Online appendix to "The US Crime Puzzle"

Holger Spamann

May 2, 2015

1 Missing data methods

This section describes in more technical detail the paper's three methods for dealing with missing data.

1.1 MI

The main method in the paper, and the only one used in tables 4 and 5, is multiple imputation (MI). MI proceeds in three steps (e.g., Little and Rubin 2002 ch. 10):

1. create D augmented data sets (I use $D = 200$) by filling in missing values of the independent variables (I do not impute the dependent variable itself[1]). Concretely, I create an individual augmented data set d with the Gibbs' sampler:
   (a) For each variable x^k in $\mathbf{x}$ with missing values, starting from the most observed to the least observed,
      i. draw (with replacement) a bootstrap sample of size n^k from the set of n^k observations that have observed values for variable x^k (this may provide some robustness to model misspecification, Little and Rubin 2002, 216);
      ii. in the bootstrap sample, regress x^k on the complete variables (including the dependent variable of the ultimate regression equation, and any year, sweep, or capital dummies) and any previously imputed variables;
      iii. replace missing values of x^k with draws from a normal distribution parameterized with the regression estimates from the previous step;

   (b) repeat step (a) ten times, now replacing the previously imputed values with newly imputed values;

[1] This restricts the sample to observations with data on the respective dependent variable.

2. estimate the regression equation (described in the previous section) separately on each augmented data set d, yielding D estimated coefficient vectors $\hat{\boldsymbol{\theta}}^{(d)} \equiv \left(\hat{\alpha}^{(d)} \quad \hat{\boldsymbol{\beta}}^{(d)\prime} \quad \hat{\gamma}^{(d)}\right)'$ and covariance matrices $\hat{\mathbf{V}}\left(\hat{\boldsymbol{\theta}}^{(d)}\right)$;[2]

3. calculate the final coefficient estimates as $\hat{\boldsymbol{\theta}}^{MI} \equiv \left(\hat{\alpha}^{MI} \quad \hat{\boldsymbol{\beta}}^{MI\prime} \quad \hat{\gamma}^{MI}\right)' = \sum_d \hat{\boldsymbol{\theta}}^{(d)}/D$ and their covariance matrix as $\hat{\mathbf{V}}\left(\hat{\boldsymbol{\theta}}^{MI}\right) = \sum_d \hat{\mathbf{V}}\left(\hat{\boldsymbol{\theta}}^{(d)}\right)/D + \sum_d \left(\hat{\boldsymbol{\theta}}^{(d)} - \hat{\boldsymbol{\theta}}^{MI}\right)^2 / (D-1)$.

Following the idea of creating a prediction model from countries other than the US, I do not use the US data (which are complete) for imputing other countries' missing values.

To make its assumptions credible, MI requires dropping a very small number of observations. First, MI assumes that the data are missing at random conditional on the non-missing data (Little and Rubin 2002). The main reasons for missing data are presumably poverty and undemocratic governments who conceal data. I therefore drop the few observations that miss data on GDP or indicators for democracy and freedom. Second, the theory of MI technically relies on multivariate normal distributions, even though it typically handles deviations from this assumption better than the alternatives (Graham 2009). Nevertheless, to avoid the most severe complications from non-normality, I exclude a very small number of observations for which categorical variables (legal origin and democracy) are missing.

1.1.1 FIML

For the homicide and incarceration rates, I also show largely identical results from full-information maximum likelihood (FIML). FIML makes the same two assumptions as MI described in the last paragraph, and I again drop observations accordingly. FIML assumes that the dependent and independent variables follow a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. It estimates $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ by maximizing

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_P L_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{1}$$

where p indexes patterns of missing data, and $L_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the log-likelihood derived from the marginal multivariate normal distribution of the variables that are not missing in pattern p and for only the observations exhibiting this pattern (e.g., Enders 2001; cf. Little and Rubin 2002 ch. 6.2, eq. 6.52). Finally, the coefficients of interest are derived from $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ by calculating the appropriate conditional expectation.

[2] When data are pooled for several years as described in the preceding section, the scalar $\hat{\alpha}$ is replaced by a vector $\hat{\boldsymbol{\alpha}}'$ of separate intercepts for each year or sweep, as the case may be.

1.1.2 Missing dummies (OLS+)

Another technique I use with the homicide and incarceration data, restricting to the same sample for comparison and again with very similar results, is to introduce a set of dummies indicating missing status for each variable that has missing data, and to fill in the missing data itself with zeroes (OLS+). That is, the regression equation becomes

$$y_{it} = \alpha + \dot{\boldsymbol{\beta}}' \dot{\mathbf{x}}_{it} + \boldsymbol{\delta}' \mathbf{m}_{it} + \gamma \mathbf{1}_{i=USA} + \varepsilon_{it}, \tag{2}$$

where $\mathbf{m}_{it}$ is a vector of zeroes and ones. The k-th element of $\mathbf{m}_{it}$ is one if the k-th element of $\mathbf{x}_{it}$ is missing and zero otherwise. $\dot{\mathbf{x}}_{it}$ is identical to $\mathbf{x}_{it}$ except that missing values are replaced with zeroes; this replacement concerns only country-year observations that have a missing value for at least one variable and hence would have been dropped in the standard approach of using only complete observations.

OLS+ slightly changes the interpretation of the coefficients on the original variables, indicated by writing $\dot{\boldsymbol{\beta}}$ instead of $\boldsymbol{\beta}$. In particular, the coefficients are now estimates of the slope for those observations that have data, which might differ from the population. The advantage, however, is that with this modified interpretation, the technique can accommodate arbitrary patterns of missing data. (I do however drop the same observations as with MI and FIML for consistency.)

2 Time trends (figure 5): regression equations

This section writes out the regression equations for the construction of the residual time trends in figure 5.

The homicide residuals plotted in the upper panel of figure 5 are coefficients γ_t from

$$IHMEhomicide_{it} = \alpha_t + \dot{\boldsymbol{\beta}}' \dot{\mathbf{x}}_{it} + \boldsymbol{\delta}' \mathbf{m}_{it} + \gamma_t \mathbf{1}_{i=USA} + \varepsilon_{it}. \tag{3}$$

Similarly, the incarceration residuals in that panel are coefficients γ_t from

$$ICPSincarceration_{it} = \alpha + \dot{\boldsymbol{\beta}}' \dot{\mathbf{x}}_{it} + \boldsymbol{\delta}' \mathbf{m}_{it} + \gamma_t \mathbf{1}_{i=USA} + \xi t + \psi t^2 + \varepsilon_{it}. \tag{4}$$

The residuals in the lower panel are coefficients γ_t from regressions of the form

$$y_{it} = \alpha + \ddot{\boldsymbol{\beta}}' \ddot{\mathbf{x}}_{it} + \ddot{\boldsymbol{\delta}}' \ddot{\mathbf{m}}_{it} + \gamma_t \mathbf{1}_{i=USA} + \xi t + \psi t^2 + \varepsilon_{it}, \tag{5}$$

where the double dots on $\ddot{\mathbf{x}}_{it}$ and $\ddot{\mathbf{m}}_{it}$ (and their coefficient vectors $\ddot{\boldsymbol{\beta}}$ and $\ddot{\boldsymbol{\delta}}$) indicate that labor laws, unemployment, and the lagged teen birth rates are missing from the set of regressors.

Appendix references

1. Enders, Craig. 2001. A Primer on Maximum Likelihood Algorithms Available for Use With Missing Data. Structural Equation Modeling 8:128-141.

2. Graham, John. 2009. Missing Data Analysis: Making it Work in the Real World. *Annual Review of Psychology* 60:549-576.

3. Little, Roderick, and Donald Rubin. 2002. *Statistical Analysis with Missing Data.* 2nd ed. Hoboken, NJ: Wiley.