**Appendix 1: Summary statistics**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Original dataset name** | **Mean and (s.d). / categories** |
| Firm-authority matches (log) | log(wincaecount) | .36 (.67) |
| Firm count (log) | lnwincount | 2.46 (2.15) |
| Authority count (log) | lncaecount | 4.85 (1.96) |
| Firm authority distance (log) | lndist | 3.95 (2.13) |
| Country | ISO\_COUNTRY\_CODE | 33 countries |
| CPV | CPVthree.ord | 317 codes |
| Authority type | CAE\_TYPE | National (.08); Local (.30); Utilities (.06); EU (.00), Int org (.00); Public body (.22); Other (.20); Natl agency (.01), Reg agency (.02), N/S (.06) |
| Number offers | NUMBER\_OFFERS | Five deciles of distribution, plus missing category. |
| Procedure type | TOP\_TYPE | Open(.79), Restricted (.07), others infrequent |
| Award value | AWARD\_VALUE\_ EURO\_FIN\_1 | Ten deciles of distribution, plus missing category. |
| EU funds | B\_EU\_FUNDS | No (.62); Yes (.08), Missing (.28) |
| Serv/supp/works | TYPE\_OF\_CONTRACT | Services (.35); Supplies (.54); Works (.10) |
| Winning criterion | CRIT\_CODE | Lowest price (.29); Most econ (.58); Missing (.11) |
| Framework agreement | FRA | No (.70); Yes (.29) |
| Subcontracted | B\_SUBCONTRACTED | No (.49); Yes (.07); Missing (.42) |
| Procurement agency | B\_ON\_BEHALF | No (.69); Yes (.07); Missing (.22) |

Dataset count: 1,467,677 after removing entries with no winner information (generally failed tenders), and sampling one transaction for each firm-authority pair.

Table 1: Summary statistics

**Appendix 2: Additional results**

**2.1 Marginal effect of contract award counts**



Figure 1: Predictive effect of authority count



Figure 2: Predictive effect of firm count

****

*Figure 3: Predictive effect of procedure type*



*Figure 4: Predictive effect of EU funding*

****

*Figure 5: Predictive effect of winning criterion*

**2.2 Results on weighted models**



Figure 1: Predictive effect of the country in the weighted model



*Figure 2: Predictive effect of competition in the weighted model*



*Figure 3: Predictive effect of buyer type in weighted models*

**2.3 Results with different levels of aggregation in the linkage procedure**

**2.3.1 Results with firms clustered at the “address-merged” level and authorities at the “cleaned” level.**



Figure 1: Variable importance plot for unweighted, transaction-count, models with less aggressive record linkage

**2.3.2 Results with firms clustered at the .10 level and authorities at the .05 level.**



Figure 1: Variable importance plot for unweighted, transaction-count, models with more aggressive record linkage

**2.4 Results using data with all transactions per firm-authority pair**

**2.4.1 Contract count models**



*Figure 1: Variable importance plot*

|  |  |
| --- | --- |
| Least diverse |  |
| 1 Pharmaceutical products | 195 Sporting services |
| 2 Aircraft and spacecraft | 196 Cybercafé services |
| 3 Drilling services | 197 Postcards, greeting cards and other printed |
| 4 Mining equipment | 198 Motion picture and video services |
| 5 Medical equipments, pharmaceuticals and personal care | 199 Recovered secondary raw materials |
| 6 Medical equipments | 200 Apiculture services |
| 7 Textile yarn and thread | 201 Equal opportunities consultancy services |
| 8 Fruit, vegetables and related products | 202 Animal husbandry services |
| 8 Internet services | 203 Machinery for paper or paperboard production |
| 10 Parts of machinery for mining, quarrying, construction | 204 Space transport services |
|  | Most diverse |

Table 1: Ranking of predicted diversity of ties by CPV-3 code, least to most diverse. Only CPV-3 codes with more than 1000 transactions.



*Figure 3: Predictive effect of the type of product*



*Figure 4: Predictive effect of the type of buyer*



*Figure 5: Predictive effect of contract value*



*Figure 6: Predictive effect of the country*

|  |  |  |  |
| --- | --- | --- | --- |
| Country coefficient | M1 | M2 | M3 |
| Governance | -.110 | -.107 | -.107 |
|  | (.00) | (.02) | (.00) |
| log(GDP/cap) |  | -.007 | -.107 |
|  |  | (.95) | (.88) |
| log(Population) |  |  | -.044 |
|  |  |  | (.00) |
| N | 28 | 28 | 28 |
| R-squared | .48 | .48 | .65 |

*Table 2: Linear regressions predicting the country coefficients. P-values in parentheses.*



*Figure 7: Predictive effect of the procedure type*



*Figure 8: Predictive effect of competition*

**2.4.2 Distance models**



*Figure 1: Variable importance plot*

|  |  |
| --- | --- |
|  |  |
| Closest | ... |
| 1 Agricultural, farming, fishing, forestry and related | 195 Medical equipments |
| 2 Miscellaneous equipment (furniture) | 196 Lifting and handling equipment and parts |
| 3 Tools, locks, keys, hinges, fasteners, chain and springs | 197 Insulated wire and cable |
| 4 Basic inorganic and organic chemicals | 198 Special clothing and accessories |
| 5 Research and development services and related | 199 Research and development consultancy services |
| 6 Furniture | 200 Electricity distribution and related services |
| 7 Petroleum, coal and oil products | 201 Electrical machinery, apparatus, equipment |
| 8 Horticultural services | 202 Software programming and consultancy services |
| 9 Travel agency, tour operator and tourist assistance | 203 Lighting equipment and electric lamps |
| 10 Computer equipment and supplies | 204 Refuse and waste related services |
| ... | Farthest |

Table 1: Ranking of predicted buyer-seller distance by CPV-3 code, closest to farthest. Only CPV-3 codes with more than 1000 transactions.



*Figure 2: Predictive effect of the type of buyer*



*Figure 3: Predictive effect of the country*

|  |  |  |  |
| --- | --- | --- | --- |
| Country coefficient | M4 | M5 | M6 |
| Governance | .215 | .201 | .178 |
|  | (.10) | (.00) | (.27) |
| sqrt(Area) |  | .002 | .002 |
|  |  | (.00) | (.00) |
| log(GDP/cap) |  |  | .071 |
|  |  |  | (.87) |
| N | 28 | 28 | 28 |
| R-squared | .10 | .65 | .65 |

*Table 2: Linear regressions predicting the country coefficients. P-values in parentheses.*



*Figure 4: Predictive effect of the procedure type*



*Figure 5: Predictive effect of competition*

**Appendix 3: Statistical methodology**

The statistical models are estimated in R, using the RandomForest package (Liaw and Wiener 2002). The command for estimating the main random forest model is:

rf1 <- randomForest(y=log(ca$wincaecount), x=ca[,xind], ntree=200, nodesize = 50, sampsize=l/10, proximity = F, importance=T, localImp = F, keep.forest = T, do.trace=T)

Here ca is the data frame containing the dataset and xind is the vector indicating which predictors to include. In the case of categorical predictors, the default behavior of the random forest algorithm is to search through all possible splits each time the categorical variable is considered. This is impossible to compute for variables with large numbers of categories (there are 2^k possibilities, where k is the number of categories). An insight which greatly simplifies computation for such diversified variables is that in case the outcome is continuous, the categorical variable can be ordered by the average outcome in each category, and then treated as continuous. This will generate the same splits as would the default procedure (Hastie et al. 2008, p 310), so this procedure is used for the CPV variable.

The RF model aggregates 200 individual trees, with convergence in terms of mean squared error achieved after about 50-100. The behavior of the error for the main RF model is presented below:



The node size of 50 allows accurate estimation of the models without becoming overly demanding on computing resources. Setting this value to a lower level does not improve classification accuracy meaningfully, but makes estimation much slower (results available on request). The “mtry” parameter, indicating how many variables to consider at each split is left at its default value, which is the number of variables divided by three. (In practice, the accuracy gain from leaving this at this level, corresponding to estimating a random forest, versus setting it equal to the number of variables, corresponding to estimating a “bagged” model appears to be minimal). The bootstrap sample to be drawn at each step is set at 1/10 of the full sample, so approximately 140,000 data points. This again allows a good balance between accuracy and computational feasibility. Larger samples bring no meaningful increase in accuracy but make estimation more difficult.

The predicted effects plots are computed using commands similar to the following:

pp.1 <- partialPlot(rf1, x.var="ISO\_COUNTRY\_CODE", pred.data= ca[sample(1:l, 20000))

The plots are obtained from random samples of 20000 data points, as there is no need to use the full dataset, which would be prohibitively computationally demanding. It can be checked that drawing repeated samples, or increasing the sample size to larger values does not change the plots in meaningful ways. The plotting command estimates the predicted value of the outcome for each level of the independent variable being plotted. As the other variables need to be kept constant, the quantity is estimated for each combination of sample values, and the results averaged. The procedure is the same as the “average partial effects” obtained in Stata.

The model accuracy (such as the 39% estimated for the first model) is estimated with respect to out-of-sample data. For each draw, the data points which are left out-of-sample are predicted using the full set of trees estimated until then.

An accessible introduction to random forest models is available in James et al (2013), and a more advanced treatment is in Hastie et al (2009).

**Appendix 4: Record linkage procedure for firms and authorities**

Inspection of the data reveals that the names of the buyers and sellers are not always recorded in a consistent manner. This is to be expected given that the recording is done by potentially thousands of different employees entering the information in the Ted system. While the winning company field requires listing the official name of the entity, this does not preclude a series of problems including inconsistent use of legal designations such as Ltd., Inc., S.A., GmbH, and so on, but also inconsistent recording of the name itself and outright misspellings. In addition, many of the languages encountered in the sample make use of diacritics, which are difficult to enter consistently and correctly on widely-used English keyboards.

In order to merge the various separate recordings of company and authority names we have broadly followed a procedure which is widely recommended by the statistics and computer science literature on record linkage (Cohen et al 2003), and also implemented in the software packages OpenRefine (Verborgh and De Wilde 2013), and RecordLinkage (Borg and Sariyar 2017). Due to the very large size of the data and various limitations in the packages listed above, we implemented the merging procedure from scratch as will be described. Additionally, we also pursued a few less successful methods, which are briefly discussed.

A recommendation of the literature on record linkage is to acknowledge the probabilistic nature of the process, and rather than searching for the right algorithm to test a series of various procedures, and evaluate each one by drawing a random sample (with a fixed sampling “seed”) from the data and checking the accuracy of the merging process by hand. We will do this by randomly selecting 100 contract awards from the full dataset and computing measures of accuracy for various procedures. Additionally, the full R computer code used for the task is make available in the replication materials.

To establish that two names are similar we made use of a measure of string distance. This procedure is common to all record linkage algorithms and solutions, and is based on the idea that while the same entity (firm, person, public institution) may be recorded under slightly different names, they are much more likely to be similar to each other than randomly chosen words.

The distance metric we settled on after some experimentation is the Jaro-Winkler (JW) distance (Jaro 1989, Winker 1990). This is considered especially appropriate for measuring distances between names of entities, as opposed to generic text, and has been shown to have the best performance among many distance metrics for named entity reconciliation by Cohen et al (2003). The Jaro distance measures the minimum number of character transpositions necessary to turn a string into another, and the JW distance adds Winkler’s key insight that often the beginning of the string is more informative than the end. The JW distance uses a parameter ranging between 0 and .25 to give more or less weight to the beginning as opposed to the end of the string. As Winkler (1990) recommends a weight of .10 to be appropriate for most tasks, we also use this weighing. The JW distance between two string ranges between 0 (completely similar) to 1 (completely dissimilar). The Stringdist R package (van der Loo 2016) is used to compute the JW distance, and in order to achieve reasonable execution speeds we employed a remote computing environment on the Microsoft Azure platform.

**Step 1: Cleaning the strings.**

The first step of all record linkage procedures is a basic “cleaning” of the data. We have therefore performed the following operations on all names of companies:

1. Removing capitalization. This is a standard procedure that is unlikely to affect substantive meaning in any significant way.

2. Removing punctuation. This ensures that, for example, S.A. and SA are the same word.

3. Removing digits. Digits usually appear in the company name field whenever a registration number is included with the company name, or in more unusual situations such as when the address is also mistakenly included.

4. Translating letters with diacritics into their “Latin” counterparts. This is a complex task, which is very well implemented in the Stringi R package (Gagolewski et al 2017). The relevant function is *stri\_trans\_general*, and the translation is into “Latin-ASCII”. This does not affect words written with Cyrillic characters (from Bulgaria), or with Greek characters (from Greece and Cyprus). This step is very useful given that English Qwerty keyboards are widely used across Europe, making it difficult to input diacritics.

5. Removing the ten most common terms that appear in company names in each country. In all of the countries in the sample, the ten most common terms are designations such as “Inc”, “SpA”, as well as possibly the name of the country. Such terms are highly unlikely to help differentiate between companies, and are a major source of variation in the recorded names. The full list of terms removed is available in the replication materials.

A similar procedure is employed to clean up the names of authority names. However, we do not remove the most common terms in this case, as they may be substantively meaningful.

In addition, we also cleaned up the winner and authority address fields, by removing capitalization, punctuation, and diacritics. We do not employ more aggressive merging methods for the address fields, as small differences in these strings may correspond to real differences (E.g. 24 Xyz St is very different from 25 Xyz St).

The names of the cities in which the winner and authority are located are cleaned in a similar manner to the addresses, but in addition we also remove any words that are written after a comma or a parenthesis, as sometimes the name of the province is written in this manner.

**Step 2: Merging on the address, and on name similarity.**

Various experiments with the data have revealed that this is the single most efficient operation for reconciling different recordings of the same entity. While there may be a few ways of writing the same company name, generally the street address is more reliably indicated. (Note that the postal code, city, website, and phone number of the entity are recorded separately.) If two different company names share the same address string, and also have a high degree of similarity, the it is highly likely that they are the same entity. (Naturally, this should be checked on a sample afterwards.) While this is not a necessary condition for two names to reflect the same entity (consider for example regional offices of the same company), it is arguably a sufficient one.

This criterion is implemented as follows: Inside each country, we consider all names that share the same address string (without the city name). For each such group we record the most frequent name as the label of the group (In case of ties, the first one in alphabetical order is the label. ) If a given name is closer than .25 on the JW metric to the label, it is then merged into the label.

This procedure will generate almost no “false positives” (groupings of names that do not belong together), but greatly reduces the name heterogeneity, and ensures that the correct name, in the sense of the most frequently used one, is applied to a large proportion of previously misclassified names (see table 1 in the body of the paper).

**Step 3: Clustering on names.**

The final step of this and most record linkage procedures is to cluster the names of the entities based on string similarity. This should reflect the idea that names such as “Siemens”, “Siemens Corp”, “Siemens Healthcare”, and so on, belong together on the basis of the fact that they are similar. When dealing with large datasets, however, this can be computationally challenging. To give a sense of the scale of the problem, even after the cleaning the data as described above and merging on the address field, there are still around 180,000 unique names in France, the largest country in the sample. The distance matrix holding the distances between all names will have a size proportional to the square of this number (on the order of 160Gb), and a simple hierarchical clustering procedure would have a length proportional to the third power of this number - and is therefore effectively non-computable even with substantial hardware resources. To get around this problem we used a “greedy” clustering algorithm, that optimizes locally rather than trying to operate on the overall distance matrix. Such greedy clustering procedures may be less accurate than non-greedy algorithms, but are computable and often provide more than adequate performance. Indeed, table 1 shows that on our sample of contract awards, the algorithm can help achieve above 95% classification accuracy in some configurations, up from around 80% on the unclustered data.

The clustering procedure used is as follows: the names of companies and authorities are sorted in decreasing order of frequency in the data (in case of ties, alphabetical order is applied). For each name, we compute the distance between itself and all names listed before it in the vector of names. (This ensures every name will be compared with every other name in the dataset for each country). If a JW distance under a certain threshold is measured, the two names are merged into the more frequent one. If multiple matches below the threshold are encountered, the closest match is the one that is merged.

The thresholds considered are .05, .10, and .15 for the JW distance. As the accepted distance for a match increases, the rate of false negatives (missed matches) should decrease, but at the same time the rate of false positives (incorrect matches), should increase. In the case of company names, we used the Jaro-Winkler parameter of .10 to give more weight to the first part of the string. In the case of authority names, we have found that it may not be the case generally that the first part of the string is more informative, so the regular Jaro distance (e.g. a parameter of p=0) is used. The R code for this procedure is available in the replication materials.

**Evaluation of the algorithm.**

To evaluate the success of the various procedures, we draw a sample of contract awards of size 100 from the full data (using the fixed sampling seed 1234). For each of the 100 firms and 100 authorities we record whether it was correctly classified at each step. In order to be correctly classified, an entry has to be both not matched with the wrong label (avoiding a false positive error), and also to not miss any potential matches in the list of names more frequently encountered than itself. Testing the first criterion is easy: for each entry, a judgement can be made on whether the label applied at each step is appropriate (E.g. “huisman muijen adviseur installaties” turning into ““huisman muijen” is appropriate, but “optimare sensorsysteme” into “optimal systems” is an error.) When in doubt, a Google search , together with a translation from Google Translate can be used for this task.

To estimate missed matches (false negatives), we perform a search of the key term or terms for each entry among the full set of names. For example, for “salus international”, we search for all names containing the string salus (even as part of another word). If an entry which we judge to be the same entity is found among those more frequently listed, then the unit fails the false negative criterion. For example, the Irish entry “dhl” was judged as inaccurate, because “dhl express” was also encountered, with a higher frequency.

Table 1 in the body of the paper presents the results of the accuracy test, and reveals three facts regarding the record linkage procedure. The first one is that the non-clustered data is of quite high quality to begin with. Around 79% of companies and 89% of authorities are correctly classified with just a basic cleaning procedure. The second fact is that the merging on address and name step is the most important one for improving classification accuracy: this increases the classification accuracy to 92% for company names and to 97% for authority names. Thirdly, as expected the false positive-false negative tradeoff shifts as the distance is increased in the clustering procedure. Clustering with a .05 distance provides the best balance for company names, but for authority names stopping at the address merging step seems to be optimal. However, as the various clustering solutions reflect different rates of false positive and false negative error, we will present results with a range of parameters, to show that the basic results do not depend on the precise clustering parameters used.

**Unsuccessful record linkage attempts.**

In the following we document a few relatively less successful attempts at performing the record linkage, which may be useful to other researchers. The first one was using the API of the OpenCorporates project, which maintains records of registered companies in countries across the world. The API attempts to match given names to companies in the OpenCorporates dataset. However, we found that it was able to match only a small subset (less than half) of our companies, even when those companies not matched are easily located with a Google search. It is hard to say why this procedure fails, but we suspect the fuzzy matching algorithm used by the API is not appropriate. The raw OpenCorporates data is not available to researchers.

The second less successful attempt was to “block” the clustering process on the city of the company or authority, by performing the clustering only inside a city. While this ensures very high accuracy in terms of avoiding false positives, we found it less well-performing than the procedure actually used in terms of avoiding missed matches in all cases.

The third unsuccessful procedure is to use the “textbook” word clustering procedure inside each country, by computing a distance matrix for all names, and then performing hierarchical clustering on that matrix. This only works on the smaller countries in our sample: While sets of up to 30,000 names can be clustered on a desktop computer, as the size increases to around 80,000 (in the case of Germany and the UK), and especially above 100,000 (Poland and France), this becomes computationally infeasible even with high-performance computing resources at our disposal. As is often the case, however, a simpler greedy clustering algorithm can provide acceptable performance and can be computed much more effectively.

**Appendix 5. Results on above-thresholds contracts**

The thresholds raise various challenges which make it difficult to identify the contracts which are truly voluntarily published. Contracts are coded as above and below the thresholds depending on whether the contract total value is above the various thresholds which were in place in the years 2009 - 2015, for various types of contracts (central government, local government, and works contracts as a separate category). Doing this tells us that 20.0% of the contract awards for which a contract price was published are potentially under the threshold. This is not a definitive estimate because the publication requirement is based on *estimated* total value, and we only have data on the realized total value. Examination of the distribution of total contract values in the overall sample and in individual countries does not reveal any obvious breaks at around 130,000 or 190,000 euros which are the most relevant thresholds, so it appears that authorities generally do not simply stop publishing contracts that come just under the thresholds. The most significant impediment, however, is that approximately 21% of the contracts do not have the total price data recorded, which makes it difficult to separate those which are under the threshold.

**5.1 Results on contract-count models**



*Figure 1: Variable importance plot*



*Figure 2: Predictive effect of the type of buyer*



*Figure 3: Predictive effect of transaction value*



*Figure 4: Predictive effect of the country*



*Figure 5: Predictive effect of competition*

**5.2 Results on distance data**



*Figure 1: Variable importance plot*



*Figure 2: Predictive effect of the country*

**Appendix 6. Linear regression results**

*Table 6. 1 Linear regression model with firm-authority matches dependent variable.*

|  |  |  |
| --- | --- | --- |
| **Variable** | **Coefficient** | **P-value** |
|  |  |  |
| **CPV code** |  |  |
| Telecommunications services | -0.4379 | .00 |
| Secondary education services | -.04194 | .00 |
| Postal and telecommunications services | -.399 | .00 |
| Installation of medical equipment | -.3982 | .00 |
|  |  |  |
| Road transport services | .0613 | .00 |
| Forestry services | .1063 | .00 |
| Reinsurance services | .1633 | .01 |
| Professional services for oil industry | .1700 | .03 |
|  |  |  |
| **Nature of the product** |  |  |
| Services (baseline) |  |  |
| Supplies | -0.002 | .43 |
| Works | -0.024 | .00 |
|  |  |  |
| **Type of authority** |  |  |
|  |  |  |
| National govt (baseline) |  |  |
| Local authorities | -0.057 | .00 |
| Utilities | 0.020 | .00 |
| EU institutions | -0.014 | .04 |
| International organizations | 0.081 | .00 |
| Body governed by public law | -0.021 | .00 |
| Other | -0.024 | .00 |
| National agency | 0.007 | .05 |
| Local agency | -0.007 | .03 |
| Not specified | -0.033 | .00 |
|  |  |  |
| **Size of contract award** |  |  |
| 871000+ | -.014 | .00 |
| <871000 | -.007 | .00 |
| <322000 (baseline) |  |  |
| <158000 | .031 | .00 |
| Missing | .040 | .00 |
| <76300 | .094 | .00 |
| <36200 | .085 | .00 |
| <16600 | .098 | .00 |
| <6940 | .109 | .00 |
| <2270 | .120 | .00 |
| <404 | .273 | .00 |
|  |  |  |
| **Framework agreement** |  |  |
| No (baseline) |  |  |
| Yes | .037 | .00 |
|  |  |  |
| **Subcontracting likely** |  |  |
| Missing (baseline) |  |  |
| No | .012 | .09 |
| Yes | -.027 | .00 |
|  |  |  |
| **Procurement agency** |  |  |
| Missing (baseline) |  |  |
| No | 0.006 | .00 |
| Yes | 0.044 | .00 |
|  |  |  |
| **Country** |  |  |
| FR | -0.201 | .00 |
| SE | -0.170 | .00 |
| UK | -0.151 | .00 |
| SI | -0.138 | .00 |
| NL | -.126 | .00 |
| NO | 0.124 | .00 |
| FI | -.114 | .00 |
| DE | -.083 | .00 |
| ES | -.072 | .00 |
| IS | -.068 | .00 |
| DK | -.057 | .00 |
| IT | -.030 | .00 |
| IE | -.029 | .00 |
| PL | -.022 | .00 |
| GR | -.017 | .00 |
| BE | .006 | .32 |
| RO | .010 | .07 |
| CZ | .023 | .00 |
| BG | .028 | .00 |
| CH | .037 | .01 |
| EE | .037 | .00 |
| LU | .058 | .00 |
| PT | .060 | .00 |
| HU | .066 | .00 |
| LI | .073 | .10 |
| SK | .081 | .00 |
| MT | .086 | .00 |
| LT | .087 | .00 |
| MK | .115 | .00 |
| CY | .130 | .00 |
| HR | .141 | .00 |
| LV | .184 | .00 |
|  |  |  |
| **Procedure type** |  |  |
| Accelerated negotiated (baseline) |  |  |
| Accelerated restricted | -0.036 | .14 |
| Awarded without publication | -0.046 | .00 |
| Competitive dialogue | -0.087 | .00 |
| Missing | -0.070 | .00 |
| Negotiated with call | -0.023 | .01 |
| Negotiated without call | 0.083 | .00 |
| Open | -0.026 | .00 |
| Restricted | -0.039 | .00 |
|  |  |  |
| **The number of offers** |  |  |
| 1 | .008 | .00 |
| 2 (baseline) |  |  |
| 3-4 | -.015 | .00 |
| 5-7 | -.033 | .00 |
| 8+ | -.074 | .00 |
| MISS | -.032 | .00 |
|  |  |  |
| **EU funding** |  |  |
| Missing(baseline) |  |  |
| No | .002 | .09 |
| Yes | -.027 | .00 |
|  |  |  |
| **Criterion for deciding winner** |  |  |
| Lowest price (baseline) |  |  |
| Most economical offer | -.010 | .00 |
| Missing | -.008 | .00 |
|  |  |  |
| **Firm contract count** | .114 | .00 |
| **Authority contract count** | .077 | .00 |
|  |  |  |
| Intercept | -.047 | .03 |
|  |  |  |
| N | 1467677 |  |

*Table 6.2. Linear regression results on distance data.*

|  |  |  |
| --- | --- | --- |
| **Variable** | **Coefficient** | **P-value** |
|  |  |  |
| **CPV code** |  |  |
| Cybercafe services | -3.014 | .13 |
| Primary education services | -2.302 | .00 |
| Recreational, cultural, services | -1.130 | .00 |
| Real estate services | -1.116 | .00 |
|  |  |  |
| Pipeline inspection services | .857 | .00 |
| Apiculture services | 1.250 | .11 |
| Leather | 1.411 | .03 |
| Trailers and semi-trailers for agriculture | 1.454 | .14 |
|  |  |  |
| **Nature of the product** |  |  |
| Services (baseline) |  |  |
| Supplies | .385 | .00 |
| Works | .058 | .00 |
|  |  |  |
| **Type of authority** |  |  |
|  |  |  |
| National govt (baseline) |  |  |
| Local authorities | 0.079 | .00 |
| Utilities | 0.415 | .00 |
| EU institutions | 0.554 | .00 |
| International organizations | 0.397 | .00 |
| Body governed by public law | 0.123 | .00 |
| Other | 0.175 | .00 |
| National agency | 0.037 | .01 |
| Local agency | 0.145 | .00 |
| Not specified | 0.193 | .00 |
|  |  |  |
| **Size of contract award** |  |  |
| 871000+ | -.024 | .00 |
| <871000 | -.025 | .00 |
| <32200 (baseline) |  |  |
| <158000 | .004 | .55 |
| <76300 | .043 | .00 |
| Missing | -.012 | .06 |
| <36200 | .096 | .00 |
| <16600 | .014 | .00 |
| <6940 | .185 | .00 |
| <2270 | .273 | .00 |
| <404 | .112 | .00 |
|  |  |  |
|  |  |  |
|  |  |  |
| **Framework agreement** |  |  |
| No (baseline) |  |  |
| Yes | -0.099 | .00 |
|  |  |  |
| **Subcontracting likely** |  |  |
| Missing (baseline) |  |  |
| No | .025 | .00 |
| Yes | .023 | .00 |
|  |  |  |
| **Procurement agency** |  |  |
| Missing (baseline) |  |  |
| No | -.032 | .00 |
| Yes | .085 | .00 |
|  |  |  |
| **Country** |  |  |
|  |  |  |
| CY | -1.806 | .00 |
| MT | -1.651 | .00 |
| BG | -0.961 | .00 |
| LV | -0.913 | .00 |
| LT | -0.741 | .00 |
| HU | -0.645 | .00 |
| HR | -0.629 | .00 |
| LU | -0.623 | .00 |
| EE | -0.473 | .00 |
| SI | -0.343 | .00 |
| IE | -0.102 | .00 |
| RO | -0.090 | .00 |
| SK | -0.023 | .40 |
| BE | 0.036 | .08 |
| CZ | 0.129 | .00 |
| GR | 0.327 | .00 |
| ES | 0.344 | .00 |
| PL | 0.426 | .00 |
| DK | 0.439 | .00 |
| FI | 0.494 | .00 |
| PT | 0.569 | .00 |
| NL | 0.763 | .00 |
| UK | 0.966 | .00 |
| IT | 0.987 | .00 |
| FR | 1.080 | .00 |
| SE | 1.164 | .00 |
| DE | 1.271 | .00 |
|  |  |  |
| **Procedure type** |  |  |
| Accelerated negotiated (baseline) |  |  |
| Accelerated restricted | -.054 | .17 |
| Awarded without publication | .132 | .00 |
| Competitive dialogue | .410 | .00 |
| Missing | .398 | .00 |
| Negotiated with call | .080 | .02 |
| Negotiated without call | .130 | .00 |
| Open | .118 | .00 |
| Restricted | .105 | .00 |
|  |  |  |
| **The number of offers** |  |  |
| 1 | -.093 | .00 |
| 2 (baseline) |  |  |
| 3-4 | .041 | .00 |
| 5-7 | .064 | .00 |
| 8+ | .084 | .00 |
| MISS | .038 | .00 |
|  |  |  |
| **EU funding** |  |  |
| Missing(baseline) |  |  |
| No | -.0101 | .00 |
| Yes | .2165 | .00 |
|  |  |  |
| **Criterion for deciding winner** |  |  |
| Lowest price (baseline) |  |  |
| Most economical offer | -.014 | .00 |
| Missing | .269 | .00 |
|  |  |  |
|  |  |  |
| Intercept | 2.937 | .00 |
|  |  |  |
| N | 1267239 |  |