Replication Instructions for "High Trade Costs and Their Consequences: An Estimated Dynamic Model of African Agricultural Storage and Trade" by Obie Porteous

1. Price, Production, and Population Data

The file PriceMaster4GAMS.xlsx contains the price data for each crop-market-month in USD/kg. Month 1 is January 2002. The months used for the estimation in the paper are months 17-136 (May 2003 – April 2013). Original price series obtained from the data sources listed in Online Appendix table A4 were converted into USD using the exchange rates in the file ExchangeRates.xlsx.

The file MktCoords.xlsx contains the latitude and longitude used for each market when calculating the market catchment areas using the methodology of Pozzi and Robinson (2008). The sub-folder "Catchments" contains the GIS shapefile for the estimated market catchments. Note: due to non-contiguity issues, there are 5 extra catchments in this file. Catchments 991 and 992 (Lagos island and Lekki peninsula) are part of the Lagos market catchment. Catchments 993 and 994 (Zanzibar and Pemba islands) are part of the Dar es Salaam market catchment. Catchment 995 (Cabinda) is part of the Luanda market catchment. The file AllocPCT.xlsx contains the estimated percentage of national population and production of each grain falling within each market's catchment area.

The file realCNH.xlsx contains the final dataset of monthly, market-level population (sheet "N," in millions) and production (sheet "H," in thousand tonnes).

2. Demand Parameter Estimation

The file trade4BSv2KO.dta contains for all country-years the annual quantity consumed of each crop in the model (sum of production and net imports where 1=maize, 2=millet, 3=rice, 4=sorghum, 5=teff, and 6=wheat), the average national annual prices, and the country population. The file trade4BSv2NEWKO.do (along with associated files meanBSnew.ado and doubleBSnew.ado) uses these data to estimate the demand parameters and bootstrapped standard errors reported in table A8 in the online appendix.

The file tradesecondINTLtauwKO.dta contains (for all country-crop-years in the model) the annual quantity consumed, average national annual price, and the landed world price (the instrument). The file tradesecondIVtauwKO.do uses these data to estimate the elasticity of substitution using both OLS and IV. See online appendix section A5 and table A7 for details.

The file tradethirdregionIVKO.dta contains for all country-years the annual population, quantity index, price index, and landed world price index (the instrument). The file tradethirdregionEPSKO.do uses these data to estimate the elasticity of demand using both OLS and IV. See online appendix section A5 and table A7 for details.

3. Inner Loop

The file gamsBS.py is a Python script calling a series of 208 successive GAMS files (.gms) used to sequentially solve for the series of 132 time-specific grain market equilibria (during harvest months, multiple files are used for tractability to solve first for regions within Africa and then for a continent-

wide equilibrium using the regional equilibria as the initial guess). Month 5 corresponds to May 2002 and month 136 to April 2013. Months 5 through 16 (May 2002 – April 2003) are included only to generate initial values for storage in month 17 (May 2003). Only the grain market equilibria in months 17 through 136 (May 2003 – April 2013) are used for the subsequent analysis.

Each GAMS file uses data from a number of Excel files as input:

- Excel files beginning with EH44 contain data on the current harvest and expectations of future harvests (calculated according to Assumption 1).
- Excel files beginning with EP44 contain data on the current population and expectations of future population (assumed to be known with perfect foresight).
- in_alpha.xlsx contains the estimated demand share parameters.
- in_mkt.xlsx contains the estimated demand shifters (A), interest rates (r), and storage costs (k).
- in_pint.xlsx contains data on the current world prices and expectations of future world prices (calculated according to Assumption 1).
- in_PopCaH.xlsx contains initial values for storage in May 2002.
- in_scalar.xlsx contains the values used for the elasticity of substitution and elasticity of demand.
- in_tau2.xlsx contains the estimated bilateral trade cost parameters.
- results6A5prior.xlsx contains an initial guess for the equilibrium in the first month (May 2002).
- path.opt is an options file letting GAMS run for sufficient time to find a grain market equilibrium.

Once the 132 time-specific grain market equilibria have been found, restab2.m is a MATLAB routine that fills them into a single file (resultsALL.xlsx) for the complete grain market equilibrium. It uses auxiliary files Smast.xlsx and links.xlsx.

4. Outer Loop for Cost Parameter Estimation

GAMS files rkest4.gms and tauest2.gms generate Excel files rkest4.xlsx and tauest2.xlsx respectively containing the price data only for the crop-market (pair)-months with storage and trade respectively in the previous inner loop's complete grain market equilibrium by combining the price data (in PriceMaster4GAMS.xlsx) with the storage data and trade data (from the previous inner loop's resultsALL.xlsx).

MATLAB file rknew2.m further processes the Excel files rkest4.xlsx and tauest2.xlsx from the previous step. The output file tauest2.csv contains the origin market, destination market, crop, time period, volume traded, origin price (from data), and destination price (from data) for crop-market pair-months with trade. The output file rknewdematlab.csv contains the market code, crop code, harvest year, minimum price during period with storage (from data), maximum price during period with storage (from data), and number of consecutive periods with storage.

Stata file rk42.do imports the data from the file rknewdematlab.csv generated above and runs the non-linear regression in equation 24 to estimate interest rates (r) and storage costs (k). Standard errors are block bootstrapped (10,000 iterations) with resampling of harvest years at the cluster (market) level. Results from the final iteration are shown in Table 3. The estimates are exported to RK4Matlab.csv. MATLAB file AARKmanipKO.m then takes RK4Matlab.csv and combines it with the demand parameters in AAlpha4MatlabKO.csv to generate in_AARK2.xlsx, which contains the demand parameters, interest rates, and storage costs at the country level. GAMS file AARKmanip2.gms then takes in_AARK2.xlsx and

uses it to update in_mkt.xlsx and in_alpha.xlsx, the files used for the inner loop, which are at the market level.

Stata file taunew.do imports the data from the file tauest2.csv generated above and estimates trade costs using equation 23. Internal standard errors are calculated by bootstrapping (10,000 iterations). Estimates, standard errors, and observations are saved in tauBS10000.dta. Stata file tauupdate.do then applies the updating rule in equation 25 to the trade cost estimates from the previous iteration (contained in Excel file Tau4Matlab.csv) and the new estimates (contained in tauBS10000.dta). Estimates from the final iteration are shown in online appendix tables A11 and A12. The output is a revised Tau4Matlab.csv. MATLAB file taumanip2.m takes trade costs reported in the format of Tau4Matlab.csv and converts them to in_tau2.xlsx, the file used for the inner loop. It uses auxiliary file Mkts4MatlabID.csv.

The versions of RK4Matlab.csv and Tau4Matlab.csv included here contain the values of my preferred initial guess, as described in the "Estimation of Cost Parameters" section of the paper.

5. Reduced-Form Regressions

The file tauKOresults.dta contains the estimated trade costs for each link (variable 'taubar') along with data on link characteristics. The file tauKOports.dta contains analogous data for the port-to-world-market links. The file countryvars4merge.dta contains country-level data from the Transparency International Corruption Perceptions Index and the World Bank Logistics Performance Index, as well as per capita GDP for 2010.

Stata file tauRF.do uses these data to run the reduced-form regressions for overland links reported in table 5. Stata file tauRFborders.do uses these data to run the reduced-form regressions for overland international links reported in table 6. Stata file tauPorts.do uses these data to run the reduced-form regressions for port-to-world-market links reported in online appendix table A13.

6. Goodness of Fit

The file fitcorr.dta contains the price data (variable 'pdat') and the simulated prices from the baseline simulation (variable 'psim'). The file fitindex.dta contains the same data, rearranged so that all crops for a single market-month are contained in a single observation (a second letter of 'd' in the variable name stands for data vs. 's' for simulated prices, while the third letter denotes the crop: 'z' for maize, 'm' for millet, 'r' for rice, 's' for sorghum, 't' for teff, and 'w' for wheat). The files fitcorrNOS.dta and fitindexNOS.dta are the analogous files for the comparison model with no storage. Stata files fitcorr.do, fitcorravg.do, and fitindex.do use these data to compute the goodness of fit statistics for prices reported in the paper.

The file AfrGrAll.dta contains annual, country-level trade flows for the relevant countries and grains from CEPII's BACI project (Gaulier and Zignago 2010). The file tradefitgams.dta contains the simulated trade flows from the baseline simulation after adding up flows from all markets and months for each country-crop-year. Stata files tradefit.do and tradefitgams.do take each of these datasets respectively and transform them into several additional .dta files, each containing net trade flows at a different level

of aggregation. Stata file tradefifit.do then compares these net trade flows from the data and the model, using them to compute the goodness of fit statistics for trade flows reported in table 7.

7. Welfare

The files resultsALL.xlsx and resultsALLcf.xlsx contain the complete grain market equilibria for the baseline model with existing high trade costs and for the main counterfactual simulation with lower trade costs. The GAMS files resultsmod.gms and resultsmodCF.gms are used to correct for a handful of crop-market-months for which extremely small equilibrium prices do not show up in the resultsALL.xlsx files.

The GAMS files welfare.gms and welfareCFtau.gms (which should be run sequentially in this order) compute the total or average values of a variety of variables of interest (e.g. total agricultural revenues, average price index) for each market over the 120 months of interest for the baseline and counterfactual equilibria respectively. Input files used that have already been described include: realCNH.xlsx, in_alpha.xlsx, in_mkt.xlsx, in_scalar.xlsx, and in_tau2.xlsx. The file in_tau2cf.xlsx contains the counterfactual low trade costs. The file YdeMatlab.xlsx contains monthly, market-level income data (in million USD) calculated as described in the "Competitive Equilibrium" section of the paper. An intermediate file (in_nprod.xlsx) containing total production of the numeraire good is generated by welfare.gms and then used in welfareCFtau.gms.

The output of the above GAMS files is written to a generated file called welfare.xlsx. A row for "total" and "average" then needs to be added manually in Excel at the bottom of each sheet. The output from both the baseline (sheet "ac") and counterfactual (sheet "cf") equilibria then need to be copied manually into Stata and saved as welfare.dta. The Stata file welfare.do can then be run to calculate the absolute and percentage changes reported in table 8.

8. Supply Response

The files resultsALLcfSRnew01.xlsx, resultsALLcfSRnew05.xlsx, and resultsALLcfSRnew.xlsx contain the complete grain market equilibria with both counterfactual low trade costs and supply response with price elasticity of supply 0.1, 0.5, and 1 respectively. The files resultsH01.xlsx, resultsH05.xlsx, and resultsH.xlsx contain the corresponding counterfactual harvests. The GAMS file resultsmodCFSRNEW.gms corrects these results as in the previous section.

The GAMS file nprodSRNEW.gms is used to compute the counterfactual production of the numeraire good (recorded in an output file called in_nprodSRNEW.xlsx). Input files used include realCNH.xlsx (already described), in_scalarb.xlsx (containing the price elasticity of supply, which must be changed manually to 0.1, 0.5, or 1 depending on the simulation), in_bee.xlsx (containing the crop-market-time-specific productivity shocks estimated as described in section II.D. of the paper), and in_nprod.xlsx (a file generated in the previous section containing baseline total production of the numeraire good).

The GAMS file welfareCFSRNEW.gms computes the total or average values of the variables of interest and writes them to welfare.xlsx as in the previous section. Refer to the folder for the previous section for the relevant input files. The same analysis in Stata done in the previous section is then used to obtain the absolute and percentage changes reported in table 10.

9. Different Elasticities

The sub-folder “Epsilon=-0.5” contains the files related to the simulations in which the elasticity of demand is set to -0.5 instead of -0.066. The sub-folder “Sigma=3” contains the files related to the simulations in which the elasticity of substitution is set to 3 instead of 1.

Each sub-folder includes revised parameter files in_scalar.xlsx (containing the new values of the elasticities), in_alpha.xlsx (containing the re-estimated demand share parameters), and in_mkt.xlsx (containing the re-estimated demand shifters). The files resultsALL.xlsx and resultsALLcf.xlsx contain the complete grain market equilibria with existing high trade costs and with counterfactual low trade costs. These files are then processed in the same way described above to obtain the absolute and percentage changes reported in table 11.

10. Full Rational Expectations

Each of the 30 small-scale models with full rational expectations has four associated files: a MATLAB program file (.m) and a YAML data file for the baseline simulation with existing high trade costs and analogous MATLAB and YAML files for the counterfactual simulation with low trade costs.

The file MatlabResults.xlsx contains the output from the simulations using the RECS solver in MATLAB. Sheet ‘Spct’ contains the percentage of harvest put into storage for the subsequent year. Sheets ‘SPMX’ and ‘SPMX2’ contain the equilibrium storage, price, imports, and exports for the one-country and two-country models respectively.

The file Stoch1.xlsx analyzes these results. Sheet ‘harvests’ contains the total harvests for each year for each country in the full rational expectations models. Sheets ‘resultsACqtty’ and ‘resultsCFqtty’ combine these data with the percentage of harvest put into storage in equilibrium in the baseline and counterfactual simulations respectively to obtain the total quantity stored each year. The total quantity stored over the ten years is then expressed as a percentage of the total harvest over the ten years, the statistic reported in the paper.

The files resultsALL.xlsx and resultsALLcf.xlsx contain the complete grain market equilibria with existing high trade costs and with counterfactual low trade costs when restricting traders’ choice of inter-annual storage to equal the results from the small-scale models with full rational expectations. These files are then processed in the same way described above to obtain the absolute and percentage changes, which are all well within the 95% confidence intervals of the changes reported in table 8.

11. Comparison Models

The file taucompare.xlsx compares the trade cost estimates using my dynamic monthly model with storage (‘baseline’) to trade cost estimates from a monthly model with no storage (‘nostorage’) and a static annual model (‘staticannual’).

The sub-folder “No Storage” corresponds to the monthly model with no storage. The file resultsALL.xlsx contains the complete grain market equilibrium with no storage using the re-estimated trade costs

(contained in in_tau2.xlsx), while the file resultsALLcf.xlsx contains the equilibrium with no storage using the counterfactual low trade costs.

The sub-folder "Static Annual" corresponds to the static annual model. The file farmgate.xlsx contains data aggregated up to the annual level. The file resultsFG.xlsx contains the complete annual grain market equilibrium using the re-estimated trade costs (contained in in_tau2.xlsx), while the file resultsFGcf.xlsx contains the annual equilibrium using the counterfactual low trade costs. GAMS files welfareFG.gms and welfareFGcf.gms (which should be run sequentially in this order) compute the total or average values of the variables of interest and write them to welfare.xlsx as in previous sections. As in previous sections, this then needs to be copied manually into Stata and saved as welfare.dta. The Stata file welfareFG.do can then be run to calculate the relevant absolute and percentage changes.

12. Trade Corridors

The file CounterfactualTau.xlsx contains the trade cost parameters used for the baseline simulation, the main counterfactual simulation in which all trade costs are lowered, and the 7 additional counterfactual simulations in which only some trade costs are lowered (tables 12 and 13). The rightmost column ('TAH90') contains a 1 for each of the 90 links corresponding to the proposed Trans-African Highways and a 0 otherwise.

For a given simulation, trade cost values need to be copied from CounterfactualTau.xlsx to Tau4Matlab.csv (included above). MATLAB file taumanip2.m (included above) then takes trade costs reported in the format of Tau4Matlab.csv and converts them to in_tau2.xlsx, the file used for the inner loop. It uses auxiliary file Mkts4MatlabID.csv (included above).

Please do not hesitate to contact me at oporteous@middlebury.edu if you need additional information for replication.