Description of data and programs used to create the final results of the manuscript entitled "Transactional Governance Structures: New Cross-Country Data and an Application to the Effect of Uncertainty", by Peter Murrell, Nona Karalashvili, and David C. Francis

The full set of data and programs are provided in one zip-file, which contains this document and has the following three subfolders:

* "MS6170Data" – where raw as well as processed data is stored
* "MS6170Programs for data processing and some tables" – where the programs that produce the processed data and some tables are stored
* "MS6170Remaining tables directly from the LatentGOLD output" – where tables that are not produced by Stata programs but are directly linked to the output from the LatentGOLD software that was used in the analysis as described in the paper (programs for which are also provided)

In what follows, we describe steps to produce each table and figure included in the paper and appendixes.

# **Initial data processing**

The Stata dofile called "pre-LatentGOLD-step1-data prep.do" takes the raw survey data, that are included in the "from WBES portal" subfolder of the "MS6170Data" folder of this replication set of folders, and that are also available on [the WBES data portal](https://login.enterprisesurveys.org/), and produces the following three processed datasets under the folder "MS6170Data":[[1]](#footnote-1)

* "LatentGold\_suppliers.dta"
* "LatentGold\_customers.dta"
* "Additional variables.dta"

The main data transformation taking place at this stage is removing special codes denoting "don't know" responses. Note that handling of item non-response is discussed in detail in the paper.

The first two datasets, i.e., "LatentGold\_suppliers.dta" and "LatentGold\_customers.dta" were turned into .sav format for the software LatentGOLD using StatTransfer. These .sav datafiles are provided in the folder "MS6170Data" in their original form as fed into LatentGOLD. The third dataset, i.e., "Additional variables.dta" is needed in the secondary data processing stage and is discussed below.

# **LatentGOLD**

Under folder "MS6170Programs for data processing and some tables", the following two programs contain the exact model specification as run on the LatentGOLD software and reported in the paper:

* "LatentGOLD\_suppliers.lgf" – which uses "LatentGold\_suppliers.sav" discussed above as its input and produces our main LatentGOLD output on firms' governance structures towards their suppliers
* "LatentGOLD\_customers.lgf" – which uses "LatentGold\_customers.sav" discussed above as its input and produces our main LatentGOLD output on firms' governance structures towards their customers

Two main types of output are obtained from LatentGOLD. First, the full estimation output for each of the two models was manually copy-pasted into excel files, as described in more detail below. Second, the estimated posterior probabilities that each firm belongs to each class are exported as datafiles: "supplier\_classification.sav" and "customers\_classification.sav", respectively for firms' relations with suppliers and customers. These output files are provided in the folder "MS6170Data" in their original form as obtained from LatentGOLD. These datasets are also publicly available on the WBES data portal as part of the files: "Landscape of Transactions\_Data on Supplier Relations.dta" and "Landscape of Transactions\_Data on Customer Relations.dta", respectively. In this transformation, the output was not modified, only renamed and re-ordered for the convenience of users.

# **Secondary data processing**

The Stata dofile called "post-LatentGOLD-step1-data prep.do" takes four inputs and produces datafile called "post-LatentGOLD-step1-data.dta" which is used in some of the secondary data analysis. The four inputs into that analysis are: (i) the estimated posterior probabilities that each firm belongs to each class in terms of their relations with suppliers (i.e., file "Landscape of Transactions\_Data on Supplier Relations.dta" mentioned above; (ii) the same for relations with customers (i.e., file "Landscape of Transactions\_Data on Customer Relations.dta"); (iii) the WBES indicators database, which is also publicly available on the WBES data portal; and (iv) the file "Additional variables.dta" from section "Initial data processing" (containing raw variables that are not included in any of the other three input files). This dofile (i.e., "post-LatentGOLD-step1-data prep.do") produces all the variables that are used to examine how governance structures vary with the characteristics of firms or their environments as presented in Section V of the paper, along with the variables used in the analysis of uncertainty as a determinant of governance structures, as presented in Section VI of the paper.

In what follows, steps to obtain each table and figure given in the paper and appendixes are provided in turn.

# **Table 1**

The Stata dofile called "Table 1.do" available under folder "MS6170Programs for data processing and some tables" takes two datafiles from the initial data processing stage called "LatentGold\_suppliers.dta" and "LatentGold\_customers.dta" and produces "Table 1.xlsx" which contains Table 1 as reported in the paper.

# **Tables 2a, 2b and 3**

The excel file "Tables 2 and 3.xlsx" available under folder "MS6170Remaining tables directly from the LatentGOLD output" is organized as follows. The full estimation output from LatentGOLD for the model estimates for the supplier-side relations was manually copy-pasted in the sheet called "4 classes, ab ce db ed cd corr". Due to the order in which the questions were asked (see Appendix A), a is 1, b is 2, d is 3, e is 4, c is 5 and f is 6. Consequently, the correlation structure "ab ce db ed cd" corresponds to 1-2, 4-5, 2-3, 3-4, and 3-5, which was the most preferred model as described in the paper. Similarly, for the customer-side relations, the full estimation output from the LatentGOLD software was manually copy-pasted in the sheet called "4 classes, ab ce fc fe corr". Due to the same question-ordering, this correlation structure corresponds to 1-2, 4-5, 5-6, and 4-6, which as the most preferred model on the customers' side as described in the paper.

The remaining sheets in the excel file use the excel commands "concatenate" and "indirect" to use the structure of the full estimation output and extract the marginal probabilities reported in Tables 2a, 2b and 3.

# **Variations in the importance of governance structures**

To analyze variations in the importance of governance structures, the so-called Step-3 analysis of the LatentGOLD software was used. This was done across many variables, as described in Section V of the paper. These analyses were performed using the structure of programs given in the files "step3\_supplier.lgf" and "step3\_customer.lgf" available under the folder "MS6170Programs for data processing and some tables". In these example programs, the covariate is a1, which is the country code. The input-file used was the same as the output file from the secondary data processing stage (i.e., the estimated posterior probabilities that each firm belongs to each class, datafiles: "supplier\_classification.sav" and "customers\_classification.sav", having dropped the observations with item non-response). The full estimation output for each of the covariates was manually copy-pasted into excel files as described in more detail below.

# **Table 4 and Figures 1-3**

The excel files "Table 4\_Figures 1-3\_suppliers.xlsx" and "Table 4\_Figures 1-3\_customers.xlsx" available under folder "MS6170Remaining tables directly from the LatentGOLD output" are organized as follows. Each sheet to the right of the sheet called "a1 results" contains the manual copy-paste of the full estimation output of the corresponding model in LatentGOLD. The left-most sheet called "Table 4\_customers" and "Table 4\_suppliers", respectively, uses the standard structure of this output, and employs excel commands "concatenate" and "indirect" to extract the relevant statistics as reported in Table 4. Similarly, sheets for Figures 1-3 use the same excel commands "concatenate" and "indirect" to extract the relevant statistics in the requried format, and plot them in figures as shown in the paper.

# **Tables 5 and E.1**

The Stata dofile called "Tables 5 and E1.do" available under folder "MS6170Programs for data processing and some tables" takes the data produced after the secondary data processing, i.e., datafile called "post-LatentGOLD-step1-data.dta" and produces three output files: (i) "Table 5.xls" which contains full output, a subset of rows of which is displayed in Table 5 of the paper; (ii) "Table 5.smcl" which contains additional statistics as reported in Table 5 of the paper, namely Cragg-Donald Wald F statistic and Kleibergen-Paap Wald F statistic. This file also contains output from the implementation of the Altonji et al. (2005)-Oster (2019) method as described in the paper. This method was implemented using the Stata dofile "psasvy.do", which is provided in the folder "MS6170Programs for data processing and some tables";[[2]](#footnote-2) And (iii) "Table E1.xlsx" which provides output reported in Appendix E, Table E1.

# **Table A.1**

The Appendix Table A.1 fully relies on external and publicly available data. The notes of the table provide exact descriptions of sources. Consequently, the program producing this output is not provided, but is available upon request.

# **Table A.2**

Information reported in Appendix Table A.2 is from the Implementation Reports for each of the six surveys used in the analysis. These reports are available on the WBES data portal as part of the documentation of each survey data.

# **Table A.3**

The Stata dofile called "Table A3.do" available under folder "MS6170Programs for data processing and some tables" takes two datafiles from the initial data processing stage called "LatentGold\_suppliers.dta" and "LatentGold\_customers.dta" and produces Table A3.xlsx, which provides the output reported in Appendix Table A.3.

# **Tables A.4 and A.5**

The Stata dofile called "Tables A4 and A5.do" available under folder "MS6170Programs for data processing and some tables" takes two datafiles from the initial data processing stage called "LatentGold\_suppliers.dta" and "LatentGold\_customers.dta" and produces Tables A4 and A5.xlsx, which provides the output reported in Appendix Tables A.4 and A.5.

# **Table C.1**

The excel file "Table C1.xlsx" available under folder "MS6170Remaining tables directly from the LatentGOLD output" is organized as follows. In the top panel, each column is a separate model estimated using LatentGOLD with the correlation structure as given in the first row. In particular, columns A-D correspond to models with 3-6 classes respectively, with no correlation structure, i.e., using the full local independence assumption as described in the paper. The columns F-I are for models with 3-6 classes respectively, with CE correlation structure, which corresponds to 4-5 correlations as described in the paper (the order in which the questions appeared was a-b-d-e-c-f as given in the Appendix A, see above). The columns K-N are for models with 3-6 classes respectively, with AB correlation structure, which corresponds to 1-2 correlations as described in the paper. And finally, columns P-S are for models with 3-6 classes respectively, with AB and CE correlation structure, which corresponds to 1-2, 4-5 correlations as described in the paper. For each column, the top panel reports the set of prominent bivariate residual correlations as described in the paper and Appendix C.

The bottom panel of each sheet of this excel file lists all the potential bivariate residual correlation structures and counts occurrence of each in the top panel. The most frequently occurring correlation structures (marked in bold borders) were used to create an additional set of models as described in the paper. The information in the bottom panel was used to create Appendix Table C.3.

# **Tables C.2a and C.2b**

The excel files "Table C2a.xlsx" and "Table C2b.xlsx" available under folder "MS6170Remaining tables directly from the LatentGOLD output" are organized as follows. Each sheet starting from the second, which is called "3 classes, no corr" contains the manual copy-paste of the full estimation output of the corresponding model from the LatentGOLD software. As noted above, the questions were asked in the order of a-b-d-e-c-f, so the correlation structures given in this excel file can be matched to the ones as reported in the paper and appendix.

The first sheets, called "Table C2a" and "Table C2b" respectively, link directly to the output to extract the relevant statistics as reported in the Appendix Tables C.2a and C.2b.

# **LatentGOLD for the second-best models**

Under the folder "MS6170Programs for data processing and some tables", the following two programs contain the exact model specifications that were applied to the LatentGOLD software for the second-best models: "LatentGOLD\_suppliers\_alternative.lgf" and   
"LatentGOLD\_customers\_alternative.lgf". Similarly to the models that were selected to be the best, two main types of output are obtained for the second-best models: (i) the full estimation output was copy-pasted into the relevant excel files, and (ii) the estimated posterior probabilities that each firm belongs to each class. The latter are in the following two datafiles respectively: "supplier\_classification\_alternative.dta" and "customers\_classification\_alternative.dta" (along with their .sav versions that were directly outputted by LatentGOLD). These output files are provided in the folder "MS6170Data" in their original form as obtained from LatentGOLD, and transferred to .dta format using StatTransfer.

# **Tables C.3a and C.3b**

Tables C.3a and C.3b are similar to Tables 2a and 2b, so the excel file "Table C3.xlsx" under folder "MS6170Remaining tables directly from the LatentGOLD output" is organized in the same way. The full estimation output for the model estimates for the supplier-side relations was manually copy-pasted in the sheet called "5 classes, ab ce db ed cd corr"; the same was done for the customer-side relations in the sheet called "5 classes, ab ce fc fe corr". As noted above, due to the questions being asked in the order a-b-d-e-c-f, these correlation structures correspond to the ones reported in the paper and appendix.

Sheets called "Table C3a" and "Table C3b" use the excel commands "concatenate" and "indirect" to use the structure of the full estimation output and extract the marginal probabilities reported in the Appendix Tables C.3a and C.3b. Note that columns AB-AD also contain information given in Tables C.4a and C.4b respectively.

# **Table C.4a and C.4b**

The Stata dofile called "Table C4.do" uses the datafiles containing the estimated posterior probabilities that each firm belongs to each class for both the first- and second-best models. Namely, for supplier-side relations, it uses datafiles: "…\ from WBES portal\Landscape of Transactions\_Data on Supplier Relations.dta" (most-preferred model) and "supplier\_classification\_alternative.dta" (second-best); and for the customers side, it uses datafiles: "…\ from WBES portal\Landscape of Transactions\_Data on Customer Relations.dta" (most-preferred model) and "customers\_classification\_alternative.dta" (second-best). The dofile creates tabulations in the Stata window that are used to produce Tables C.4a and C.4b.

The bottom panels of Tables C.4a and C.4b report estimates of the prevalence of each class. For the most-preferred models, these are reported in Tables 3a and 3b, while for the second-best models, these are available in excel file "Table C3.xlsx", columns AB-AD of the first two sheets.

# **Tables C.5, C.6a and C.6b**

The excel file "Tables C5 and C6.xlsx" under folder "MS6170Remaining tables directly from the LatentGOLD output" is organized as follows. The full estimation output from LatentGOLD is reported in sheets "4 classes, ab ce db ed cd corr" for the suppliers' side and "4 classes, ab ce fc fe corr" for the customers' side. These are the same as used in other files containing the same output. The sheets "Tables C5-suppliers and C6a" and "Tables C5-customers and C6b" extract the relevant statistics. Additionally, the AvePPk and mcaPk values were manually copy-pasted from the LatentGOLD in these sheets. This information is then used, through excel formulas, to create Tables C.5, C.6a, and C.6b, with the relevant portions highlighted with bold borders in the excel file.

# **Table D.1**

The Stata dofile called "Table D1.do" available under folder "MS6170Programs for data processing and some tables" takes the data produced after the secondary data processing, i.e., the datafile called "post-LatentGOLD-step1-data.dta" and produces Table D1.xlsx, which provides the output reported in Appendix Table D.1.

# **Tables D.2a and D.2b, Figures D.1-D.6**

The excel files "Table D2a and Figures D1a-D6-suppliers.xlsx" and "Table D2b and Figures D1a-D6-customers.xlsx" available under folder "MS6170Remaining tables directly from the LatentGOLD output" are organized similarly to the files for Table 4 and Figure 1-3. In particular, each sheet to the right of the sheet called "a1 results" contains the manual copy-paste of the full estimation output of the corresponding model in LatentGOLD. The left-most sheet called "Table D2a " and "Table D2b", respectively, use the standard structure of this output, and employ excel commands "concatenate" and "indirect" to extract the relevant statistics as reported in Appendix Tables D.2a and D.2b. FDR and FWER statistics are obtained via relevant formulas as given in these files.

Similarly, sheets for Figures D.1-D.6 use the same excel commands "concatenate" and "indirect" to extract the relevant statistics in a necessary format, and plot them in figures as shown in the Appendix D. The names of the sheets facilitate matching with the respective figures in the appendix.

# **File navigation**

* **Folder "MS6170Data\from WBES portal"** contains the following files:
  + Raw data from the six surveys used in the analysis:
    - "Argentina-2017-full data.dta"
    - "Bolivia-2017-full data.dta"
    - "Ecuador-2017-full data.dta"
    - "Paraguay-2017-full data.dta"
    - "Peru-2017-full data.dta"
    - "Uruguay-2017-full data.dta"
  + WBES indicators: "ES-Indicators-Database-Global-Methodology.dta"
  + Estimated posteriors from the most preferred models:
    - "Landscape of Transactions\_Data on Supplier Relations.dta"
    - "Landscape of Transactions\_Data on Customer Relations.dta"
* **Folder "MS6170Data"** contains the following files:

| **File** | **Note** |
| --- | --- |
| Additional variables.dta | Produced as part of the "Initial data processing" described above |
| customers\_classification.sav | The estimated posterior probabilities that each firm belongs to each class using the most preferred model, directly from LatentGOLD, corresponds to "Landscape of Transactions\_Data on Customer Relations.dta" |
| customers\_classification\_alternative.dta | The estimated posterior probabilities that each firm belongs to each class using the second-best model, produced from the .sav file of the same name using StatTransfer |
| customers\_classification\_alternative.sav | The estimated posterior probabilities that each firm belongs to each class using the second-best model, directly from LatentGOLD |
| LatentGold\_customers.dta | Produced as part of the "Initial data processing" described above |
| LatentGold\_customers.sav | Produced from LatentGold\_customers.dta using StatTransfer to feed to LatentGOLD |
| LatentGold\_suppliers.dta | Produced as part of the "Initial data processing" described above |
| LatentGold\_suppliers.sav | Produced from LatentGold\_ suppliers.dta using StatTransfer to feed to LatentGOLD |
| post-LatentGOLD-step1-data.dta | Produced as part of the "Secondary data processing" described above |
| supplier\_classification.sav | The estimated posterior probabilities that each firm belongs to each class using the most preferred model, directly from LatentGOLD, corresponds to "Landscape of Transactions\_Data on Supplier Relations.dta" |
| supplier\_classification\_alternative.dta | The estimated posterior probabilities that each firm belongs to each class using the second-best model, produces from the .sav file of the same name using StatTransfer |
| supplier\_classification\_alternative.sav | The estimated posterior probabilities that each firm belongs to each class using the second-best model, directly from LatentGOLD |

* **Folder " Programs for data processing and some tables"** contains the following files:

| **File** | **Note** |
| --- | --- |
| LatentGOLD\_customers.lgf | LatentGOLD specification for the most-preferred model (customers side) |
| LatentGOLD\_customers\_alternative.lgf | LatentGOLD specification for the second-best model (customers side) |
| LatentGOLD\_suppliers.lgf | LatentGOLD specification for the most-preferred model (suppliers side) |
| LatentGOLD\_suppliers\_alternative.lgf | LatentGOLD specification for the second-best model (suppliers side) |
| post-LatentGOLD-step1-data prep.do | Stata dofile for "Secondary data processing" described above |
| pre-LatentGOLD-step1-data prep.do | Stata dofile for the "Initial data processing" described above |
| psasvy.do | Stata dofile to implement Altonji et al. (2005)-Oster (2019) method, called by dofile called "Tables 5 and E1.do" |
| step3\_customer.lgf | LatentGOLD specification for so-called step-3 analysis for customers' side, using a1 (country) as covariate. Similar models were run across many variables as described in Section V of the paper. |
| step3\_supplier.lgf | LatentGOLD specification for so-called step-3 analysis for suppliers' side, using a1 (country) as covariate. Similar models were run across many variables as described in Section V of the paper. |
| Table 1.do | Stata dofile producing Table 1 |
| Table A3.do | Stata dofile producing Appendix Table A.3 |
| Table C4.do | Stata dofile producing Appendix Tables C.4a and C.4b |
| Table D1.do | Stata dofile producing Appendix Table D.1 |
| Tables 5 and E1.do | Stata dofile producing Table 5 and Appendix Table E.1 |
| Tables A4 and A5.do | Stata dofile producing Appendix Tables A.4 and A.5 |

* **Folder " Programs for data processing and some tables"** contains the following files that contain the relevant full estimation output from LatentGOLD and produce the corresponding tables and figures as given in the file names:
  + Table 4\_Figures 1-3\_customers.xlsx
  + Table 4\_Figures 1-3\_suppliers.xlsx
  + Table C1.xlsx
  + Table C2a.xlsx
  + Table C2b.xlsx
  + Table C3.xlsx
  + Table D2a and Figures D1a-D6-suppliers.xlsx
  + Table D2b and Figures D1a-D6-customers.xlsx
  + Tables 2 and 3.xlsx
  + Tables C5 and C6.xlsx

1. Users need an account (available free of charge) to gain access to the WBES data portal. [↑](#footnote-ref-1)
2. The dofile was obtained from <https://selmawalther.weebly.com/uploads/5/0/2/0/50204335/psasvy.do> in July 2022. It was edited in two ways: (i) the entry `mcontrol' was added in regressions in lines: 23, 87, and 91, without which the output from this method does not match the relevant corresponding output from regressions; (ii) "cap" was added in line 12 to prevent stumbling when the corresponding installation is not needed. [↑](#footnote-ref-2)