Essay 1: Critical Reading of Donaldson and Hornbeck (2016)

Etienne Comperat

Part I: Executive Summary

A) Introduction

In "Railroads and American Economic Growth: a "Market Access" Approach", Donaldson and Hornbeck revisit the seminal question posed by Robert Fogel in "Railroads and American Economic Growth" (1964): what was the economic contribution of railroads to 19th-century U.S. development? Departing from Fogel's "social savings" approach, the authors instead adopt a "market access" framework derived from spatial general equilibrium trade theory. Their innovation lies in quantifying how changes in trade costs, induced by railroad expansion between 1870 and 1890, affected county-level agricultural land values via shifts in general equilibrium market access. Thus, they draw on Eaton and Kortum (2002) to define market access as a county's access to other counties' markets weighted by population and adjusted for trade costs, capturing a location's trade potential.

To implement their framework, the authors construct a comprehensive GIS-based network of railroads and waterways in the late nineteenth-century United States and compute lowest-cost freight routes between counties. These cost-minimizing routes are then used to derive a theoretically grounded measure of each county's market access and to estimate the aggregate impact of railroad expansion. In other words, they seek to go beyond Fogel's approach to capture both direct and indirect spillover effects of infrastructure improvements and to measure how these improvements affected the American economy. Overall, the paper demonstrates that the expansion of the railroad network between 1870 and 1890 substantially increased counties' market access and that these increases were strongly capitalized into higher agricultural land values. Finally, the authors run a series of counterfactual simulations in which they estimate the effect of removing all railroads from the 1890 U.S. network on the total value of agricultural land. They find that the mitigation effects of extending canals or improving wagon infrastructure would have been limited, confirming the key role of the railroad network in shaping the 19th-century U.S. economic landscape.

B) Research Question and Intuition

Beyond studying the impact of railroad network expansion on the American economy, the paper addresses a core methodological issue in spatial economics and economic history: how can we estimate the aggregate effect of place-based infrastructure investments, like the 19th-century U.S. railroad network, when such interventions generate extensive general equilibrium spillovers beyond directly treated units?

While much of the early historical literature presumed railroads were key to U.S. development, this assumption began to face quantitative scrutiny with Fogel's "social saving" methodology. In Fogel (1964), the author proposed to estimate the "social saving" of railroads by comparing freight costs across modes for selected routes, an approach that would become dominant in the historical literature. By focusing on a counterfactual, the author argued that in the absence of railroads, river and canal transportation could have substituted effectively along most major freight routes. From this, he concludes that the aggregate impact of railroads on the agricultural sector was modest. However, this approach relied on strong assumptions, such as fixed quantities and fixed locations. In that sense, this approach has been criticized for overlooking how infrastructure transforms broader trade patterns, land use, and regional productivity. Indeed, as David (1969) and Lebergott (1966) noted, such methods may miss the broader spatial reorganization and overlook spillovers.

Donaldson and Hornbeck maintain Fogel's focus on the agricultural sector and build on his intuition that the value of agricultural land, as an immobile factor, reflects the costs of getting agricultural goods to market. However, they introduce the concept of market access, a reduced-form expression derived from structural trade models (e.g., Eaton and Kortum, 2002), as a sufficient statistic to capture the total exposure of a county to changes in the transportation network. The logic is that improvements in market access reduce the cost to reach other markets and raise the net revenue from agricultural production, which is then capitalized into land values. Indeed, trade infrastructure affects not only the cost of shipping existing goods but also which markets an area can profitably serve. When a county gains access to more or larger markets at lower cost, its products fetch higher net prices, which raises land demand and value. Hence, by measuring how expansion in the railroad network affected counties' market access, the authors infer both direct and indirect effects of railroads on economic development and welfare.

C) Empirical Challenge: Methodology and Results

A key empirical challenge that arises stems from the spatially diffuse nature of the treatment: railroads may affect all counties, either directly or indirectly, through interlinked trade network-induced changes in accessibility. This raises concerns about identification, measurement error, and attenuation bias if one were to

use only the presence of local railroad as the treatment variable. As a result, the authors need to implement a framework to estimate aggregate treatment effects in an empirical setting with important treatment spillover effects.

To address this issue, Donaldson and Hornbeck construct a detailed geographic information system (GIS)-based network database encompassing all navigable waterways, canals, railroads, and major wagon routes in the contiguous United States in 1870 and 1890.

Freight costs are assigned to each mode of transportation using parameters from Fogel, enabling the computation of minimum-cost freight routes between all county pairs. These data are combined with digitized county-level information on agricultural land values and population from the U.S. Censuses of Agriculture and Population. The logic is that increased access to markets raises the marginal value of output and in a competitive land market, this is capitalized into higher land prices. The key explanatory variable is a county's market access, computed as a weighted sum of other counties' populations, weighted inversely by bilateral trade costs, derived from gravity-based trade models (Eaton and Kortum, 2002). By doing so, the authors aim to capture the ease with which a county can trade with other markets. Estimations are conducted using both OLS and IV strategies, with historical waterway access as an instrument for market access.

The empirical results suggest that increases in market access between 1870 and 1890 are capitalized into substantially higher agricultural land values. The elasticity of land value with respect to market access is estimated to lie between 0.3 and 0.5. Note that these estimates are robust to restricting identification to distant market access, excluding own-county railroad tracks, and varying instrument specifications.

The advantages of this methodology lie in its alignment with general equilibrium trade theory, while the model maintains the framework underlying Fogel's "social saving" approach. Instead of estimating reduced-form effects of local railroad density, which may be endogenous or mismeasured, the authors estimate the aggregate impact of "market access", and therefore each county is influenced by changes elsewhere in the network. Moreover, by calculating lowest-cost county-to-county freight routes, the authors avoid relying on "shortest path" estimates that could be misleading. Overall, this paper constitutes a significant methodological and empirical advancement in the economic history of American infrastructure.

D) Counterfactual Simulations

The other important contribution of the paper lies in its counterfactual analysis. Using the estimated elasticity and the structural relationship between trade costs and market access, the authors simulate the impact of removing the railroad network in 1890. In the baseline scenario, under a fixed population distribution,

eliminating railroads reduces total agricultural land value by 60.2%, equivalent to 3.22% of GNP. This stands in contrast to Fogel's estimate of 2.7%, which considered only direct freight savings along select routes. The intuition is that railroads enabled spatial reallocation of production and expanded market access nationwide, not just along a few high-volume corridors.

To test the robustness of these findings, the authors vary the counterfactual population distribution using historical values from 1830, 1850, and 1870, finding little difference in aggregate losses. They also solve for the counterfactual equilibrium population allocation endogenously, holding total U.S. population fixed. The main results remain unchanged, suggesting that the effects are not driven solely by static location assumptions.

They also examine scenarios with endogenous utility. Holding total population fixed but allowing worker utility to fall in the absence of railroads, land values fall by less, but the total economic loss remains large. These results emphasize that land values may only partially reflect the broader welfare consequences of transportation infrastructure. In both cases, the counterfactual effects on population and welfare reflect additional aggregate losses from removing railroads. Neither the baseline estimates nor Fogel's estimates, based on losses in agricultural land value only, account for.

Finally, the authors explore whether alternative transportation investments could have substituted for railroads. They find that Fogel's proposed canal extensions would have mitigated only 13% of the losses and lowered the cost of wagon transportation only 21%. Even these partial gains would have required enormous investments and would not have matched the spatial coverage or cost-efficiency of railroads. Moreover, in the absence of railroads, waterway congestion could have increased, amplifying losses (Holmes and Schmitz 2001).

E) Contributions

This paper contributes to the literature through both its findings and its approach. In economic history, it provides a major revision of one of the field's canonical estimates, demonstrating that the aggregate effects of railroads were much larger than previously thought. By grounding the analysis in general equilibrium trade theory and using historical data in a spatially explicit framework, the authors address critiques of the social saving methodology (David, 1969; Lebergott, 1966). In spatial economics, the paper operationalizes the concept of market access as a sufficient statistic for regional economic potential, providing empirical validation for theoretical models in the New Economic Geography tradition (Krugman, 1991; Redding and Venables, 2004). The results confirm that trade cost reductions reconfigure spatial equilibria in ways that go far beyond direct freight savings. The paper also contributes to the empirical toolkit for evaluating

infrastructure by showing how to integrate structural trade models with GIS transport data and land value outcomes. This approach is widely applicable to contemporary settings, from highway systems to digital infrastructure.

In sum, Donaldson and Hornbeck (2016) offer a foundational analysis of the aggregate effects of transportation infrastructure, by proposing a methodology for empirical spatial analysis, in both historical and modern contexts.

Part II: Comparative Approach

In this section, we discuss how the findings and methodology of Donaldson and Hornbeck (2016) relate to the broader literature in urban, regional, and spatial economics. The paper hinges on three fundamental concepts-trade costs, market access and agricultural land value, and explores how their interaction drives regional economic transformation. It provides a rigorous analysis of how declining trade costs, driven by transportation network expansion, alter market access and reshape the spatial economy through changes in land values.

A) Falling Trade Costs and Land Use

Trade costs, broadly defined, refer to any extra costs incurred when selling in a location different from where production occurs. These include transportation costs, information frictions, and trade policy barriers. Donaldson and Hornbeck explicitly focus on the first component and how its fall affected the 19th-century American economy.

Historically, trade costs have declined dramatically over the past two centuries, a trend widely documented in the literature. Bairoch (1989), for example, shows that the share of transport costs in the final price of goods such as wheat, bar iron, cotton thread, and textiles fell significantly between 1830 and 1910. This decline is precisely the phenomenon Donaldson and Hornbeck study in the context of U.S. railroads: the impact of falling transportation costs for agricultural goods in remote areas.

In the urban economics literature, the monocentric city model interprets declining transportation costs as falling commuting costs. In the standard model (without constraint), lower commuting costs then translate into urban expansion and higher land prices at the former urban fringe, due to relaxed spatial constraints on labor mobility. However, this logic does not seem to apply to the setting studied by Donaldson and Hornbeck. Indeed, the 19th-century railroad system was primarily designed to reduce the cost of shipping goods, not people, as the authors point out; "the effect of railroads was mainly to reduce distances of expensive wagon transportation." Thus, the increase in land values observed in their study is not the result of pressure along

the bid-rent curve in agricultural areas, but rather a result of expanded "market access" for peripheral regions. Moreover, their framework does not model changes in land use (e.g. urban, rural); we will delve more into this aspect in the second essay.

Hence, the focus of the paper is not on within-city dynamics but on how remote agricultural counties, beyond the urban fringe, were transformed by improved access to distant markets. The land price increases are attributed here to expanded trade opportunities and the revaluation of economic geography, rather than urbanization or density effects.

B) Market Access and the NEG Framework

Drawing on Eaton and Kortum (2002) and Fujita, Krugman, and Venables (1999), Donaldson and Hornbeck define "market access" using a structural gravity model grounded in the trade literature. Their definition captures both local and network-wide effects of infrastructure on a county's ability to engage in trade, not only through direct connection, but also indirect effects via the network. Hence, market access becomes a weighted sum of other counties' economic size, with weights declining in bilateral trade costs. The novelty lies in their empirically grounded construction of this measure using historical GIS data and transport cost estimates, avoiding the use of geographic or Euclidean proxies for distance.

In the New Economic Geography (NEG) framework, first articulated by Krugman (1991), a firm's "market potential" is defined as the spatially discounted sum of demand across all locations. This spatial discount factor ϕ corresponds to the "freeness" of trade: $\phi = \tau - (\sigma - 1)$, which increases as trade costs τ fall. Accordingly, the expansion of the railroad network in Donaldson and Hornbeck increases the effective "freeness" of trade between counties, raising their market potential. The paper essentially shifts the focus from the firm in Krugman's model to the county, showing how improved trade linkages due to railroads raise economic value as proxied by land prices.

NEG models also suggest that trade costs are a central determinant of spatial disparities, especially across different phases of integration. Before the railroad boom, trade costs over land were prohibitively high. Most counties depended on waterways and rudimentary roads, leaving interior regions economically isolated. Railroads sharply reduced these costs, allowing such areas to connect with distant, densely populated markets. Donaldson and Hornbeck empirically document this phase transition: from a fragmented, high-cost economy toward a more integrated spatial structure.

Overall, their findings provide clear evidence of spatial integration via market access. Land values increased the most in counties where trade costs declined the most, consistent with the NEG prediction that reductions in iceberg trade costs contribute to economic development in newly accessible regions.. This

could be understood as the "convergence" phase of the NEG bell-shaped effect. In Donaldson and Hornbeck's framework, land values act as proxies for location-specific welfare or productivity, akin to wages or utility in traditional urban models, offering an alternative lens on the formation of core–periphery structures.

Finally, the paper's counterfactual simulations illustrate the essential role of trade costs in shaping spatial equilibria. Removing the railroad network effectively reverses the phase transition, pushing the economy back to a dispersed, high-cost configuration. In this scenario, land values collapse, especially in interior and non-coastal regions. The authors show that alternative investments in canals or roads would not have reached the cost-reduction threshold necessary to trigger full spatial integration, confirming the railroad system's role as a phase-shifting technology in economic geography.

C) NEG Extensions: Land Use and Trade Costs in Agriculture

The standard New Economic Geography (NEG) framework, as derived from Krugman (1991), provides a powerful account of how increasing returns to scale, trade costs, and factor mobility shape spatial economic outcomes. However, this baseline model includes several simplifying assumptions that limit its applicability to rural and agricultural settings; the most important in this context is zero trade costs on agricultural goods. Altogether, this assumption is not suited to explaining spatial dynamics in economies where agriculture remains a major sector and where land use and agricultural trade costs are spatially heterogeneous and important, such as the 19th-century American countryside. In this section, we discuss two extensions that try to address these shortcomings, related to Donaldson and Hornbeck (2016) finding: Picard and Zeng (2005) and Puga (1999), focusing respectively on trade costs and land use in agriculture, both providing a rationale for further dispersion forces.

Picard and Zeng (2005) extend the NEG framework by introducing agricultural trade costs as a key structural friction in spatial equilibrium. By relaxing the standard assumption of costless trade in agricultural goods, they show how transport frictions can create persistent regional inequalities, making peripheral areas appear unproductive when they are in fact under-connected. This introduces a powerful dispersion force: improvements in agricultural connectivity reduce trade costs, help interior regions become more competitive, and mitigate the tendency toward industrial agglomeration.

Their key insight is that when agricultural trade costs are high, rural regions suffer from lower wages and declining competitiveness, making them unattractive for manufacturing firms. But as agricultural trade costs fall, through improvements in transport infrastructure or connectivity, interior regions become more economically viable, encouraging the dispersion of industrial activity. In this way, agriculture itself becomes a source of spatial dispersion, moderating the core–periphery outcome emphasized in standard NEG. This

theoretical result closely mirrors the empirical mechanism documented in Donaldson and Hornbeck (2016): railroads lowered agricultural trade costs and enhanced rural market access, which was capitalized here into higher land values. While Picard and Zeng focus on the resulting shifts in wage structures and firm locations, Donaldson and Hornbeck measure how these general equilibrium forces affect the value of land as a fixed production factor. Both studies affirm that spatial frictions in agriculture play a critical role in determining equilibrium location outcomes, and both demonstrate that connectivity in the agricultural sector can be a key lever for achieving more balanced spatial development. Thus, it would be interesting to see if Picard's and Zeng's predictions on manufacturing spatial allocation hold in the context of the 19th-century American railway network expansion.

Finally, Puga (1999) extends the NEG framework by putting the emphasis on the non-monotonic relationship between trade costs and the spatial concentration, allowing for both agglomeration and re-dispersion as trade costs fall. A key feature of his two-sectors model is that arable land is a fixed, immobile factor in agricultural production. Agriculture, using land and labor under constant returns to scale, introduces spatial immobility and heterogeneity into the equilibrium. Although agricultural goods are traded without cost in the model, landowners consume locally, and differences in land endowment and wages create variation in agricultural returns. This structure ensures that agricultural regions maintain economic mass even when manufacturing activity centralizes, providing a spatial anchor that moderates full agglomeration.

Within this framework, Puga identifies three distinct phases of spatial equilibrium in response to falling trade costs: at high trade costs, economic activity remains dispersed to meet local demand; at intermediate levels, agglomeration emerges due to increasing returns and market-size effects; and at very low trade costs, re-dispersion sets in as cost sensitivities, immobile inputs, and congestion reduce the relative advantages of centralization.

While Puga's model is theoretical and abstracts from agricultural trade costs, it offers a compelling conceptual lens through which to interpret the findings of Donaldson and Hornbeck (2016). Their analysis can be viewed as an empirical application of Puga's framework to the historical U.S. setting. The paper's counterfactuals mirror the phases identified by Puga: before the railroad expansion, the U.S. was in a high trade cost and spatially dispersed phase. The arrival of railroads shifted the country into the intermediate trade cost phase, marked by spatial integration and gains from agglomeration through improved access to markets. The authors also simulate the removal of the railroad network, finding a sharp collapse in agricultural land values and a return to economic dispersion, which replicates Puga's low-access, high-cost equilibrium. Thus, like Puga, Donaldson and Hornbeck emphasize that infrastructure investments fundamentally alter the "economic geography phase" of a country and that changes in market access can reorganize the spatial distribution of productivity and welfare.

References

Allen, Treb, and Dave Donaldson. *Persistence and Path Dependence in the Spatial Economy.* Cambridge, Mass: National Bureau of Economic Research, 2020.

Bairoch, Paul. *European Trade Policy, 1815–1914.* In Mathias, P. and Pollard, S., editors, *Economic History of Europe*, vol. 7. Cambridge University Press, Cambridge, 1989.

Caliendo, Lorenzo, Maximiliano Dvorkin, and Fernando Parro. "Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock." *Econometrica* 87, no. 3 (2019): 741–835. doi:10.3982/ECTA13758.

Chambru, Cédric, Emeric Henry, and Benjamin Marx. "The Dynamic Consequences of State Building: Evidence from the French Revolution." *American Economic Review* 114, no. 11 (2024): 3578–3622. doi:10.1257/aer.20220110.

David, Paul A. "Transport Innovation and Economic Growth: Professor Fogel on and off the Rails." *Economic History Review* 22 (1969): 506–524.

Donaldson, Dave, and Richard Hornbeck. "Railroads and American Economic Growth: A 'Market Access' Approach." *Quarterly Journal of Economics* 131, no. 2 (2016): 799–858. doi:10.1093/qje/qjw002.

Eaton, Jonathan, and Samuel Kortum. "Technology, Geography, and Trade." *Econometrica* 70, no. 5 (2002): 1741–79. doi:10.1111/1468-0262.00352.

Fajgelbaum, Pablo D., and Edouard Schaal. "Optimal Transport Networks in Spatial Equilibrium." *Econometrica* 88, no. 4 (2020): 1411–52. doi:10.3982/ECTA15213.

Fogel, Robert William. *Railroads and American Economic Growth: Essays in Econometric History.* Baltimore: Johns Hopkins Press, 1964.

Fujita, Masahisa, Paul R. Krugman, and Anthony J. Venables. *The Spatial Economy: Cities, Regions, and International Trade.* Cambridge, Mass. and London: MIT Press, 1999.

Holmes, Thomas J., and James A. Schmitz. "Competition at Work: Railroads vs. Monopoly in the U.S. Shipping Industry." *Federal Reserve Bank of Minneapolis Quarterly Review* 25 (2001): 3–29.

Jedwab, Remi, and Alexander Moradi. "The Permanent Effects of Transportation Revolutions in Poor Countries: Evidence from Africa." *Review of Economics and Statistics* 98, no. 2 (2016): 268–84. doi:10.1162/rest_a_00540.

Krugman, Paul. "Increasing Returns and Economic Geography." *Journal of Political Economy* 99(3): 484–499, 1991.

Lebergott, Stanley. "United States Transportation Advance and Externalities." *Journal of Economic History* 26 (1966): 437–461.

Mayer, Thierry, and Soledad Zignago. "Market Access in Global and Regional Trade." HAL CCSD, 2005.

Picard, Pierre M., and Dao-Zhi Zeng. "Agricultural Sector and Industrial Agglomeration." *Journal of Development Economics* 77, no. 1 (2005): 75–106. doi:10.1016/j.jdeveco.2004.03.002.

Puga, Diego. "The Rise and Fall of Regional Inequalities." *European Economic Review* 43, no. 2 (1999): 303–34. doi:10.1016/S0014-2921(98)00061-0.

Redding, Stephen J., and Anthony J. Venables. "Economic Geography and International Inequality." *Journal of International Economics* 62, no. 1 (2004): 53–82. doi:10.1016/j.jinteco.2003.07.001.