

## **Notes on the Programs and Data used for:**

### **“The Costs of Remoteness: Evidence from German Division and Reunification”**

Stephen J. Redding and Daniel M. Sturm  
London School of Economics

March 2008

## **Overview**

This readme file discusses the data and files contained in the RemotenessAER.zip archive for “The Costs of Remoteness: Evidence from German Division and Reunification” (Redding and Sturm 2008). To re-create all the results in the paper run the following Stata do file:

```
runalldofiles.do
```

The code runs with version 8.0 or later of Stata.

## **Directory Structure**

When extracting the files from the zip archive, select the “use folder names” option. Once the zip archive has been unzipped, the directory structure is as follows:

regressions

This directory contains two subdirectories including the data and Stata code used to create all the regression tables in the paper, Figures 3 and 4, and the additional regression results reported in the paper and the web-based technical appendix.

calibration\_simulation

This directory contains the Matlab code used for the calibration and simulation of the model reported in Section II.C of the paper. The directory also contains the Stata code that creates Figures 1 and 2.

grid\_search

This directory contains the Matlab code used for the quantitative analysis of the model discussed in the Section V.C of the paper. The directory also contains the Stata code that creates Figures 5 and 6 and Table 3.

## Directory regressions\main\_results

Run “main\_results.do” to create Figures 1-4, the results in Tables 1-2 and 4-7, and the non-parametric estimation results discussed in Section IV.C of the paper. This Stata do file runs a number of subroutines discussed below.

The main\_results directory has subdirectories with the following contents:

data	RemotenessAER_Main.dta is the main Stata dataset containing data for our sample of 119 West German cities
graphs	Graphs for Figures 3 and 4
logs	Stata log files
subroutines	“border_graphs.do” generates Figures 3 and 4  “matching.do” creates matched pairs of treatment and control cities to be used in the matching estimation  “nonparametric_regressions.do” generates the non-parametric estimation results
tables	Output files with results reported in Tables 2-7
temp	Temporary datasets generated by “main_results.do” and its subroutines

## Directory regressions\robustness

This directory has the following two subdirectories, each of which has a similar structure to the main\_results directory discussed above:

techappendix_robustness	Generates the results in the robustness tests discussed in Section IV.B of the paper and in more detail in the web-based technical appendix  “2002_sample_robustness.do” generates the robustness results for the sample of West German cities with a population greater than 50,000 in 2002. This file uses RemotenessAER_Robustness_2002sample.dta from the data subdirectory.  “techappendix_other_robustness.do” generates the other robustness tests discussed in Section IV.B of the paper and in more detail in the web-based technical appendix. This file uses the dataset RemotenessAER_Main.dta from the main_results directory discussed above.
-------------------------	---

other\_robustness                      Generates the Fulda gap and East Germany robustness tests discussed in Section V.E of the paper

“east\_germany.do” generates the East Germany robustness results. This file uses RemotenessAER\_Robustness\_DDR.dta from the data subdirectory.

“fulda\_gap.do” generates the Fulda Gap robustness results. This file uses the dataset RemotenessAER\_Main.dta from the main\_results directory discussed above.

## **Directory calibration\_simulation**

This directory reproduces the results from the calibration and simulation of the model using central values for parameters from the existing literature as discussed in Section II.C of the paper.

The directory has the following sub-directories:

data                                      Contains all of the csv datasets that are inputs or outputs for the Matlab and Stata programs in this directory.

graphs                                    Contains Figures 1 and 2 from the paper.

### **Model Calibration**

To calibrate the model to the distribution of population across cities in pre-war Germany run the following Matlab file:

Calibrate\_original.m

This file uses the following inputs:

MatlabPopData.csv                      Contains the data on the 206 pre-war German cities used to calibrate the model on the 1939 distribution of population across these cities. There are 15 columns of variables. See the comments in “Calibrate\_original.m” for a description of these variables.

Matlabtrans.csv                          Contains data on the inverse of bilateral distance between 206 pre-war German cities ( $\text{distance\_ij}^{-1}$ ). Distance is the shortest Great Circle Distance between cities. The value of a city’s distance with itself is set equal to 1.

Matlabbrdbrd.csv	Contains data on a bilateral dummy variable which is equal to 1 if pairs of pre-war German cities both lie within the boundaries of the future West Germany.
------------------	--

## Model Simulation

To simulate the impact of division on the distribution of population across West German cities run the following Matlab file:

Simulate\_original.m

This file uses the following inputs:

OrigCalibData.csv	This file is an output file from the Matlab program “Calibrate_original.m”. The file contains data on the 206 pre-war German cities. There are 23 columns of variables. See the comments in “Simulate_original.m” for a description of these variables.
-------------------	---

Matlabtrans.csv	As defined above.
-----------------	-------------------

Matlabbrdbrd.csv	As defined above.
------------------	-------------------

## To Create Figures 1 and 2

To create Figures 1 and 2 in the paper run the following Stata file:

graph\_orig\_calsim.do

This file uses the following inputs:

OrigSimData.csv	This file is an output file created by the Matlab program “Simulate_original.m”. The file contains data on the 119 West German cities. There are 32 columns of variables. See the comments in “graph_orig_calsim.do” for a description of these variables.
-----------------	--

MatlabLegend.csv	This file contains the numeric identifier for each city and the city name.
------------------	--

## Directory grid\_search

This directory reproduces the results from the quantitative analysis of the model as discussed in Section V.A of the paper.

Note that the grid search undertaken in the quantitative analysis of the model is computationally intensive, as it involves calibrating and simulating the model for the 51,152 parameter configurations for which the model has a unique equilibrium. As a result the grid search takes several days to run on current generations of computers.

The directory `grid_search` has the following subdirectories:

<code>data</code>	Contains all of the csv datasets that are inputs or outputs for the Matlab and Stata programs in this directory.
<code>graphs</code>	Contains Figures 5 and 6 from the paper.

## **Grid Search**

To calibrate the model and simulate the impact of division for alternative possible parameter configurations, run the following Matlab file:

`grid_search.m`

This file using the following inputs

<code>MatlabPopData.csv</code>	As defined above.
<code>Matlabtrans.csv</code>	As defined above.
<code>Matlabbrdbrd.csv</code>	As defined above.

The file “`grid_search.m`” creates the following outputs

<code>ConvTest.csv</code>	Contains data for the 51,152 parameter configurations for which the model has a unique stable equilibrium. Column 1 is a dummy variable that equals 1 if the calibration converged to equilibrium. Column 2 is a dummy variable that equals 1 if the simulation converged to equilibrium.
<code>output.csv</code>	Contains the simulated moments from the calibration and simulation of the model for small and large cities for the 51,152 parameter configurations for which the model has a unique equilibrium. There are 18 columns of variables. See the comments in the Stata do file “ <code>gridsearch_results_table3_figure5.do</code> ” for a description of these variables.
<code>BestfitData.csv</code>	Contains the results for our sample of 119 West German cities of the calibration and simulation of the model for the parameter configuration with the smallest sum of squared deviations between the simulated and estimated division treatments for small and large cities. There are 19 columns of variables. See the comments in the Stata do file “ <code>bestfit_parameters_figure6.do</code> ” for a description of these variables.

### **To Create Figures 5 and 6**

To create Figures 5 and 6 in the paper run the following Stata files:

gridsearch_results_table3_figure5.do	Reads in output.csv and creates Table 3 and Figure 5
bestfit_parameters_figure6.do	Reads in BestfitData.csv and creates Figure 6