A Tractable Model of Transshipment*

Jingting Fan† Wenlan Luo‡

July 2020

Abstract

This note presents a simple model of transshipment, under which bilateral transport costs are an explicit and differentiable function of route and transshipment costs. We characterize bilateral transport costs and their elasticities with respect to transshipment costs.

Movements of goods and people usually combine multiple modes of transportation. Most merchandise exports are first shipped via ground transport to ports and then loaded to cargo ships. Often times, they are initially loaded to small ships and then, together with goods from other small ships, transferred to large ships in a regional hub for the long haul. Within a country, logistic companies usually connect flights with truckers in delivering packages. For passengers whose origin and destination are not major airline hubs, transfers between flights, railway, and cars are also all but unavoidable.

Such transfers incur significant time and monetary costs to both business and passengers. To reduce these costs, economic activities gravitate towards transshipment points. The clustering of activities around transshipment points has given rise to many great cities historically. For example, in the U.S., many of the 18th century portage sites along the east coast fall line grew into major cities (Bleakley and Lin, 2012); Chinese city Zhengzhou, which sits on the intersections of multiple major railways and national roads, grew from a small county to a city with 10 million population in less than 60 years.

This note presents a tractable model of transshipment. The model builds on the 'round-about' transport model of Allen and Arkolakis (2019), which we extend to allow for multiple coexisting modes and costly transfer between modes.[1] The advantage of the model is in its tractability: bilateral travel costs are an analytical and differentiable function of network parameters; moreover, the elasticity of bilateral costs with respect to the cost of switching modes at any city can also be derived. This makes possible analytical characterizations of how improvements in efficiency of transfer in a city affect shipment flows and the aggregate welfare.

*We thank Costas Arkolakis as our discussant at SUFE-IESR Trade Workshop for inspiring this note.

†Pennsylvania State University, PA, USA; jxf524@psu.edu

‡Tsinghua University, Beijing, China; luowenlan@gmail.com

[1] Fan, Lu and Luo (2019) extends Allen and Arkolakis (2019) to incorporate two co-existing modes of transportation. This note shows that the extension can be readily modified to allow for transfer costs.

In the rest of this note, we start by briefly describing the setup in Fan, Lu and Luo (2019). We then add transshipment costs to the model.

Two-mode transport without transfer costs. Consider an economy with N locations, connected by two modes of transportation. We label the two modes as H and L, standing for high- and low-capacity, respectively. These two modes can be entirely different categories, such as railway and truck, or different types of vehicles within a category, such as small and large cargo ships. Let t_{ij}^H be the ad-valorem cost of going from i to j along edge $i \to j$ on network H and t_{ij}^L the corresponding cost on network L. If i and j are not directly connected on a network, the edge cost is infinity.

We assume that the drivers or travelers from o to d visit sequentially nodes on the networks before arriving at the destination d. With detours and repeated edges on a path allowed, travelers have infinitely many possible paths to choose from. Travelers choose the one with the lowest cost, subject to their idiosyncratic preference over routes, given by a Frechet distribution with dispersion parameter θ. Modeling idiosyncratic preference with Frechet shock enables analytical characterization of the trade cost as a function of the transport network.[2] Specifically, slightly abusing notations, define two $N \times N$ matrices, H and L, such that the i, jth element of these matrices, $[H]_{ij}$, and $[L]_{ij}$ are given by the following:

$$[H]_{ij} = (t_{ij}^H)^{-\theta}$$
$$[L]_{ij} = (t_{ij}^L)^{-\theta}.$$

Define X to be the following

$$X = (I - H - L)^{-1}.$$

Fan et al. (2019) shows that the result in Allen and Arkolakis (2019) extends to this setting: after taking into account travelers' idiosyncratic preference, the average effective cost between o and d, denoted by τ_{od} is

$$\tau_{od} = c \cdot [X]_{od}^{-\theta},$$

where $[X]_{od}$ is the o, dth entry of X and c is the constant $\Gamma(\frac{\theta-1}{\theta})$. This average is taken across all possible paths from o to d with arbitrarily many edges. It allows a traveler to combine different segments of H and L to form a path to their own taste. For example, let H be air transport and L be car transport; a traveler from State College, PA to Athens, GA, can first take a flight to Atlanta, transferring in D.C., and then drive from Atlanta to Athens.

Introducing transshipment. Despite the flexibility in combining multiple modes of transport, the above routing model overlooks that mode transfers are costly for both goods and people.

[2] Notice because of this idiosyncratic preference, a small fraction of drivers might end up taking a detour; an even smaller fraction might take a roundabout.

Below we add such cost to the model.

We introduce two additional matrices, $\tilde{H}$ and $\tilde{L}$. The i,jth entry of $\tilde{H}$ is $[\tilde{H}]_{i,j} \equiv (\tilde{t}_{ij}^H)^{-\theta}$, where $\tilde{t}_{ij}^H$ is the cost of going along the edge $i \to j$ of network H *when a change of mode is needed in* i. For example, if the traveler arrives in i via network L but continues from i to j via H, then the cost along the edge $i \to j$ is $\tilde{t}_{ij}^H$ instead of t_{ij}^H, the cost for when no mode change is needed. $\tilde{t}_{ij}^H - t_{ij}^H$ is therefore the transshipment cost. Similarly, $\tilde{L}$ is defined such that $[\tilde{L}]_{ij} \equiv (\tilde{t}_{ij}^L)^{-\theta}$, with $(\tilde{t}_{ij}^L)$ being the cost of moving along the edge $i \to j$ when a mode switch is needed in i.

The following proposition characterizes the dependence of bilateral trade costs on transshipment costs.

Proposition 1. *Define* $A \equiv \begin{bmatrix} H & \tilde{L} \\ \tilde{H} & L \end{bmatrix}$, $B \equiv (I - A)^{-1} \equiv \begin{bmatrix} B^1 & B^2 \\ B^3 & B^4 \end{bmatrix}$,[3] *then*

a) The trade cost between o and d for $o \neq d$ *is given by:*

$$\tau_{od} = c \cdot \left(\left[[H, L](I - A)^{-1} \begin{bmatrix} I \\ I \end{bmatrix} \right]_{od} \right)^{-\frac{1}{\theta}}$$

where c is the constant $\Gamma(\frac{\theta-1}{\theta})$ *and* Γ *is the gamma function.*

b) The elasticity of trade cost between o and d w.r.t. the transshipment cost from i to j for mode H is

$$\frac{\partial \log(\tau_{od})}{\partial \log(\tilde{t}_{ij}^H)} = \frac{\sum_k (h_{ok} b_{ki}^2 + l_{ik} b_{ki}^4) \times \tilde{h}_{ij} \times (b_{jd}^1 + b_{jd}^2)}{\sum_k (h_{ok} b_{kd}^1 + l_{ok} b_{kd}^3) + \sum_k (h_{ok} b_{kd}^2 + l_{ik} b_{kd}^4)},$$

and similarly,

$$\frac{\partial \log(\tau_{od})}{\partial \log(\tilde{t}_{ij}^L)} = \frac{\sum_k (h_{ok} b_{ki}^1 + l_{ik} b_{ki}^3) \times \tilde{l}_{ij} \times (b_{jd}^3 + b_{jd}^4)}{\sum_k (h_{ok} b_{kd}^1 + l_{ok} b_{kd}^3) + \sum_k (h_{ok} b_{kd}^2 + l_{ik} b_{kd}^4)},$$

where $y_{ok} = [Y]_{od}$, $\forall\, Y \in \{H, L, \tilde{H}, \tilde{L}, B^1, B^2, B^3, B^4\}$

Proof. See the appendix. □

Three remarks on this proposition are in order. The first is on the geometric interpretations of the expressions. In part a) of the proposition,

$$[H, L](I - A)^{-1} = [HB^1 + LB^3, HB^2 + LB^4].$$

The first block of this, $HB^1 + LB^3$, is the sum of (the $-\theta$ power of) travel costs across all paths from o to d that ends in d with mode H, which consists of those starting from o with mode H ($H \times B^1$) and those starting from o with mode L ($L \times B^3$). The second block is the sum across all

[3] A sufficient condition for $I - A$ to be invertible is that the spectral radius of L is less than one (Allen and Arkolakis, 2019). This will be case if the road network adjacency matrix is sparse and the routing elasticity θ is large.

paths ending in d in mode L. $\Big[[H,L](I-A)^{-1}\begin{bmatrix} I \\ I \end{bmatrix}\Big]_{od}$ is thus the sum across all paths from o to d.

The elasticity $\frac{\partial \log(\tau_{od})}{\partial \log(\tilde{t}_{ij}^{H})}$ has an intuitive meaning, too. The denominator is simply the sum of the $-\theta$ power of travel costs across all paths from o to d, of which there are four types based on whether a path leaves o in H or L and whether it arrives in d in H or L. The numerator is the sum across all paths that arrive in i in L and transfer to H before the next stop j, and then from j using any mode to finally arrive at d, i.e., all routes that uses the transfer $L \to i \to H \to j$. As in Allen and Arkolakis (2019), this ratio corresponds to the fraction of trade costs spent on these routes. Under the limit case of $\theta \to \infty$, this elasticity converges to the fraction of all trips from o to d that goes through the $L \to i \to H \to j$ transfer.

Second, while in practice an improvement in transfer efficiency in location i likely affects the goods and passengers transferring through i to all nearby locations, part b) of the proposition focuses only on efficiency improvements that are 'directional', i.e., those that are directed to only shipments to j from i. We note that this restriction is without loss of generality because any comprehensive change can be written as a collection of 'directional' changes. For example, if we assume that $\tilde{t}_{ij}^{H} = t_{ij}^{H}\delta_i$, where δ_i is the efficiency of transshipment at location i, then we can build on part b) of the proposition to derive $\frac{\partial \log \tau_{od}}{\partial \log \delta_i}$.

Third, this proposition offers a direct mapping from the network fundamental, given by H, $\tilde{H}$, L, and $\tilde{L}$, to bilateral trade costs. Using the tools developed by Allen and Arkolakis (2019), this module could be easily extended to allow for congestion along an edge $i \to j$, or congestion within a transshipment point i. This module can then be embedded in trade and economic geographic frameworks to analyze how shipment flows and the aggregate welfare react to changes in transfer costs. As an example, in the final part of this note, we illustrate how we can use the proposition to derive analytically the welfare gains from an improvement in transshipment efficiency.

The welfare effects of transshipment efficiency. The type of policy evaluations we are interested in is a change in transshipment efficiency in i that simultaneously reduces $\tilde{t}_{ij}^{H}$ for all destination $j \neq i$. For simplicity, we focus on a case where the first welfare theorem holds as in Fan et al. (2019)

Let the aggregate welfare be W. The trade or economic geography block of the model usually implicitly determine W as a function of trade cost matrix τ. The first order effect of an improvement in efficiency in transferring from L to H in location o is thus:

$$\begin{aligned}\Delta W &= \sum_{o,d} \frac{\partial \log(W(\tau))}{\partial \log(\tau_{od})} \Delta \log(\tau_{od}) \\ &= \sum_{o,d} \Big(\frac{\partial \log(W(\tau))}{\partial \log(\tau_{od})} \cdot \sum_{j \neq i} \big(\frac{\partial \log(\tau_{od})}{\partial \log(\tilde{t}_{ij}^{H})} \Delta \log(\tilde{t}_{ij}^{H}) \big) \Big),\end{aligned}$$

in which $\frac{\partial \log(\tau_{od})}{\partial \log(\tilde{t}^H_{ij})}$ is characterized in Proposition 1, and $\frac{\partial \log(W(\tau))}{\partial \log(\tau_{od})}$ is associated with some observable variables in the trade model (e.g., in an efficient model, this is associated with the value of trade flows from o to d).

Using tools in Allen and Arkolakis (2019), it is also possible to derive the response in shipment flows through rerouting in response to such a change transshipment efficiency.

References

Allen, Treb and Costas Arkolakis, "The Welfare Effects of Transportation Infrastructure Improvements," Working Paper 25487, National Bureau of Economic Research January 2019.

Bleakley, Hoyt and Jeffrey Lin, "Portage and path dependence," *The quarterly journal of economics*, 2012, *127* (2), 587–644.

Fan, Jingting, Yi Lu, and Wenlan Luo, "Valuing Domestic Transport Infrastructure: A View from the Route Choice of Exporters," *Available at SSRN 3528448*, 2019.

A Proof of the Proposition

Proof. Define $X^1 = \begin{bmatrix} H & L \end{bmatrix} \equiv \begin{bmatrix} X^1_H & X^1_L \end{bmatrix}$. Then $[X^1_H]_{od}$ is the $-\theta$ power of the cost of going from o to d in exactly one step, arriving in d via the H mode. Similarly, X^1_L is the collection of power of the cost of going from o to d in exactly one step, arriving in d via the L mode.

Suppose that $X^n \equiv \begin{bmatrix} X^n_H & X^n_L \end{bmatrix}$ to be such that $[X^n_H]_{od}$ and $[X^n_L]_{od}$ is the sum of the $-\theta$ power of cost of all paths going from o to d in exactly n step, arriving at d via mode H and mode L, respectively. Then the sum across all paths going from o to d in exactly $n+1$ step, arriving in d via mode H is: $X^n_H H + X^n_L \tilde{H}$. The first term in this sum captures the paths arriving in one step away from d via mode H and then continue via H to d; the second component captures the paths arriving in one step away from d in L and then continue through H, thus incurring a transfer cost. Similarly, the collection of all paths arriving at d via mode L is $X^n_H \tilde{L} + X^n_L \tilde{L}$.

We can therefore express $\mathbb{X}^n$ recursively as:

$$X^{n+1} = \begin{bmatrix} X^n_H & X^n_L \end{bmatrix} \cdot \begin{bmatrix} H & \tilde{L} \\ \tilde{H} & \tilde{L} \end{bmatrix} = X^n \cdot \begin{bmatrix} H & \tilde{L} \\ \tilde{H} & \tilde{L} \end{bmatrix}$$

The sum across all paths with any length $n \geq 1$ is:

$$\begin{aligned} X &\equiv \sum_{n=1}^{\infty} X^n \cdot \begin{bmatrix} I \\ I \end{bmatrix} \\ &= [H, L](I - A)^{-1} \begin{bmatrix} I \\ I \end{bmatrix}, \text{where } A = \begin{bmatrix} H & \tilde{L} \\ \tilde{H} & L \end{bmatrix}. \end{aligned}$$

This proves part a) of the proposition.

To prove part b) of the proposition, denote $B \equiv (I-A)^{-1}$. Then we have for the expanded routing matrix A

$$
\begin{aligned}
\frac{\partial A}{\partial \log \tilde{t}^H_{ij}} &= -\theta (\tilde{t}^H_{ij})^{-\theta} \begin{bmatrix} 0 & 0 \\ E_{ij} & 0 \end{bmatrix} \\
\Rightarrow \frac{\partial B}{\partial \log \tilde{t}^H_{ij}} &= \frac{\partial (I-A)^{-1}}{\partial \log \tilde{t}^H_{ij}} = -(I-A)^{-1} \frac{\partial (I-A)}{\partial \log \tilde{t}^H_{ij}} (I-A)^{-1} = -\theta (\tilde{t}^H_{ij})^{-\theta} B \begin{bmatrix} 0 & 0 \\ E_{ij} & 0 \end{bmatrix} B \\
\Rightarrow \frac{\partial X}{\partial \log \tilde{t}^H_{ij}} &= -\theta (\tilde{t}^H_{ij})^{-\theta} [H, L] B \begin{bmatrix} 0 & 0 \\ E_{ij} & 0 \end{bmatrix} B \begin{bmatrix} I \\ I \end{bmatrix}.
\end{aligned}
$$

Now consider

$$
\begin{aligned}
B \begin{bmatrix} 0 & 0 \\ E_{ij} & 0 \end{bmatrix} B = \begin{bmatrix} B^1 & B^2 \\ B^3 & B^4 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ E_{ij} & 0 \end{bmatrix} \begin{bmatrix} B^1 & B^2 \\ B^3 & B^4 \end{bmatrix} \\
= \begin{bmatrix} B^2 E_{ij} B^1 & B^2 E_{ij} B^2 \\ B^4 E_{ij} B^1 & B^4 E_{ij} B^2 \end{bmatrix},
\end{aligned}
$$

and for the o, dth entry of X,

$$
x_{od} = \sum_k (h_{ok} b^1_{kd} + l_{ok} b^3_{kd}) + \sum_k (h_{ok} b^2_{kd} + l_{ok} b^4_{kd}).
$$

Therefore,

$$
\begin{aligned}
\Big[\frac{\partial X}{\partial \log \tilde{t}^H_{ij}} \Big]_{od} &= \Big[-\theta (\tilde{t}^H_{ij})^{-\theta} [H, L] \begin{bmatrix} B^2 E_{ij} B^1 & B^2 E_{ij} B^2 \\ B^4 E_{ij} B^1 & B^4 E_{ij} B^2 \end{bmatrix} \begin{bmatrix} I \\ I \end{bmatrix} \Big]_{od} \\
\Rightarrow \frac{\partial \log x_{od}}{\partial \log \tilde{t}^H_{ij}} &= -\theta (\tilde{t}^H_{ij})^{-\theta} \frac{\sum_k (h_{ok} b^2_{ki} + l_{ik} b^4_{ki}) \times (b^1_{jd} + b^2_{jd})}{\sum_k (h_{ok} b^1_{kd} + l_{ok} b^3_{kd}) + \sum_k (h_{ok} b^2_{kd} + l_{ok} b^4_{kd})} \\
&= -\theta \frac{\sum_k (h_{ok} b^2_{ki} + l_{ik} b^4_{ki}) \times \tilde{h}_{ij} \times (b^1_{jd} + b^2_{jd})}{\sum_k (h_{ok} b^1_{kd} + l_{ok} b^3_{kd}) + \sum_k (h_{ok} b^2_{kd} + l_{ok} b^4_{kd})}.
\end{aligned}
$$

And trade cost $\tau_{od} = c \cdot x_{od}^{-1/\theta}$. Therefore

$$
\frac{\partial \log \tau_{od}}{\partial \log \tilde{t}^H_{ij}} = -\frac{1}{\theta} \frac{\partial \log x_{od}}{\partial \log \tilde{t}^H_{ij}} = \frac{\sum_k (h_{ok} b^2_{ki} + l_{ik} b^4_{ki}) \times \tilde{h}_{ij} \times (b^1_{jd} + b^2_{jd})}{\sum_k (h_{ok} b^1_{kd} + l_{ok} b^3_{kd}) + \sum_k (h_{ok} b^2_{kd} + l_{ok} b^4_{kd})}.
$$

We can similarly derive $\frac{\partial \log(\tau_{od})}{\partial \log(\tilde{t}^L_{ij})}$. □