

ENTERED
Office of Proceedings
May 16, 2025
Part of
Public Record
May 7, 2025



Subject: Request for access to the Confidential Carload Waybill Sample

Dr. Kristen Monaco
Director, Office of Economics
Surface Transportation Board
395 E Street SW
Washington, DC 20423

Dear Dr. Monaco,

I am writing to request access to the full, unmasked Confidential Carload Waybill Sample (CCWS) from 1990-2020 for the purposes of academic research. I am pursuing a project studying the economic implications of environmental risks to transportation infrastructure, with a particular focus on the US rail network. The project will involve a substantial empirical component in which I intend to use the CCWS to evaluate how disruptions to rail service influence the flow of goods and transportation costs in the short- and long-run.

For my project, I require access to fields pertaining to waybill origin and termination nodes, date, commodity, states travelled, freight revenue, transit and miscellaneous charges, variable cost, weight, and carrier. The public CWS is insufficient for my purposes as it does not contain sufficiently geographically granular information on shipment routing characteristics, carrier, nor unmasked revenue and cost information. To my knowledge, there exists no other data source containing this information.

Researcher Qualifications

I am a Ph.D. student within the Department of Economics at the University of Toronto, currently in the second year of a six year program. I previously received my M.A. and H.B.A. in Economics from the University of Toronto. Through my economics training, I have gained extensive experience analyzing data, and formulating and estimating econometric models. Additionally, as a research assistant for academic faculty, I gained experience working with confidential data at the Statistics Canada Research and Data Center. I am familiar with protocols for handling and storing confidential data, and ensuring that analysis output does not reveal sensitive information.

My research will be done under the supervision of Dr. Stephan Heblich and Dr. Jeffrey Sun, respectively Professor and Assistant Professor within the Department of Economics at the University of Toronto.

Research Question

Extreme weather events can have profound effects on transportation infrastructure, particularly rail. It is well-established that extreme weather events can cause significant damage to infrastructure and lead to transportation disruptions and delays.

However, what is less clear are the broader economic costs of disruption risk, and its implications for the spatial distribution of economic activity and optimal investments. Rail is essential for the movement of commodities. The cost of transporting goods has consequences for decisions made by firms and workers about where to concentrate economic activity, and for overall consumer welfare.

Firms and policymakers should make choices about infrastructure allocation with disruption risk in mind. Several options are available: the level of infrastructure along a segment can be maintained or improved; investments can be made to protect against environmental disruptions, such as elevated tracks to protect against flooding; or, a segment could be re-routed altogether to avoid the risk. However, the returns to the various types of investment are heterogeneous across space, and will be jointly determined by a particular segment's disruption risk and economic importance.

My research will aim to address the following question: *How should railroads and policymakers respond to environmental disruption risk through the spatial allocation of investments in rail infrastructure?*

Preliminary Analysis

In the appendix, I have included the results of some preliminary analysis which highlight the need for access to CWS data. Identifying environmental-related railroad disruptions is a non-trivial task, and could be identified/defined in different ways. I start by looking at the inundation of railroad tracks from major flooding events. Tables 1 and 2 illustrate that the flooding of railroads is a relatively rare event, with typically only a few major floods per year, affecting only a small proportion of overall railroad track.

The second way I identify disruptions is looking at train derailments caused by environmental events. Figure 1 shows that train derailments are a relatively rare event, with at most only a few dozen occurring within a single year. It also shows that such derailments have increased over time, especially since the early 1990s. Figure 2 shows that such derailments are not geographically concentrated — they occur across most of the continental US.

Additionally, I use the public CWS to run some exploratory regressions about how disruptions affect rail flows. Table 3 reports the results. I find that reported flows and the flooding of railroads have a statistically significant correlation. These preliminary results suggest that flooding, and possibly other types of environmental disruptions, do affect rail movements, and this shows up in the waybill data.

From these preliminary analyses, I have demonstrated that environmental-related railroad disruptions are a geographically widespread but relatively rare phenomenon, with an impact on railroad flows that can be shown statistically.

Research Plan

To address the question at hand, I intend to embed a transportation flows model within an economic geography model, and use the CCWS to help quantify this model and estimate disruption costs. I will then use the quantified model to perform economic analysis, shedding light on how rail infrastructure should be optimally allocated in light of disruption risk along each rail link, and associated construction costs.

Transportation Flows Model Component: The transportation flows model will capture the US Class I rail network. The transportation cost of using a particular segment of rail will be determined by the underlying level of infrastructure investment, disruption risk, and total usage of the link (congestion). The transportation cost of shipping goods between a pair of locations will be a function of the underlying transportation network, based on the cost of the links used along possible routes. Railroad companies/carriers set transportation charges to maximize profits. This model will rely on recent theoretical advancements in the transportation economics literature¹ to obtain tractability and estimatability.

Economic Geography Model Component: Nesting the transportation flows model will be a model of economic geography.² This model will capture the fact that different goods are produced in different locations, and since consumers demand a variety of goods, this creates a need for trade. The cost of trading goods between locations will be determined by the transportation flows model. This will be a general equilibrium setting, encapsulating the idea that spatial linkages and the distribution of economic activity will adapt in response to changes to the trade cost structure. I define a location as a core-based statistical area (CBSA), which is a geographically smaller unit of analysis than a BEA, warranting the granularity of the CCWS.

Model Estimation: To quantify the model, I will require a sense of the spatial economic linkages between locations in the US, as well as the rail transportation costs, and contribution of disruption risk to these long-run average costs. Based on the model, I derive a structural relationship between the value of traded goods between location pairs, and the bilateral transportation cost — commonly referred to as a gravity equation.³ I can infer how disruptions to rail affect the transportation cost by running a gravity regression that includes information on disruptions as a regressor. One example of a disruption to analyze is the Great Flood of 1993, which inundated over 500 miles of track and caused millions of dollars of damage to rail infrastructure in the midwestern US. However, I anticipate there will be many more disruptions to analyze over the 30 year window.

To effectively assess the impact of temporary disruptions on trade flows, it is crucial to access temporally and geographically granular data. Disruptions are typically localized, short-lived, and infrequent, necessitating information on the flow of commodities at fine geographic and temporal levels over extended periods. Although the Commodity Flows Survey (CFS) provides data on commodity shipments between US locations, it is collected only every five years, which

¹ See Allen, Treb and Arkolakis, Costas (2022). The Welfare Effects of Transportation Infrastructure Improvements. *The Review of Economic Studies*, Volume 89, Issue 6, Pages 2911–2957, <https://doi.org/10.1093/restud/rdac001>.

² See, for instance, Redding, Stephen (2016). Goods trade, factor mobility and welfare. *Journal of International Economics*, Volume 101, Pages 148–167, <https://doi.org/10.1016/j.jinteco.2016.04.003>.

³ See Chaney, Thomas (2018). The Gravity Equation in International Trade: An Explanation. *Journal of Political Economy*, Volume 126, Issue 1, <https://doi.org/10.1086/694292>.

is insufficient for capturing the effects of temporary disruptions. Instead, data from the confidential Carload Waybill Sample (CCWS), which offers detailed insights into commodities shipped by rail between origin-destination pairs at small time intervals, is required to adequately monitor and analyze the impacts of rare events like flooding on trade flows.

I require information on all origin-destination pairs, not just a restriction to disruption-prone areas, to exploit broad spatial and temporal variation to identify disruption effects. It is also crucial that I have broad data on rail shipments covering the entirety of the US Class I rail network since I not only aim to identify the direct effect of disruptions on affected segments, but also the indirect network and broader general equilibrium effects. For example, if a particular route is temporarily disrupted due to flooding, this may lead to re-routing which increases flows elsewhere. It may also lead to an economic reallocation as changes in the structure of transportation costs lead to supply and demand responses.

I am also interested in the flows across various routes in addition to the overall origin-destination pairs. Knowing how frequent various routes and segments are utilized would allow me to identify the extent to which a disruption of a particular rail segment will impact the overall flow of commodities between two locations. It will also be useful in model calibration.

Addressing Data Limitations: I acknowledge that there are significant limitations to the CCWS data that will have to be addressed in my research design. First, reported waybills alone do not allow for the calculation of trade flows. To remedy this, I intend to infer commodity prices from the CFS data, and scale up the flows from the waybill data to match the CFS reported values for years in which the survey was conducted. Second, the CCWS contains no direct information on rail disruptions, so I will have to collect this information from other sources, such as newspaper archives and datasets of geocoded environmental disasters. Third, the CCWS may oversample frequent events and undersample rare events including the types of disruptions of interest. However, using stratification may help remedy this problem. Fourth, the waybill date does not necessarily indicate the date of travel, and associated rates are often set ahead of time, before a disruption is realized. While prices may be fixed in the short run, this does not preclude the possibility of a long-run response. Hence, I intend to analyze the leads and lags surrounding the event date. Fifth, the CCWS does not actually attribute routes to waybills. However, in some instances, a route can be probabilistically inferred using the procedure developed by Degiovanni and Yang (2023)⁴ specifically for the CCWS.

Model Application: To answer my research question, I intend to use the model to perform cost-benefit analyses of various infrastructure allocation strategies. I will consider general improvements to a rail link (e.g. adding an additional set of tracks), as well as improvements aimed at strengthening a link against disruption risk. The model will be important to estimating policy benefits, as it captures not only the direct impact of improved infrastructure, but also broader network and economic effects that cannot be captured by more simplistic approaches. The benefit of a policy will depend on: the extent of disruption risk, how central the link is in the transportation network, and the extent to which the link services economically important spatial linkages. The costs will be considered to be the approximate construction and maintenance costs associated with link improvements.

⁴ Degiovanni, Pedro and Yang, Ron (2023). Economics of Scale and Scope in Railroading. Working Paper.

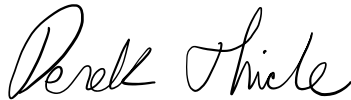
Confidentiality

I will comply with the confidentiality terms laid out in 49 CFR § 1244.9. At the Department of Economics, I will have access to an IT solution that will enable for secure storage of the confidential data. The planned research will measure aggregate effects which may be differentiated by network characteristics, but will contain no firm- or commodity-level differentiation, nor any reported information on specific prices or revenues for a particular firm, station, or railroad. Prior to the release of results, I will submit all of my preliminary results to the STB Office of Economics to ensure they comply with confidentiality standards.

Request for CWS

I request access to CWS data from 1990 to 2020. Access will be provided for myself (Derek Thiele). The CWS will be used solely for academic research. The results of confidentiality-compliant analyses may be published online and in an economics journal.

Sincerely,

A handwritten signature in black ink, reading "Derek Thiele". The signature is written in a cursive, flowing style.

Derek Thiele

Economics PhD Student
Department of Economics
University of Toronto
150 St George St, Toronto, ON M5S 3G7

Appendix

<i>Table 1: Number of Unique Flooding Events Affecting US Railroads, By Year</i>			
Year	Number of Unique Flooding Events	Year	Number of Unique Flooding Events
2001	2	2010	4
2002	7	2011	3
2003	11	2012	2
2004	10	2013	3
2005	4	2014	1
2006	5	2015	2
2007	3	2016	4
2008	5	2017	3
2009	4	2018	4
The flooding extent of major flooding events from 2000-2018 from the Global Flood Database intersected with railroads as defined by the NTAD North American Rail Network Lines dataset.			

<i>Table 2: Total Amount of Railroad Track Flooded (km), By Year</i>			
Year	Total Flooded Track (km)	Year	Total Flooded Track (km)
2001	4	2010	181.09
2002	64.54	2011	69.05
2003	265.4	2012	3.76
2004	145.29	2013	47.37
2005	49.56	2014	10
2006	71.78	2015	44.24
2007	14.34	2016	9.83
2008	26.23	2017	36.23
2009	63.83	2018	31.80
The flooding extent of major flooding events from 2000-2018 from the Global Flood Database intersected with railroads as defined by the NTAD North American Rail Network Lines dataset.			

<i>Table 3: Regression Results — Impact of Flooding Event on Rail Flows</i>			
	(1) Aggregate Flows (tons)	(2) Aggregate Flows (tons)	(3) Aggregate Flows (tons)
Flooding of least-cost railroad route (indicator)	-1.564*** (-40.37)		-0.159*** (-4.90)
Period in which a flood occurred (indicator)		-0.393*** (-7.68)	
Fixed effects	N	N	Origin-destination + time effects
Number of observations	6,775,816	6,775,816	2,415,815
Unit of observation is an origin-destination-year-month. Origin and destination locations are BEA areas. Aggregate flows are taken as the sum of the expanded tons for all waybills associated with the origin-destination pair for the particular time period. The regressions are all poisson, with the outcome variable being aggregate flows. For models (1) and (3), the regressor/independent variable is an indicator for whether a flood occurred or not on the least-cost path (on the freight I rail network) between the origin and destination nodes. For model (2), the regressor is an indicator for whether a flood occurred anywhere on the network during the time period. Model (3) includes fixed effects for the origin-destination pair, as well as the time period. Data sources: NTAD North American Rail Network Lines, Global Flood Database, Public Use Carload Waybill Sample (2000-2018) <i>t</i> statistics in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$			

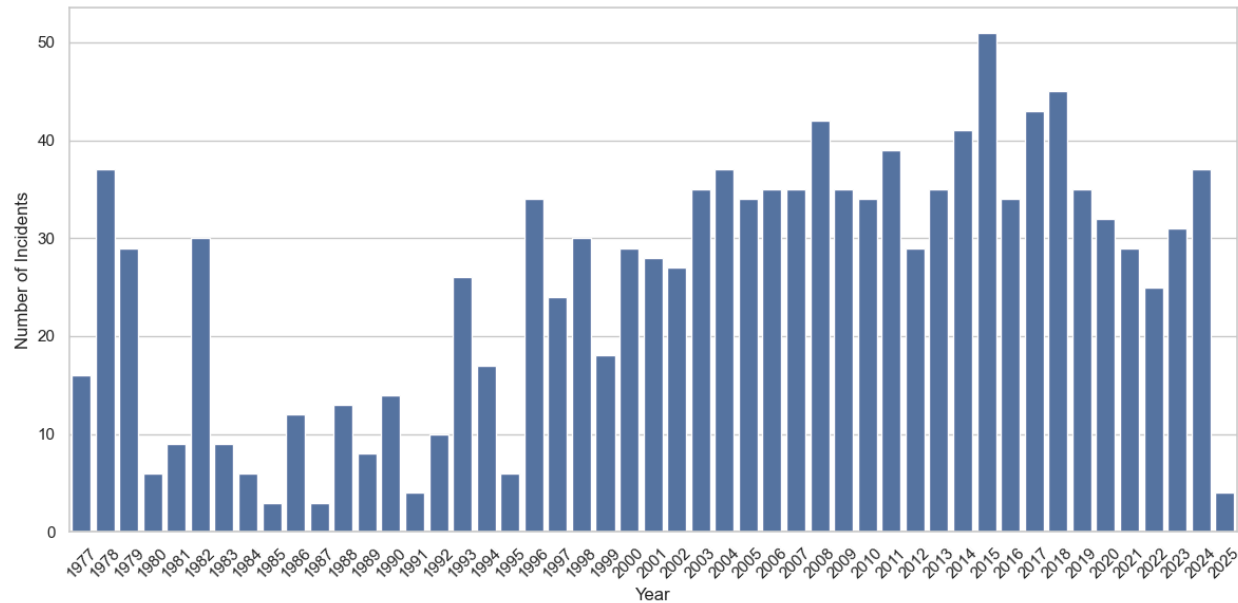


Figure 1: Number of Environmental-Related Train Derailments Over Time, 1977-2025

Source: Rail Equipment Accident/Incident Data (Form 54) from the Federal Railroad Administration



Figure 2: Map of Environmental-Related Train Derailments, 1977-2025
 Source: Rail Equipment Accident/Incident Data (Form 54) from the Federal Railroad Administration

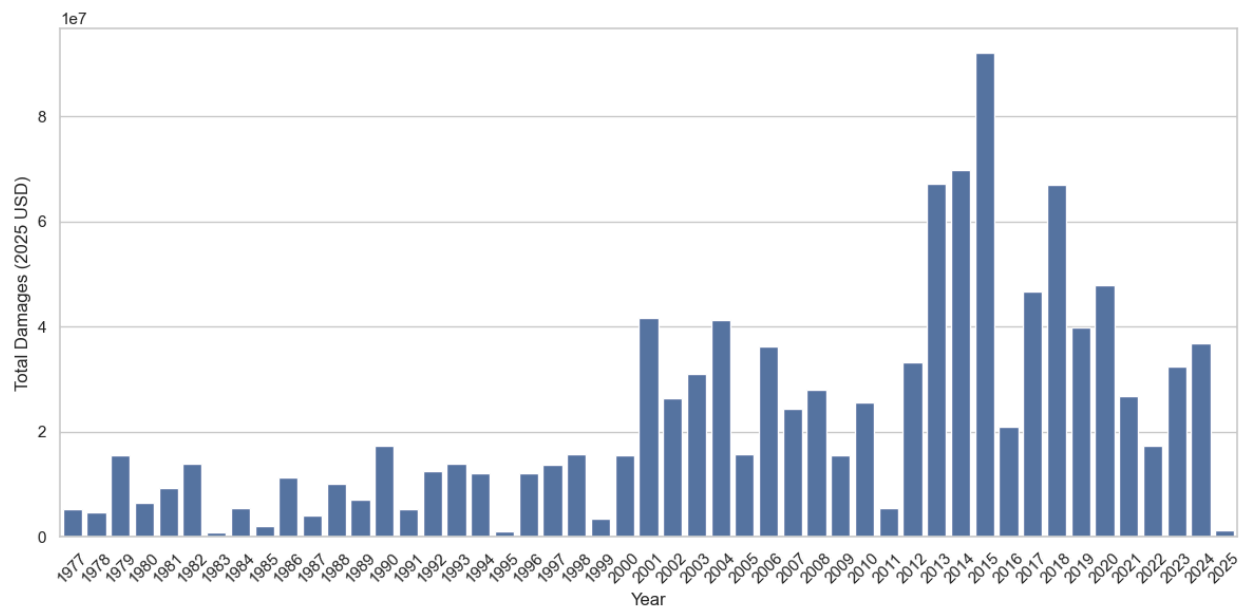


Figure 3: Total Damages (2025 USD) Associated With Environmental-Induced Derailments
 Source: Rail Equipment Accident/Incident Data (Form 54) from the Federal Railroad Administration