

Confidential Waybill Sample Application

Ron Yang

PhD Candidate, Business Economics

Department of Economics

Harvard University | Harvard Business School

Pedro Degiovanni

PhD Candidate, Economics

Department of Economics

Harvard University

Project Objectives

- Our general research question is: In a transportation network with multiple modes (i.e. rail, truck, etc.), how do the effects of infrastructure improvements spill across modes and lanes? How does market power among shippers and carriers affect these spillovers?
- Applied questions which we seek to answer include:
 - How can we measure and decrease the cost of transporting products from port of entry to final consumers?
 - Should infrastructure investments target high or low concentration modes?
 - Do the benefits from local infrastructure decisions come from increased efficiency, or "business stealing" from other localities?
- The expected outcomes of our research project will include:
 - One or more papers for submission & publication in a top Economics journal

- Presentations at academic economics conferences and seminar series
- This research project will contribute to the economics literature in several ways:
 - We will contribute to the international trade literature by measuring the contribution of within-country infrastructure to international trade costs.
 - We will contribute to the spatial economics literature by investigating the varying role of market power across transportation modes.
 - We will contribute to the public economics literature by evaluating the spillover effects of potential infrastructure projects.

Scope of Work

Our research will have three components:

Reduced Form & Data Aggregation

We will begin by constructing a merged dataset which combines three main sources: the confidential Waybill dataset, the Census Commodity Flow Survey, and trucking freight prices from a private firm.

We will analyze general patterns observed in this merged dataset. Among others, we will study whether there is a correlation between changes in market concentration and prices; whether mergers have contributed to the reduction in rates observed during the last 20 years; whether prices vary systematically depending on the direction of travel (i.e. West-East vs. East-West); whether notable infrastructure projects, such as the Alameda Corridor and the Chicago Region Environmental and Transportation Efficiency Program have been associated with increases in volume or reductions of prices; how changes in the relative competitiveness between trucks and rails, given by shocks to diesel prices, are reflected in market outcomes, etc.

Naïve Structural Estimation

As a first step towards the full structural estimation, we will use a naïve model which assumes that pricing decisions along each lane are independent.

We will estimate the preferences of shippers using data on the market shares and prices of transportation along different modes. Using these preference estimates, with a model of how rail firms compete in price setting, we can infer operating costs along each lane. This allows us to decompose variation in rail prices into several sources: (1) differences in shipper demand for rail, (2) differences in concentration of rail within lane, and (3) differences in operating costs.

For these two first components, we need the confidential Waybill data for two reasons: first, the Public Waybill Sample does not identify the origin and destination of 57% of the observations, rendering them unusable for our purposes.

Full Structural Estimation

The most complicated component of our project will entail a full structural model where shippers can make both mode and route decisions, and rail firms can set prices given their entire network of rail. In this model, for example, a shipper can respond to an increase in the price of rail leaving Los Angeles by choosing to ship through another port; alternatively, an infrastructure improvement such as CREATE, can redirect flows from land-bridging and sea shipping towards overland transportation.

In a first paper, we will consider the strategic pricing and quantity decisions of railroads and trucks, paying attention to their exercise of market power across the present transportation network. We will use the results from the Naïve Structural Estimation to inform the features that our network will need. At a minimum, we will take into account the location and quality of railways, roads, ports and navigable waterways, which will influence the above decisions.

Given this model, we will estimate the economic parameters governing shipper route / mode decisions and rail firm pricing decisions. Armed with these parameter estimates, we may simulate a variety of counterfactual scenarios. For example, if operating costs for rail fell along a particular lane, and we allowed rail and trucking firms to re-optimize across the country, how would prices, quantities, profits, and consumer welfare look differently?

In follow-up projects, we will make the model more flexible, and study how the locations of population, production and consumption depend on the transportation network, and we will allow the network to vary endogenously.

To econometrically estimate this model, we need to observe origins, destinations, the railroads involved, the commodity, the shipment size, the distance, the number of interchanges (including the railroads involved as well as the distances and revenues to each), the size of the shipment, whether the railroad cars are owned by the railroad or the shipper and other variables that may influence rates or differentiate between small and large consumers. Without these variables, we would be unable to identify the model parameters and / or compute counterfactuals. These essential variables for our research can only be provided by the confidential Waybill dataset.

Detailed Data Requested

We request all data from 1996 to the present.

Our project is fundamentally about the spillovers across diverse regions and markets across the United States. As a result, we could not answer our economic question of interest with a dataset which is limited in geography or commodity.

Data Management Plan

We will adhere to the full confidentiality and data privacy standards of the waybill data. We have access to a secure Research Computing Grid at Harvard Business School which is equipped to store, handle, and protect Confidential and/or Personally-Identifiable level data. We also have access to legal resources at the Harvard Business School Library for signing and/or enforcing our confidentiality contract.

The waybill data will be used only for the academic research project described above, and only by the two principal investigators: Ron Yang and Pedro Degiovanni.