**Online Appendix**

Endogeneity between imports and AD filings is a theoretical possibility, and is potentially a difficult empirical problem. Instrumenting highly over-dispersed data with zero-inflated negative binomial models is technically very challenging and somewhat beyond the current econometric frontier.

In our case, the AD filings data are very overdispersed count data (i.e., zeroes dominate the count). In overdispersed count data, we face a tradeoff between instrumenting for potentially endogenous variables and accounting for excess zeroes in the model. In zero inflated negative binominal (ZINB) models, which we show are most appropriate for our data, the first stage is used to account for the excess zeroes problem, and excess zeroes dominate these data. Currently, to our knowledge, there are no Instrumental Variable zero inflated negative binomial models.

Two alternative instrumental variables (IV) strategies are available. The control function IV Poisson procedure is outlined in Cameron and Trivedi (2013; 2010, 607-610), and was due originally to Windmeijer and Santos Silva (1997). The commands can be implemented in STATA 13 (the “ivpoisson” command).[[1]](#footnote-1) The IV Poisson procedure contains a test for exogeneity of the regressors. The IV Poisson procedure, however, has difficulty in estimating models with more than two plausibly endogenous variables.

We truncate the sample for countries with sustained undervaluation[[2]](#footnote-2), and test for the exogeneity of arm’s length and related party imports to trade disputes. We find valid instruments for both types of imports (details available from the author). In the online appendix Table below (model 2), we report the results. The test for exogeneity fails to reject the null hypothesis of exogeneity, which indicates that trade is plausibly exogenous to filings. The results in model 2, moreover, shows that arm’s length and related party coefficient estimates have similar signs and levels of statistical significance as in the ZINB model, though the coefficient estimates are much larger.

The zero-inflated negative binomial models, however, have much better estimation properties than the identical IV Poisson models on identical samples. The sum of squared errors for the ZINB models are 20% the size of the SSEs for the IV Poisson models. For comparison purposes, the full ZINB models with all relevant regressors for the undervalued samples are reported below (online appendix Table models 3 and 4). The signs and levels of statistical significance of the ZINB parameter estimates on the truncated undervaluation samples are close to those reported in the paper.

An alternative estimation strategy is to use GMM-system estimations, which allow for the instrumenting of multiple endogenous regressors. In the reviewers’ table below, model 5 is a GMM-system (two step estimator). The estimates are “difference transformed.” The Sargan statistic suggests that we have valid instruments, and the AB1 and AB2 tests jointly reject the presence of serial correlation.

The GMM-system estimates, however, provide poor predicted values of these highly overdispersed count data, giving rise to negative predictions for trade disputes, which are of course impossible. Forty percent of the predicted trade disputes from the GMM-system estimator in the sample have a negative value. Neither the ZINB models nor the IV Poisson models, by contrast, ever predicts a negative trade dispute, which is reassuring. Moreover, the sum of squared errors for the ZINB predictions is half of that for the GMM-system estimators. The ZINB estimator has much better estimation properties and gives a much better approximation of the data.

We note that the signs, levels of statistical significance, and the size of the parameter estimates of the GMM-system model are relatively similar to the ZINB estimates reported in the paper. The one difference is that the arms-length trade\*undervaluation interaction is not statistically significant and is not reported. The estimated BIT coefficient is also not statistically significant.

In summary, the possible endogeneity between imports and AD filings is a difficult problem. Instrumenting over-dispersed data is technically very challenging and beyond the current econometric frontier.

Cameron, A. C., and P. K. Trivedi. 2013. *Regression Analysis of Count Data*. 2nd ed. New York: Cambridge University Press

Windmeijer, F., and J. M. C. Santos Silva. 1997. Endogeneity in count data models: An application to demand for health care. *Journal of Applied Econometrics* 12: 281–294.

Online Appendix Table

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) | (5) |  |
| *Sample & estimator:* | RP imports  Orthogonalized  ZINB | Undervalued (>20%) IV Poisson | Undervalued (>20%) ZINB | Undervalued (>10%) ZINB | GMM-2 Step (difference-transformed) |  |
| Disputest-1 | 0.152\*\* |  | -0.099 | -0.25 | 0.359\*\*\* |  |
|  | (0.072) |  | (0.105) | (0.13) | (0.097) |  |
| Disputest-3 |  |  |  |  | -0.435\*\*\* |  |
|  |  |  |  |  | (0.115) |  |
| Undervaluationt-1 | 0.002 |  |  |  | 0.070 |  |
|  | (0.403) |  |  |  | (0.195) |  |
| Related-party Importst-1 | -0.74 | -103.228\* | -45.392\*\*\* | -13.123\*\*\* | -3.530 |  |
|  | (0.602) | (54.032) | (7.66) | (2.835) | (2.6334) |  |
| Undervaluationt-1 x | -7.497\*\*\* |  |  |  | -6.885\* |  |
| U.S. Related-party Importst-1 | (2.52) |  |  |  | (3.532) |  |
| Arm's-length Importst-1 | 4.312\*\*\* | 84.018\*\* | 55.312\*\*\* | 48.081\*\*\* | 6.214\*\* |  |
|  | (7.367) | (34.262) | (9.662) | (7.501) | (2.503) |  |
| Undervaluationt-1 x | 16.42\*\*\* |  |  |  |  |  |
| U.S. Arm’s-length Importst-1 | (4.877) |  |  |  |  |  |
| Capital Account Opennesst-1 | -0.019\*\*\* |  | -0.035\*\*\* | -0.02\*\* | -0.007\*\* |  |
|  | (0.005) |  | (0.013) | (0.011) | (0.003) |  |
| Polityt-1 | 0.063\*\* |  | -0.018 | -0.044\* | 0.005 |  |
|  | (0.03) |  | (0.058) | (0.023) | (0.012) |  |
| GDP/capitat-3 | 0.523\*\*\* |  | 0.951\*\*\* | 0.521\* | 0.069 |  |
|  | (0.12) |  | (0.303) | (0.309) | (0.047) |  |
| Bilateral Investment Treatyt-1 | -0.468\*\* |  | 0.852\* | 1.087\*\*\* | 0.404 |  |
|  | (0.308) |  | (0.452) | (0.404) | (0.416) |  |
| RP Imports  Exogeneity Test [p-value for the null |  | [1.03] |  |  |  |  |
| hypothesis of exogeneity] |  |  |  |  |  |  |
| AL Imports  Exogeneity Test [p-value for the null |  | [-1.03] |  |  |  |  |
| hypothesis of exogeneity] |  |  |  |  |  |  |
| Observations | 1338 | 393 | 395 | 534 | 1375 |  |
| Countries | 108 | 52 | 50 | 57 | 109 |  |
| Sargan Test p- value |  |  |  |  | [0.58] |  |
| AB(1) p-value |  |  |  |  | [0.001\*\*\*] |  |
| AB(2) p-value |  |  |  |  | [0.292] |  |

Notes: Model 1 estimates a model with the residuals of a model of related party trade from using lagged values of U.S. FDI into the host country plus all the regressors in Table 4, model 5. The interpretation is that, controlling for the predicted effects of FDI and other covariates, the estimate of the related party trade coefficient is the unique identifying variance of related party imports, controlling for the imports’ dependence on FDI, undervaluation, and other covariates. Sustained undervaluation is defined as three consecutive years (t-1, t-2, t-3) with a currency that is 20% (model 3) or 10% (model 4) or more under-valued using the indicators described in the appendix.

1. , See [www.stata.com/manuals13/rivpoisson.pdf](http://www.stata.com/manuals13/rivpoisson.pdf) for details. [↑](#footnote-ref-1)
2. , Where undervaluation (overvaluation) for three consecutive years was under 20%. Similar results are obtained using a 10% threshold. [↑](#footnote-ref-2)