Reviewer Appendix: Exporters and the environment

J. Scott Holladay   *University of Tennessee*

1.   Imputed Sales and Employment Data

Dunn and Bradstreet is unable to collect actual sales and employment data for each establishment in its database. They use two procedures to fill in the missing data. First, many establishments report ranges rather than a single value for either sales, employment or both. In those cases D&B reports the lower end of the range. If no range is provided D&B uses a proprietary method to impute the missing data. The method takes advantage of the cross sectional relationship between establishments in the same industry to predict missing values. Walls and Associates, who provides the NETS, uses a different technique to impute missing data employing both cross-sectional and time series techniques. If the predictions are consistent with D&B's predictions the NETS reports that imputed value. If Walls and Associates predicted value is significantly different than that of D&B, the Walls and Associates value is reported. For each year-establishment observation there is a flag indicating the data category. Table 1 summarizes the fraction of establishments in each category for both sales and employment. Nearly 90% of establishments report

actual employment data, but only 16% of establishments report sales actual sales data. The 89% of establishments account for 91% of total employment in the data.

To assess the sensitivity of the results to the imputation procedure, I re-estimate the baseline equation using only employment data that was directly reported by establishments to Dunn and Bradstreet. Specifically, I re-run regressions from Table 2 of the paper using the estimating equation:

$$E_{ijt} = \alpha + \pi W_{ijt} + \beta Ex_{ijt} + \gamma_j + \delta_t + \epsilon_{ijt}, \tag{1}$$

on the subset of observations that report actual employment numbers replacing log sales with log employment where appropriate. Table 2 summaries the results of these regressions which . The results are consistent with those from the full data set. Before controlling for firm size exporters pollute significantly more than non-exporters. Conditioning on firm size by introducing logged employment flips the exporter coefficient sign. Exporters pollute around 10% less then non-exporters with the same number of employees. The magnitude of the exporter coefficient changes slightly from the full data, but the sign and statistical significance of the coefficients remains unchanged.

The results of these regressions suggest that limiting the sample to observations with directly reported employment data and using employees rather than sales as the proxy for firm size does not materially affect the results. The

2

results are also robust to using the sales data from firms that directly report employees as a proxy for firm size. Relying on firms that report actual sales and using sales as the proxy for firm size produces similar point estimates, but the results are no longer statistically significant due to the much smaller sample size.

2. Data Quality

The NETS data is compiled by Walls and Associates from annual snapshots of the Dunn and Bradstreet's Dunns Marketing Information (DMI) database. Dunn and Bradstreet takes goes to considerable lengths to maintain the quality of data used to create their establishment level credit ratings. There is no legal requirement for establishments to report, or to provide accurate data, but the credit ratings are widely used and the vast majority of establishments in the United States obtain a DUNS number from Dunn and Bradstreet and provide data. Providing inaccurate data could adversely affect an establishment's credit score in the future and Dunn and Bradstreet carefully evaluates all new applicants for credit scoring to confirm that they are not existing establishments attempting to evade their existing credit rating. Dunn and Bradstreet uses phone interviews, mail surveys and internet searches to maintain and update the data. See Walls & Associates "Understanding Data in the NETS Database" for more information on NETS data quality.

The EPA uses data reported to the Toxic Release Inventory (TRI) to pro-

duce the Risk Screening Environmental Indicators. Establishments that hold more than a threshold amount of toxic chemicals are required to report the quantities of chemicals they use and how they were disposed of or released into the environment. TRI reporting is mandatory and firms that fail to report or misreport are fined. Individuals face potential criminal charges for submitting false information to the government, although though prosecutions appear to be rare. EPA employs a variety of quality assurance procedures to ensure that data reported by firms is correct including cross checks with other firms in the same industry and previous submissions from the same facility. Facilities are informed of any discrepancies and required to provide evidence of the accuracy of their data or revise their submission.

3. Data Appendix

Table 3 lists each variable, its source and a brief description. The National Establishments Time Series (NETS) and EPA's Risk Screening Environmental Indicators (RSEI) both contain DUNS numbers and year, which combine to identify unique observations in the panel. Unfortunately, the DUNS number field is optional for establishments in the EPA's database. Because the field is optional it does not receive the same level of quality control and auditing as required fields and contains missing or invalid DUNS numbers in some cases. To match establishments with missing or invalid DUNS numbers I use location and SIC industry variables that are common across both data

sets.

Due to incomplete data on location (and DUNS numbers) in the RSEI dataset, matching every polluting establishment is impossible. 74.7% of the establishments identified by the EPA match with observations in the NETS each year. The RSEI observations that were matched are smaller as measured by pounds of emissions and hazard score, but there are no significant differences in the risk generated by those emissions. Unmatched establishments could not be assigned plant characteristic data from the NETS so I cannot compare the matched and unmatched establishments along those dimensions. While there are differences in the level of emissions between the two groups, there are no differences in the ratios of any measure of emissions.[1] Table 4 below summarizes the pollution data for matched and unmatched firms. Some of the matched establishments do not appear in the final data set due to missing observations for particular variables in the NETS or RSEI.

---

[1]The difference between the matched and unmatched groups in pounds and hazard are significant at the 1% level. The differences between the groups risk scores and the ratio of pounds to hazard, pounds to risk and risk to hazard are not significant at the 10% level.

Table 1: Data Collection Category: Sales and Employment

Sales	Employment Direct Report	Range Report	D&B Imputed	Walls Imputed	Total
Direct Report	0.15	0.01	0	0	0.16
Range Report	0.02	0	0	0	0.02
D&B Impute	0.04	0	0	0	0.04
Walls Impute	0.69	0.02	0	0.07	0.78
Total	0.89	0.03	0	0.07	

Note: Direct report is data reported directly by the establishment (category 0 is the NETS). Range report is the bottom of a range of values reported by an establishment (category 1 in the NETS). D&B impute is imputed by Dunn and Bradstreet using primarily cross-sectional techniques (category 3). Walls impute is imputed by Walls and Associates relying on both time series and cross sectional data (category 4).

Table 2: Exporters' Environmental Performance with Directly Reported Employment Data

	(A1) Log Hazard	(A2) Log Hazard	(A3) Log Hazard	(A4) Log Hazard
Log Employees		0.875***	0.865***	0.868***
		(107.649)	(107.738)	(107.605)
Relocations				0.074***
				(3.329)
Foreign Owned				0.104*
				(1.919)
Credit Rating				-0.002**
				(-2.311)
Female CEO				-0.208***
				(-4.607)
Export	0.048**	-0.130***	-0.099***	-0.105***
	(2.375)	(-6.650)	(-5.075)	(-5.355)
SIC6 FE	Y	Y	Y	Y
State FE	N	N	Y	Y
Year FE	N	N	Y	Y
R^2	0.305	0.349	0.371	0.372
N	174162	174162	174162	174162

Note: The dependent variable in each regression is the log of reported TRI hazard score. Export is a dummy variable that takes the value of 1 if the establishment has reported exporting in the NETS. Sample size consists of 174,162 observations for which actual employment is reported. Additional controls include number of employees, number of relocations, minimum credit rating in the past year and indicators for female CEO and foreign owned. All standard errors are clustered at the establishment level. *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level. The results are consistent with those of the full sample.

Table 3: Variable List

Variable Name	Frequency	Source	Description
DunsNumber	Static	NETS	D-U-N-S Establishment Number
Company	Static	NETS	Business Name
Address	Static	NETS	Street Address
City	Static	NETS	City Name
State	Static	NETS	State Postal Abbreviation
Zip Code	Static	NETS	5-Digit Postal Zip Code
Zip +4	Static	NETS	4-Digit Zip Code Extension
Latitude	Static	NETS	Establishment Latitude
Longitude	Static	NETS	Establishment Longitude
Year	Annual	NETS	Reporting year for annual data
Emp	Annual	NETS	Establishment Employees
EmpC	Annual	NETS	Establishment Employee Code (0 = Actual, 1 = Bottom of Range, 2 = D&B Estimate, 3 = Walls Estimate)
Sales	Annual	NETS	Establishment Sales
SalesC	Annual	NETS	Establishment Sales Code (0 = Actual, 1 = Bottom of Range, 2 = D&B Estimate, 3 = Walls Estimate)
SIC	Annual	NETS	8-digit Standard Industrial Classification Number
Im/Ex/Both	Static	NETS	Import/Export Indicator (B = Both, E = Export, I = Import, Space = Neither)
HQDuns	Static	NETS	Ultimate/Parent/HQ D-U-N-S Number
PayDexMin	Annual	NETS	Maximum Dun & Bradstreet PayDex Score for year
PayDexMax	Annual	NETS	Minimum Dun & Bradstreet PayDex Score for year
D&B Rating	Annual	NETS	Dunn & Bradstreet Credit Rating
TRIFID	Static	RSEI	EPA's establishment identifier
DunsNumber	Static	RSEI	D-U-N-S Establishment Number (Optional Field)
Company	Static	RSEI	Business Name
Address	Static	RSEI	Street Address
City	Static	RSEI	City Name
State	Static	RSEI	State Postal Abbreviation
Zip Code	Static	RSEI	5-Digit Postal Zip Code
Zip + 4	Static	RSEI	4-Digit Zip Code Extension
Latitude	Static	RSEI	Establishment Latitude
Longitude	Static	RSEI	Establishment Longitude
SIC Codes	Static	RSEI	8-digit Standard Industrial Classification Number
Year	Annual	NETS	Reporting year for annual data
Pounds	Annual	RSEI	Amount of hazard chemicals reported by TRI facilities as released or transferred. (measured in pounds)
Hazard	Annual	RSEI	Pounds released multiplied by the chemical's toxicity weight
Risk	Annual	RSEI	Risk-related results combine dose with toxicity weight & pop. estimate, producing a unit-less value proportional to impact

Table 4: Comparing Matched and Unmatched Firms in the TRI

	Matched	Unmatched
Pounds	328,591	439,374
Hazard	21,728	30,990
Risk	2,873	3,075
Hazard/Pounds	0.066	0.071
Risk/Hazard	0.13	0.10
N	202,666	68,743

Note: Pounds are the quantity of emissions, hazard is a score that measures the quantity and toxicity of emissions and risk measures the quantity, toxicity and location of emissions. Matched firms appear in both the TRI and NETS databases and make up the dataset used in the analysis. Unmatched firms exist in the TRI, but cannot be matched to a NETS observation typically due to missing DUNS numbers and incomplete location data in the TRI data. Matched establishments generate significantly fewer pounds of pollution and lower hazard scores. The other differences are not statistically significant at the five percent level.