ReadMe file for "Face-to-Face Communication in Organisations"

by Diego Battiston, Jordi Blanes i Vidal and Tom Kirchmaier

Data Access

The data used in this paper is proprietary and therefore not available to researchers at large. However, access to the data may be obtained by directly contacting the Greater Manchester Police.

Contact information is as follows:

Dr Peter Langmead-Jones
External Relations and Performance
Greater Manchester Police
Central Park, M40 5BP
Manchester, United Kingdom
Peter.LangmeadJones@gmp.police.uk

Researchers gaining access to this data (including every person in the research team with potential access to the data) will be required to pass a vetting process conducted by the GMP. This process requires the researchers to be UK residents[1], and can involve delays of several months.

The interested researchers will also need to comply with strict data protection and security requirements, including investing in server (and physical) infrastructure suitable to store highly confidential data. These requirements are set by GMP and specified in the negotiated Data Use Agreement.

The researcher must request and negotiate access to the different datasets described in this document. The authors' experience is that this negotiation can take some months and involve additional delays.

As described in the do-files provided in this replication package, the data is extracted directly by GMP from its internal records and made available in different formats and input files depending on the timespan of the sample, the extraction method and the type of database containing the information. These datasets are checked and anonymized by GMP before being accessible to a researcher and some variables existing in the GMP records remain inaccessible to any researcher (including the authors of this paper) due to confidentially levels above Data Use Agreements. GMP has therefore examined and cleaned all requested datasets from these variables before providing the files to the authors. This process is quite time/resource demanding for GMP and additional delays are expected even after the Data Use Agreements have been signed.

[1] Additional restrictions on minimum time residing in the country can also apply.

Description of Datasets in the Analysis

The data analysed was created by merging three separate sources of information: a 'Command and Control' dataset (cctl), a 'Call Handler Daily Logs' dataset (dailylogs) and a 'Crime' dataset (crime).[2] We now describe the main characteristics of these datasets, together with the main estimation variables. The names of the variables as provided to us varied sometimes depending on the extraction method and timespan of the sample. Therefore, we provide a description of these variables rather than an actual name for them.[3] All the incident and individual identifiers were anonymised by the Greater Manchester Police prior to the researchers accessing the data.

The cctl is a dataset of incidents registered by the system, for the years April 2008 - October 2015. The initial size of this dataset is 9,712,380 observations (many incidents do not reach the system through phone calls -e.g. reported to police stations directly or found by police officers on duty - and/or don't require any police response). Our main analysis is restricted to the period October 2009 to December 2011 when we observe handlers and operators in the same location. We additionally include the period January 2012 to December 2013 for regressions using the 'placebo period'. The cctl dataset is constructed by aggregating four separate datasets (pst, grd, ref, and all, details in provided do file).[4]

The main variables required for the estimation are:

- An incident identifier
- Call Handler identifier
- Radio Operator identifier
- The exact date and time at which the incident was created
- An identifier for the division and subdivision where the incident occurred
- The eastings and northings of the incident location (from which we manually calculated the latitude and longitude).
- The grade and opening code
- The response and allocation time
- An indicator of the incoming phone line (999 or 101)

The crime dataset is a dataset of crimes in the Greater Manchester Police for the years 2009-2014. The initial size of this dataset is 898,467 (many crimes do not reach the system through the OCB and therefore they are not dealt by handlers or operators). This dataset only registers incidents that end up being classified as crimes. Incidents and crimes have different identifiers, but the dataset includes a crosswalk to match across them.

The main variables used in the analysis are:

- An incident identifier
- A crime number identifier
- A dummy for whether the crime was cleared
- The date on which the crime was cleared
- The date on which the crime was reported

[2] GMP (2015a, 2016, 2019). These databases are similar to the ones described in detail in Blanes i Vidal and Kirchmaier (2018) with the addition of some variables required for our analysis (e.g. information on handlers' and operators' location). For more information, we refer the reader to the replication notes of that article.

[3] In order to ease reproducibility, we reference the date of the latest version of the databases provided to the authors' and used in the analysis.

[4] Division/sub-division information is obtained from incident characteristics (GMP, 2019).

- The ending time of the crime

The call handler daily logs contain information of events inputted by call handlers in the computing system or changes in her status (e.g. creation of the incident, switching to not ready status). The initial size of this dataset is 23,282,090 events for the period October 2009 to November 2014. The dataset also contains information on the call handling position from which the event was created.

The main variables for our analysis are:

- A handler identifier
- Event category
- Time and duration of the event
- Call Handling Position

Also, the following auxiliary datasets were used for some of the analysis:

- A 'Human Resources' dataset (base) containing monthly information on:[5]
    - An identifier for OCB staff (mainly call handlers, radio operators and supervisors)
    - The job title of the worker (call handler, radio operator, supervisor, etc.)
    - The division and branch that the individual is associated with
    - Age, gender, ethnicity, etc.
    - The date when the worker joined the GMP
    - The rank or grade of the individual

- 'Victim Satisfaction Surveys' (VSS):[6] it is a dataset containing the responses to phone interviews (commissioned by the GMP) to randomly chosen crime victims, in order to measure satisfaction with the police handling of the incident. The dataset covers the period 2011-2016 (with different coverage over time) and consists of 26,402 observations. The dataset is created by harmonizing a series of xls files provided by the GMP containing the questionnaires and answers of all respondents. The main variables of analysis include:
    - A crime number identifier
    - The date of the interview
    - A variable indicating (in a scale of 1 to 7) the satisfaction of the person regarding how the police managed the incident.
    - A variable indicating if the opinion about the police improved, decreased, or remain the same after the incident handling.
    - A variable measuring the response time (reported by the person in a scale of 1 to 7)

- 'Floorplans' (positions_xy_year): These are datasets for every year (before 2012) with information of the x-y coordinates of each position (call handling, radio unit, supervision) within rooms. The datasets are constructed directly from a series of xls files provided by the GMP where each position is associated to column and a row within a rectangle representing the room.[7]

[5] GMP(2015b)
[6] GMP(2017)
[7] GMP(2016)

- 'Evaluations' (audits): it is a dataset with the scores assigned by the supervisor to the handler and the date of the audit.[8]

- 'Text measures' (text_words_lenght): a dataset with the identifier of the incident, the word and character counts of the first line of the description of the incident in the GMPICS system.[9]

Replication Files and Instructions

We provide four sets of STATA do files that, together with the original data files, will allow the interested researcher to replicate the tables and figures in the paper. The do files are ordered chronologically, according to the first number in their name. The do files have been (very) lightly edited for clarity, with added comments and description of the intermediate steps.

(1) Do files that provide an initial clean of the separate datasets (1_cctl, 1_crime, 1_dailylogs, 1_base, 1_evaudits, 1_vss)
(2) Do files that merge all the datasets together and prepare the estimation sample (2_dataset).
(3) Do files that run and store the regressions and the descriptive statistics (3_regressions, 3_balance, 3_descriptive)
(4) Do files that create the tables and figures in the paper (4_tablesfigures)

All figures and tables in the paper are created in the do-file *4_tablesfigures.do* (the do-file creates the whole figure/table including titles and footnotes) with the exception of appendix Figures A5.A and A5.B that are created in *3_descriptive.do*. All the tables and figures produced by *4_tablesfigures.do* use as input the regressions created and stored by *3_regressions.do*, with the exception of Table 2 which uses as input the file *summary.dta* file created in *3_decriptive.do*. The do-file *4_tablesfigures.do* creates the tables and figures in the same order as they appear in the paper. We have included separators indicating the Table/Figure number and the description of it. The title of the table/figure is generated in the same code, so it is straightforward to match the tables/figures from the paper with the part of the code generating it. The do-file export the tables to format *.tex* directly.

Software:

All the data processing and results have been produced using Stata (version 16.1). Regressions in do file *3_regressions* require installing the package *reghdfe* (Correia, 2017). We used the version 5.7.2 dated 29th July 2019.

The do-file that creates the tables (*4_tablesfigures*) requires installing the package *listtab* (Newson, 2012) and we have used the version dated 4th November 2012 (installed on 4th March 2017).

Both packages *reghdfe* and *listtab* are free and publicly available from the Statistical Software Components (SSC) and can be installed by the user from the Stata command line. Current and older versions of each package are also maintained by their authors in their corresponding websites (see Correia (2017) and Newson (2012) for details).

[8] GMP(2016)
[9] GMP(2015a)

Map of Tables/Figures Generated by the Code to File Names

We provide the mapping between the tables/figures in the text and the file generated by the code. We omit Figures 1-5 that are not based on data analysis.

Table/Figure	File	Do-File
Figure 6	**score.png**	*4_tablesfigures.do*
Figure 7	**quantiles.png**	*4_tablesfigures.do*
Table 1	**DescriptiveCareer.tex**	*4_tablesfigures.do*
Table 2	**Summary.tex**	*4_tablesfigures.do* (based on summary.dta created in *3_decriptive.do*)
Table 3	**Baseline.tex**	*4_tablesfigures.do*
Table 4	**SpilloversH.tex**	*4_tablesfigures.do*
Table 5	**DisplacementWin.tex**	*4_tablesfigures.do*
Table 6	**Residuals.tex**	*4_tablesfigures.do*
Table 7	**Incentives.tex**	*4_tablesfigures.do*
Table 8	**HeterogeneityTechnology.tex**	*4_tablesfigures.do*
Table 9	**Distance.tex**	*4_tablesfigures.do*
Table 10	**HeterogeneitySlack.tex**	*4_tablesfigures.do*
Table 11	**Welfare.tex**	*4_tablesfigures.do*
Table 12	**Welfare2.tex**	*4_tablesfigures.do*

References

Blanes i Vidal, Jordi, and Kirchmaier T., (2018), "The Effect of Police Response Time on Crime Clearance Rates", Review of Economic Studies, 2018, vol. 85, issue 2, 855-891

Correia, Sergio, (2017), "reghdfe: Stata module for linear and instrumental-variable/gmm regression absorbing multiple levels of fixed effects." Statistical Software Components s457874, Boston College Department of Economics. https://ideas.repec.org/c/boc/bocode/s457874.html

GMP (2015a), "Command and Control Databases 2008-2015". Multiple electronic files. Retrieved 2015/10/01. Provided to the authors under Data Processing Agreement 2014/07/16.

GMP (2015b), "Human Resources Records 2005-2015". Multiple electronic files. Retrieved 2015/10/11. Provided to the authors under Data Processing Agreement 2014/07/16.

GMP (2016), "Workforce Management System and OCB Records 2009-2014". Multiple electronic files. Retrieved 2016/02/24. Provided to the authors under Data Processing Agreement 2014/07/16.

GMP (2017), "Victim Satisfaction Surveys 2011-2016". Multiple electronic files. Retrieved 2017/05/06. Provided to the authors under Data Processing Agreement 2014/07/16.

GMP (2019), "Crime and Incident Data 2009-2014". Multiple electronic files. Retrieved 2019/07/01. Provided to the authors under Data Processing Agreement 2014/07/16.

Newson, Roger, (2012), "From resultssets to resultstables in Stata." The Stata Journal 12 (2): 191-213. https://doi.org/10.1177/1536867X1201200203