utilized to perform OLS whose regressors are completely uncorrelated with one another, thereby disposing of the issue of collinearity. Then we have the following structural equation

$$y = P_r \beta_r + 1 \gamma_r + v \quad (10)$$

where  $\beta_r : (r \times 1)$ ,  $\gamma_r$  is a scalar, and  $v : (n \times 1) \sim I.I.D.(0, \sigma_v)$ .

While the omitted-variable bias disappears together with collinearity, endogeneity given by  $P_r' v \neq 0$  may still be a problem. In fact while the principal components are orthogonal, measurement with error and simultaneity may persist. IVE must therefore be imposed upon eq. (10) to obtain consistent estimates. Of the different IVEs, GMM is

---

<sup>2</sup> The cited authors adopt a proportional hazards model. Ng and Bai (2008) use the critical  $t$  statistic as above, posited to be 2.50.

thus far the most renown since particularly useful in the presence of autocorrelation and heteroskedasticity (Newey and West, 1987), although it exhibits small-sample shortcomings (Newey and Windmeijer, 2008).

## 2.2. Factor Instrumental Variables (FIV)

FIV has been recently popularized by several authors (Bai and Ng, 2006, 2008; Kapeitanos and Marcellino, 2007), motivated by the shortcoming of traditional PCA whose selection criterion is typically inconsistent for a large number of regressors and observations (Bai and Ng, 2002). In its simplest form, FIV builds upon the following static factor model written in common-component form (Chamberlain and Rothschild, 1983)

$$X = F\Lambda' + e \quad (11)$$

where, given knowledge of  $p$  and of the true number of factors  $r \ll p$  (Bai and Ng, 2002),  $X$  is a given data matrix of random or actual-data, its size being  $n \times p$ , and  $F : (n \times r)$ ,  $\Lambda : (p \times r)$ ,  $e : (n \times p)$ , such that for  $j = 1, \dots, p$ :

$$X = \begin{bmatrix} x_{1,1} & \cdots & x_{1,p} \\ \vdots & & \vdots \\ x_{n,1} & \cdots & x_{n,p} \end{bmatrix}; e = \begin{bmatrix} e_{1,1} & \cdots & e_{1,p} \\ \vdots & & \vdots \\ e_{n,1} & \cdots & e_{n,p} \end{bmatrix}; F = \begin{bmatrix} f_{1,1} & \cdots & f_{1,r} \\ \vdots & & \vdots \\ f_{n,1} & \cdots & f_{n,r} \end{bmatrix}; \Lambda = \begin{bmatrix} \lambda_{1,1} & \cdots & \lambda_{1,r} \\ \vdots & & \vdots \\ \lambda_{p,1} & \cdots & \lambda_{p,r} \end{bmatrix}$$

where  $X$  is observed and may be a stationary or non-stationary panel dataset, while all the other elements are unobserved.  $F$  is the matrix of common factors,  $\Lambda$  the matrix of factor loadings and  $e$  the matrix of the idiosyncratic component.

We suppose that  $E(F) = 0$ ,  $E(FF') = I_r$ ,  $E(e) = 0$ , and  $E(F'e) = 0$ , where  $I_r$  is the  $(r \times r)$  identity matrix, which applies only in the case of orthogonal rotation, and  $e$  might be auto- or cross-correlated (Bai and Ng, 2002). Assumptions A and B of Bai and Ng (2002) are assumed to hold throughout.

For ease of reading, the model above described may be specifically expressed for  $i = 1, \dots, n$  and  $\ell = 1, \dots, r$ , and for each  $j$  as follows

$$x_{i,j} = f_i \cdot \lambda_\ell + e_{i,j} \quad (12)$$

where  $x_{i,j} : (n \times 1)$ ,  $f_i : (n \times r)$ ,  $\lambda_\ell : (r \times 1)$  and  $e_{i,j} : (n \times 1)$ . Obviously, as pointed out also by some authors (Bai and Ng, 2002),  $\lambda_\ell$  corresponds to the OLS estimate of the coefficients of the regression of  $x_{i,j}$  over the observed  $f_i$ .

Depending on whether  $n < p$  or  $n > p$ , two sample covariance specifications of eq. (11) may be derived:

$$XX' : (n \times n) = F\Lambda' \Lambda F' + \Phi; \Phi = ee' \quad (13)$$

$$X'X : (p \times p) = \Lambda F' F \Lambda' + \Psi; \Psi = e'e. \quad (14)$$

Both (13) and (14) can undergo Singular Value Decomposition (SVD) whereby the unobservables may be approximated in the following manner. For the first case, let  $SVD(XX'/np) = F_r S_r V_r'$ , where  $F_r$  and  $V_r$  are unitary matrices and  $S_r$  is a diagonal matrix of singular values. All are of size  $n \times n$ . Then  $\hat{F}_r = \sqrt{n} F_r$  and subsequently  $\Lambda_r = \hat{F}_r' X / n$ , such that, finally, the common components and the estimated errors respectively are  $\hat{C}_r : (n \times r) = \hat{F}_r \Lambda_r$  and  $\varepsilon : (n \times p) = X - \hat{C}_r$ . Finally,  $S_r = \Lambda \Lambda' / p$  and  $FF' / n = I_r$ . For the second case we have:  $SVD(X'X/np) = F_r S_r V_r'$ , whereby  $\hat{F}_r = \sqrt{p} F_r$  and  $\Lambda_r = \hat{F}_r' X' / p$  such that  $\hat{C}_r : (r \times n) = \hat{F}_r \Lambda_r$  and  $\varepsilon : (n \times p) = X - \hat{C}_r'$ .

Finally, similar to above,  $S_r = \Lambda \Lambda' / n$  and  $F'F / p = I_r$  and, since  $\Lambda_r \sim N(0, S_r)$  and  $\hat{F}_r \sim N(0, I_r)$ , the variance of the common components is always smaller than that of the other variables, independent of their statistical distribution (e.g. Normal, I.I.D.). Therefore,  $Var(\hat{C}_r) < Var(X)$  and  $Var(\hat{C}_r) < Var(Z_m)$ .

The rationale for using factor analysis – as advanced in the Introduction – is that large datasets ( $p \rightarrow \infty$ ), can be size-reduced to a few common components which explain most of the behavior contained in  $X$ . The true number of components  $r$  is unknown ex ante and can be determined by formal statistical procedures (Bai and Ng,



2002, 2007; Onatski, 2009; Ahn and Horenstein, 2009) some of which are reported in the Technical Appendix. The factors so obtained are supplemented – if necessary – by additional instruments to form the entire instrument dataset, and utilized in estimating a classical regression equation with preselected exogenous variables.

### 2.3. Generalized Method of Moments (GMM)

Let the following equation system

$$y = X\beta + \varepsilon \quad (15)$$

$$X = Z_m\Pi_1 + \hat{C}_r\Pi_2 + y\Pi_3 + V \quad (16)$$

where eq. (15) is a structural-form regression equation with the matrix of the regressor variables  $X:(n \times p)$  and the disturbance vector  $\varepsilon:(n \times 1) \sim I.I.D.(0, \sigma_\varepsilon)$ , and eq. (16) is a generalized reduced form (RF) that endogenizes  $X$  under the assumption that  $X'\varepsilon \neq 0$ .  $\Pi_1$  and  $\Pi_2$  are the RF parameter matrices, respectively, of the instruments included in  $Z_m:(n \times m)$ , ( $m \geq p$ ) and of the  $r$  common components  $\hat{C}_r$ , while  $V$  is a  $n \times p$  disturbance matrix, posited to be  $E(V'\varepsilon) = 0$ . Specifically, for  $XX'$  ( $n < p$ ),  $\hat{C}_r = \hat{F}_r\Lambda_r$  while for  $X'X$  ( $n > p$ ),  $\hat{C}_r = (\hat{F}_r\Lambda_r)'$ . Finally,  $\Pi_3:(m \times p)$ ,  $Cov(\hat{C}_r, X) < Cov(Z_m, X)$ , and  $\Pi_3:(1 \times p)$ . The following matrices are set to be:  $\Pi_2 = 0$  in SPCA and  $\Pi_3 = 0$  in FIV.

It is indifferent for the working of GMM whether  $X$  is preselected by SPCA or by a theoretical rationale that accommodates FIV. What is relevant is that in SPCA  $E(\Pi_1, \Pi_3) \neq 0$ , while in FIV, if the data have a factor model structure,  $E(\Pi_2) \neq 0$  and if in addition a mixture of both  $Z_m$  and  $\hat{C}_r$  as instruments is used as in the majority of empirical applications (Bai and Ng, 2007, 2008; Kapeitanos and Marcellino, 2007),  $E(\Pi_1) \neq 0$ .

Standard GMM utilizes instruments and moment matrices (Hansen, 1982). Again, it is indifferent whether the instrument matrix is made up of the least relevant regressors dropped out of eqs. (5) and (6) in the SPCA procedure or of other variables, e.g. the mixture of common factors and lagged regressors as required by FIV.

Let for simplicity the set of instruments be only  $Z_m$  for the moment, such that  $\Pi_2, \Pi_3 = 0$ , and  $E(Z_m' V) = 0$  for exogeneity and  $E(Z_m' X) \neq 0$  for relevance, and let the corresponding set of moments be

$$u_m : (n \times m) \equiv Z_m'(y - X\hat{\beta}); E[u_m(\hat{\beta})] = 0 \quad (17)$$

where  $\hat{\beta}$  is the first-stage estimator, and the covariance matrix of eq. (16)

$$W_m : (m \times m) = u_m' u_m \quad (18)$$

the GMM estimator, for the weight matrix  $\Omega^{-1} : (m \times m) = (W/n)^{-1}$ , is the following:

$$\hat{\beta}_{GMM} = (G' \Omega^{-1} G)^{-1} G' \Omega^{-1} Z' y \quad (19)$$

where  $G : (m \times p) = \partial u_m / \partial \hat{\beta}$  is the Jacobian matrix and  $Z' y : (m \times 1)$ . The asymptotic distribution of the GMM estimator is

$$n^{1/2}(\hat{\beta}_{GMM} - \beta_0) \sim N\left[0, (G' W^{-1} G)^{-1}\right] \quad (20)$$

where  $\beta_0$  is the population parameter and the related asymptotic variance is defined as

$$A \text{ var } GMM = (G' \Omega^{-1} G)^{-1}. \quad (21)$$

The GMM estimator of eq. (14) is asymptotically highly biased of order  $O_p(m)$  if there is endogeneity, i.e. correlation among the RHS variables of eq. (11) (Newey and Windmeijer, 2008). It is also asymptotically biased, although to a lesser degree, if the moments of eq. (18) are far from normally distributed (Altonji and Segal, 1996) and are not variance minimizers because the first-stage estimator  $\beta_r$  is inconsistent (Newey and Smith, 2004).

In addition, the size of  $Z_m$  may be quite large and well beyond the bounds suggested by the literature (Andrews and Stock, 2007) and in practice it may be very difficult to achieve instrument validity even if  $m$  and  $n$  are very large (Andrews and Stock, 2007), since instruments that are exogenous may be weak and viceversa.

In the case of SPCA, given  $\widehat{G} = -(Z_m' Z_m \Pi_1 + Z_m' y \Pi_3)/n$  the asymptotic variance of the estimator is

$$A \text{ var } SPCA = (\widehat{G}' \widehat{\Omega}^{-1} \widehat{G})^{-1} \quad (22)$$

while in the case of FIV for both  $\Pi_1 \neq 0$  and  $\Pi_2 \neq 0$ ,  $\check{G} = -(Z_m' Z_m \Pi_1 + Z_m' \hat{C}_r \Pi_2)/n$ , and the asymptotic variance of the estimator is

$$A \text{ var } FIV = (\check{G}' \check{\Omega}^{-1} \check{G})^{-1}. \quad (23)$$

After denoting  $Est$  and  $G^*$  respectively any of the three estimators and corresponding Jacobians considered, we know that in general  $\lim_{\Omega^{-1} \rightarrow \infty} A \text{ var } Est = 0$  and that  $\lim_{\check{G} \rightarrow \infty} A \text{ var } Est = 0$ . Then, independent of the values taken by the parameter matrices in eq. (16), if the following conditions hold:  $\check{G} > \widehat{G} > G$  and  $\check{\Omega}^{-1} > \Omega^{-1} > \widehat{\Omega}^{-1}$ , we automatically expect that

$$A \text{ var } SPCA > A \text{ var } GMM > A \text{ var } FIV \quad (24)$$

which implies, for  $n, p \rightarrow \infty$ , a greater consistency of the SPCA estimator, as claimed by its authors (Bair et al., 2006) and a higher degree of efficiency obtained with the FIV estimator (Bai and Ng, 2006; Kapeitanos and Marcellino, 2007).

This conclusion and its related causes, i.e. the magnitude of the components of the parameter asymptotic variance – the weight matrix and the Jacobians – are put to empirical testing via Montecarlo experimentation to assess the truth of eq. (24). Specifically, the goal is to check whether eq. (24) holds for any value of  $n$  and  $p$  and also for  $n, p \rightarrow \infty$ , namely, for limited and asymptotic sample length and size. In Table 1, the variables of eqs. (15) and (16) are all simulated random N.I.D. values and the number of the common components is fixed to 3, while that of the endogenous regressors and of the instruments is set to be 3,7 and 5,25 (panes A and B, respectively), two reasonable value couples for common practitioners.

The magnitude of the asymptotic variance and of its components is assessed by means of their smallest singular value (SSV). In the table, for select different sample

length ( $n$ ) and panel size ( $p$ ) of N.I.D.-distributed artificial series that replicate the data matrix  $X$  of eq. (11), the sample means and standard deviations of 1,000 replications of the SSVs of the weight matrix ( $\Omega^{-1}$ ,  $\widehat{\Omega}^{-1}$  and  $\check{\Omega}^{-1}$ ), of the Jacobians ( $G$ ,  $\widehat{G}$  and  $\check{G}$ ;  $G'G$ ,  $\widehat{G}'\widehat{G}$  and  $\check{G}'\check{G}$ ), and of the asymptotic parameter variances are reported. To simplify the reading of the table, the batches are respectively denoted as OMINV, JAC, JAC2 and AVAR.

A quick glance at Table 1 shows that the asymptotic parameter variance AVAR (columns 4 and 8, panes A and B) is always the largest in the case of SPCA and the smallest in the case of FIV independent on whether  $n < p$  or  $n > p$ , but obviously dependent, as the other variances, on the number of instruments. Its magnitude stands between 1/3 and 1/6 of the AVAR of standard GMM for both few and many instruments, and is many more times smaller in all cases than that of SPCA.

The weight matrix OMINV (columns 1 and 5 of both panes) is the smallest in the case of SPCA, while JAC and JAC2 are the largest in FIV for few instruments and the smallest with many instruments. This implies that in the former case the correlation between the instruments and the common components is high relative to the other estimators, while the situation is reverted with many instruments.

To conclude, the AVAR of FIV is always the smallest of the given estimators, irrespective of the length and size of the random sample, but its magnitude does not necessarily depend on the Jacobians especially in the case of many instruments. Therefore, it must be made to depend on the coefficient matrix  $\Pi_2$  of eq. (16) which ties the matrix  $X$  with its common components. For this reason, and especially with small samples, FIV is largely preferable to the other two techniques, and this approach will be pursued in the empirical Sections that follow.

### **3. Violent crime statistics in the US and modern crime theories**

This section describes the main statistical features of criminal trends and of several related variables in the US during the last quarter century. It introduces also the modern theories on the determinants of aggregate criminal behavior and of the ensuing debates.

### 3.1. Trends in violent crime, law and order, and demographics

Trends occurred in the last quarter century in violent crime, in law-and-order, and in the major preselect demographic variables are shown in the Figure plots 1-3, 4-7 and 8, respectively.

All the data exhibited are on an annual basis since at least 1976, and are issued by the US Bureau of Justice (BOJ) based on the National Crime Victimization Survey in conjunction with the Uniform Crime Report (UCR). Violent crime, according to FBI, includes in descending order of severity murder and nonnegligent manslaughter (merged as homicide), forcible rape, robbery and aggravated assault BOJ adds to the latter also simple assault. All of these crimes involve, according to the UCR, force or threat of force.

The trends during the period 1976-2005 of the two major crime classifications – property and violent crime<sup>3</sup> – are exhibited in the upper panels of Figure 1. They both show a declining pattern, the former steeper than the latter. In fact the drop in property crime is nearly fourfold, while that in violent crime is roughly twofold. The ratio of violent with respect to property crime passes from nearly 9% to over 12%.

The trend of the rate (per 100,000 inhabitants) of total violent crimes committed with firearms is shown in the left bottom panel of Figure 1 and shows a growth from 1976 to 1993 and then a substantial drop until the year 2000, later to stabilize. A similar pattern is followed by the use of firearms in total violent crime, which is exhibited in the right bottom panel of Figure 1.

The trend of total homicide rate is shown in Figure 2, upper left panel. It nearly doubled from the mid 1960's to the late 1970's, then peaked in 1980 at 10.2 and subsequently fell off to 7.9 in 1984. It rose again in the late 1980's and early 1990's to another peak in 1991 of 9.8. From 1992 to 2000, the rate declined sharply. Since then, the rate

---

<sup>3</sup> The victims of the 9/11/01 terrorist attacks are not included in this analysis.

has been stable. Its share over total violent crime, however, has diminished substantially, passing from 1.8% in 1976 to 1.2% in 2005.

The ethnic characteristics of homicide victims and offenders differ, as blacks are disproportionately represented as both homicide victims and offenders. From the BOJ reports, the victimization and the offending rates for blacks respectively were 6 and 7 times higher than those for whites. The other panels of Fig.2 put to evidence these patterns which show that, in practice, the black-total homicide ratio hovers at around 80%.

Fig. 3 shows the trends of the other four violent crime rates: rape, robbery and the two categories of assault. They all follow a similar pattern, consistent with total violent crime of which, respectively, they represent a fairly constant share close to 6%, a falling share (from 42% to 31%) and a surging share (from 50% to over 60%). In practice, assault has slowly but steadily replaced robbery. Whether this is evidence of improved private deterrence against the supply of loot crime it is an open question that will be pursued in Sect. 5.

Figs. 4 to 7 show law and order statistics. Total arrest rates and arrest rates per age group show in Fig. 4 a pattern that is on average consistent with violent crime committed with firearms (Fig.1). They follow in fact the classic inverted U-shape trend, peaking in the early-mid 1990's. However, the arrest ratio of young teens (age 14 and below) presents a rather worrisome bulge in the years 2000, not shared by the other age groups. Correctional statistics (available from 1980), persons under capital punishment (available from 1976) and crime-related public expenses (available from 1982) are shown in the last three Figures provided.

Fig. 5 shows the number of persons in custody of state correctional authorities. The numbers are total and by most serious offense, and are available from 1980 onwards. The total number of persons in correctional custody rises by nearly fourfold during the last quarter century of reported data. Of these, the slowest growing is the number of persons that have committed property crimes, while the fastest growing is the number of drug-related inmates which grows by a whopping amount of 12 times. In fact its ratio over the total number of persons in correctional custody rises from 1% to 4%, while the others remain fairly constant.

While Fig. 6 is quite self explanatory at showing the rise in absolute terms of persons under capital punishment and executed, Fig. 7 shows the financial effort of state, federal and local authorities toward the police, the judicial and the correctional system. The figures are available only since 1982 and are expressed in current Dollar terms (mn.). Expressed in percapita real Dollar terms (by using the GDP deflator, 2000=100) the public expenses that have received the largest share have been those earmarked to the police, although declining from 53% to 46%. The share of the correctional system, instead, has risen from 25% to 32%, consistent with the rise of prison inmates shown in Fig.5. Finally the share received by the judicial system has been mostly constant at 22%.

Finally, Fig. 8 shows the trends of the major preselect demographic variables: abortion and birth rates (per 1,000) women and by two age groups (15-19, 20-24). All four peak almost simultaneously in 1989-91 after a swift rise, especially in abortions. This rise is most likely connected with the effects of the *Roe vs. Wade* ruling that legalized abortion in January 1973, and rapidly adopted as state law all over the nation. Observationally speaking, one generation after the ruling the abortion rates have started falling, and together with them also the birth rates of the youngest group of women. The birth rates of the second group, instead, show a rather independent pattern. While repeating needless pursuing the matter further at this juncture, a clearer picture of the phenomenon may be conveyed by analyzing the contents of the modern crime theories.

### **3.2. Modern crime theories**

The statistical evidence supplied above has prompted macro-criminologists to delve into the determinants of crime. Three major hypotheses have been lately laid down in the academic literature: the “Roe-Wade effect”, the “More-guns-less-crime” hypothesis, and the “Death-penalty effect”. All of them are tested by means of large panel datasets, provided by the FBI and other official sources in terms of tabulations of states, metropolitan statistical areas, cities with over 10,000 inhabitants, suburban and rural counties, and colleges and universities. The amount of data ranges from over 500 to over 5,000.

The first testable hypothesis the “Roe-Wade effect”, known also as the abortion-crime link (Donohue and Levitt, 2001, 2004, 2006). It is therein posited and proved that widespread legalized abortion introduced in 1973 has a significant causal relationship

with the observed crime trends. In fact their reduction since 1987 onwards (see Figs. 1-3) is supposedly due to the reduced number of potential criminals by age cohorts, especially within the context of low-income classes and/or racial minorities (Levitt, 2004).

The second hypothesis posits that behind crime trends some major deterrent causes may loom behind, like state laws on Concealed-Carry Weapons (Lott and Mustard, 1997) or improved policing and sentencing. The “More guns less crime” hypothesis, however, is jeopardized by Ayres and Donohue (2003) who use more complete county data by adding five years of county data and seven years of state data, and reinforce the abortion-crime link, while at the same time demonstrate the adoption of shall-issue laws in general increases crime. Lott and Whitley (2004) prove instead that abortion explains very little in the U.S. crime rate reduction. Other hypotheses on this line are advanced, such as the crack cocaine epidemic (Joyce, 2004) and the death penalty deterrence effect on homicide.

The death penalty hypothesis rests on Ehrlich’s original finding (1975) which predicts that an increase in perceived probabilities of apprehension, conviction given apprehension, or execution given conviction will reduce an individual’s incentive to commit murder. Hence, death penalty is shown to produce a significant deterrent effect upon crime. This hypothesis has been recently corroborated by the finding that, on average, one execution deters 18 murders (Dezhbakhsh et al., 2003). The authors utilize a county-level panel dataset that covers the post-moratorium period: 1977-present (moratorium was established in 1972 by the Supreme Court in *Furman vs. Georgia*). This is the most disaggregate and detailed data used in this literature, that is addressed at overcoming aggregation bias emerging from using state or national data, but is highly criticized by Donohue and Wolfers (2005).

#### **4. The Available Dataset and Time Disaggregation Techniques.**

This section exhibits the crime and crime-related variables that can be used for estimation in testing for the determinants of aggregate criminal behavior in the US. All figures, except three, are supplied by official sources in annual terms, while quarterlization is needed for estimation purposes. The most popular benchmark-based temporal disaggregation techniques are presented and, among these, the Santos Silva-Cardoso is selected.



#### **4.1. The Available Dataset.**

The dataset available for the purpose of testing the determinants of violent crime in the US during the years 1982-2006 is exhibited in the Data and Source Appendix. The time series of total violent crime and its FBI categorized components, homicide, rape, robbery and assault, represent the endogenous variable singly tested (eq. 10) for its determinants. The regressor/instrument list (eq. 1) comprises 70 time series, subdivided into ten different categories, ranging from crime to educational variables, and from demographic to law-and-order variables, and more.

All of the figures are provided on an annual basis, some commencing in 1976, some in 1980, and others in 1982 (R2 and R8 of the Appendix). Exclusion is made for three time series which are provided with quarterly frequency: Personal Disposable Real Income, total unemployment rate and the GDP deflator.

While the entirety of the cited crime analysts uses panel data to broaden the field of observations by simultaneously avoiding the limits imposed by aggregation bias, scanty degrees of freedom, and unobserved heterogeneity across individuals, this paper uses the classic aggregate time-series analysis approach, where the (very) aggregate annual dataset may be extended in time length through time disaggregation procedures. The purpose is to provide econometric results that may be comparable to those obtained by the authors cited in Sect. 3.2 although, obviously, not free from the limits described above.

#### **4.2. Time Disaggregation Methodology**

Several procedures can perform temporal disaggregation, i.e. low-frequency (LF) to high-frequency (HF) transformation. The most utilized benchmark-based procedures are the following: Denton (1971), Chow-Lin (1971), Fernández (1981), and Santos Silva-Cardoso (2001). In spite of the aggregation bias involved and of their synthetic description of the actual – but unobserved – data, they are widely used with different variants in many compilations of quarterly national accounts (Chen, 2007). The available benchmarks utilized in this paper are the three time series provided with quarterly frequency and described above. Let alone these, the entirety of the dataset contained in the Appendix is subject to the select temporal disaggregation techniques.

These techniques share the use of one or more HF benchmarks to produce estimators by Generalized Least Squares (GLS) over each of the LF variables. The ensuing LF errors are then transformed into HF errors under different hypotheses about their structural behavior (e.g. AR(1) in the Chow-Lin and Santos Silva-Cardoso procedures, and I(1) in Fernández). All variables included in the dataset are expressed in logarithms, as this transformation appears more suitable to the assumptions underlying the disaggregated model (additivity of effects, normality and homoscedasticity of errors). Moreover, they are all flows, being expressed as rates, ratios or in percapita terms (see Data and Source Appendix).

Details of the interpolating methods are supplied in Di Fonzo (2003), Proietti (2004), Chen (2007) and Quilis (2005, 2006). Basically, each LF series is transformed into the corresponding HF series by respecting the annual constraint given by its observed value, and by exploiting the coefficient estimate that ties it to the benchmark variable(s) used. The resulting LF disturbance is subsequently HF-transformed by appropriate filtering and subject to the structural behavior found or assumed by the different procedures.

The formal steps required to implement the time disaggregation procedures whose target, in the present context, is to quarterlize annual figures, are exhibited by means of the following formulas. Briefly, the steps involved commence from letting

$$\tilde{y}_a = \tilde{W}_a \delta + u_a; \quad a = 1, \dots, M \quad (25)$$

where  $\tilde{y}_a$  is each of the LF series annually observed for  $a \subseteq M$ ,  $\tilde{W}_a : (M \times K)$  is the matrix of the  $K \geq 1$  annualized benchmark series used<sup>4</sup>,  $\delta$  the estimated coefficient(s), and  $u_a \sim N.I.D.(0, \sigma_u^2)$  the LF residuals. The constant term is included in the estimation. Then let

$$C\tilde{y}_a = C\tilde{W}_a \delta + Cu_a \quad (26)$$

---

<sup>4</sup> By common practice, the HF benchmarks are LF-transformed by means of their sum or average over each subperiod (Di Fonzo, 2003; Chen, 2007).

w here  $C$  is the  $(n \times M)$  constraint matrix that performs the LF-HF transformation of  $\tilde{y}_a$ , namely,  $C = I_M \otimes c_s$  where  $I_M : (M \times M)$  is an identity matrix and  $c_s$  is a row vector of size  $s$  which is the number of HF data points contained in each LF data point (four in the case of annual to quarterly transformation), and  $c_s$  is a vector of ones in the case of the temporal aggregation of a flow, and of  $1/s$  in the case of averaging (Quilis, 2005).

Finally, the HF series are constructed as follows

$$\tilde{y}_i = \tilde{W}_i \hat{\delta} + v_i; \quad i = 1, \dots, n \quad (27)$$

where

$$\hat{\delta} = \left( \tilde{W}_a' C' V^{-1} C \tilde{W}_a \right)^{-1} \left( \tilde{W}_a' C' V^{-1} \tilde{y}_a \right) \quad (28)$$

is the GLS coefficient estimate from eq. (25),  $\tilde{W}_i : (n \times K)$  the matrix of the observed benchmark series, and  $V : (M \times M) = u_a u_a'$ . In eq. (27),  $v_i : (n \times 1)$  are the HF residuals obtained by transforming the LF residuals of eq. (25) into HF, namely  $L u_a = v_i$ , where  $L : (n \times M) = V_{HF} C' V^{-1}$ , and  $V_{HF} : (n \times n) = v_i v_i'$  is the HF matrix of residuals selected or found by Maximum Likelihood Estimation (MLE) to follow an AR(1) or I(1) structure.

In this paper, the Quilis (2006) Matlab codes are applied with one benchmark series (GDP) to the Denton procedure and with two benchmark series (GDP and the unemployment rate) to the other procedures: Chow-Lin, Fernández and Santos Silva-Cardoso. In most cases, since the available LF series are mid-year observations,  $c_s$  is a vector of  $1/s$ . Elsewise,  $c_s$  is a vector of ones. All procedures, except the last, follow the static modelization of eqs. (25)-(27). The last (Santos Silva-Cardoso, 2001) treats  $\tilde{y}_a$  as an innovation by replacing it with  $\phi \tilde{y}_{a-1}$  in eq. (25), where the parameter  $\phi$  is chosen by grid search.

The four time disaggregated datasets are compared in a run-up for selecting the originating procedure that ‘best’ interpolates the figures, i.e. the one whose HF series is closest in terms of mean squared distance (MSD) from the LF available data, where

$$\text{MSD} = \frac{1}{n} \sum_{i=1}^n (C\tilde{y}_a - \tilde{y}_i)^2 \quad (29) \quad (36)$$

Among the competing procedures, the Santos Silva-Cardoso is chosen as it scores the smallest MSD in over 70% of the cases. For this reason it can be defined the best interpolator.

## 5. Empirical Results

Once the Santos Silva-Cardoso procedure has been selected to time disaggregate the logged dataset introduced in Sect. 4.1, the transformed figures are all subject to centering and scaling which is required for homogenization, being originally expressed with different measures (e.g. Dollars, ratios, inhabitants). The purpose is to obtain a dataset with all time series  $N.I.D.(0,1)$ .

Because some annual series begin in 1982 (see Sect. 4.1), the time-span covered by the quarterly disaggregated figures is forced to be 1982:Q1-2005:Q4, a total of 96 observations for each series of the dataset. Table 2 displays the empirical results of the two major violent-crime equations: total violent crime and homicide rates. The dependent variable of each equation corresponds to eq. (10) while the regressor/instrument set is represented by eq. (11). All variables are described in the Appendix. Before estimation they are all made  $I(0)$  to prevent spurious effects on coefficients and  $t$  statistics (Granger and Newbold, 1974).

Each crime equation is estimated by exploiting the main (if not sole) advantage of SPCA by establishing in a first instance the regressor list by means of the first three steps of this procedure, which selects the appropriate variables from the data contained in the Appendix, according to the methodology described in Sect. 2.1. By consequence, each crime equation is characterized by a different set of determinants: the Principal Components derived from SPCA which are mutually orthogonal and automatically ensure zero collinearity.

Even after extracting the regressors (18 and 11 for each equation, respectively), the dimension of the potential instrument set remains very large, and may include quite a few weakness cases. For what expressed in Sect. 2.2, FIV is the best candidate to achieve appropriate instrumenting after combining actual data variables with common components. However, the reader is warned that, by a quick glance at Table 2, some  $t$  statistics are bloated, as a consequence of the inability of FIV to entirely remove endogeneity. Notoriously, the method of Empirical Likelihood (EL) (Owen, 1988, 2001) outsmarts all other estimators in achieving optimal instrument detection (Newey and Smith, 2004) because it is based on empirical probabilities regarding the moments. However, the issue cannot be pursued here as it requires excess explanations in the present context.

The remaining variables are candidates for instrumenting: 96 for the equation which regards total violent crime rate and 48 for the equation of the homicide rate. In the first equation they are further reduced to 13 by detection of strong instruments (those with an absolute correlation magnitude greater than .6) and by inclusion of one true single common component, as detected by the majority of the formal tests (where  $k$  max =8) described in the Technical appendix. By means of the same procedures, candidate instruments are respectively reduced to 7 and 6 thanks to strong instrumenting and the above-mentioned formal testing.

Total violent crime is known to be a combination of its categorized components and is by consequence expected to merge and average out their respective determinants. A few regressors stand out as significant determinants: total legalized abortion (AGIP and AGIW, Alan Guttmacher source), and public expenditures on police (EXPOLICE), both carrying the correct expected negative slope. Total juvenile legal and illegal abortion (ABRATE\_2024) and birth rates (BIRATE\_1519), and also the juvenile and more mature populations (POPUNDER14 and POP\_25ANDOVER) carry all the correct expected positive slope. Only two significant determinants are found to exhibit the wrong sign: incarcerations and alcohol consumption (TOTINC and BEER). In practice, the Levitt hypothesis, which is founded on demographics, finds substantial support to the significant and preponderant accompaniment – however – of crime prevention.

Homicide rates significantly depend with the right sign on the following determinants: the share of violent crimes and, more specifically, the shares of murders and rob-

beries committed with firearms (FIRE\_TCRIMERATIO, FIRE\_MURDRATIO, FIRE\_ROBBRATIO) . Felonies and the murder rate committed with firearms (FELONY\_GUN\_RATE and FIRE\_MURDRATE) have instead a negative impact, while other variables like prevention and repression, including death penalty, do not appear as determinants as commonly expected by large sections of public sentiment and of the academia. The significant effect of the ratios is not simply a definitional matter of crime statistics since it implies the weight that firearms play in committing homicides. In other words, the use and peruse of firearms in committing violent crimes heavily affects the variable under scrutiny and lends significant support to the theories contrary to free firearms circulation (see the Levitt-Lott debate, Section 3.2).

## **6. Conclusion**

After proving the advantages of the FIV procedure over SPCA and standard GMM in terms of efficiency, FIV is adopted to test the more recent theories on the determinants of violent crime in the United States for the period 1982-2005. The available dataset is very large and includes criminal, law-and-order, socio-economic and demographic variables provided by official sources mostly on an annual basis. To avoid the renown small-sample doldrums, the dataset is quarterlized by means of the Santos Silva-Cardoso method, found to be the best interpolator among several time-disaggregation procedures. The transformed dataset is utilized to estimate single-equation estimation of total violent crime and its major categorized component: the homicide rate. The regressors of each equation are established to be the principal components computed via SPCA, while the remaining series represent the potential instruments, in all cases reduced in number by appropriate factor selection procedures . For total violent crime, the most significant determinants are found to be most demographic variables together with law enforcement, while unchecked gun availability chiefly affects homicides.

## References

- Ahn S.C. and Horenstein A.R. (2009) *Eigenvalue Ratio Test for the Number of Factors*, mimeo, Arizona State University and Instituto Autónomo Tecnológico de México.
- Altonji J. G. and Segal L.M. (1996) *Small-Sample Bias in GMM Estimation of Covariance Structures*, Journal of Business and Economic Statistics, 4, 353-366.
- Andrews D.W K. and Stock J. (2007) *Testing with Many Weak Instruments*, Journal of Econometrics, 138, 24-46.
- Ayres I. and Donohue J.J. (2003) *Shooting Down the "More Guns, Less Crime" Hypothesis*, Stanford Law Review, 55, 1193-1312.
- Bai J. and Ng S. (2002) *Determining the Number of Factors in Approximate Factor Models*, Econometrica, 70, 191-221.
- Bai J. and Ng S. (2006) *Instrumental Variable Estimation in a Data Rich Environment*, NYU mimeo.
- Bai J. and Ng S. (2007) *Determining the Number of Primitive Shocks in Factor Models*, Journal of Business and Economic Statistics, 26, 52- 60.
- Bai J. and Ng S. (2008) *Selecting Instrumental Variable Estimation in a Data Rich Environment*, Journal of Time Series Econometrics, 1, 1-32.
- Bair E., Hastie T., Paul D. and Tibshirani R. (2006) *Prediction by Supervised Principal Components*, Journal of the American Statistical Association, 101, 119-137.
- Cattell, R.B. (1966) *The scree test for the number of factors*, Multivariate Behavioral Research, 1, 245-276.
- Chamberlain G. and Rothschild M. (1983) *Arbitrage, factor structure, and mean-variance analysis on large asset markets*, Econometrica, 51, 1281-1304.
- Chen B. (2007) *An Empirical Comparison of Methods for Temporal Distribution and Interpolation at the National Accounts*, Office of Directors, Bureau of Economic Analysis, Washington D.C..

- Chow, G. and Lin, A.L. (1971) *Best Linear Unbiased Distribution and Extrapolation of Economic Time Series by Related Series*, Review of Economic and Statistics, 53, 372-375.
- Denton, F.T. (1971) *Adjustment of Monthly or Quarterly Series to Annual Totals: an Approach Based on Quadratic Minimization*, Journal of the American Statistical Society, 66, 99-102.
- Dezhbakhsh H., Rubin P. and Shepherd J.P. (2003) *Does Capital Punishment Have a Deterrent Effect?: New Evidence from Post-moratorium Panel Data*, American Law and Economics Review, 5, 344-376.
- Di Fonzo T. (2003) *Temporal Disaggregation of Economic Time Series: Towards a Dynamic Extension*, EUROSTAT.
- Donohue J. and Levitt S. (2001) *The Impact of Legalized Abortion on Crime*, Quarterly Journal of Economics, 116, 379-420.
- Donohue J. and Levitt S. (2004) *Further Evidence that Legalized Abortion Lowered Crime: A Reply to Joyce*, Journal of Human Resources, 39, 29-49.
- Donohue J. and Levitt S. (2006) *Measurement Error, Legalized Abortion, and the Decline in Crime: A Response to Foote and Goetz (2005)*, National Bureau of Economic Research Working Paper 11987.
- Donohue J. and Wolfers J. (2005) *Uses and Abuses of Empirical Evidence in the Death Penalty Debate*, Stanford Law Review, 58, 791-846.
- Ehrlich, I. (1975) *The Deterrent Effect of Capital Punishment: A Question of Life and Death*, American Economic Review, 65, 397-417.
- Fernández R.B. (1981) *A Methodological Note on the Estimation of Time Series*, Review of Economics and Statistics, 63, 471-476.
- Granger C.W.J. and Newbold P. (1974) *Spurious Regressions in Econometrics*, Journal of Econometrics, 2, 111-120.



- Groen J. J. and Kapeitanos G. (2009) *Parsimonious Estimation with Many Instruments*, Federal Reserve Bank of New York Staff Report no. 136.
- Hansen L.P. (1982) *Large Sample Properties of Generalized Method of Moments Estimator*, *Econometrica*, 50, 1029-1054.
- Jolliffe, I. (1982) *A note on the use of Principal Components in Regression*, *Applied Statistics*, 31, 300-303.
- Joyce T. (2004) *Did Legalized Abortion Lower Crime?*, *Journal of Human Resources*, 39, 1-28.
- Kapeitanos G. and Marcellino M. (2007) *Factor-GMM Estimation with Large Sets of Possibly Weak Instruments*, Working Papers 577, Queen Mary University of London, Department of Economics.
- Koch I. and Naito K. (2008) *Prediction of Multivariate Responses with a Select Number of Principal Components*, *Electronic Journal of Statistics*, 1, 1-24.
- Levitt S.D. (2004) *Understanding why Crime Fell in the 1990s: Four Factors that Explain the Decline and Six that Do not*, *Journal of Economic Perspectives*, 18, 163-190.
- Lott J.R. and Mustard D.B. (1997) *Crime, Deterrence, and Right-to-carry Concealed handguns*, *Journal of Legal Studies*, 26, 1-68.
- Lott J.R. and Whitley J. (2004) *Abortion and Crime: Unwanted Children and Out-of-Wedlock Births*, *Economic Inquiry*, 45, 304-324.
- Newey W. K. and West K. D. (1987) *A Simple Positive Semi-Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix*, *Econometrica*, 55, p.703-708.
- Newey W. and Smith R.J. (2004) *Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators*, *Econometrica*, 72, 219-255.
- Newey W. and Windmeijer F. (2008) *GMM with Many Weak Moment Conditions*, MIT Department of Economics, manuscript.

Onatski A. (2009) *Determining the Number of Factors from Empirical Distribution of Eigenvalues*, manuscript, Economics Department, Columbia University.

Owen A. B. (1988) *Empirical Likelihood Ratio Confidence Intervals for a Single Functional*, *Biometrika*, 75, 237-249.

Owen A. B. (2001) *Empirical Likelihood*, Chapman and Hall, New York.

Proietti T. (2004) *Temporal Disaggregation by State Space Methods: Dynamic Regression Methods Revisited*, Working Papers, EUROSTAT.

Quilis E.M. (2005) *Benchmarking Techniques in the Spanish Quarterly National Accounts*, Eurostat-OECD Workshop on Frontiers in Benchmarking Techniques and Their Application to Official Statistics, Luxembourg.

Quilis E.M. (2006) *A MATLAB Library of Temporal Disaggregation and Interpolation Methods: Summary*, D.G. del Tesoro y Política Financiera, Madrid, Spain.

Santos Silva, J.M. and Cardoso, F.N. (2001) *The Chow-Lin Method Using Dynamic Models*, *Economic Modeling*, 18, 269-280.

## Technical Appendix

This appendix exhibits the formulas of the formal statistical procedures utilized to compute the true number of common factors ( $r$ ), namely the three Panel Criteria (PC) and the three Information Criteria (IC), together with the Akaike and Bayesian Information Criteria (AIC and BIC) reported by Bai and Ng (2002). The Eigenvalue Ratio Estimator (ERE) and the Eigenvalue Ratio Test proposed by Ahn and Horenstein (2009) are included.

Let  $k_{\max}$  be the maximum number of factors admitted, usually 8, and the sequence  $k = 1, \dots, k_{\max}$ . After replacing  $\ell$  by  $k$  and solving eq. (12) for the disturbances

$\hat{e}_{i,j} = x_{i,j} - f_i \cdot \hat{\lambda}_k$ , define  $V(k) = \min \left( np^{-1} \sum_{i=1}^n \sum_{j=1}^p \hat{e}_{i,j}^2 \right)$  the sum of the squared residuals –

divided by  $np$  – obtained by the regressions of the observed  $X$  on the  $k$  common components. Finally define  $\sigma^2$  as  $V(k \max)$  and  $M = \min(n, p)$ .

The information criteria reported are all based on detecting the minimum variance plus a penalty for overfitting within the sequence  $k = 1, \dots, k \max$ :

$$PC_1 = V(k) + k\sigma^2 \left( \frac{n+p}{np} \right) \ln \left( \frac{np}{n+p} \right);$$

$$PC_2 = V(k) + k\sigma^2 \left( \frac{n+p}{np} \right) \ln(M);$$

$$PC_3 = V(k) + k\sigma^2 \left( \frac{\ln M}{M} \right);$$

$$IC_1 = V(k) + k \left( \frac{n+p}{np} \right) \ln \left( \frac{np}{n+p} \right);$$

$$IC_2 = V(k) + k \left( \frac{n+p}{np} \right) \ln(M);$$

$$IC_3 = V(k) + k \left( \frac{\ln M}{M} \right);$$

$$AIC = V(k) + k\sigma^2 \left( 2 \frac{n+p-k}{np} \right); \quad BIC = V(k) + k\sigma^2 \left( \frac{(n+p-k) \ln(np)}{np} \right).$$

The other two statistics are:

$$ERE(k) = \frac{V(k-1) - V(k)}{V(k) - V(k+1)}; \quad ERT(k) = S_k - S_{k-1}$$

where the first is the ratio of changes in the sum of squared residuals reported above, and the second is a *screeplot* test (Cattell, 1966) wherein  $S_k$  is the  $k$ .th eigenvalue of the

SVD of the covariance matrix (eqs. 13 and 14). The true number of common factors ( $r$ ) is detected whenever a maximum is achieved in either of them.

## **Data and Source Appendix**

### **A. List of endogenous variables**

Total violent crime and categories per 1,000 persons aged 12 and above:

TOTAL\_VIOLENT\_CRIME,

Homicide offending rate: HOMOFF\_RATE\_TOTAL,

*Source: Bureau of Justice and FBI Uniform Crime Reports.*

### **B. List of regressors/instruments (Number of variables in brackets)\*.**

R1) Public spending on law and order, percent and corrected by GDP Deflator (3).

EXPOLICE, EXPJUD, EXPCORR.

*Source: Historical Tables, Budget of the United States Government, Fiscal Year 2008 (mn. Dollars).*

R2) Correctional population (7).

Percent of persons under capital punishment: TDEATHROW, TOTEXEC.

Percent of persons in custody of state correctional authorities, total and by category: TOTINC, TPROBATION, TJAIL, TPRISON, TPAROLE.

*Source: FBI Uniform Crime Reports, annual.*

R3) Educational variables (6).

Percent of persons (aged 25 and over) who have completed high school or college, total and by gender:

ALLHISCHOOL\_25, MALEHISCHOOL\_25, FEMALEHISCHOOL\_25.

Percent of people (aged 25 and over) who have completed elementary school, high-school percent dropout rates (age 16- 24), percent of both sexes enrolled in all school levels (age 3-34): COMPLETED\_25, TOTALDROP, ALL\_ENROLRATIO\_334.

*Source: US Census Bureau, Educational Attainment Historical Tables.*

R4) Demographic variables (12).

Total abortion rates (percent and per 1,000 women aged 15-24): AGIP, AGIW.

*Source: Alan Guttmacher Institute.*

Abortion, birth and pregnancy rates (per 1,000 women) by age groups (15-19, 20-24):

ABRATE\_1519, ABRATE\_2024, BIRATE\_1519, BIRATE\_2024, PREGRATE\_1519, PREGRATE\_2024.

*Source: Alan Guttmacher Institute, "U.S. Teenage Pregnancy Statistics. National and State Trends and Trends by Race and Ethnicity", 2006. Table 2.1 for ages 15-19, Table 2.6 for ages 20-24.*

Total abortion rate, abortions per 1,000 women aged 15-24, per 1,000 live births, and per 1,000 women aged 15 to 44: CDCP, CDCW, CDCRATIO, CDCRATE.

*Source: Centers for Disease Control.*

R5) Gun-related crime rates by circumstance (5).

FELONY\_GUN\_RATE, ARGUMENT\_GUN\_RATE, GANG\_GUN\_RATE, OTHER\_GUN\_RATE, UNKNOWN\_GUN\_RATE.

*Source: Bureau of Justice and FBI Uniform Crime Reports.*

R6) Homicide offending rates by age group (6).

Homicide rates (under14, 14-17, 18-24, 25-34, 35-49, 50 and over): ORU14, OR1417, OR1824, OR2534, OR3549, ORO50.

*Source: Bureau of Justice and FBI Uniform Crime Reports*

R7) Total and disaggregate violent crime rates committed with firearms (4).

Firearm rates (total, murder, robbery, aggravated assault): FIRE\_TCRIMERATE, FIRE\_MURDRATE, FIRE\_ROBBRATE, FIRE\_ASSRATE,

*Source: Bureau of Justice and FBI Uniform Crime Reports.*

R8) Arrest rates, total and by age group. Homicide clearance ratio (7).

Total arrest rate, arrest rates by age groups (under14, 14-17, 18-20, 21-24, 25 and over): TOT\_ARRATE, UNDER14\_ARRATE, A1417\_ARRATE, A1820\_ARRATE, A2124\_ARRATE, OVER25\_ARRATE,

*Source: Bureau of Justice and FBI Uniform Crime Reports*

Homicide clearance ratio (percent): HOM\_CLEARRATIO,

*Source: FBI, Supplementary Homicide Reports, 1976-2005.*

R9) Drug-related arrest rates (5).

Total drug arrest rate and arrest rates for heroin and cocaine, marijuana, synthetic drugs, and other drugs: TOT\_DRUG\_ARR, HERO\_COCA\_ARR, MARIJ\_ARR, SYNTH\_DRUG\_ARR, OTHER\_DRUG\_ARR.

*Source: Bureau of Justice and FBI Uniform Crime Reports*

R10) Other indicators (15).

Total property crime ratio (percent over total crimes committed): TOT-PROP\_CRIMERATIO.

*Source: Bureau of Justice and FBI Uniform Crime Reports.*

Gini Ratios for Households: GINI\_ALL

*Source: US Census Bureau, Historical Income Tables - Households, Table H-4.*

Total yearly production of beer, tones: BEER,

*Source: Brewer's Almanac.*

Real Disposable Personal Income Quarterly, 2000 chained USD, Seasonally Adjusted.: REAL\_INCOME.

*Source: U.S. Department of Commerce: Bureau of Economic Analysis.*

GDP Deflator Quarterly, Seasonally Adjusted: GDP\_DEFLATOR.

*Source: Federal Reserve Bank of Saint Louis Database (FRED).*

Real income lowest 5%, all races: ALL\_LO5%.

*Source: US Census Bureau, Historical Income Tables - Households, Table H-1.*

Total unemployment rate Quarterly : TOTAL\_UNEMP.

*Source: Bureau of Labor Statistics.*

Total resident population and by age group (under14, 14-17, 18-20, 21-24, 25 and over):

TOTAL\_POP, POP\_UNDER14, POP\_1417, POP\_1820, POP\_2124,  
POP\_25ANDOVER.

*Source: US Census Bureau.*

Percent of all families below poverty levels with and without children under age of 18:  
ALL\_WW, ALL\_W.

*Source: US Census Bureau.*

---

\* Percent, unless otherwise stated, refers to the percent rate of the variable over total resident population. Rate is intended per 100,000 inhabitants.

**Table 1.**

Mean (M.) and standard deviation (S.D.) of Montecarlo simulated values of the minimum singular values of the weight matrix (OMINV), of the Jacobians (JAC and JAC2) and of the asymptotic parameter variance (AVAR) of the GMM, SPCA and FIV estimators\*. Select different sample lengths ( $n$ ) and sizes ( $p$ ) of random normal variates, 1,000 replications each.

	A. 3 Regressors and 7 instruments								B. 5 Regressors and 25 instruments							
	n<p				n>p				n<p				n>p			
	OMINV	JAC	JAC2	AVAR	OMINV	JAC	JAC2	AVAR	OMINV	JAC	JAC2	AVAR	OMINV	JAC	JAC2	AVAR
Column	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
	n=150, p=200				n=100, p=30				n=150, p=200				n=100, p=30			
GMM M.	0.0021	0.0861	0.0084	0.2078	0.0030	0.1061	0.0126	0.1771	0.0020	0.2369	0.0571	0.0186	0.0026	0.2855	0.0831	0.0149
GMM S.D.	0.0010	0.0311	0.0059	0.2879	0.0015	0.0368	0.0084	0.2763	0.0004	0.0321	0.0152	0.0056	0.0006	0.0396	0.0226	0.0048
SPCA M.	0.0002	0.0869	0.0085	30.203	0.0005	0.1073	0.0129	17.199	0.0010	0.2367	0.0571	0.0367	0.0017	0.2861	0.0834	0.0223
SPCA S.D.	0.0002	0.0311	0.0060	99.154	0.0005	0.0370	0.0085	34.277	0.0004	0.0330	0.0156	0.0191	0.0006	0.0397	0.0227	0.0102
FIV M.	0.0030	0.0886	0.0088	0.0466	0.0041	0.1266	0.0182	0.0249	0.0021	0.2326	0.0552	0.0123	0.0027	0.2832	0.0819	0.0056



FIV S.D.	0.0010	0.0257	0.0050	0.0483	0.0013	0.0372	0.0102	0.0298	0.0004	0.0316	0.0148	0.0036	0.0006	0.0402	0.0229	0.0016
	n=70, p=100				n=100, p=70				n=70, p=100				n=100, p=70			
GMM M.	0.0042	0.1268	0.0180	0.1628	0.0031	0.1031	0.0120	0.1810	0.0034	0.3321	0.1129	0.0109	0.0026	0.2845	0.0826	0.0151
GMM S.D.	0.0022	0.0439	0.0123	0.2955	0.0016	0.0369	0.0083	0.2778	0.0009	0.0508	0.0341	0.0039	0.0006	0.0414	0.0237	0.0047
SPCA M.	0.0010	0.1264	0.0179	10.283	0.0005	0.1037	0.0121	20.757	0.0026	0.3329	0.1133	0.0136	0.0017	0.2849	0.0828	0.0225
SPCA S.D.	0.0008	0.0438	0.0121	15.690	0.0004	0.0367	0.0085	45.384	0.0008	0.0503	0.0338	0.0059	0.0005	0.0413	0.0238	0.0104
FIV M.	0.0059	0.1364	0.0211	0.0453	0.0043	0.1172	0.0155	0.0366	0.0036	0.3260	0.1089	0.0084	0.0027	0.2802	0.0803	0.0090
FIV S.D.	0.0022	0.0416	0.0127	0.0610	0.0015	0.0336	0.0086	0.0333	0.0009	0.0495	0.0326	0.0029	0.0006	0.0409	0.0232	0.0029
	n=30, p=50				n=150, p=100				n=30, p=50				n=150, p=100			
GMM M.	0.0093	0.1843	0.0384	0.1048	0.0021	0.0862	0.0083	0.1969	0.0072	0.4528	0.2105	0.0017	0.0019	0.2381	0.0578	0.0187
GMM S.D.	0.0055	0.0665	0.0271	0.1321	0.0010	0.0291	0.0053	0.2779	0.0028	0.0741	0.0674	0.0014	0.0004	0.0330	0.0156	0.0057

SPCA M.	0.0045	0.1804	0.0365	0.3003	0.0002	0.0869	0.0085	28.855	0.0068	0.4530	0.2108	0.0016	0.0010	0.2381	0.0578	0.0365
SPCA S.D.	0.0036	0.0631	0.0250	0.4721	0.0003	0.0305	0.0057	54.370	0.0030	0.0743	0.0673	0.0013	0.0004	0.0327	0.0155	0.0186
FIV M.	0.0128	0.2011	0.0459	0.0383	0.0030	0.0958	0.0103	0.0361	0.0074	0.4393	0.1987	0.0016	0.0021	0.2354	0.0566	0.0103
FIV S.D.	0.0060	0.0614	0.0276	0.0450	0.0010	0.0278	0.0057	0.0335	0.0029	0.0738	0.0655	0.0012	0.0004	0.0327	0.0153	0.0032

\* For each estimator, respectively GMM, SPCA and FIV, OMINV:  $\Omega^{-1}$ ,  $\widehat{\Omega}^{-1}$ ,  $\check{\Omega}^{-1}$ , JAC:  $G$ ,  $\widehat{G}$ ,  $\check{G}$ , JAC2:  $G'G$ ,  $\widehat{G}'\widehat{G}$ ,  $\check{G}'\check{G}$ , and AVAR:  $(G'\Omega^{-1}G)^{-1}$ ,  $(\widehat{G}'\widehat{\Omega}^{-1}\widehat{G})^{-1}$ ,  $(\check{G}'\check{\Omega}^{-1}\check{G})^{-1}$ . See eqs. (15)-(23) of the text.

**Table 2.**

Empirical results of FIV estimation of Total violent crime and Homicide rates in the United States, 1982-2005.

Regressors	Slope and t statistic	
1) Total violent crime rate		
EXPOLICE	-7.3601	13.6196
TOTINC	1.5128	3.4910
AGIP	-1.3968	3.3597
AGIW	-2.9090	6.9300
BIRATE_1519	0.7004	2.0425
ABRATE_1519	-0.1514	0.3045
BIRATE_2024	-0.4675	1.2625
ABRATE_2024	0.9323	4.8260
PREGRATE_1519	-0.2824	1.1550
PREGRATE_2024	0.1471	0.9903
A1820_ARRATE	-1.9014	4.9229
HOM_CLEARRATIO	-0.2309	0.5220
BEER	-0.7346	2.7706
TOTAL_POP	-0.1823	0.6775

POP_UNDER14	0.9871	3.5773
POP_1820	0.0944	0.2844
POP_2124	0.1359	0.3511
POP_25ANDOVER	1.5478	6.6852
CONSTANT	0.0000	6.1169
2) Homicide rate		
FELONY_GUN_RATE	-7.6521	20.2230
FIRE_TCRIMERATE	0.4601	1.0987
FIRE_MURDRATE	-1.6821	4.8206
FIRE_ROBBRATE	-0.4119	0.4588
FIRE_ASSRATE	-0.3923	0.9471
FIRE_TCRIMERATIO	2.7849	5.9144
FIRE_MURDRATIO	3.3467	4.6188
FIRE_ROBBRATIO	0.6314	1.8787
FIRE_ASSRATIO	-0.1735	0.2380
MARIJ_ARR	-0.0778	0.1178
ALL_LO5%	-1.0519	1.7331
CONSTANT	0.0000	4.9933

FIGURE 1. Major crime rates. Years 1976–2005.

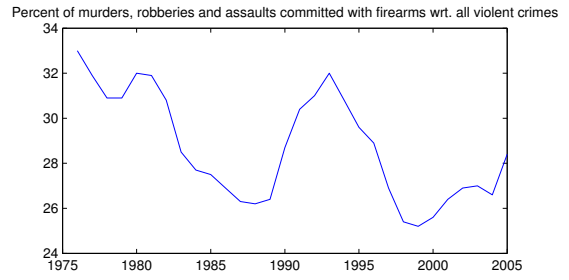
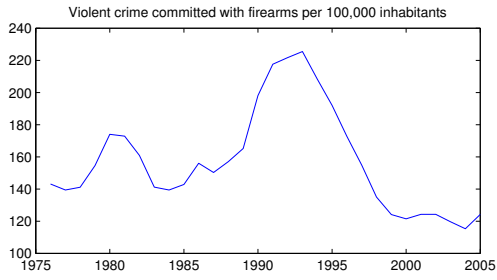
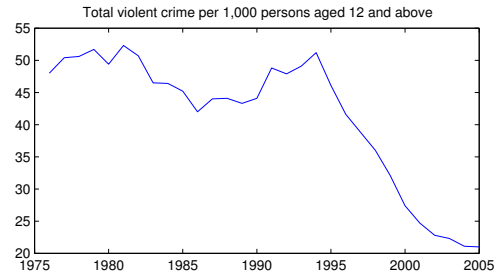
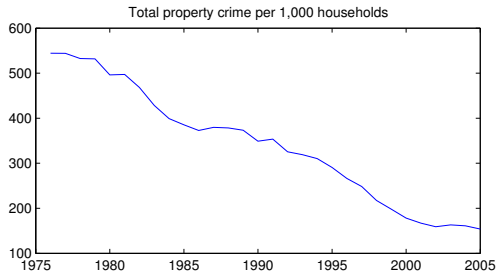


FIGURE 2. Homicide rates per 100,000 inhabitants. Years 1976–2005.

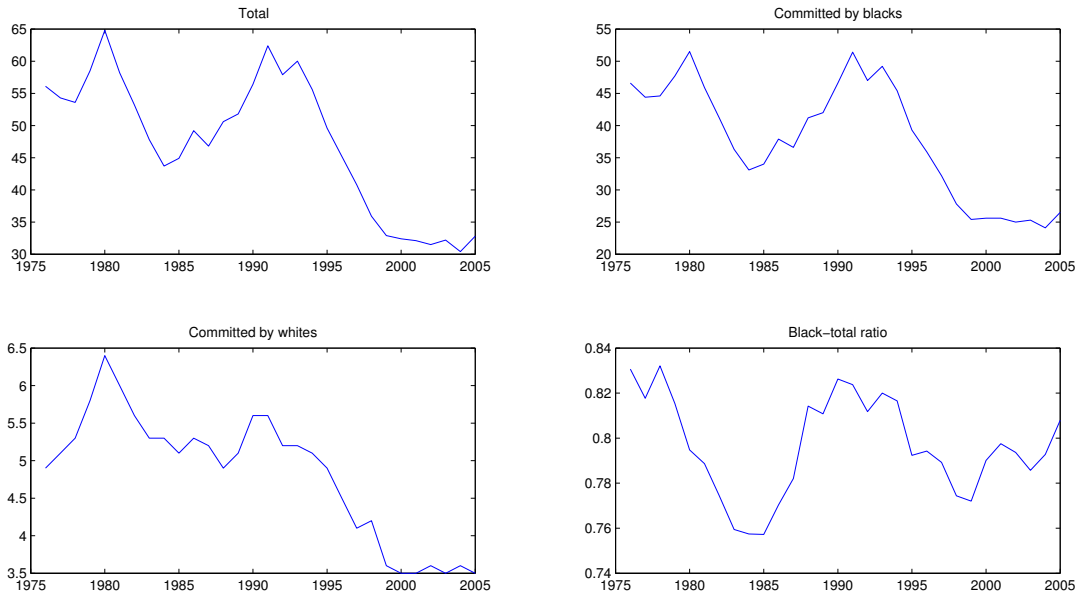


FIGURE 3. Other violent crime rates per 1,000 persons aged 12 and above. Years 1976–2005.

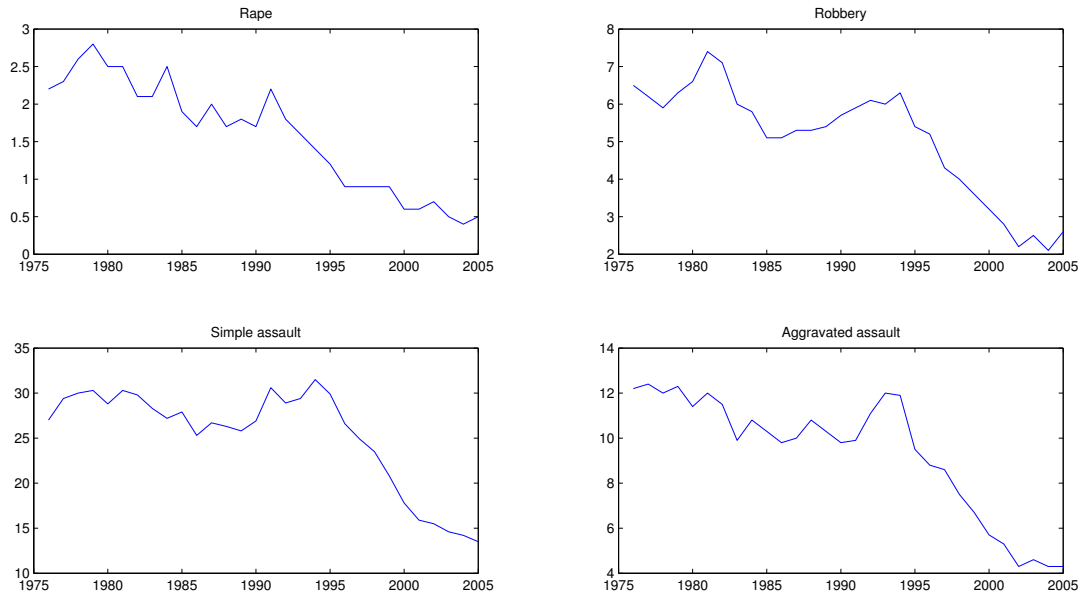


FIGURE 4. Total and select arrest rates per 100,000 population. Years 1976–2005.

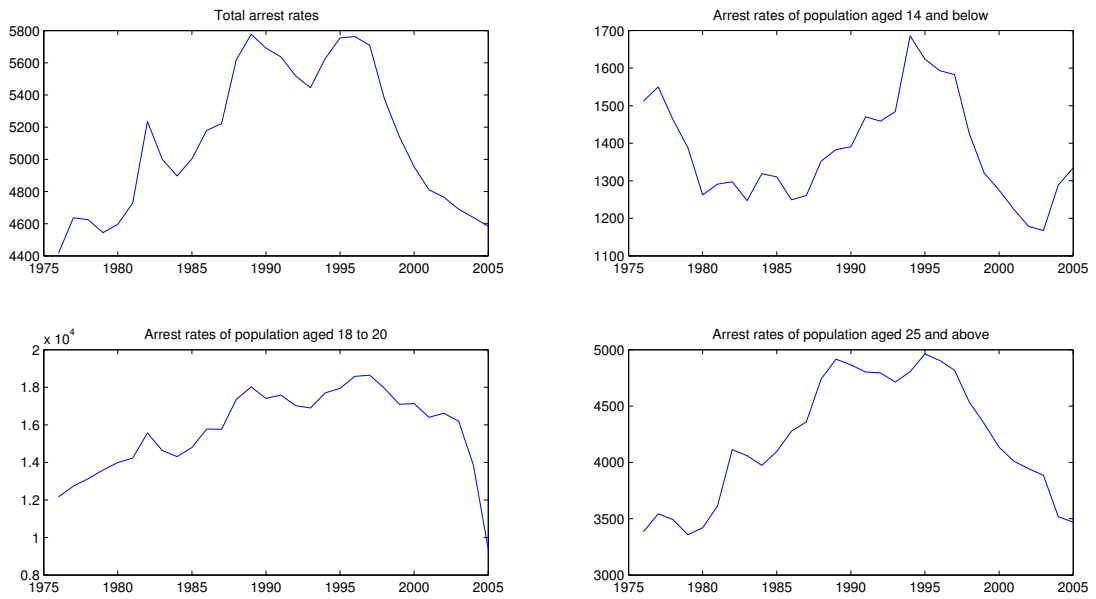


FIGURE 5. Number of persons in correctional custody. Total and by most serious offense. Years 1980–2005.

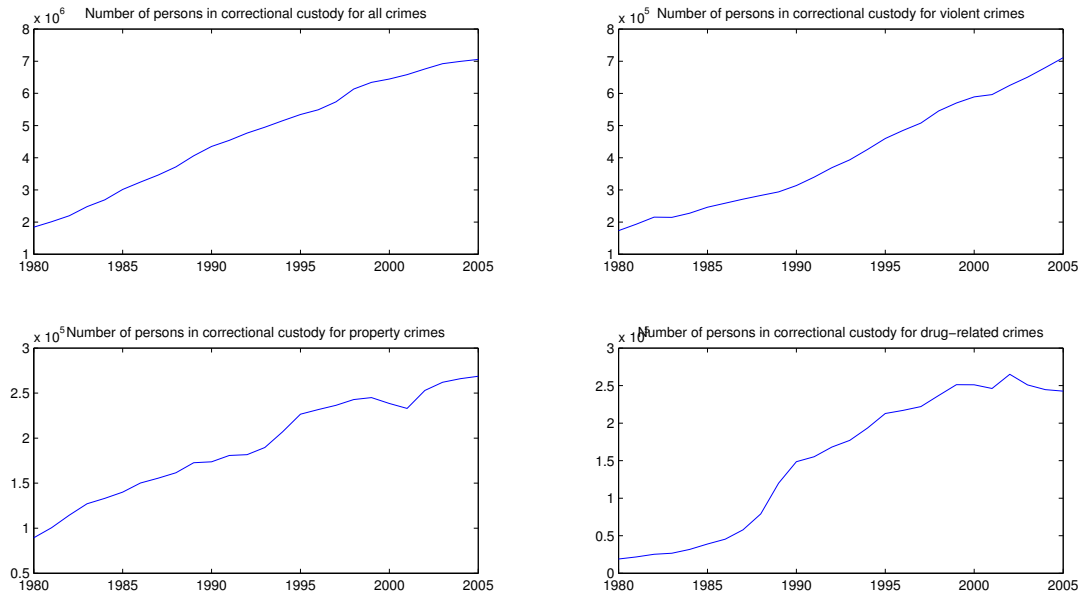


FIGURE 6. Number of persons under capital punishment and number of executions. Years 1976–2005.

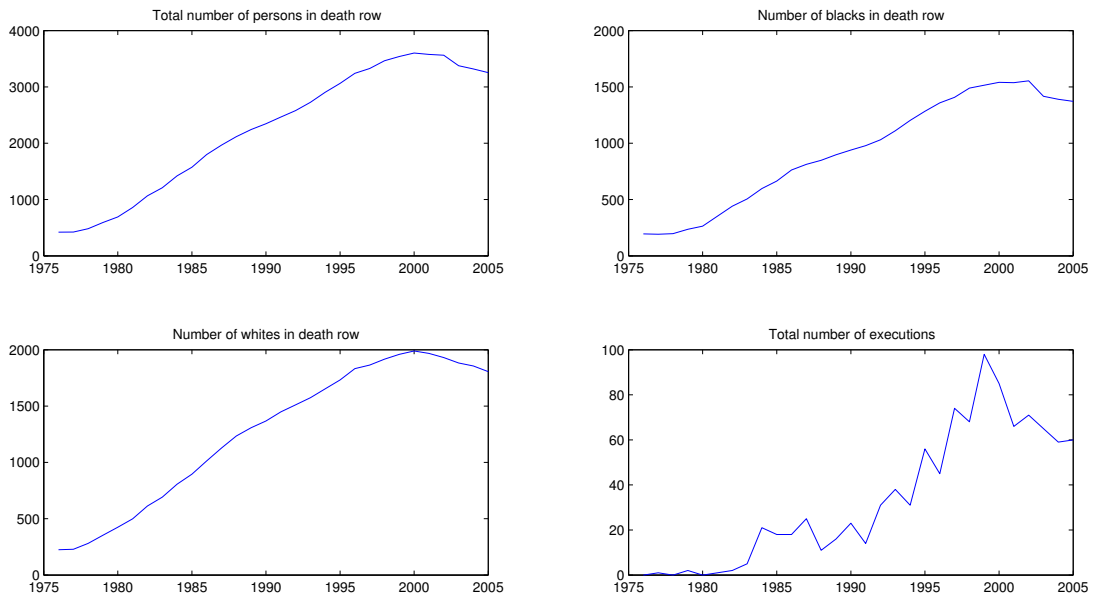




FIGURE 7. Public expenses, total and by function, in current US \$. Years 1982–2005.

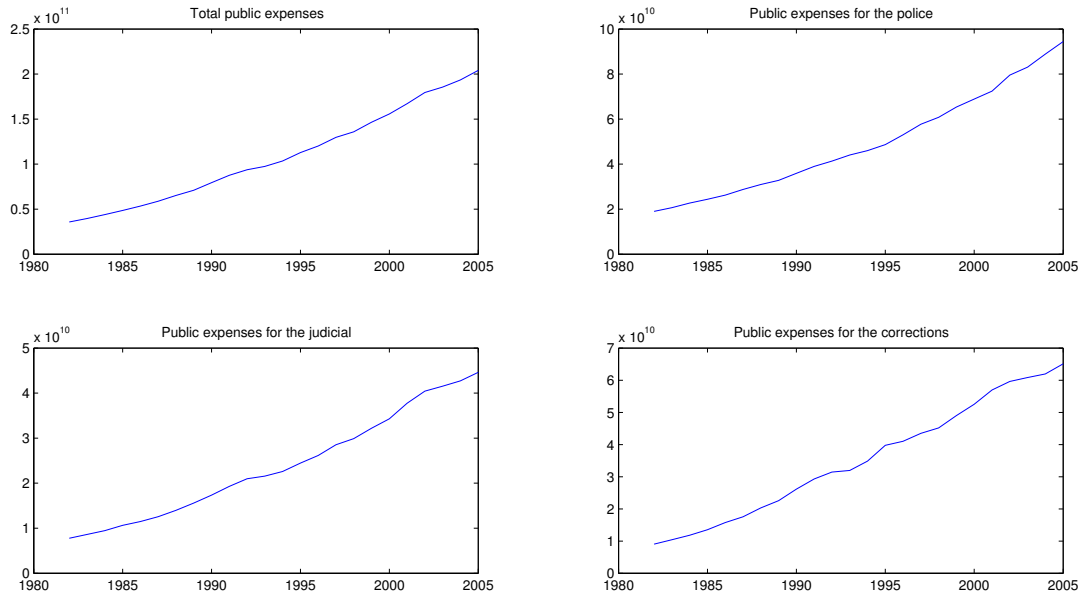


FIGURE 8. Abortion and birth rates (per 1,000 women) by age group. Years 1976–2005.

